

Quantifying the dynamics of viral recombination during free virus and cell-to-cell transmission in HIV-1 infection

Jesse Kreger,^{1,*} Josephine Garcia,² Hongtao Zhang,² Natalia L. Komarova,¹ Dominik Wodarz,^{1,3} and David N. Levy²

¹Department of Mathematics, Rowland Hall, University of California, Irvine, CA 92697, USA, ²Department of Basic Science, New York University College of Dentistry, 921 Schwartz Building, 345 East 24th Street, New York, NY 10010–9403, USA and ³Department of Population Health and Disease Prevention, Program in Public Health, Susan and Henry Samueli College of Health Sciences, University of California, Irvine, CA 92697, USA

*Corresponding author: E-mail: kregerj@uci.edu

†<https://orcid.org/0000-0001-6438-171X>

Abstract

Recombination has been shown to contribute to human immunodeficiency virus-1 (HIV-1) evolution *in vivo*, but the underlying dynamics are extremely complex, depending on the nature of the fitness landscapes and of epistatic interactions. A less well-studied determinant of recombinant evolution is the mode of virus transmission in the cell population. HIV-1 can spread by free virus transmission, resulting largely in singly infected cells, and also by direct cell-to-cell transmission, resulting in the simultaneous infection of cells with multiple viruses. We investigate the contribution of these two transmission pathways to recombinant evolution, by applying mathematical models to *in vitro* experimental data on the growth of fluorescent reporter viruses under static conditions (where both transmission pathways operate), and under gentle shaking conditions, where cell-to-cell transmission is largely inhibited. The parameterized mathematical models are then used to extrapolate the viral evolutionary dynamics beyond the experimental settings. Assuming a fixed basic reproductive ratio of the virus (independent of transmission pathway), we find that recombinant evolution is fastest if virus spread is driven only by cell-to-cell transmission and slows down if both transmission pathways operate. Recombinant evolution is slowest if all virus spread occurs through free virus transmission. This is due to cell-to-cell transmission 1, increasing infection multiplicity; 2, promoting the co-transmission of different virus strains from cell to cell; and 3, increasing the rate at which point mutations are generated as a result of more reverse transcription events. This study further resulted in the estimation of various parameters that characterize these evolutionary processes. For example, we estimate that during cell-to-cell transmission, an average of three viruses successfully integrated into the target cell, which can significantly raise the infection multiplicity compared to free virus transmission. In general, our study points towards the importance of infection multiplicity and cell-to-cell transmission for HIV evolution.

Key words: recombination; cell-to-cell transmission; multiplicity of infection; mathematical models; evolution.

1. Introduction

Human immunodeficiency virus-1 (HIV-1) infection eventually results in the development of AIDS, typically after several years. Viral evolution is thought to be a major contributor to disease progression, and poses important challenges to antiviral treatments and to the development of protective vaccines (Bonhoeffer et al. 1997; Hirsch 1999; Kimata et al. 1999). HIV-1 is characterized by a relatively high mutation rate (3×10^{-5} per base pair per generation) (Mansky and Temin 1995), which together with the fast turnover of the virus population (Wei et al. 1995; Ho et al. 1995; Perelson et al. 1996) contributes to the large evolutionary potential of the virus. Besides mutations, however, HIV-1 can also undergo recombination because the virions are diploid (Moutouh, Corbeil, and Richman 1996; Levy et al. 2004). If a cell is infected with two different virus strains and two different genomes are packaged into the offspring virus, recombination can occur between these two strains when the virus infects a new cell and undergoes reverse transcription. Recombination can accelerate the generation of two-hit mutant viruses (virus strains with point mutations at two relevant locations) from two different one-hit mutants (virus strains with a single relevant point mutation), which is likely faster than the generation of two-hit mutants by point mutations alone. These processes can be especially important for the evolution of viral variants that simultaneously escape multiple immune cell clones or drugs.

While recombination has been shown to significantly contribute to viral evolution *in vivo* (Moutouh, Corbeil, and Richman 1996), mathematical models have demonstrated that the dynamics of recombination can be extremely complex (Bretscher et al. 2004; Althaus and Bonhoeffer 2005; Fraser 2005; Carvajal-Rodríguez et al. 2006; Kouyos, Fouchet, and Bonhoeffer 2009; Kreger, Komarova, and Wodarz 2020). Recombination can not only help the generation of multi-hit mutants but also break existing mutant combinations apart. The net effect of recombination is difficult to predict and depends on underlying assumptions about fitness landscapes, the nature and magnitude of epistatic interactions, and the relative balance of free virus and direct cell-to-cell transmission. Efforts have been made to quantify some of those processes, such as the fitness landscapes and the nature of epistasis in the evolution of drug resistance (Bonhoeffer et al. 2004).

Here, we seek to quantify in detail how the different virus transmission pathways impact the evolution of recombinants. The nature of virus transmission during viral spread through the cell population is likely crucial for the rate of recombinant evolution, through variations in infection multiplicity (Hubner et al. 2009; Del Portillo et al. 2011; Law et al. 2016). Free virus transmission typically results in the infection of cells containing one virus, while direct cell-to-cell transmission through virological synapses typically results in the multiple infection of cells, and additionally is likely to result in the co-transmission of different virus strains from one cell to the next (Del Portillo et al. 2011; Law et al. 2016). Therefore, cell-to-cell (or synaptic) transmission is likely to be beneficial for the rate at which viral recombinants emerge.

Here, we combine mathematical models with *in vitro* experiments that utilize fluorescent reporter viruses (Levy et al. 2004) to test this hypothesis, to quantify the dynamical processes that lead to recombinant generation, and to estimate underlying parameters. Viruses can carry genes for cyan fluorescent protein (eCFP) and yellow fluorescent protein (eYFP). Recombination between these two variants results in viruses

characterized by green fluorescence in target cells (Levy et al. 2004). Additionally, recombination of the green fluorescent virus with a non-glowing virus can break the recombinant apart, resulting in yellow and cyan fluorescent viruses. The viral transmission pathway can be modulated by placing the culture on a gentle rocking platform (Sourisseau et al. 2007; Komarova et al. 2013), which disrupts synaptic transmission. The frequency of infected cells displaying cyan, yellow and/or green is measured by flow cytometry over several days after infection. Mathematical models are applied to these experimental data to quantitatively characterize the dynamics.

2. Materials and methods

2.1 Experimental setup

2×10^6 CEM-SS cells were infected with 25 μ l NLENY1-IRES (YFP) or NLENC1-IRES (CFP), representing nineteen virions/cell, by spinoculation at 1,200 g for 2 hours at 37°C as described previously (Trinite et al. 2014). Cells were cultured in Gibco Advanced RPMI-1640 with 5 per cent FBS plus penicillin and streptomycin and 50 μ M β -mercaptoethanol. The next day these cells were washed and YFP- and CFP-infected cells were mixed with uninfected CEM-SS cells.

