VIRUS EVOLUTION

Virus Evolution, 2024, **10(1)**, 1–13
DOI: https://doi.org/10.1093/ve/vead082
Advance Access Publication 28 December 2023
Research Article

Computational multigene interactions in virus growth and infection spread

Bradley Schwab and John Yin*,†

Wisconsin Institute for Discovery, Chemical and Biological Engineering, University of Wisconsin-Madison, 330 N. Orchard Street, Madison, WI 53715, USA
†https://orcid.org/0000-0001-6146-0594
*Corresponding author: E-mail: john.yin@wisc.edu

Abstract

Viruses persist in nature owing to their extreme genetic heterogeneity and large population sizes, which enable them to evade host immune defenses, escape antiviral drugs, and adapt to new hosts. The persistence of viruses is challenging to study because mutations affect multiple virus genes, interactions among genes in their impacts on virus growth are seldom known, and measures of viral fitness are yet to be standardized. To address these challenges, we employed a data-driven computational model of cell infection by a virus. The infection model accounted for the kinetics of viral gene expression, functional gene–gene interactions, genome replication, and allocation of host cellular resources to produce progeny of vesicular stomatitis virus, a prototype RNA virus. We used this model to computationally probe how interactions among genes carrying up to eleven deleterious mutations affect different measures of virus fitness: single-cycle growth yields and multicycle rates of infection spread. Individual mutations were implemented by perturbing biophysical parameters associated with individual gene functions of the wild-type model. Our analysis revealed synergistic epistasis among deleterious mutations in their effects on virus yield; so adverse effects of single deleterious mutations were amplified by interaction. For the same mutations, multicycle infection spread indicated weak or negligible epistasis, where single mutations act alone in their effects on infection spread. These results were robust to simulation in high- and low-host resource environments. Our work highlights how different types and magnitudes of epistasis can arise for genetically identical virus variants, depending on the fitness measure. More broadly, gene–gene interactions can differently affect how viruses grow and spread.

Keywords: gene-gene interactions; epistasis; genotype-phenotype; biophysical modeling; virus; network kinetics; growth; spread; infection; fitness; vesicular stomatitis virus; VSV.

Introduction

Epistasis describes how gene–gene interactions contribute to traits or behaviors of biological processes or organisms. Traits of complex diseases such as diabetes, asthma, or hypertension may depend on contributions and interactions from multiple genes (Cordell 2002). An understanding of epistasis in complex traits, including disease, has to a large extent advanced by systematically studying how genetic variations in model organisms contribute to quantifiable traits (Moore and Williams 2009; Mackay 2013). Approaches of systems biology, which combine mechanistic understanding of individual gene functions with mathematical and computational representations, can forecast behaviors or traits of processes in living cells, tissues, and organisms; such approaches offer a potentially useful way to computationally quantify epistasis (Segrè et al. 2005; Phillips 2008). More specifically, kinetic models of virus reproduction within their host cells can provide a foundation for probing how gene–gene interactions impact virus growth (You and Yin 2002; Yin and Redovich 2018).

It is well-known that genetic variation within individual genes of a virus can have critical impacts on its growth and spread; noteworthily, single-gene examples that impact virus fitness are the reverse transcriptase of human immunodeficiency virus 1 (HIV-1) (Goudsmit et al. 1997), the neuraminidase of influenza A virus (Yen et al. 2005), and the spike protein of severe acute respiratory syndrome coronavirus 2 (Harvey et al. 2021). Despite the depth of knowledge and understanding of nucleotide-level variation on these and other gene functions, and ultimately virus growth and spread, our understanding of how variants of two or more genes interact to influence such behaviors is limited. Studies of nucleotide substitution mutations and interactions between mutations have provided evidence for adaptation of virus fitness by multiple small steps for phage ϕ6 (Burch and Chao 1999), they have shown how pairs of deleterious or beneficial mutations can exhibit positive or negative epistasis for vesicular stomatitis virus (VSV) (Sanjuán, Moya, and Elena 2004), and they have suggested how variation across clinical isolates of HIV-1 exhibits positive epistasis (Bonhoeffer et al. 2004). However, such studies have neglected to account for how mutations map onto gene functions and interactions in their impacts on virus growth or correlates of fitness. Here, we show how a data-driven computational model of virus growth can be adapted to explicitly account for functional gene–gene interactions in their impacts on quantifiable proxies for virus growth and infection spread.

It is feasible to perform 'computational experiments' in epistasis on virus growth and infection spread. Studies of such virus behaviors are enabled by several features: (1) virus genomes are small and are in many cases well-studied; they have a manageable or countable number of often well-characterized molecular functions (Fields, Knipe, and Howley 2013); (2) key environmental factors, specifically, host cell resources are relatively well understood with respect to how they can contribute to virus growth (You, Suthers, and Yin 2002; Kim and Yin 2004; Lim et al. 2006; Birch, Ruggero, and Covert 2012); (3) molecular mechanisms of virus growth have a significant overlap with known mechanisms of gene expression (transcription and translation), genome replication, packaging, and progeny release (Fields, Knipe, and Howley 2013); and (4) mathematical and computational models or representations of how such processes individually and collectively contribute to growth and spread behaviors are now feasible (Yin and Redovich 2018). Here, we focus on VSV, a prototypical RNA virus, where the mechanisms of its transcriptional regulation, gene expression, and genome replication are relatively well understood (Lyles, Kuzmin, and Rupprecht 2013), as highlighted in Fig. 1. These mechanisms have been incorporated into a biophysical model of VSV growth kinetics during infection of a host cell (Lim et al. 2006). In this work, the VSV model is employed to simulate how one or more mutations in VSV functions interact to affect the kinetics of VSV production.

Modeling epistasis

To quantitatively characterize different types of epistasis, consider a simple case that describes how a single deleterious mutation, x, and a more deleterious mutation, y, interact in their effects on organismal fitness. Specifically, suppose that the wild type has fitness, $w = 1$, and variants carrying x or y have fitnesses 0.9 or 0.6, respectively. In the simplest case, a variant genome with both mutations x and y will behave in a way where mutations each contribute independently, without interaction, yielding a fitness that is their product, $(x)(y)$, or 0.54. Alternatively, if the mutations interact in a way that exacerbates their deleterious effects, then synergistic epistasis results and the fitness of the organism is less than 0.54; and if the interaction between mutations softens or buffers their deleterious effects, then the result is antagonistic epistasis with organismal fitness above 0.54. This example can be generalized to quantify how more than two mutations interact in their effects on fitness. Specifically, by using a power function as previously proposed (Lenski et al. 1999), one may regress measures of mutant fitness (w_i) versus number of deleterious mutations (n) for computational fitness variants:

$$(w_i) = -\alpha n^{\beta}, \quad \text{(a)}$$

where different metrics of fitness (i) are all relative to wild type, so $0 < w_i < 1$ for variants that are less fit than wild type. The parameter α is positive, and for deleterious mutations, its magnitude is larger for mutations that have greater deleterious effects on fitness; the parameter β characterizes the type of epistasis among mutations, described in Fig. 2.

During a single generation or cycle of virus growth, a common measure or correlate of virus fitness is the 'burst size' or average yield of virus particles per infected cell (Y); the average yield is experimentally measured by simultaneously exposing a large population of susceptible cells ($\sim 10^6$) to a population of virus particles that are in excess, with multiplicity of infection 3–10, where 3–10 virus particles are available to infect each cell (You, Suthers, and Yin 2002; Lam, Duca, and Yin 2005). After allowing time for adsorption, typically an hour, particles that have not adsorbed are washed away, culture medium is replenished, and the titer or level of released progeny particles is estimated by sampling the supernatant solution at different times and performing the plaque assay. Division of the virus titer at each time by the number of infected cells provides the average number of virus particles released per cell, $N(t)$; a plot of these values over the time course of infection provides the one-step growth curve (Fig. 3(A)), where the total particles released per cell is the growth yield (Y). The one-step growth behavior of $N(t)$ versus time (t) may also be used to estimate a correlate of fitness associated with infection spread. Specifically, in an environment with plentiful host cell resources, one may anticipate fitness that is not limited to a single cycle or generation of virus production, but instead accounts for potentially multiple cycles that occur during the spread of infection. Then, it would be advantageous for the virus to not only make as many particles as possible but also do so as quickly as possible and move on to infect further rounds of susceptible host cells. After multiple cycles of growth spanning time (t), the number of viral particles produced will be approximately $Y^{\frac{t}{\tau}}$ where Y is the average yield per cycle, (τ) is the time within a cycle that maximizes production over multiple cycles, and $\frac{t}{\tau}$ is the number of infection cycles. This number may be written as $\left[Y^{\frac{1}{\tau}}\right]^t$, so a fitness correlate for infection spread should maximize $Y^{\frac{1}{\tau}}$ over multiple cycles of infection. One may then define a time-dependent correlate of fitness, rate (R) that equals $\max\left\{N(t)^{\frac{1}{t}}\right\}$, which may be estimated from the $N(t)$ versus time (t) behavior over a single cycle, as shown in Fig. 3(B); in short, we find the value of t that maximizes the rate (R).

To regress the simulated behavior of single-cycle yields from variant virus infection relative to wild type, we can simulate one-step growth curves, as in Fig. 3(A), for wild-type and variant virus, estimate their respective average yields per cell, Y_{wildtype} and Y_{variant}, respectively, and define w_{growth} $(= Y_{\text{variant}}/Y_{\text{wildtype}})$, which we determine for diverse variants carrying different numbers of deleterious mutations (n) and fit to Equation 1. Similarly, for a correlate of fitness based on infection spread, which maximizes the rate of virus production; we can estimate rates for wild-type and variant viruses, R_{wildtype} and R_{variant}, respectively, and define w_{spread} $(= R_{\text{variant}}/R_{\text{wildtype}})$, which can be fit to Equation 1.

Here, we study VSV, a negative-sense, single-stranded RNA virus that has served as a model system for laboratory studies of virus evolution (Holland et al. 1982) and for computational kinetic modeling of virus intracellular growth (Lim et al. 2006; Yin and Redovich 2018). Computational models of virus intracellular growth link mechanistic mathematical representations, typically ordinary differential equations, of the steps that define a single cycle of host cell infection: virus particle adsorption to their cell surface receptors, endocytosis, membrane fusion, expression of viral genomes to be delivered to the cells, transcription and translation of virus functions, assembly of progeny virus particles from genomes and structural proteins, and release of progeny particles into the extracellular environment. Parameters of the model, which are determined from independent experiments, have in most cases been extracted from the biophysical and molecular biological literature and, in some cases, have been estimated from growth or other quantitative experiments (Lim et al. 2006). Specific details of the VSV model and its implementation in the current work are given in the Methods section.