Three independent experiments were performed at the following cell numbers in 3 ml of culture medium in T-75 flasks:

Expt	Subset	YFP + CFP $\times 10^6$	Uninfected $\times 10^6$
1, 2, 3	A	0.75	2
1, 2, 3	B	0.3	2
1, 2	C	0.075	2

These cultures were established in duplicate, one for the stationary condition and one for the rocking condition. The stationary set of cultures was placed on a shelf of a TC incubator at a 10° angle to concentrate the cells at one end. The rocking set was placed on a rocking platform in the incubator set at twelve tilts per minute. On each subsequent day 1/8–1/4 of the culture was collected for flow cytometry for YFP, CFP and GFP as described previously (Levy et al. 2004). New culture medium was added to maintain a consistent volume in each culture. Details about the fluorescent reporter viruses used in this study can be found in Levy et al. (2004).

2.2 Mathematical models

We consider an ordinary differential equation (ODE) model of virus dynamics that tracks the populations of uninfected cells, as well as different types of infected cell populations. It incorporates both free virus and direct cell-to-cell transmission and includes recombination processes. A basic schematic of the model is shown in Fig. 1. Due to the complexity of the equations, they are displayed in the Supplementary Materials, Section S1. Basic model assumptions and their relation to the experimental data are summarized in Section 3. The different models are fit to the experimental data with standard methods, and the model that is most powerful at explaining the data is selected with the F-test for nested models. Details of the data fitting procedures are provided in the Supplementary Materials, Section S2.

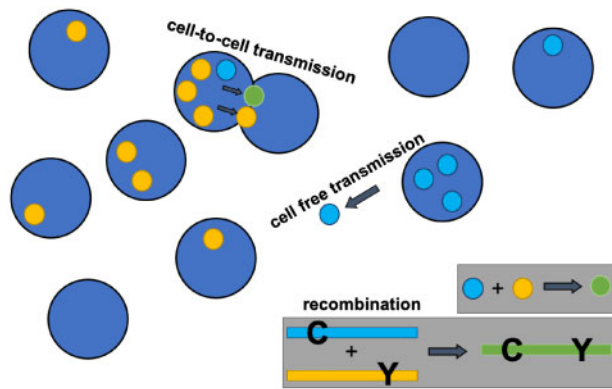


Figure 1. Basic model schematic. Cell free and synaptic cell-to-cell transmission are shown. In the ODE model, both processes are non-spatial. An example of a recombination event is also shown.

3. Results

3.1 Basic dynamics of in vitro virus growth

The experimental system and the mathematical models are described in Section 2, with additional details given in the [Supplementary Materials](#). Based on previous work by others and us ([Sourisseau et al. 2007](#); [Komarova et al. 2013](#)), in vitro virus growth experiments were performed under two conditions: 1, under static conditions in which both free virus and synaptic transmission operate and 2, under shaking conditions where cultures were placed on a gentle rocking platform, which prevents most synaptic transmission events from taking place. For each experimental condition, eight experiments were performed that differed in viral inoculum size. Time-series were obtained for each condition, where we tracked cells infected with two types of single-hit mutant: those labeled with cyan fluorescent protein, denoted as C, and those labeled with yellow fluorescent protein, denoted as Y. In addition, we followed the population of cells infected with double-hit mutants labeled with green fluorescent protein, denoted as G, as well as populations of cells that contained combinations of virus types, such as YC, YG, CG and YCG. Examples of this can be seen in [Figs. 2](#) (shaking condition) and [3](#) (static condition).

As expected, exponential growth was observed for the total population of infected cells. This growth was faster under static compared to shaking conditions, resulting in higher virus levels for static conditions during the time frame of the experiment ([Fig. 3](#)).

3.2 The mathematical model and data fitting

We used an extension of basic virus dynamics models ([Nowak and May 2000](#); [Perelson and Ribeiro 2013](#)) that includes the multiple infection of cells as well as the occurrence of both free virus and synaptic transmission, based on our previous work ([Komarova, Levy, and Wodarz 2012](#); [Komarova et al. 2013](#)). Due to the need to track reporter viruses that glow in multiple colors, the equations for the model are rather complicated and are described in detail in the [Supplementary Materials](#) Section S1. Here, basic assumptions are summarized and also shown schematically in [Fig. 1](#). Uninfected target cells are assumed to proliferate with a rate r and die with a rate d . Infected cells are generated through free virus transmission with a rate β , and through synaptic transmission with a rate γ . Free virus transmission results in the infection of the target cells with one virus.

Synaptic transmission is assumed to lead to the simultaneous infection of the target cell with S copies of the virus, where the value of S is estimated from parameter fitting procedures. Infected cells are assumed to die with a rate a . Cells can be infected with 1, single-mutant viruses that carry cyan fluorescent protein, C, or yellow fluorescent protein, Y; 2, the recombinant virus strain that shows green fluorescence (G); and 3, the non-glowing virus strain that does not show any fluorescence, W. Hence, the model tracks cell populations that are infected with any combination of these virus strains, at defined multiplicities (i.e. cells can contain i copies of C, j copies of Y, k copies of G, and l copies of W viruses). The fitness of all virus types is assumed to be identical. During the infection process, the model assumes that recombination can occur with a certain probability if the infecting virus carries two genetically different genomes. All possible recombination events are described in the [Supplementary Materials](#). The model is given by a set of ODEs and is hence non-spatial.

Besides these basic processes, an accurate model fit to all the data required the incorporation of two more processes into the model.

(i) It has been demonstrated that infection can result in the generation of a latently infected cell, but that super-infection/multiple infection can result in the activation of the latent virus. The reason is that superinfection results in TAT complementation and thus activation of the latent genome in the cell ([Bregnard et al. 2012](#)). For example, if a cell becomes infected with Y virus, it can become latent and the infected cell will not glow. Subsequent superinfection with C virus, however, will result in an infected cell that glows in both colors. As a consequence, the observed (glowing) population of singly infected cells is smaller than the actual population. This applies mostly to free virus transmission because it occurs mostly in singly infected cells. Because we assume synaptic transmission to result in the infection of cells with S viruses ($S > 1$), this effect can be ignored in this context. Hence, we assumed that upon infection of a cell with free virus, there is a probability ϵ that the infection becomes latent and that the virus consequently does not glow. Upon superinfection, however, we assume that the latent virus becomes activated and glows. Without this addition, the model under-predicts the double color YC cell numbers, and the model with latency can predict the YC population more accurately, as determined statistically by the F-test for nested models ([Supplementary Fig. S1](#)).