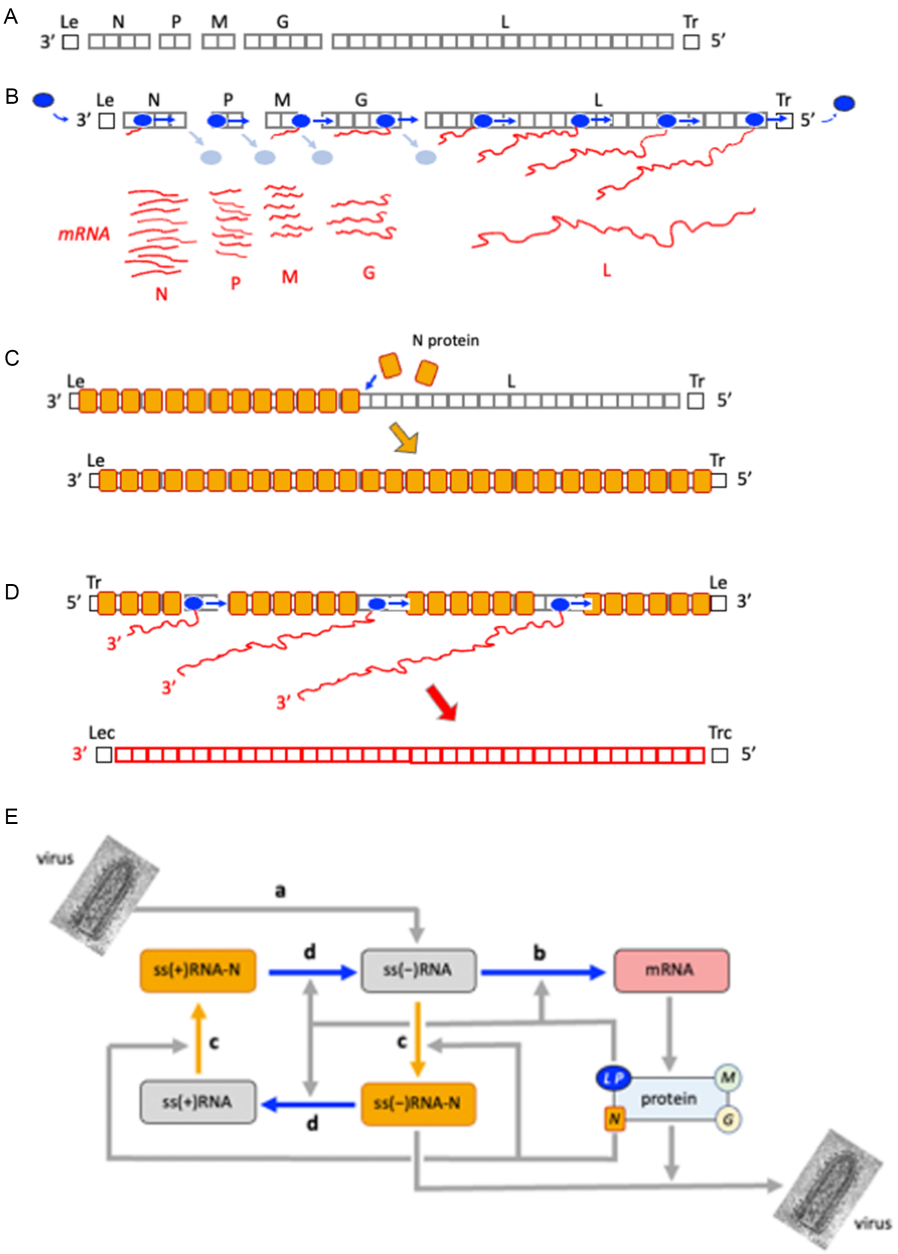

Figure 1. Essential genome structure and mechanisms of VSV growth during infection. (A) The genome of VSV is a single-stranded negative-sense RNA of 11,161 nucleotides, encoding a leader sequence (Le), five genes (nucleocapsid protein, N; phosphoprotein, P; matrix protein, M; glycoprotein, G; large protein, L), and a trailer sequence (Tr). Gaps between genes are intergenic sequences that encode partial transcription termination sites. (B) The negative-sense RNA genome is used as a template for transcription by the viral polymerase (filled oval), a complex of L and P proteins. Transcription is initiated when the polymerase binds a single promoter at the genomic 3′ end; synthesis of mRNA proceeds for each gene, and fraction of polymerases terminates at each intergenic site, causing a gradient in mRNA (and subsequent protein) expression from genes N through L. (C) High levels of N protein expression drive full encapsidation of the viral genome, and (D) the fully encapsidated genome then serves as a template for the polymerase to make the complementary full-length anti-genome, which carries complementary leader (Lec) and complementary trailer (Trc) sequences. Similarly, the N protein encapsidates the full-length anti-genome, which then serves as a template for replication of the full-length genome (not shown). (E) The underlying genetic architecture of VSV shows how expression of virus genes and emergence of their functions play out in a sequence of processes that culminate in the release of progeny virus particles. Here, the image of a VSV particle was adapted from Zhou et al. (2022). A VSV particle binds to its host cell receptor, enters the cell cytoplasm, and uncoats the virus genome or ss(-)RNA, described in (a), which then serves as a template for production of virus transcripts, by the mechanism shown in (b). The transcripts serve as templates for synthesis of the five proteins, and a complex of L and P proteins forms the RNA-dependent RNA polymerase (filled oval). Accumulation of N protein (filled rectangle) leads to encapsidation of the virus genome, as in (c), and the encapsidated genome serves as a template for synthesis of the full-length anti-genome, as in (d); likewise, encapsidation of the full-length anti-genome by N protein is followed by replication of the full-length virus genome. Finally, encapsidated genomes combine with other viral proteins to assemble and bud from the host membrane, yielding progeny virus particles.

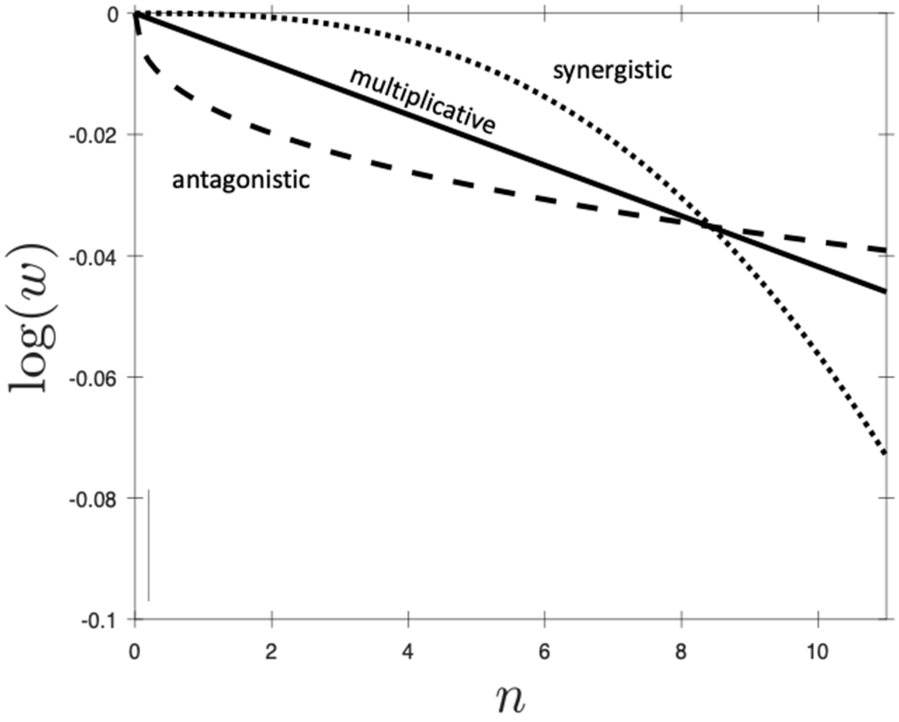

Figure 2. Interactions among deleterious mutations reflect different forms of epistasis. The average fitness of a variant (w) drops as its deleterious mutations increase in number (n). In the simplest case, mutations act independently in their effects on fitness, so the fitness of a variant carrying two or more mutations can be estimated by multiplying the fitness associated with variants carrying their single mutations; epistasis is multiplicative, $\beta = 1$. Alternatively, interactions can synergize or exacerbate effects of individual mutations ($\beta > 1$), or interactions can antagonize each other or buffer the effects of individual mutations ($\beta < 1$), where β refers to the exponent in Equation (a). The average fitness of a variant is defined relative to the wild-type fitness, so w for variants with deleterious mutations is generally less than unity, and $\log(w)$ is negative.

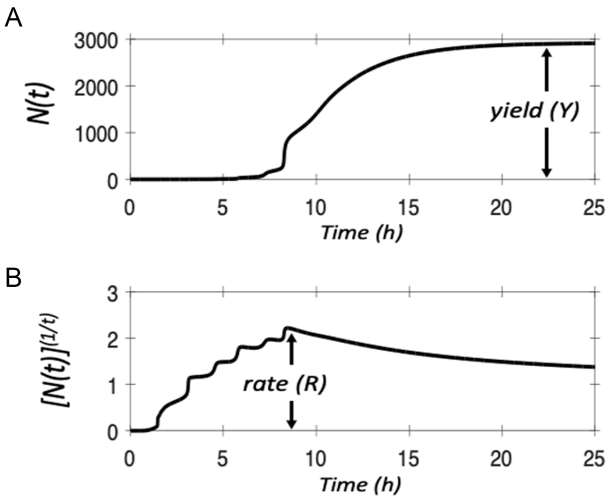

Figure 3. Correlates of fitness for virus growth and infection spread. The release of virus progeny during a single cycle of growth enables estimates of (A) virus growth (yield, Y) and (B) infection spread (rate, R). Details are given in the main text.