(ii) We found that for the shaking experiments, the model constructed so far consistently underpredicted multiply infected cells that contained the recombinant virus, that is the triple color GYC cells and double color cells GY and GC, although the number of cells infected with only GFP viruses (G) cells could be predicted more accurately ([Supplementary Fig. S2](#)). We hypothesized that although shaking results in the mixing of cells, the perfect mixing dynamics assumed by our ODE model might be responsible for this discrepancy. In particular, we propose that when, for example, a YC-infected cell releases virus particles, they are likely to re-enter the same cell and generate a G virus upon recombination, thus explaining the higher than expected number of GYC cells. The reason is that as viruses are released from a cell, this cell is unlikely to immediately move away from its present location, but rather remains in the current vicinity for a while, which makes re-infection a likely event. Similarly, if a cell is infected with GW (where 'W' stands for non-glowing virus) and only glows green, recombination between G and W can give rise to a C or a Y virus. Re-infection of the same cell with these viruses can then yield GCW and GYW

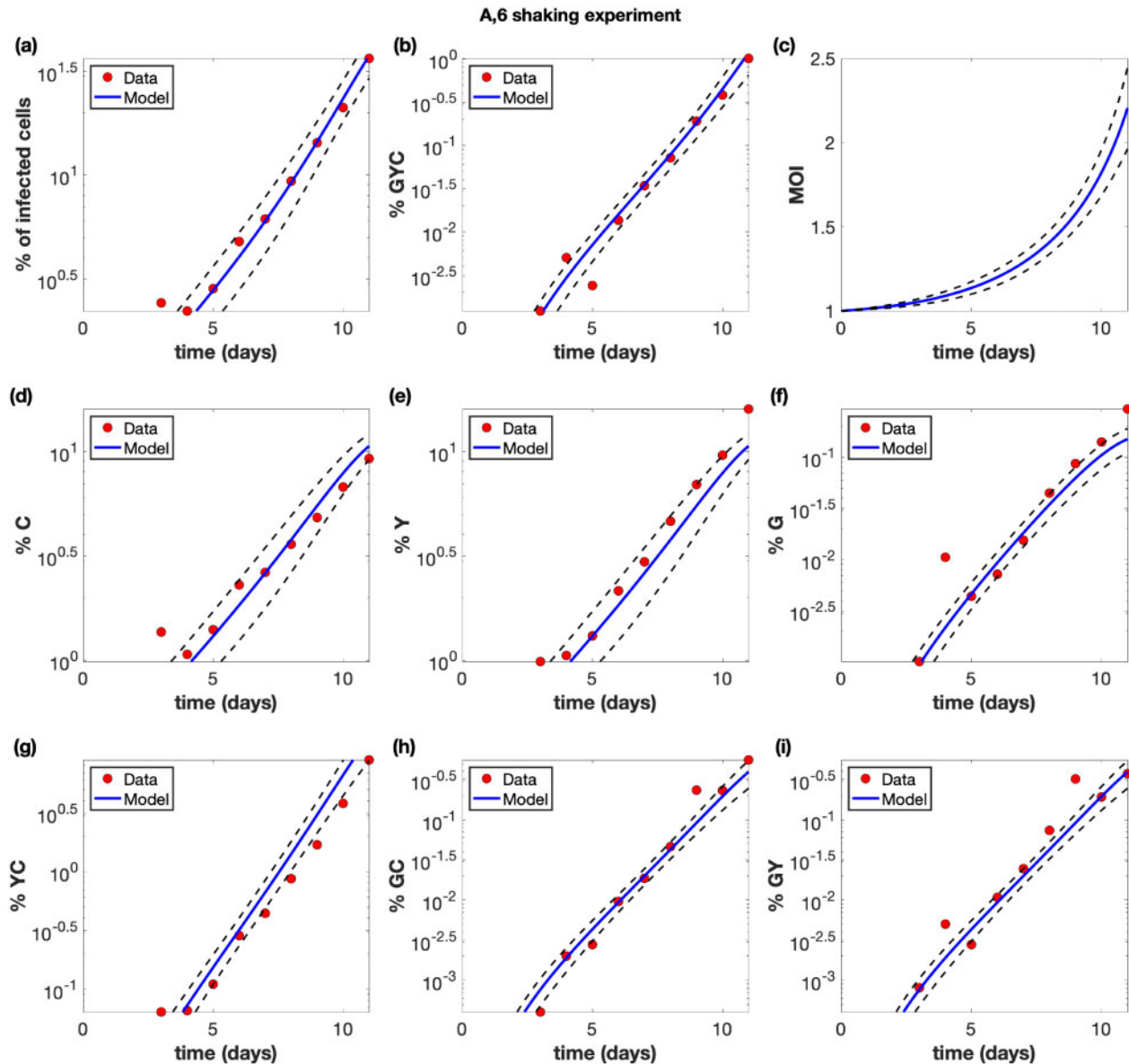


Figure 2. Example of a shaking experiment (experiment A6). The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in the Materials and Methods section with details in the [Supplementary Materials](#), and the fitting procedure is described in the [Supplementary Materials](#). Best fit parameters are included in [Supplementary Table S2](#) in the [Supplementary Materials](#). The horizontal axis for all panels represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one active copy of C. (e) The percentage of cells infected with at least one active copy of Y. (f) The percentage of cells infected with at least one active copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent pointwise 95% prediction confidence bands.

cells (which glow in two colors). We first tested this idea with an agent-based model that tracked the spatial location of cells; preliminary explorations indicated that such a model could account for the data more accurately, but due to the complexity of the model and the corresponding data fitting procedures, we decided not to proceed with the agent-based model. Instead, we modified the ODE model to capture the assumption of a higher frequency of self-infection phenomenologically (see [Supplementary Materials](#) Section S2). This model, while still compatible with a straightforward fitting procedure, could more accurately describe the GYC, GY, and GC data than the simpler model without this assumption ([Supplementary Fig. S2](#)) and

was determined to be a more powerful model (despite containing an extra ‘self-infection’ parameter) by the F-test for nested models.

These results indicate that the straightforward perfect mixing ODE models that are typically used to describe virus dynamics are not accurate descriptions of these *in vitro* dynamics at the level of detail presented here (this is discussed further below).