Results

Epistatic interactions for virus growth are synergistic

Interactions between deleterious mutations in their effects on single-cycle virus growth exhibited synergistic epistasis, as evident from the concave down patterns of $\log(w_{growth})$, or fitness relative to wild type, versus number of mutations across a range of mutation severity (Fig. 4A). For the most severe class of mutations in their effects on fitness, estimation of the exponent, β, was larger than three and decreased as severity effects were reduced; but in all cases, β was greater than unity, consistent with synergistic interactions (Fig. 4B). When the intracellular growth model was parametrized for a host cell that provides lower-resource levels (delayed-brain tumor (DBT) cells) for virus growth (Lim et al. 2006) than the base case (baby hamster kidney (BHK) cells), qualitatively similar results were found for implementation of the same mutations across five severity classes (Fig. 4C and D).

Epistatic interactions for infection spread are weak or negligible

Interactions between deleterious mutations in their effects on multicycle virus growth (infection spread) exhibited weak or negligible epistasis, as evident from the effectively linear patterns of $\log(w_{spread})$, or fitness relative to wild type, versus the number of mutations across a range of mutation severity (Fig. 5A). Accordingly, estimation of the exponent, β, was close to unity, consistent with mutations acting independent of each other (Fig. 5B). When the intracellular growth model was parametrized for a host cell that provides lower-resource levels (DBT cells) for virus growth than the base case (BHK cells), qualitatively similar results (minimal epistasis) were found for implementation of the same mutations across five severity classes (Fig. 5C and D).

Discussion

At a basic level, life history traits of viruses describe how single and multiple cycles of host cell invasion enable virus reproduction and infection spread. Such traits can be quantified by employing a

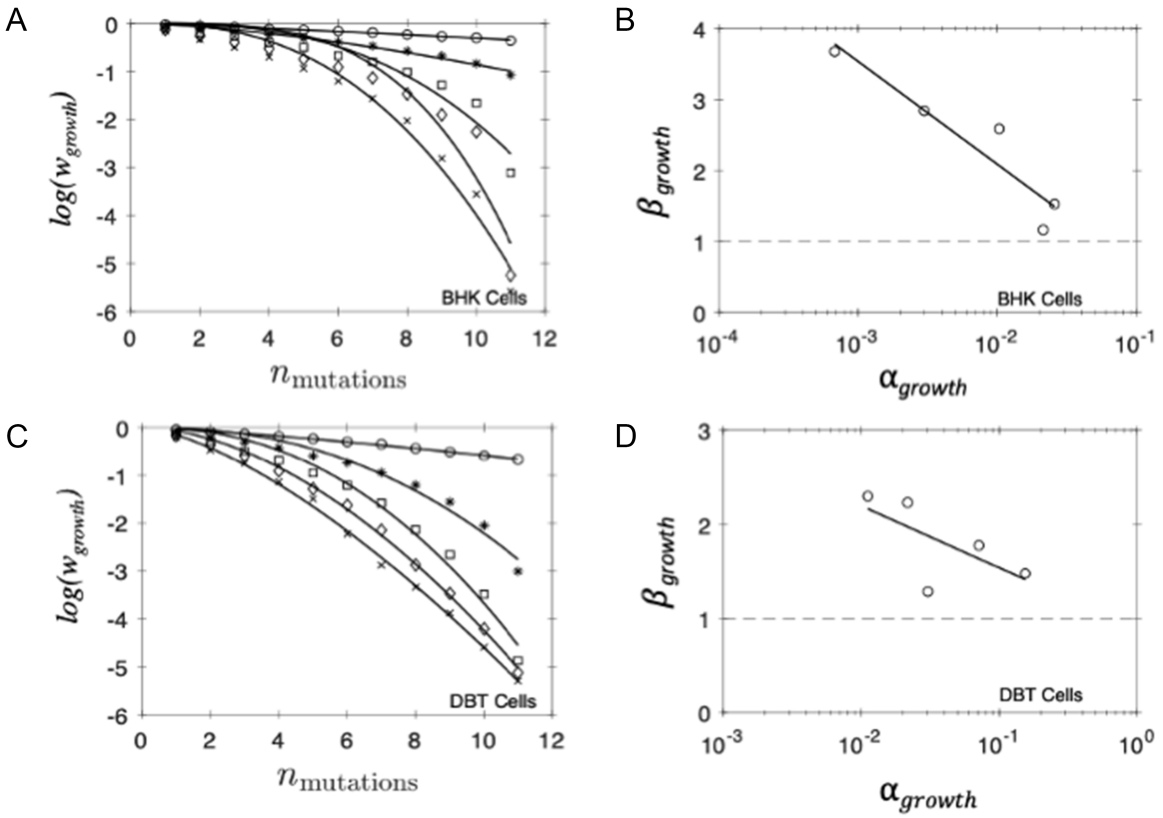

Figure 4. Interactions among deleterious mutations are synergistic in their effects on virus growth. (A) Virus fitness drops as the number of deleterious mutations increases; the concave down dependence is indicative of synergistic epistasis. Fitness here was based on the yield (Y) of virus progeny from simulations of virus one-step growth for different severity and number of mutations. Five classes of severity are shown, and simulations were parameterized for growth on BHK host cells. (B) Dependence of β on α as defined by Equation (a) and Fig. 2. Synergistic epistasis is associated with $\beta > 1$, and larger α are associated with more deleterious effects of mutation on fitness. (C) and (D) are like (A) and (B), except they have been parameterized for growth on DBT cells, which provide fewer resources for virus growth than BHK cells.

kinetic model for virus intracellular growth; the model can enable one to compute the time-dependent release of virus progeny from an average infected host cell, commonly known as the one-step growth behavior of the virus. Two features of this behavior can be mapped onto correlates of fitness for single- and multicycle growth.

When a single cell is infected, the average yield or number of virus progeny produced can be readily estimated from the simulated one-step growth. In addition, when multiple cells are available and an initial cell is infected, then we expect a fitness correlate for multicycle growth to depend on the timing of virus progeny production and release; earlier or more rapid release of progeny particles from the initial and subsequent infected cells would have a selective advantage over later or more slowly released particles. We have shown here how the simulated one-step growth behavior can also be used to estimate such a fitness correlate for multicycle growth associated with infection spread.

We have simulated the effects of mutations on virus growth by simply altering biophysical parameters of the model from their base-case (or wild-type) values and confirmed when they have a deleterious effect on virus growth. However, the mechanistic link between a mutation and its effect on growth is not straightforward. For example, a single mutation in the VSV polymerase could affect multiple interactions or processes; these could include the elongation rate of the polymerase, the strength of promoters, or the degree of intergenic attenuation, affecting multiple parameters of the model. Alternatively, mutations in different genes could have the same or similar effects on a process characterized by a single parameter; for example, mutations in the viral nucleocapsid and matrix protein could be described by a change in the rate of association between these proteins. Our approach has been to note the extensive literature linking mutations in proteins to changes in their biophysical properties and parameters (DePristo, Weinreich, and Hartl 2005), so the changes we implemented in biophysical parameters of VSV proteins could be plausibly linked to specific mutation. We did not consider what specific mutations would be needed to enable such changes, which would be beyond the scope of the current study.

Comparable results were found when a kinetic model of bacteriophage T7 growth in host *Escherichia coli* was used to probe gene–gene interactions (You and Yin 2002). In the T7 model, correlates of fitness were defined based on calculated one-step growth behavior in 'poor' and 'rich' host resource environments. Fitness in a poor host resource environment for phage T7 was calculated in the same way as *growth* yield of virus progeny in the current work for VSV from a single cycle of infection; considering the extreme case where a virus infects a single cell, and no other cells are available for subsequent infection, a plausible measure of fitness is the overall yield of virus produced, without an associated penalty or benefit for slower or faster production. Moreover, fitness in a rich resource environment for phage T7 is

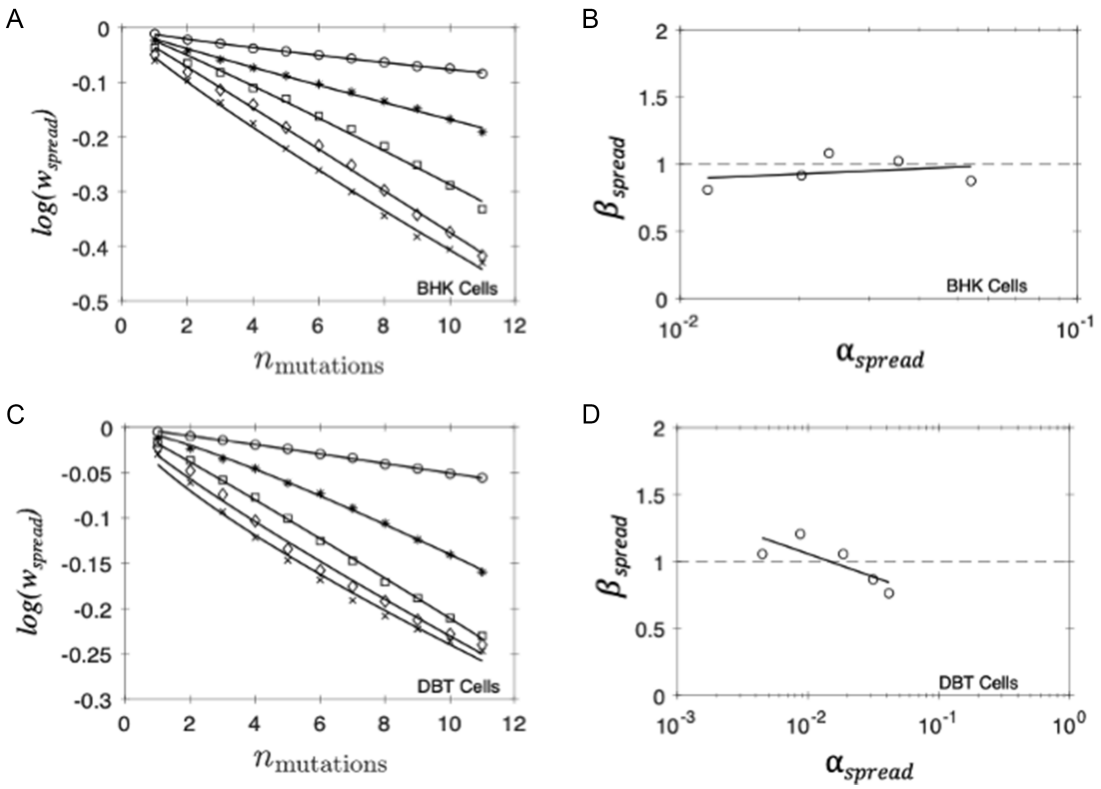

Figure 5. Interactions among deleterious mutations are weak or negligible in their effects on virus infection spread. (A) Virus fitness drops as the number of deleterious mutations increases; the near-linear dependence is indicative of non-interacting mutations or negligible epistasis. Fitness was based on the optimal rate (R) of virus progeny production from repeated one-step growth, a metric for infection across a population of host cells; simulations are shown for virus production rate for different severity and number of mutations. Five classes of severity are shown, and simulations were parameterized for growth on BHK host cells. (B) Dependence of β on α as defined by Equation (a) and Fig. 2. Negligible epistasis is associated with $\beta \simeq 1$, and larger α are associated with more deleterious effects of mutation on fitness. (C) and (D) are like (A) and (B), except that they have been parameterized for growth on DBT cells, which provide fewer resources for virus growth than BHK cells.

analogous to the fitness of infection *spread* over multiple cycles in the current work for VSV; here, the initial infection of a susceptible cell occurs in an environment where many cells are available to serve as hosts for subsequent generations or cycles of growth, so faster initial and subsequent progeny production plausibly correlates with higher fitness. For analogous cases, phage T7 production on poor resources or VSV single-cycle growth, synergistic epistasis was found for interactions among deleterious mutations in their effects on fitness. However, for phage T7 production on rich resources or VSV multicycle spread, T7 exhibited purely antagonistic epistasis, while epistasis for VSV was weak or negligible.

In the current work, differences in VSV growth on susceptible (BHK) versus relatively resistant (DBT) host cells yielded similar results: single-cycle growth exhibited synergistic epistasis, while multicycle spread gave weak or negligible epistasis. However, a subtle difference was revealed by comparing the plots of β versus α for spread on BHK or DBT host cells. On BHK cells, the deleterious mutations acted independent of each other (or $\beta \simeq 1$), independent of the mutational severity (α), while on DBT host cells, β versus α exhibited a negative slope, indicating a transition from synergistic to antagonistic epistasis as mutational severity increased. A similar synergistic-to-antagonistic transition with increasing mutational severity was observed for phage T7 under poor host resource (or single-cycle) growth conditions. Antagonistic epistasis between deleterious mutations was dominant in a theoretical model of virus-inspired RNA replication (Elena, Solé, and Sardanyés 2010); however, the model neglected entry, gene expression, and other essential processes for virus intracellular growth. To the best of our knowledge, no other systems beyond VSV and T7 have exhibited such transitions between forms of epistasis, as driven by the severity of mutations in their effects on fitness.

The key result of this work is that interactions among the same mutations can give different magnitudes and signs of epistasis, depending on how organismal fitness is defined. For virus fitness defined by the fecundity or progeny yield from infected host cells, interactions among deleterious mutations were synergistic; but for virus fitness defined by the yield-per-time or fecundity rate of infected cells, the extent of interaction between mutations was reduced, allowing for transition between weakly positive or negative epistasis. These results could be tested in the laboratory by identifying mutations that cause detectable changes in the average fecundity or production rate of virus progeny from infected cells. For example, deep mutational scanning (DMS) can be used to make all amino acid variants at each residue of a target protein (Fowler and Fields 2014), and DMS has been used to make variants of viral proteins that bind to cellular receptors (Haddox, Dingens, and Bloom 2016; Starr et al. 2020), replicate viral genomes (Soh et al. 2019), and perform other functions important for virus growth (Burton and Eyre 2021). Variant viruses carrying one or more mutations, generated by DMS or related methods, could be characterized in the laboratory based on features of their one-step growth, such as yield or production rate (Jin and Yin 2021). Such

variants could also be assessed using correlates of virus fitness, including effects of infection on host density (Turner, Draghi, and Wilpiszeski 2012), live-cell reporters of virus infection (Akpinar, Timm, and Yin 2015), or rates of infection spread (Yin 1991; Lee and Yin 1996; Akpinar, Inankur, and Yin 2016). For each measure of fitness, different variants carrying one or more deleterious mutations could be analyzed, as we have done here. Differences in fecundity, measured by progeny yield, extent of host lysis, or reporter level, would exhibit synergistic epistasis; by comparison, differences in fitness based on rates, measured by progeny production rate, the rate of host lysis, or the rate of reporter expression, would be relatively little if any epistasis is present.

Here, we have taken a first step toward considering how gene–gene interactions might exacerbate or buffer effects of deleterious mutations on virus behaviors, focusing on virus growth and infection spread. Future work will give a more comprehensive picture; in particular, such studies might also account for the effects of genetic variation of eukaryotic viruses on activating host innate and adaptive immune responses to infection. Moreover, future work might elucidate how epistasis impacts transmission between hosts as well as the stability of virus particles in aerosols, on surfaces, and other environments inhabited by their hosts.

Methods

Modeling virus growth

Using algebraic and differential forms of equations, our mathematical model aims to account for established molecular processing steps in the development of VSV. Most model parameters were extracted from the literature. However, five parameters were obtained by fitting our simulation results to experimental data that were from the literature and our own experiments. Key model parameters are provided in Table S1 (Supporting Information), and details of the parameter estimation process have been published (Lim et al. 2006).

Virus binding and penetration

VSV initiates an infection by binding to a receptor such as phosphatidylserine, a lipid component in the plasma membrane (Schlegel et al. 1983; Rose and Whitt 2001). After the binding step, the VSV particle is endocytosed via a clathrin-coated pit and then penetrates intracellular vesicles such as endosomes by membrane fusion (Matlin et al. 1982; Florkiewicz and Rose 1984; Superti et al. 1987; Rose and Whitt 2001). The penetration leads to the release of the encapsidated negative-sense viral genome and virus proteins into the cytoplasm of the host cell. By assuming that the binding step is rate-determining (Miller and Lenard 1980), we lump these early steps from the binding to the penetration into a first-order expression:

$$\frac{dV_b}{dt} = k_b V_{ex}, \tag{1}$$

where V_b and V_{ex} are the concentrations of bound and extracellular virus particles, respectively, t is time, and k_b is the apparent rate constant for virus binding. After binding, we assume that the bound virus is immediately endocytosed and fused, and its genome and protein components are instantaneously released into the cytoplasm at the expense of the fused virus particle. The protein stoichiometry of a single VSV particle and the lengths of each viral gene and protein are summarized in Table S2 (Supporting Information) (Thomas et al. 1985; Wagner 1987).

Population distribution of polymerases and nucleocapsids

Following the release of the encapsidated genome and proteins into the cytoplasm, VSV transcription is initiated. The viral transcription was assumed to be independent of host cell functions such as replication (Simonsen, Batt-Humphries, and Summers 1979). Instead, the viral complex of L and P proteins, with a stoichiometry of 1–3.6, was taken to function as polymerase in transcription and replication (Gao and Lenard 1995). In the absence of P protein, L protein cannot bind to the genome or anti-genome (Mellon and Emerson 1978). After binding to the 3′ promoter regions of the genomic and anti-genomic templates, the viral polymerase starts to synthesize its own RNA transcription and replication products by elongating along the templates. During transcription, a fraction of elongating polymerases terminates transcription by dissociating from the templates as they encounter regulatory signals in intergenic regions (Iverson and Rose 1981; Rose and Whitt 2001). In addition to the regulated polymerase dissociation, time-dependent concentration changes of the polymerases and the viral templates in the cytoplasm influence the distribution of polymerases on the viral templates during transcription and replication. Hence, the distribution of polymerases continuously varies over the viral templates, ultimately determining the relative synthesis levels of mRNAs and genome-size RNAs.

We simulate the transcription and replication processes by considering the spatial–temporal distribution of template-associated polymerases. We first partition the viral genome and anti-genome templates into forty segments, excluding their 3′ and 5′ end regions, which are the leader (Le) and trailer regions (Tr) for the genome and the complementary trailer (Trc) and complementary leader regions (Lec) for the anti-genome, respectively. For the genome template that is used for transcription as well as replication, we specially grouped the segments into five genes. We chose forty as a minimum number of total segments that allows each gene to be split into a specific integer number of segments, proportional to the length of the gene. By considering the mechanisms for the interactions between polymerase and the intergenic regulatory sequences of the templates, as described in the Transcription section, we simulated the polymerase flux into each segment over the time elapsed from the initiation of transcription on each template. Then, we correlated the level of polymerase occupying each gene-encoding section of the template with the synthesis rate of each corresponding viral mRNA. In a similar way, the distribution of polymerases on the replication templates was correlated with the synthesis rate of viral genome-sized RNA. Such explicit treatment of polymerase spatial distributions on the viral genome and anti-genome templates was central to modeling the growth of wild-type and gene-rearranged virus strains. This treatment systematically accounts for polymerization-associated time delays and the polymerase fluxes into each template segment.