For data fitting, we simultaneously fit the model to the corresponding shaking and static experiments and to all infected subpopulations that were experimentally quantified. The detailed methodology is available in Section S2 of the

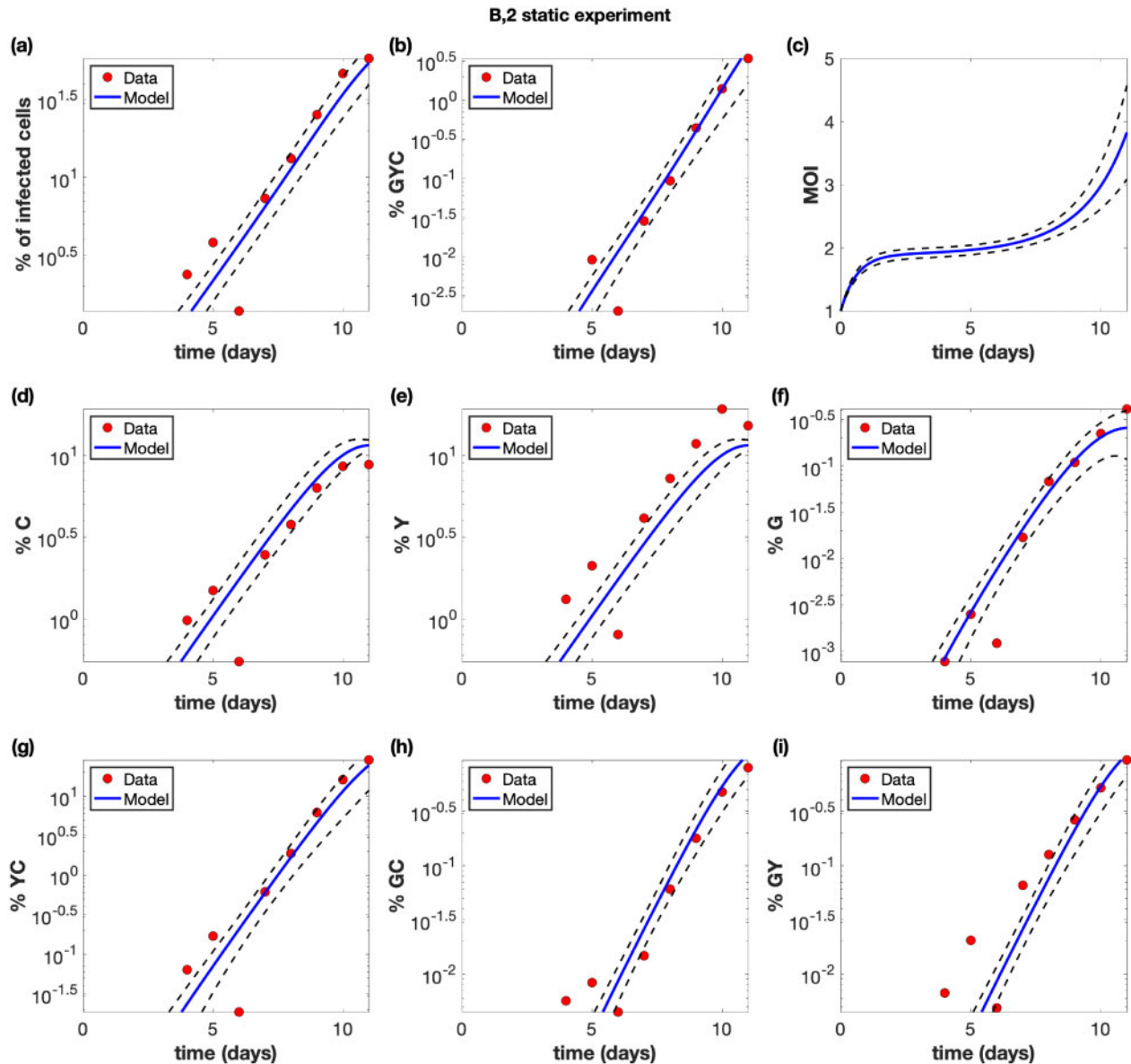


Figure 3. Example of a static, non-shaking experiment (experiment B2). The experimental data (red circles) are presented with best fit curves from the model (blue lines). Best fit parameters are included in [Supplementary Table S2](#) in the [Supplementary Materials](#). The horizontal axis for all panels represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one active copy of C. (e) The percentage of cells infected with at least one active copy of Y. (f) The percentage of cells infected with at least one active copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent pointwise 95% prediction confidence bands.

Supplementary Materials. Some parameters were fixed according to the information in literature, and these values are shown in [Supplementary Table S1](#) in the [Supplementary Materials](#). The rest of the parameter values were estimated by model fitting, and the best fit parameters are given in [Supplementary Table S2](#) in the [Supplementary Materials](#). The model fits are shown for select experiments in [Figs. 2](#) and [3](#), and for all experiments in the [Supplementary Figs. S13–S28](#). We note that the data are typically noisier for the shaking compared to the static experiment, due to stochastic effects, especially if the experiments start with a low initial percentage of infected cells.

In what follows, we describe different parameter estimates and their implications for infection dynamics.

3.3 Quantifying the relative contribution of free virus and synaptic transmission to virus spread

On the most basic level, we estimated the growth rate of the virus in the shaking and static conditions, and hence estimated the rate of free virus transmission and the rate of synaptic transmission. Previous work showed that shaking can increase the rate of free virus spread compared to static cultures due to mixing the virus more efficiently among the cells, by a factor of approximately $f = 1.33$ ([Komarova et al. 2013](#)). Therefore, we corrected for this when calculating the rate of free virus spread from the shaking experiments (full details of the methodology are given in the [Supplementary Materials](#)). According to our calculations, the rate of free virus spread was on average

1.80 ± 0.31 fold higher than the rate of synaptic transmission (see Section S3 of the [Supplementary Materials](#)). In our previous study, these rates were more equal, with the rate of free virus transmission being on average 1.1 ± 0.1 -fold higher than the rate of synaptic transmission ([Komarova et al. 2013](#)). This is further elaborated on in the Section 4.

3.4 Further parameter estimates

The model fitting to the experimental data allowed us to also estimate a variety of other parameters that characterize this system. 1, The probability for the virus to become latent upon single infection was rather consistent across the different experiments, with an average of $\varepsilon = 0.41 \pm 0.084$. In other words, about 40 per cent of all infection events of uninfected cells (via free virus transmission) are estimated to result in viral latency in this system. 2, We estimated the recombination probability to be $\rho = 0.30 \pm 0.14$. This is relatively close to the maximally possible recombination probability of $\rho = 0.5$ in this model (see [Supplementary Materials](#)). 3, An important question concerns the number of viruses that are transferred from one cell to another through virological synapses during cell-to-cell transmission. Experimental data indicate that virus transfer through virological synapses is a very efficient process, with tens to hundreds of viruses transferred ([Chen et al. 2007](#); [Hubner et al. 2009](#)). Not all of these viruses, however, are likely to successfully integrate into the genome of the target cell. In our experiments, we estimated that on average $S = 3.0 \pm 0.35$ viruses successfully infect a target cell per synaptic transmission event. 4, The parameterized model further allowed us to estimate the average infection multiplicities in the experiments. As expected, the estimated multiplicities increased as the virus population grew. For the shaking experiments, multiplicities ranged between 1 and 2.2 during the time frame of the experiments ([Fig. 2c](#)). For the static experiments, the average infection multiplicity ranged from 1 to 4 during the phase of virus growth ([Fig. 3c](#)).