Before estimating the polymerase flux, we need to figure out how the polymerase complex and M protein compete for binding to the genomic nucleocapsid as well as how the polymerases bound to nucleocapsids are subsequently distributed to one of three possible tasks: transcription, replication of genome, or replication of anti-genome. In our model, we assume that the genomic templates (negative-sense nucleocapsid) whose promoters (leader regions) are free of polymerases are available for association with free polymerase or M protein. We further assume that the associations of the free genomic templates by M proteins or polymerases

take place instantaneously:

$$(-)\mathrm{nc} - \mathrm{pol_l}\frac{S_{pol}}{l_l} = (-)\mathrm{nc}^{M,new} + (-)\mathrm{nc}^{pol,new}, \quad (2)$$

where (–)nc, $(-)\mathrm{nc}^{M,new}$ and $(-)\mathrm{nc}^{pol,new}$ are the concentrations of total genomic nucleocapsids and subsets of genomic nucleocapsids whose promoters are newly occupied by M protein and polymerase, respectively. S_{pol} is the spacing between neighboring polymerases on the genomic or anti-genomic template, $\mathrm{pol_l}$ is the concentration of polymerases bound to the promoter region (Le) of the genomic template, and l_l is the length of the promoter region. Specifically, the second term in the left-hand side of the equation denotes the concentration of the genomic templates whose promoters are currently occupied by polymerases. In our model, the concentration of the genomic nucleocapsids whose promoters are bound to polymerases and the concentration of the polymerases bound to the promoters of the genomic nucleocapsids are interchangeable with each other by the factors (l_l/S_{pol}) and (S_{pol}/l_l), respectively. The binding of M protein or polymerase initiates reactions leading to virion assembly or RNA synthesis, respectively. Because the initiation of RNA synthesis by the polymerase requires a finite time, a space between adjacent polymerases on the template (S_{pol}) would be maintained during infection, assuming a fixed elongation rate. With these considerations, one may expect that at any time, the concentration of nucleocapsids available for the new binding of the free proteins is inversely proportional to the concentration of polymerases currently bound to the leader region of the genomic nucleocapsids ($\mathrm{pol_l}$) and the polymerase spacing (S_{pol}) as shown in the second term of Equation 2. The ratio of $(-)\mathrm{nc}^{M,new}$ to $(-)\mathrm{nc}^{pol,new}$ is determined by the ratio of the association rates of M protein and polymerase with the genomic nucleocapsid, which is further a function of the rate constants and relative amounts of the corresponding free components in the cytosol:

$$\frac{(-)\mathrm{nc}^{M,new}}{(-)\mathrm{nc}^{pol,new}} = \frac{r_{asso,M}}{r_{asso,pol}} = \frac{k_M(1-\mathrm{cond_M})M}{k_{pol}(L\cdot\mathrm{trans}-\mathrm{pol_{total}})} = \frac{S_{cond}(1-\mathrm{cond_M})M}{(L\cdot\mathrm{trans}-\mathrm{pol_{total}})}, \quad (3)$$

where $r_{asso,M}$ and $r_{asso,pol}$ are the rates of the associations of M protein and polymerase with the genomic nucleocapsid, respectively, and k_M and k_{pol} are the rate constants for each association reaction, respectively. S_{cond} denotes the ratio of the two rate constants ($=k_M/k_{pol}$). Unlike L protein, 10 per cent of synthesized M proteins are associated with the plasma membrane (Chong and Rose 1993). In the Equation (3), $\mathrm{cond_M}$ is the fraction of M proteins associated with the plasma membrane, trans is the fraction of L proteins satisfying the polymerase stoichiometry with P protein, $\mathrm{pol_{total}}$ is the total concentration of polymerases associated at the time with nucleocapsids, and M and L are the total concentrations of M and L proteins not assembled into viral progeny. If the concentration of P protein (P) is larger than 3.6-fold concentration of L protein, then trans is equal to 1. Otherwise, trans is equal to $P/(3.6L)$. In our model, M and L proteins compete for free genomic nucleocapsids, and the condensed nucleocapsids, owing to their association with M proteins, cannot be utilized for transcription or replication (Rose and Whitt 2001). From Equations 2 and 3, the newly occupied nucleocapsids by polymerases ($(-)\mathrm{nc}^{pol,new}$) can be calculated:

$$(-)\mathrm{nc}^{pol,new} = \left((-)\mathrm{nc} - \mathrm{pol_l}\frac{S_{pol}}{l_l}\right)\frac{1}{\frac{S_{cond}(1-\mathrm{cond_M})M}{(L\cdot\mathrm{trans}-\mathrm{pol_{total}})}+1}, \quad (4)$$

In the same way, given S_{pol}, the concentration of positive-sense anti-genomic nucleocapsids available for binding to polymerases would be $((+)\mathrm{nc} - \mathrm{pol_{trc}}(S_{pol}/l_{trc}))$, where (+)nc is the total concentration of anti-genomic nucleocapsids, $\mathrm{pol_{trc}}$ is the concentration of the polymerases bound to the promoter region (Trc) of the anti-genomes, and l_{trc} is the length of the promoter region. Because the anti-genome has a stronger promoter than the genome (Rose and Whitt 2001; Le Mercier et al. 2002), which is quantified by S_{prom} in our model, S_{prom}-fold more polymerases bind to the promoter of the anti-genome than to that of the genome. Under the limitation of free polymerase complexes, the concentration of the polymerases newly binding to the promoters of the genomes or the anti-genomes ($\mathrm{pol_{term}}^{new}$) could be described as follows:

$$\mathrm{pol_{term}}^{new} = \min\left((-)\mathrm{nc}^{pol,new}\frac{l_l}{S_{pol}} + \left((+)\mathrm{nc}\frac{l_{trc}}{S_{pol}}S_{prom} - \mathrm{pol_{trc}}\right), L\cdot\mathrm{trans} - \mathrm{pol_{total}}\right), \quad (5)$$

where S_{prom} is the strength of the anti-genomic promoter relative to that of the genomic promoter, and min(A, B) equals the smaller of A and B. The distribution of newly bound polymerases between genomic and anti-genomic templates is determined by the concentrations of the available free templates of each type and the relative strengths of their promoters:

$$\mathrm{pol_l}^{new} = \frac{(-)\mathrm{nc}^{pol,new}\frac{l_l}{S_{pol}}}{(-)\mathrm{nc}^{pol,new}\frac{l_l}{S_{pol}} + \left[(+)\mathrm{nc}\frac{l_{trc}}{S_{pol}}S_{prom} - \mathrm{pol_{trc}}\right]}\mathrm{pol_{term}}^{new}, \quad (6.1)$$

$$\mathrm{pol_{trc}}^{new} = \mathrm{pol_{term}}^{new} - \mathrm{pol_l}^{new}, \quad (6.2)$$

where $\mathrm{pol_{trc}}^{new}$ is the concentration of the polymerases newly binding to the complementary trailer region (promoter) of the anti-genome. The polymerases newly binding or already bound to the promoters of the genomes and anti-genomes start viral RNA synthesis as a transcription or replication process.

Transcription

The viral polymerase in the leader region of the genome starts either transcription or replication. If there are sufficient N proteins, transcription is inhibited by the encapsidation of nascent positive-sense RNAs by N proteins, and then replication dominates transcription (Blumberg, Leppert, and Kolakofsky 1981; Banerjee and Barik 1992; Plumet, Duprex, and Gerlier 2005). In contrast, if there are insufficient free N proteins, then transcription dominates replication. In the model, we correlate the extent of transcription dominance with the availability of N proteins by introducing a factor, encap. This factor is defined as the ratio of the number of free N proteins to the number required to encapsidate all available nascent genome-sized viral RNAs. Only nocap (= 1 – encap) of the polymerases bound to the genomic promoters can start the transcription:

$$\frac{d\mathrm{pol_{N,1}}}{dt} = k_{e,pol}\left[(1-\phi_N)\,\mathrm{nocap}\,\frac{\mathrm{pol_l}}{l_l} - \frac{n_{sec,N}}{l_{mRNA,N}}\mathrm{pol_{N,1}}\right], \quad (7)$$

where $\mathrm{pol_{N,1}}$ is the concentration of the polymerases located at the first segment of the N gene, $k_{e,pol}$ is the elongation rate of polymerase, ϕ_N is the attenuation factor for N gene, $n_{sec,N}$ is the total number of segments of N gene, and $l_{mRNA,N}$ is the length of N mRNA (Table S2 (Supporting Information) (Thomas et al. 1985; Wagner 1987)). The genome segments are continuously charged

with incoming polymerases and discharged with outgoing polymerases with a rate of $k_{e,pol}$ (Equations 7–9). If the polymerase input to the leader region of the genome is decreased owing to a lack of free polymerases, then the polymerase concentrations downstream of the leader region will be subsequently reduced. There are conserved intergenic sequences involved in letting a fraction of viral polymerases release from the genome template at intergenic sections during transcription, which is so-called partial transcription termination or attenuation (Ball et al. 1999; Rose and Whitt 2001). Because the transcription is initiated from the 3′ end promoter, the attenuation mechanism causes genes more proximal to the 3′ end to be more highly expressed, which ultimately leads to an unequal concentration distribution of viral mRNAs. The extents of partial transcription termination are quantified by the attenuation factors, ϕ_i, in our model. These are 0/0.25/0.25/0.25/0.05 for leader–N/N–P/P–M/M–G/G–L intergenic regions, respectively (Iverson and Rose 1981; Ball et al. 1999; Rose and Whitt 2001; Barr, Whelan, and Wertz 2002). ϕ_i fraction of polymerases are released in intergenic region i. With Equations 7–9, we simulate the polymerase flux into each gene segment, which is proportional to the elongation rate of polymerase, but inversely proportional to the extent of attenuation:

$$\frac{\mathrm{dpol}_{i,j}}{\mathrm{dt}} = k_{e,pol}\left[(1-\phi_N)\,\frac{n_{sec,i-1}}{l_{mRNA,i-1}}\mathrm{pol}_{i-1,nsec,j-1} - \frac{n_{sec,i}}{l_{mRNA,i}}\mathrm{pol}_{i,j}\right] j=1,\quad i=\mathrm{P,M,G,\ L}, \tag{8}$$

$$\frac{\mathrm{dpol}_{i,j}}{\mathrm{dt}} = k_{e,pol}\frac{n_{sec,i}}{l_{mRNA,i}}\left(\mathrm{pol}_{i,j-1}-\mathrm{pol}_{i,j}\right) j\neq 1,\ i=\mathrm{N,P,M,G,L}, \tag{9}$$

where $\mathrm{pol}_{i,j}$ is the concentration of the polymerases located at the jth segment of gene i and $i-1$ indicates the prior gene of gene i. The amount of newly synthesized mRNAs for each gene is determined by the concentration of polymerases occupying each gene section on the genome template and the decay rates of the mRNAs:

$$\frac{\mathrm{dmRNA}_i}{\mathrm{dt}} = \frac{k_{e,pol}}{l_{mRNA,i}}\mathrm{pol}_{t,i} - k_{d,mRNA}\mathrm{mRNA}_i, \tag{10}$$

where mRNA_i is the concentration of mRNAs for gene i, $k_{d,mRNA}$ is the decay rate constant of mRNA that is the same for all five viral mRNAs (Pennica et al. 1979), and $\mathrm{pol}_{t,i}$ is the total concentration of the polymerases occupying on the ith gene.

Our formulation for transcription assumes that the synthesis of viral mRNAs is rate-controlled by the transcription initiation as well as the elongation of polymerase. Transcription initiation rate is parameterized by the spacing between neighboring polymerases in our model. At a given polymerase elongation rate, the larger polymerase spacing indicates the lower rate of transcription initiation. Transcription initiation modulates the input of polymerases to the leader region of the genome.