3.5 Comparing the effect of viral transmission pathways on recombinant generation: beyond the experimental setup

The experimental setup discussed above represents an artificial system to parameterize an evolutionary model of viral recombination. Here, we use the parameterized model to simulate the dynamics in more general terms and to go beyond the experimental setup in the following ways.

- i. Shaking conditions resulted in largely free virus transmission, while static conditions allowed both free virus and synaptic transmission to proceed. Using the estimated parameters in computer simulations, we further predict the dynamics for a scenario where only synaptic transmission takes place, which was not feasible experimentally.
- ii. The inherent differences in viral growth rates in the shaking and static experiments make a direct comparison of recombinant evolution under the different transmission pathways difficult to interpret. Hence, we adjusted the infection rates such that they are the same for free virus transmission only, synaptic transmission only, and a combination of the two.
- iii. The experimental setup is an artificially constructed system that was used to measure key parameters connected to recombinant evolution. We can use the parameterized equations to move beyond this particular system and describe a

more complete evolutionary picture. That is, we can start with a ‘wild-type’ virus (which was not part of the experimental setup) and model the generation of single-hit mutants by point mutation (which would technically correspond to the Y and C viruses, even though these reporter viruses are not generated by single mutation events in the experiments), and the subsequent evolution of double mutants through recombination (which would correspond to G viruses).

In the first set of simulations, we started with an equal number of single-mutant viruses, as done in the experiments, and ignored point mutations. We started the simulations with 10^9 uninfected cells, and in contrast to the experiments, we did not allow for exponential expansion of the uninfected cell population, because lack of extensive cell expansion corresponds better to the conditions in which virus grows during acute HIV infection *in vivo*. We let the total number of infected cells grow until a population size of 4×10^8 cells was reached. This was done assuming that only synaptic transmission (not possible in experiments), only free virus transmission, or both types of transmission occur. As a first step, we used unadjusted transmission parameters that were identical to the ones estimated from the experimental data. The number of cells that contain the recombinant virus was quantified as a function of time ([Fig. 4a](#)). In these simulations, recombinants rise the fastest if both transmission pathways operate, compared to only a single active transmission pathway. This is largely due to both transmission types resulting in the fastest overall rate of viral transmission.

To compare the effect of the transmission pathways themselves on recombinant generation (independent of the differences in viral transmission rates), we corrected for the differences in viral transmission rates associated with the different pathways, such that the basic reproductive ratio of the virus ($R_0 = \frac{\beta(1-\varepsilon)+\gamma}{\alpha}$) was identical for free virus only, synaptic only, and for a combination of both transmission pathways ([Fig. 4b](#)). When starting from the experimentally implemented initial conditions (infected cells that contain either one or the other virus type), we observe that the transmission pathways do not make a significant difference for the rate at which recombinants evolve. The reason is that we start with an infected cell population that contains either one or the other virus. Recombination requires the two virus types to come together in the same cell, which is likely to happen only at larger virus loads both for free virus and for synaptic transmission, explaining the lack of a difference in these simulations. Further, we note that at high virus loads, the number of recombinants in simulations with only free virus transmission slightly overtakes those in simulations with synaptic transmission; at this stage in the dynamics, the superinfection of latently infected cells becomes a common event in free virus transmission, which elevates the rate of productive free virus infection due to TAT complementation, and results in a higher effective reproductive number for the free-virus pathway compared to the synaptic pathway.

A different picture is observed if we start with a low number of coinfecting cells that contain both types of single-mutant virus ([Fig. 4c](#)). Now, recombinant evolution is significantly faster for simulations with only synaptic transmission than for simulations that include free virus transmission. The reason is that synaptic transmission allows the repeated co-transmission of genetically different virus strains, which facilitates recombination processes in the growing virus population, highlighting the importance of this mechanism. Free virus transmission, in

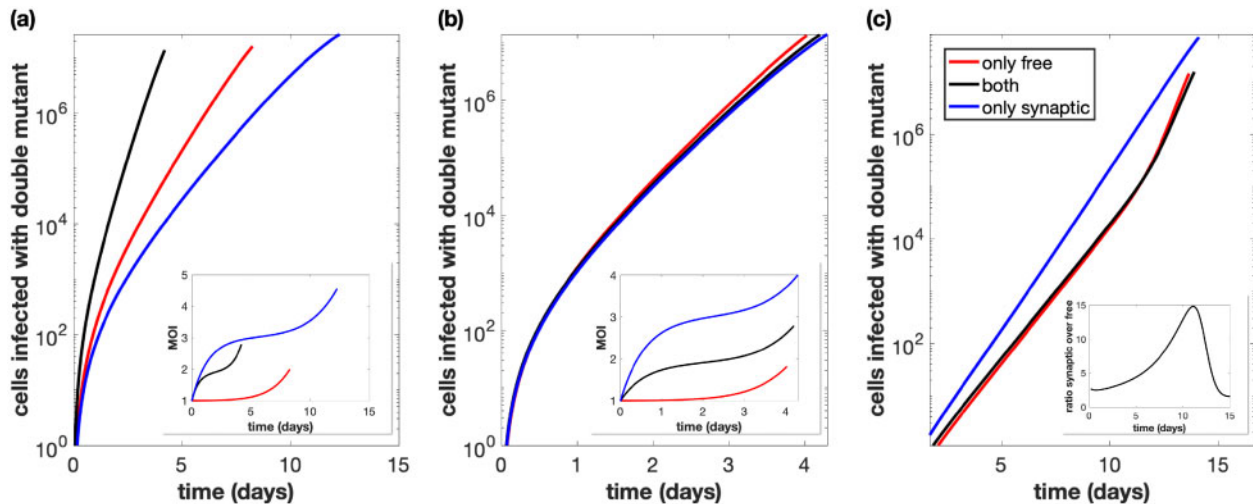


Figure 4. Total number of cells infected with at least one active copy of double-mutant virus plotted against time. The multiplicity of infection (MOI) is included in the inset. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40 per cent of this initial amount, while also assuming that uninfected cells do not divide. Parameters are the best fit parameters from static experiment B2, which are listed in [Supplementary Table S2](#) in the [Supplementary Materials](#). The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single-mutant strains.

contrast, contributes to the separation of the two single-mutant virus strains into separate cells, which slows down recombinant generation.

Next, we consider a more complete evolutionary scenario. We start with a wild-type virus only and introduce point mutations into the model. The wild-type virus can thus give rise to two different kind of single-mutant viruses, and double-mutant viruses can be generated either by recombination between the single mutants, or by additional point mutations in those single mutants (Fig. 5). We start with basic simulations (unadjusted parameters) and consider the number of cells that contain two different single-mutant viruses over time, since these cells form the basis for recombination. Despite the fact that the estimated rate of synaptic transmission is slower than the estimated rate of free virus transmission, in our experimental system, the number of cells containing both single mutants rises sharply at low viral load in simulations that take into account only synaptic transmission, compared to simulations that only take into account free virus transmission (Fig. 5a). The reason is that 1, more reverse transcription events occur with synaptic transmission, thus generating more single-mutant viruses (Fig. 5a, inset), and 2, once the two different single-mutant viruses have come together in the same cell, synaptic transmission enables their repeated co-transfer to target cells, explaining the sharp rise. As viral load grows to higher levels in the simulations, the number of cells coinfecting with both single mutants under free virus transmission overtakes those under synaptic transmission (Fig. 5a) because at high viral loads, the rate of coinfection becomes relatively high even for free virus transmission, and the estimated rate of free virus transmission is faster than the rate of synaptic transmission.

Looking at the number of double mutants (Fig. 5b), we observe similar patterns. Figure 5c and d repeat these plots with adjusted parameters, such that the basic reproductive ratio of the virus is the same, independent of the transmission pathways that are assumed to occur. We see that the synaptic pathway contributes most to double-mutant evolution, due to the repeated co-transmission of the two single-mutant

strains. This is most pronounced when considering the number of cells that are coinfecting with both single mutants (Fig. 5c). The effect of synaptic transmission on the evolution of double mutants is qualitatively the same, but less pronounced because apart from recombination (Fig. 5d), mutation processes also occur in these simulations. Therefore, although synaptic transmission significantly enhances the chances for recombination to occur (due to the co-transmission of the different single mutants), the mutation processes in the much more abundant singly infected cell population mask this effect to an extent.

4. Discussion and conclusion

In this study, we used mathematical models in combination with experimental data to determine the contribution of free virus transmission and direct cell-to-cell transmission to the evolution of recombinant viruses. This was possible through the use of experimental techniques to separate the two viral transmission pathways, and the use of fluorescent reporter viruses that allowed us to track recombinant evolution. Fitting of mathematical models to the experimental data allowed the estimation of important parameters that characterize the recombination process, and the models were further used to run evolutionary simulations to go beyond the experimental system and to quantify how the two transmission pathways influence the number of recombinants generated under a more complete set of evolutionary processes.

We found that direct cell-to-cell transmission through virological synapses promotes the evolution of recombinants, due to the following mechanisms: 1, synaptic transmission increases the infection multiplicity of cells, which is the basis for recombinant generation; 2, synaptic transmission promotes the repeated co-transmission of two different virus strains from cell to cell, which increases the chances to eventually generate a recombinant virus; 3, synaptic transmission increases point mutation generation because the simultaneous transfer of multiple viruses to the target cell increases

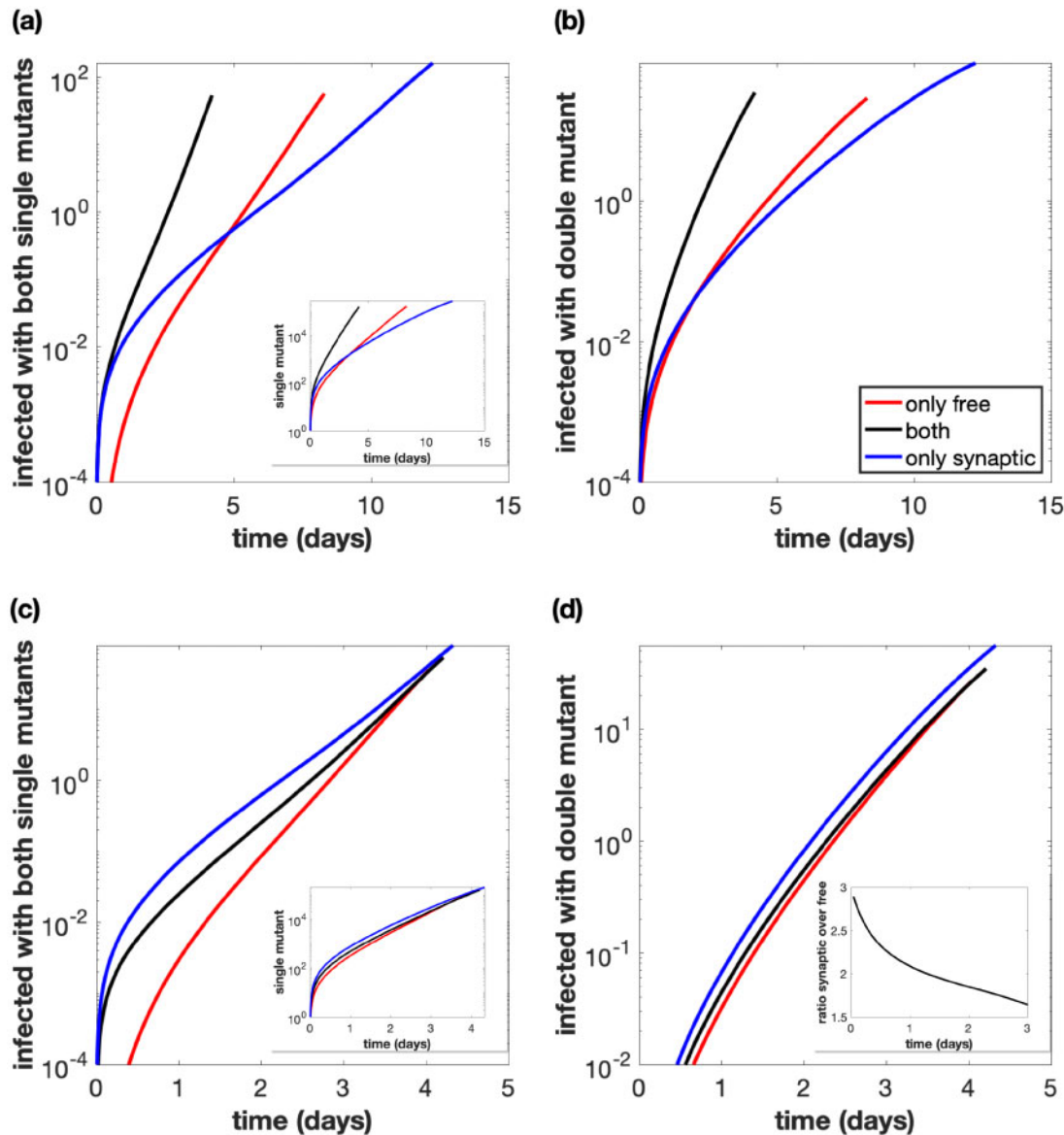


Figure 5. Same as in Fig. 4, but the simulations start with a small equal amount of cells singly infected with the wild-type, and mutations are included. (a) Total number of cells infected with at least one copy of both single-mutant strains. The number of cells infected with at least one copy of one of the single mutants is included in the inset. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double-mutant virus plotted against time. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases. The ratio of the number of cells infected with the double mutant for only synaptic transmission versus only free virus transmission is included in the inset.

the number of reverse transcription events, during which point mutations are mostly likely to occur. We estimated cell-to-cell transmission to result in the successful integration of about three viruses per synapse in the target cell. Experiments have shown that tens to hundreds of virus particles can be simultaneously transmitted per synapse (Chen et al. 2007; Hubner et al. 2009), but it is likely that only a subset of these result in successful integration.