Translation

We consider that both translation initiation and polypeptide chain elongation contribute to the rate of viral protein synthesis. The translation initiation rate is parameterized by the ribosomal spacing. In our model, we first calculated the number of ribosomes involved in viral translation by considering the maximum concentration of the ribosomes bound to viral mRNAs at a fixed ribosomal spacing:

$$\mathrm{rib} = \min\left(n_{rib,avail}, \frac{\sum_{j=N}^{L} l_{mRNA,j}\mathrm{mRNA}_j}{S_{rib}}\right), \tag{11}$$

where rib and $n_{rib,avail}$ are the concentrations of the ribosomes actually involved in viral translation and the ribosomes available for viral translation, respectively, and S_{rib} is the spacing between neighboring ribosomes.

The ribosomes involved in viral translation (rib) are allocated to the five types of viral mRNAs according to their length and abundance, assuming that each viral mRNA has the same efficiency of translation initiation (Lodish and Froshauer 1977):

$$\mathrm{rib}_i = \frac{l_{mRNA,i}\mathrm{mRNA}_i}{\sum_{j=N}^{L} l_{mRNA,j}\mathrm{mRNA}_j}(\mathrm{rib}), \tag{12}$$

where rib_i is the concentration of the ribosomes assigned to mRNA_i.

The synthesis rate of each viral protein depends on the elongation rate of the ribosome, linear density of ribosomes on its corresponding mRNA, and its first-order decay rate:

$$\frac{dp_i}{dt} = \frac{k_{e,rib}}{l_{p,i}}\mathrm{rib}_i - k_{dp,i}p_i \quad i=\mathrm{P,M,\ G,L}, \tag{13}$$

where p_i is the concentration of protein i, $k_{e,rib}$ is the elongation rate of the ribosome, $l_{p,i}$ is the length of protein i, and $k_{dp,i}$ is the decay rate constant of protein i.

We also accounted for the consumption of free N proteins during the encapsidation of genome-length nascent RNAs and assumed that the degradation of nucleocapsids yielded intact N proteins:

$$\frac{dp_i}{dt} = \frac{k_{e,rib}}{l_{p,i}}\mathrm{rib}_i - k_{dp,i}p_i - n_i\left(\mathrm{encap}\cdot k_{e,pol}\left(\frac{\mathrm{pol}_{tr}}{l_{tr}}+\frac{\mathrm{pol}_{lec}}{l_{lec}}\right) - k_{d,nc}\left((+)\mathrm{nc}+(-)\mathrm{nc}\right)\right), \quad i=\mathrm{N}, \tag{14}$$

where n_N is the stoichiometry of N protein in a single nucleocapsid or virion progeny (Table S2 (Supporting Information) (Thomas et al. 1985; Wagner 1987)), pol_{tr} and pol_{lec} are the concentrations of the polymerases located in the trailer and complementary leader regions of the genomes and the anti-genomes, respectively, l_{lec} ($=l_l$) is the length of the complementary leader region, and $k_{d,nc}$ is the decay rate constant of nucleocapsid. As progeny virions are assembled, the concentration of each protein is reduced by the amount corresponding to its stoichiometry in a single virion particle.

Replication

We assumed that N protein regulates the switch of the role of polymerase between transcription and replication by encapsidating the newly synthesized RNAs (Banerjee and Barik 1992; Plumet, Duprex, and Gerlier 2005). The polymerase that starts the replication in the leader region of the genome requires further supply of N proteins to skip the attenuation signals at each gene junction and thereby to complete each round of replication. Depending on the availability of N proteins, nocap ($=1-$encap) fraction of polymerases terminates the replication at each gene junction in our model:

$$\frac{\mathrm{dpol}_{r,n,N,1}}{\mathrm{dt}} = k_{e,pol}\left(\mathrm{encap}\,\frac{\mathrm{pol}_l}{l_l} - \frac{n_{sec,N}}{l_{mRNA,N}}\mathrm{pol}_{r,n,N,1}\right), \tag{15}$$

where $\mathrm{pol}_{r,n,N,1}$ is the concentration of the replicating polymerases on the first segment of the N gene section in the negative-sense

genomic nucleocapsid,

$$\frac{\mathrm{dpol}_{\mathrm{r},n,\mathrm{i},1}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\left(\mathrm{encap}\frac{n_{\mathrm{sec,i-1}}}{l_{\mathrm{mRNA,i-1}}}\mathrm{pol}_{\mathrm{r},n,\mathrm{i-1},n\mathrm{sec,i-1}} - \frac{n_{\mathrm{sec,i}}}{l_{\mathrm{mRNA,i}}}\mathrm{pol}_{\mathrm{r},n,\mathrm{i},1}\right), \quad \mathrm{i} = \mathrm{P,M,G,L}, \tag{16}$$

where $\mathrm{pol}_{\mathrm{r},n,\mathrm{i},1}$ and $\mathrm{pol}_{\mathrm{r},n,\mathrm{i-1},n\mathrm{sec},\mathrm{i-1}}$ are the concentrations of the replicating polymerases on the first segment of gene i and on the last segment of gene i – 1, respectively.

The level of polymerases that scan through the whole genome ($\mathrm{pol}_{\mathrm{tr}}$) determines the amount of newly synthesized anti-genomic nucleocapsids, (+)nc:

$$\frac{\mathrm{dpol}_{\mathrm{tr}}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\left(\frac{n_{\mathrm{sec,i}}}{l_{\mathrm{mRNA,i}}}\mathrm{pol}_{\mathrm{r},n,\mathrm{i},j} - \frac{\mathrm{pol}_{\mathrm{tr}}}{l_{\mathrm{tr}}}\right), \mathrm{i} = \mathrm{L},\ j = n_{\mathrm{sec}},\mathrm{L}, \tag{17}$$

$$\frac{\mathrm{d}(+)\mathrm{nc}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\mathrm{encap}\left(\frac{\mathrm{pol}_{\mathrm{tr}}}{l_{\mathrm{tr}}}\right) - k_{\mathrm{d,nc}}(+)\mathrm{nc}, \tag{18}$$

where l_{tr} ($=l_{\mathrm{trc}}$) is the length of the trailer region of the genome. We also considered the first-order kinetics for the decay of anti-genomic nucleocapsid.

The synthesis and decay of genomic nucleocapsids are described in the same way as for those of the anti-genomic nucleocapsids except that the polymerases on the anti-genomic templates are not released in intergenic regions:

$$\frac{\mathrm{dpol}_{\mathrm{r,p},j}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\left(\mathrm{encap}\frac{\mathrm{pol}_{\mathrm{trc}}}{l_{\mathrm{trc}}} - \frac{n_{\mathrm{sec}}}{l - l_{\mathrm{trc}} - l_{\mathrm{lec}}}\mathrm{pol}_{\mathrm{r,p},j}\right), j = 1, \tag{19}$$

where $\mathrm{pol}_{\mathrm{r,p},j}$ is the concentration of the replicating polymerases on the jth segment of the positive-sense anti-genomic nucleocapsids, l is the total length of the genome, and n_{sec} is the total number of segments of the genome.

$$\frac{\mathrm{dpol}_{\mathrm{r,p},j}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\left(\frac{n_{\mathrm{sec}}}{l - l_{\mathrm{trc}} - l_{\mathrm{lec}}}\right)\left(\mathrm{pol}_{\mathrm{r,p},j-1} - \mathrm{pol}_{\mathrm{r,p},j}\right), j = 2,\ldots n_{\mathrm{sec}}, \tag{20}$$

$$\frac{\mathrm{dpol}_{\mathrm{lec}}}{\mathrm{dt}} = k_{\mathrm{e,pol}}\left(\frac{n_{\mathrm{sec}}}{l - l_{\mathrm{trc}} - l_{\mathrm{lec}}}\mathrm{pol}_{\mathrm{r,p},n_{\mathrm{sec}}} - \frac{\mathrm{pol}_{\mathrm{lec}}}{l_{\mathrm{lec}}}\right), \tag{21}$$

$$\frac{\mathrm{d}(-)\mathrm{nc}}{\mathrm{dt}} = k_{\mathrm{e,pol}}(\mathrm{encap})\left(\frac{\mathrm{pol}_{\mathrm{lec}}}{l_{\mathrm{lec}}}\right) - k_{\mathrm{d,nc}}(-)\mathrm{nc}. \tag{22}$$

In our model, non-encapsidated nascent genome and anti-genome fragments are released from polymerases and immediately degraded.

As polymerases leave the promoter regions by moving toward the downstream sequences, the concentration of polymerases on the promoters will decrease. The dynamic changes in the polymerase concentrations on the promoters of the genomic and the anti-genomic templates are finally described, respectively:

$$\mathrm{pol}_{\mathrm{l},n+1} = \mathrm{pol}_{\mathrm{l},n} + {\mathrm{pol}_{\mathrm{l},n}}^{\mathrm{new}} - \mathrm{pol}_{\mathrm{l-leave},n}, \tag{23}$$

where $\mathrm{pol}_{\mathrm{l-leave}}$ is the concentration of the polymerases leaving the genomic promoters, O_n and O_{n+1} are the concentrations of a component (O) at time n and time $n+1$ (in our numerical integration, time $n+1$ – time $n = \Delta t$), respectively.

$$\mathrm{pol}_{\mathrm{trc},n+1} = \mathrm{pol}_{\mathrm{trc},n} + {\mathrm{pol}_{\mathrm{trc},n}}^{\mathrm{new}} - \mathrm{pol}_{\mathrm{trc-leave},n}, \tag{24}$$

where $\mathrm{pol}_{\mathrm{trc-leave}}$ is the concentration of the polymerases leaving the anti-genomic promoters.

Assembly and budding

We assume that the condensation of negative-sense nucleocapsid by M protein initiates the virion assembly and the condensed nucleocapsids are not degraded in the same manner as virion progeny. Whenever the requirement for the stoichiometric amounts of proteins is satisfied, progeny virions are instantaneously assembled and released to the extracellular space. The time required for the condensation of the negative-sense nucleocapsid, the assembly, and the budding of progeny virion was assumed to be negligible relative to the preceding steps.