We note that while for a fixed infection rate, purely synaptic transmission results in fastest recombinant evolution, free virus transmission can yield other advantages for the virus that might be equally important for its success, such as more efficient mixing of viruses among cells, which can promote virus spread and dissemination. The rate of recombinant evolution studied here is only one component that determines the

success of the virus, and the presence of the two viral transmission pathways has to be interpreted in this light.

Our analysis also repeated some of our previous work (Komarova et al. 2013) where we estimated the relative contribution of free virus and synaptic transmission to virus growth. In this study, it was estimated that the two transmission pathways contributed more or less equally to virus spread, where the rate of free virus transmission was approximately 1.1-fold faster than the rate of synaptic transmission. A subsequent study repeated our work and came to the same conclusion (Iwami et al. 2015). In the present study, we estimated the rate of free virus transmission to be approximately 1.8-fold faster than the rate of synaptic transmission, which is a larger difference. The reason for this discrepancy is that in the present study, we took into account the generation of latently infected

cells upon free virus transmission, while the previous work did not. During synaptic transmission, latency is not established in our model because this is very unlikely due to the phenomenon of TAT complementation in multiply infected cells (Bregnard et al. 2012). If we ignore free virus infection events that result in latency and re-calculate the rate of free virus transmission (see Section S3 of the [Supplementary Materials](#)), it is only about 1.06-fold faster than the rate of synaptic transmission, that is the two transmission pathways contribute about equally to virus spread, as in previous studies (Komarova et al. 2013; Iwami et al. 2015). Note that while the rate of productive free virus infection is the measure that is of immediate importance for the expansion of the virus population, the latently infected cells that are being generated can become relevant with a time delay if the virus spontaneously activates or becomes activated through TAT complementation upon superinfection. Hence, it is useful to consider both the total rate of free virus infection and the rate of productive infection.

Other work indicates that the relative importance of the two transmission pathways can depend on the microenvironment. For instance, in (Imle et al. 2019), HIV-1 spread in suspension was driven completely by free virus transmission, whereas virus spread in 3D collagen was driven by synaptic transmission, with approximately a 22 per cent (36 per cent) contribution of free virus transmission in loose (dense) collagen. This also brings up the effect of spatially restricted virus spread for the evolutionary dynamics explored here. The dynamics of virus growth under spatially restricted direct cell-to-cell transmission has been explored in the context of Hepatitis C virus infection (Graw et al. 2015) and more generally (Kumberger et al. 2018). In the context of HIV infection, evidence for spatially clustered virus growth has been provided from experiments with HIV-infected humanized mice (Law et al. 2016), and the consequences of spatially restricted synaptic transmission for the evolution through recombination has been studied recently with mathematical models (Kreger, Komarova, and Wodarz 2020). It is unlikely that extensive spatial restrictions apply to the experimental virus growth cultures investigated here, in which the virus population grew exponentially rather than according to growth laws that are more typical for infected cell clusters. An extension of the currently described evolutionary dynamics to a spatial setting, however, will be important for future work.

An interesting finding in our study was that at the resolution of the data presented here, standard ODEs of virus dynamics failed to adequately describe several multiply infected cell subpopulations. We found that incorporating the assumption of limited mixing and the consequent re-infection of cells by their own offspring virus resulted in better model fits to the data. While this can be relevant to our *in vitro* system with transformed cell lines as infection targets, the relevance for virus replication *in vivo* might be less due to different infected cell life-spans and reverse transcription kinetics *in vivo* compared to the *in vitro* setting. Spatially restricted virus spread to new target cells (rather than self-infection), however, could have a similar effect since several viruses from the source cell likely get passed on to the same target cell due to the limited number of cells in the spatial neighborhood.

Our models and data analysis provide an important link between direct cell-to-cell transmission, heightened infection multiplicity, and an increased rate of recombinant evolution. While this was established within the framework of *in vitro* experimentation, these notions are likely also relevant *in vivo*. Synaptic transmission has been documented *in vivo* and infection multiplicities of cells in the tissue from HIV-infected

patients have been shown to be between three and four or even higher on average (Jung et al. 2002). Other studies reported the average infection multiplicities in the blood and tissue of HIV-infected patients to be closer to one (Josefsson et al. 2011; Josefsson et al. 2013), although the restriction of the analysis to cells that express the CD4 receptor could have artificially lowered the estimate of infection multiplicities. The CD4 receptor becomes eventually down-regulated following the infection of the cell with the first virus. Ignoring such cells in the analysis could miss those cells with the highest infection multiplicities. The spatial modeling approaches discussed above might provide additional insights for these evolutionary dynamics *in vivo*. Future studies should further examine how these processes affect viral evolution under the assumption that mutants differ in fitness, and that the coinfection of cells with virus strains of different fitness can result in complementary and inhibitory interactions. The modeling framework and parameter estimates provided here form a basis for future investigation.

Supplementary data

[Supplementary data](#) and experimental data are available at [Virus Evolution](#) online.

Conflict of interest: None declared.

Funding

This work was supported by NSF National Science Foundation grant no. DMS 1662146/1662096. We also acknowledge support from the NSF National Science Foundation-Simons Center for Multiscale Cell Fate Research.