Host cell

In our model, the host cell provides unlimited building blocks such as nucleoside triphosphates and amino acids for the growth of virus. However, as viral components accumulate during the course of infection, some key host components for translation such as initiation and elongation factors may be depleted (Centrella and Lucas-Lenard 1982; Wagner 1987). Two main viral products, leader mRNA and M protein, contribute to the deficiency by inhibiting the synthesis of host macromolecules at the transcription level (McGowan, Emerson, and Wagner 1982; Wagner 1987; Lyles et al. 1996; Rose and Whitt 2001). Because leader mRNA starts to accumulate soon after the initiation of infection, and a small amount of the component is enough to trigger the inhibition (McGowan, Emerson, and Wagner 1982; Wagner 1987), the pool of host factors is continuously reduced from the onset of infection. We quantify this reduction with a single decay rate constant specific to the type of host cell:

$$f_{\mathrm{dec}} = \exp\left(-k_{\mathrm{d,host}}t\right), \tag{25}$$

where f_{dec} is the level of host translation factors at time t, relative to that of the initial state of the cell before infection (at $t = 0$), and $k_{\mathrm{d,host}}$ is the decay rate constant.

The inhibition by the leader mRNA causes a first-order decay of the host factors, resulting in a shortage of the ribosomes equipped with the accessory factors for viral translation in the late infection stage in our model. Unlike viral transcription and replication, viral translation is directly affected by the decay of host factors since it depends entirely on host machinery. In the early infection, host mRNAs outnumber viral mRNAs and thereby successfully compete for the host translation machinery. However, the newly synthesized M proteins inhibit the host transcription initiation and the export of host mRNAs from the nucleus to the cytoplasm (Lyles 2000), thereby causing a gradual shift in translation from host mRNAs to viral mRNAs. For our model, we assumed that the potency of the inhibition by the M protein was independent of the type of cell and its differentiation state (Lyles et al. 1996), and we developed an empirical formula using available experimental data from the literature (Lyles et al. 1996) to account for the competition between host and viral mRNAs for ribosomes. Lyles et al. co-transfected the host cells with VSV M mRNA and chloramphenicol acetyl transferase (CAT) plasmid DNA, and then they quantified the expression of CAT based on its activity, as a function of the expression of VSV M protein (Lyles et al. 1996). In their experiment, the gene expression of CAT was more reduced at higher M protein expression levels. We assume that the decrease in the expression of CAT (or its activity decrease) is proportional to the decrease in the occupancy of host mRNAs by the translation machinery. Using their experimental data, the occupancy of host mRNAs by the translation machinery is correlated with the number of newly synthesized M proteins in the cytoplasm:

$$\mathrm{rib}_{\mathrm{host}} = 1.88 \times 10^3 \left(M_{\mathrm{cell}}\left(1 - \mathrm{cond}_{\mathrm{M}}\right)\right)^{-0.9582}, \tag{26}$$

where rib_host is the fraction of the translation machinery associated with host mRNAs and M_{cell} is the total number of newly synthesized M proteins per cell.

Considering the decay of host factors and the competition between host and viral mRNAs, we could derive a formula to quantify the number of fully functional ribosomes that are available for the viral protein synthesis over time post-infection ($nrib_{avail}$):

$$nrib_{avail} = nrib\,(1 - rib\_host) f_{dec}, \tag{27}$$

where nrib denotes the total concentration of ribosomes whether they incorporate all the required accessory factors for their translation function.

Although the ribosomes distribute into membrane-bound and cytoplasmic forms, each class supporting the syntheses of the viral G protein and the other four viral proteins (N, P, M, and L), respectively, we treated the ribosomes in our model as one population.

Initial condition for simulation

The initial condition for our simulation is set by a fixed number of infectious extracellular virus particles per cell ($V_{ex}(0)$). At time zero ($t = 0$), the number of bound virus particles and the level of all viral components within cells are zero. In our model, binding of extracellular virus particles to cells reduces their level (Equation 1), and an encapsidated genome and stoichiometric amounts of viral proteins (Table S2 (Supporting Information) (Thomas et al. 1985; Wagner 1987)) are then immediately released from each bound virus particle to the cytoplasm. Specifically, we assume that all N proteins from a bound virus particle are released as a form of encapsidated genome complex. Downstream processes, beginning with transcription, are then initiated. In our simulation, viral infection starts with rib_host = 0.99 (with rib_host = 0.9, ∼0.9995 simulations showed the same results). Other key model parameters for simulation are summarized in Table S1 (Supporting Information). In addition, a nomenclature list is shown in Table S3 (Supporting Information).

Computation of fitness dependence on mutations

In the base-case or wild-type infection setting, we simulated the one-step growth of VSV on BHK cells as previously done (Lim et al. 2006). Eighteen parameters of the model associated with molecular processes and mechanisms were candidates for perturbation to simulate virus variants arising from mutation. These include five attenuation factors that characterize the fractional release of the viral RNA-dependent RNA polymerases from the negative-sense single-stranded RNA genome as it transcribes viral mRNA, creating a gradient in time and level of viral gene expression; five degradation constants associated with the translated viral proteins of the five encode genes (nucleocapsid N, polymerase P, matrix M, glycoprotein G, and large L); two degradation rates for the mRNA (one constant for all genes) and the nucleocapsids; an elongation rate for the viral polymerase; and a kinetic constant for the binding and entry of the initial infecting virus particle. Additional parameters were the spacing between viral polymerases, the fraction of M proteins bound to the plasma membrane, the ratio of rate constants for the associations of M proteins and polymerases with genomic nucleocapsids, and the relative strength of the anti-genomic promoter to the genomic promoter. The model is a system of differential and algebraic equations that uses a fourth-order Runge–Kutta method with a timestep of 15 s to simulate a 25-h infection. The model further accounts for consumption of essential host resources over the course of infection.

We selected eleven of the eighteen virus-associated parameters of the VSV model, considering how their perturbations could be plausibly linked to a deleterious mutation; for example, increasing an individual intergenic attenuation factor or protein degradation rate had a negative impact on both the virus yields and production rates from the simulated one-step growth. To create different classes of mutation, each parameter range was linearly divided into five subranges. For example, $\eta_{n,wt} = 1$, so the parameter range was (0, 1). A Class 1 severity, the lightest, would be (0.8, 1). A class 5 severity, where the largest reduction in fitness was seen, would be (0, 0.2). Initially, we followed the standard set by You et al., where all five classes were included in the simulations and subsequent analysis. But initial testing gave more severe reductions in yield and rate than those seen by You et al., such that a mutant with several Class 3 mutations would likely have low to zero fitness; to avoid such potentially lethal mutations, we narrowed the parameter ranges to class 1 and 2 mutations, dividing these two subranges into five new ones. For example, $\eta_{n,wt} = 1$ would have the five ranges between 1.0 and 0.60, with each range spanning 0.08.

A simulation of a VSV particle carrying a genome that had n mutations in the specified k class of severity was built by randomly choosing n parameters from the available set and then multiplying the wild-type parameter by a randomly chosen constant from the k severity level. The constant was selected from a uniform distribution that spanned the k severity subrange. For each unique pair (n, k), where $1 \leq k \leq 5$ and $1 \leq n \leq 11$, 500 simulations were run using a distributed network of computers. The random number generation was seeded by identifiers unique to every set of trials, ensuring that sets of parameters and constants were not repeated. Each simulated mutant growth was used to estimate mutant yields and rates, which were normalized to their wild-type values to give w_{growth} and w_{spread}, and Equation 1 was fit to the logarithm of each fitness correlate by minimizing the sum-squared-error to determine the α and β parameters for each fitness type at each severity level. Each mutation was tested prior to the simulations to ensure that $w_i \leq w_{i,wt}$, so when applying the power law from Equation 1, $\log(w_i) \leq 0$. The model and subsequent analysis was performed in MATLAB using the computed resources and assistance of the University of Wisconsin-Madison Center for High Throughput Computing in the Department of Computer Sciences. All codes written in support of this work are publicly available at https://github.com/bfschwab/Epistasis-VSV.git

Supplementary data

Supplementary data is available at *VEVOLU* Journal online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding

This work was supported by the US National Science Foundation (DMS-2151959) and the National Institutes of Health (OT2OD030524).

Acknowledgements

We thank Hayley Boigenzahn for test running the model and for thoughtful suggestions.

Conflict of interest: The authors declare that they have no competing interests.

Author contributions

B.S. and J.Y. conceived study and designed methods. B.S. performed simulations and analysis. J.Y. and B.S. wrote the paper.

References

Akpinar, F., Inankur, B., and Yin, J. (2016) 'Spatial-Temporal Patterns of Viral Amplification and Interference Initiated by a Single Infected Cell', *Journal of Virology*, 90: 7552–66.

Akpinar, F., Timm, A., and Yin, J. (2015) 'High-Throughput Single-Cell Kinetics of Virus Infections in the Presence of Defective Interfering Particles', *Journal of Virology*, 90: 1599–612.

Ball, L. A. et al. (1999) 'Phenotypic Consequences of Rearranging the P, M, and G Genes of Vesicular Stomatitis Virus', *Journal of Virology*, 73: 4705–12.

Banerjee, A. K., and Barik, S. (1992) 'Gene Expression of Vesicular Stomatitis Virus Genome RNA', *Virology*, 188: 417–28.

Barr, J. N., Whelan, S. P. J., and Wertz, G. W. (2002) 'Transcriptional Control of the RNA-Dependent RNA Polymerase of Vesicular Stomatitis Virus', *Biochimica Et Biophysica Acta (BBA) – Gene Structure and Expression*, 1577: 337–53.

Birch, E. W., Ruggero, N. A., and Covert, M. W. (2012) 'Determining Host Metabolic Limitations on Viral Replication via Integrated Modeling and Experimental Perturbation', *PLoS Computational Biology*, 8: e1002746.

Blumberg, B. M., Leppert, M., and Kolakofsky, D. (1981) 'Interaction of VSV Leader RNA and Nucleocapsid Protein May Control VSV Genome Replication', *Cell*, 23: 837–45.

Bonhoeffer, S. et al. (2004) 'Evidence for Positive Epistasis in HIV-1', *Science*, 306: 1547–50.

Burch, C. L., and Chao, L. (1999) 'Evolution by Small Steps and Rugged Landscapes in the RNA Virus Φ6', *Genetics*, 151: 921–7.

Burton, T. D., and Eyre, N. S. (2021) 'Applications of Deep Mutational Scanning in Virology', *Viruses*, 13: 1020.

Centrella, M., and Lucas-Lenard, J. (1982) 'Regulation of Protein Synthesis in Vesicular Stomatitis Virus-Infected Mouse L-929 Cells by Decreased Protein Synthesis Initiation Factor 2 Activity', *Journal of Virology*, 41: 781–91.

Chong, L. D., and Rose, J. K. (1993) 'Membrane Association of Functional Vesicular Stomatitis Virus Matrix Protein in Vivo', *Journal of Virology*, 67: 407–14.

Cordell, H. J. (2002) 'Epistasis: What It Means, What It Doesn't Mean, and Statistical Methods to Detect It in Humans', *Human Molecular Genetics*, 11: 2463–8.

DePristo, M. A., Weinreich, D. M., and Hartl, D. L. (2005) 'Missense Meanderings in Sequence Space: A Biophysical View of Protein Evolution', *Nature Reviews Genetics*, 6: 678–87.

Elena, S. F., Solé, R. V., and Sardanyés, J. (2010) 'Simple Genomes, Complex Interactions: Epistasis in RNA Virus', *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20: 319–30.

Fields, B. N., Knipe, D. M., and Howley, P. M. (2013) *Fields Virology*, 6th edn. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Florkiewicz, R. Z., and Rose, J. K. (1984) 'A Cell Line Expressing Vesicular Stomatitis Virus Glycoprotein Fuses at Low pH', *Science*, 225: 721–3.

Fowler, D. M., and Fields, S. (2014) 'Deep Mutational Scanning: A New Style of Protein Science', *Nature Methods*, 11: 801–7.

Gao, Y., and Lenard, J. (1995) 'Cooperative Binding of Multimeric Phosphoprotein (P) of Vesicular Stomatitis Virus to Polymerase (L) and Template: Pathways of Assembly', *Journal of Virology*, 69: 7718–23.

Goudsmit, J. et al. (1997) 'Broad Spectrum of in Vivo Fitness of Human Immunodeficiency Virus Type 1 Subpopulations Differing at Reverse Transcriptase Codons 41 and 215', *Journal of Virology*, 71: 4479–84.

Haddox, H. K., Dingens, A. S., and Bloom, J. D. (2016) 'Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture', *PLoS pathogens* 12

Harvey, W. T. et al. (2021) 'SARS-CoV-2 Variants, Spike Mutations and Immune Escape', *Nature Reviews, Microbiology*, 19: 409–24.

Holland, J. J. et al. (1982) 'Rapid Evolution of RNA Genomes', *Science*, 215: 1577–85.

Iverson, L. E., and Rose, J. K. (1981) 'Localized Attenuation and Discontinuous Synthesis during Vesicular Stomatitis Virus Transcription', *Cell*, 23: 477–84.

Jin, T., and Yin, J. (2021) 'Patterns of Virus Growth across the Diversity of Life', *Integrative Biology*, 13: 44–59.

Kim, H., and Yin, J. (2004) 'Energy-Efficient Growth of Phage Q Beta in Escherichia coli', *Biotechnology and Bioengineering*, 88: 148–56.

Lam, V., Duca, K. A., and Yin, J. (2005) 'Arrested Spread of Vesicular Stomatitis Virus Infections in Vitro Depends on Interferon-Mediated Antiviral Activity', *Biotechnology and Bioengineering*, 90: 793–804.

Lee, Y., and Yin, J. (1996) 'Detection of Evolving Viruses', *Nature Biotechnology*, 14: 491–3.

Le Mercier, P. et al. (2002) 'Ambisense Sendai Viruses Are Inherently Unstable but Are Useful to Study Viral RNA Synthesis', *Journal of Virology*, 76: 5492–502.

Lenski, R. E. et al. (1999) 'Genome Complexity, Robustness and Genetic Interactions in Digital Organisms', *Nature*, 400: 661–4.

Lim, K. I. et al. (2006) 'Model-Based Design of Growth-Attenuated Viruses', *PLoS Computational Biology*, 2: e116.

Lodish, H. F., and Froshauer, S. (1977) 'Rates of Initiation of Protein Synthesis by Two Purified Species of Vesicular Stomatitis Virus Messenger RNA', *Journal of Biological Chemistry*, 252: 8804–11.

Lyles, D. S. (2000) 'Cytopathogenesis and Inhibition of Host Gene Expression by RNA Viruses', *Microbiology and Molecular Biology Reviews*, 64: 709–24.

Lyles, D. S. et al. (1996) 'Potency of Wild-Type and Temperature-Sensitive Vesicular Stomatitis Virus Matrix Protein in the Inhibition of Host-Directed Gene Expression', *Virology*, 225: 172–80.

Lyles, D. S., Kuzmin, I. V., and Rupprecht, C. E. (2013) 'Rhabdoviridae', In: D. M. Knipe and P. M. Howley (eds) *Fields Virology*, 6th edn. pp. 885–922. Philadelphia, PA: Lippincott Williams & Wilkins.

Mackay, T. F. C. (2013) 'Epistasis and Quantitative Traits: Using Model Organisms to Study Gene–Gene Interactions', *Nature Reviews Genetics*, 15: 22–33. Nature Publishing Group.

Matlin, K. S. et al. (1982) 'Pathway of Vesicular Stomatitis Virus Entry Leading to Infection', *Journal of Molecular Biology*, 156: 609–31.

McGowan, J. J., Emerson, S. U., and Wagner, R. R. (1982) 'The Plus-Strand Leader RNA of VSV Inhibits DNA-Dependent Transcription of Adenovirus and SV40 Genes in a Soluble Whole-Cell Extract', *Cell*, 28: 325–33.

Mellon, M. G., and Emerson, S. U. (1978) 'Rebinding of Transcriptase Components (L and NS Proteins) to the Nucleocapsid Template of Vesicular Stomatitis Virus', *Journal of Virology*, 27: 560–7.

Miller, D. K., and Lenard, J. (1980) 'Inhibition of Vesicular Stomatitis Virus Infection by Spike Glycoprotein. Evidence for an Intracellular, G Protein-Requiring Step', *The Journal of Cell Biology*, 84: 430–7.

Moore, J. H., and Williams, S. M. (2009) 'Epistasis and Its Implications for Personal Genetics', *The American Journal of Human Genetics*, 85: 309–20.

Pennica, D. et al. (1979) 'Decay of Vesicular Stomatitis Virus mRNAs in Vivo', *Virology*, 94: 484–7.

Phillips, P. C. (2008) 'Epistasis—the Essential Role of Gene Interactions in the Structure and Evolution of Genetic Systems', *Nature Reviews Genetics*, 9: 855–67. Springer Science and Business Media LLC.

Plumet, S., Duprex, W. P., and Gerlier, D. (2005) 'Dynamics of Viral RNA Synthesis during Measles Virus Infection', *Journal of Virology*, 79: 6900–8.

Rose, J. K., and Whitt, M. A. (2001) 'Rhabdoviridae: The Viruses and Their Replication', in Fields, B. N., Knipe, D. M. and Howley, P. M. (eds) *Fields Virology*, Vol. 1, 4th edn. pp. 1221–44. Philadelphia, PA, USA: Lippincott Williams & Wilkins Publishers.

Sanjuán, R., Moya, A., and Elena, S. F. (2004) 'The Contribution of Epistasis to the Architecture of Fitness in an RNA Virus', *Proceedings of the National Academy of Sciences*, 101: 15376–9.

Schlegel, R. et al. (1983) 'Inhibition of VSV Binding and Infectivity by Phosphatidylserine: Is Phosphatidylserine a VSV-Binding Site?', *Cell*, 32: 639–46.

Segrè, D. et al. (2005) 'Modular Epistasis in Yeast Metabolism', *Nature Genetics*, 37: 77–83.

Simonsen, C. C., Batt-Humphries, S., and Summers, D. F. (1979) 'RNA Synthesis of Vesicular Stomatitis Virus-Infected Cells: In Vivo Regulation of Replication', *Journal of Virology*, 31: 124–32.

Soh, Y. Q. S. et al. (2019) 'Comprehensive Mapping of Adaptation of the Avian Influenza Polymerase Protein PB2 to Humans', *Elife*, 8: e45079.

Starr, T. N. et al. (2020) 'Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding', *Cell*, 182: 1295–1310.e20.

Superti, F. et al. (1987) 'Entry Pathway of Vesicular Stomatitis Virus into Different Host Cells', *Journal of General Virology*, 68: 387–99.

Thomas, D. et al. (1985) 'Mass and Molecular Composition of Vesicular Stomatitis Virus: A Scanning Transmission Electron Microscopy Analysis', *Journal of Virology*, 54: 598–607.

Turner, P. E., Draghi, J. A., and Wilpiszeski, R. (2012) 'High-Throughput Analysis of Growth Differences among Phage Strains', *Journal of Microbiological Methods*, 88: 117–21.

Wagner, R. R. (1987) *The Rhabdoviruses*. New York, NY: Plenum Press.

Yen, H.-L. et al. (2005) 'Neuraminidase Inhibitor-Resistant Influenza Viruses May Differ Substantially in Fitness and Transmissibility', *Antimicrobial Agents and Chemotherapy*, 49: 4075–84.

Yin, J. (1991) 'A Quantifiable Phenotype of Viral Propagation', *Biochemical and Biophysical Research Communications*, 174: 1009–14.

Yin, J., and Redovich, J. (2018) 'Kinetic Modeling of Virus Growth in Cells', *Microbiology and Molecular Biology Reviews*, 82: e00066–17.

You, L., Suthers, P. F., and Yin, J. (2002) 'Effects of Escherichia coli Physiology on the Growth of Phage T7 in Vivo and in Silico', *Journal of Bacteriology*, 184: 1888–94.

You, L., and Yin, J. (2002) 'Dependence of Epistasis on Environment and Mutation Severity as Revealed by *in Silico* Mutagenesis of Phage T7', *Genetics*, 160: 1273–81.

Zhou, K. et al. (2022) 'Atomic Model of Vesicular Stomatitis Virus and Mechanism of Assembly', *Nature Communications*, 13: 5980.

Virus Evolution, 2024, **10(1)**, 1–13
DOI: https://doi.org/10.1093/ve/vead082
Advance Access Publication 28 December 2023
Research Article