References

- Althaus, C. L., and Bonhoeffer, S. (2005) 'Stochastic Interplay between Mutation and Recombination during the Acquisition of Drug Resistance Mutations in Human Immunodeficiency Virus Type 1', *Journal of Virology*, 79: 13572–8.
- Bonhoeffer, S. et al. (1997) 'Virus Dynamics and Drug Therapy', *Proceedings of the National Academy of Sciences of the United States of America*, 94: 6971–6.
- et al. (2004) 'Evidence for Positive Epistasis in HIV-1', *Science (New York, N.Y.)*, 306: 1547–50.
- Bregnard, C. et al. (2012) 'Suboptimal Provirus Expression Explains Apparent Nonrandom Cell Coinfection with HIV-1', *Journal of Virology*, 86: 8810–20.
- Bretscher, M. T. et al. (2004) 'Recombination in HIV and the Evolution of Drug Resistance: For Better or for Worse?', *Bioessays : News and Reviews in Molecular, Cellular and Developmental Biology*, 26: 180–8.
- Carvajal-Rodríguez, A., Crandall, K. A., and Posada, D. (2006) 'Recombination Estimation under Complex Evolutionary Models with the Coalescent Composite-Likelihood Method', *Molecular Biology and Evolution*, 23: 817–27.
- Chen, P. et al. (2007) 'Predominant Mode of Human Immunodeficiency Virus Transfer between T Cells Is Mediated by Sustained Env-Dependent Neutralization-Resistant Virological Synapses', *Journal of Virology*, 81: 12582–95.
- Del Portillo, A. et al. (2011) 'Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection', *Journal of Virology*, 85: 7169–76.
- Fraser, C. (2005) 'HIV Recombination: What Is the Impact on Antiretroviral Therapy?', *Journal of the Royal Society, Interface*, 2: 489–503.

- Graw, F. et al. (2015) 'Quantification of Hepatitis C Virus Cell-to-Cell Spread Using a Stochastic Modeling Approach', *Journal of Virology*, 89: 6551–61.
- Hirsch, V. M. (1999) 'Evolution of the Fittest Ends in Tragedy', *Nature Medicine*, 5: 488–9.
- Ho, D. D. et al. (1995) 'Rapid Turnover of Plasma Virions and Cd4 Lymphocytes in Hiv-1 Infection', *Nature*, 373: 123–6.
- Hubner, W. et al. (2009) 'Quantitative 3D Video Microscopy of HIV Transfer across T Cell Virological Synapses', *Science*, 323: 1743–7.
- Imle, A. et al. (2019) 'Experimental and Computational Analyses Reveal That Environmental Restrictions Shape HIV-1 Spread in 3D Cultures', *Nature Communications*, 10: 1–18.
- Iwami, S. et al. (2015) 'Cell-to-Cell Infection by HIV Contributes over Half of Virus Infection', *eLife*, 4:
- Josefsson, L. et al. (2013) 'Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals That the Majority of CD4+ T-Cells Contain One HIV-1 DNA Molecule', *PLoS Pathogens*, 9: e1003432.
- et al. (2011) 'Majority of CD4+ T Cells from Peripheral Blood of HIV-1-Infected Individuals Contain Only One HIV DNA Molecule', *Proceedings of the National Academy of Sciences of the United States of America*, 108: 11199–204.
- Jung, A. et al. (2002) 'Multiply Infected Spleen Cells in HIV Patients', *Nature*, 418: 144.
- Kimata, J. T. et al. (1999) 'Emerging Cytopathic and Antigenic Simian Immunodeficiency Virus Variants Influence AIDS Progression', *Nature Medicine*, 5: 535–41.
- Komarova, N. L., Levy, D. N., and Wodarz, D. (2012) 'Effect of Synaptic Transmission on Viral Fitness in HIV Infection', *PLoS One*, 7: e48361.
- et al. (2013) 'Relative Contribution of Free-Virus and Synaptic Transmission to the Spread of HIV-1 through Target Cell Populations', *Biology Letters*, 9: 20121049.
- Kouyos, R. D., Fouchet, D., and Bonhoeffer, S. (2009) 'Recombination and Drug Resistance in HIV: Population Dynamics and Stochasticity', *Epidemics*, 1: 58–69.
- Kreger, J., Komarova, N. L., and Wodarz, D. (2020) 'Effect of Synaptic Cell-to-Cell Transmission and Recombination on the Evolution of Double Mutants in HIV', *Journal of the Royal Society, Interface*, 17: 20190832.
- Kumberger, P. et al. (2018) 'Accounting for Space—Quantification of Cell-to-Cell Transmission Kinetics Using Virus Dynamics Models', *Viruses*, 10: 200.
- Law, K. M. et al. (2016) 'In Vivo HIV-1 Cell-to-Cell Transmission Promotes Multicopy Micro-Compartmentalized Infection', *Cell Reports*, 15: 2771–83.
- Levy, D. N. et al. (2004) 'Dynamics of HIV-1 Recombination in Its Natural Target Cells', *Proceedings of the National Academy of Sciences of the United States of America*, 101: 4204–9.
- Mansky, L. M., and Temin, H. M. (1995) 'Lower in Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase', *Journal of Virology*, 69: 5087–94.
- Moutouh, L., Corbeil, J., and Richman, D. D. (1996) 'Recombination Leads to the Rapid Emergence of HIV-1 Dually Resistant Mutants under Selective Drug Pressure', *Proceedings of the National Academy of Sciences of the United States of America*, 93: 6106–11.
- Nowak, M. A., and May, R. M. (2000) *Virus Dynamics. Mathematical Principles of Immunology and Virology*. Oxford: Oxford University Press.
- Perelson, A. S., and Ribeiro, R. M. (2013) 'Modeling the within-Host Dynamics of HIV Infection', *BMC Biology*, 11: 96.
- et al. (1996) 'Hiv-1 Dynamics in-Vivo - Virion Clearance Rate, Infected Cell Life- Span, and Viral Generation Time', *Science (New York, N.Y.)*, 271: 1582–86.
- Sourisseau, M. et al. (2007) 'Inefficient Human Immunodeficiency Virus Replication in Mobile Lymphocytes', *Journal of Virology*, 81: 1000–12.
- Trinite, B. et al. (2014) 'Suppression of Foxo1 Activity and down-Modulation of CD62L (L-Selectin) in HIV-1 Infected Resting CD4 T Cells', *PLoS One*, 9: e110719.
- Wei, X. P. et al. (1995) 'Viral Dynamics in Human-Immunodeficiency-Virus Type-1 Infection', *Nature*, 373: 117–22.

Supplementary Materials

Jesse Kreger^a, Josephine Garcia^c, Hongtao Zhang^c, Natalia L. Komarova^a,
Dominik Wodarz^{b,a} and David N. Levy^c

a) Department of Mathematics, University of California Irvine, Irvine CA 92697

b) Department of Population Health and Disease Prevention Program in Public Health

Susan and Henry Samueli College of Health Sciences, University of California, Irvine CA 92697

c) Department of Basic Science, New York University College of Dentistry, New York NY 10010

Contents

1	Mathematical model	1
2	Fitting procedures and best fit parameters	4
3	Relative contribution of free virus and synaptic transmission	10
4	Role of synaptic transmission in evolution	13
5	Further theoretical considerations	16
6	Further justification of parameter estimates	18
7	Best fit parameters and plots	23
8	Comparing the effect of viral transmission pathways on recombinant dynamics for all best fit parameter values	23

1 Mathematical model

Mathematical models have been used previously to describe the relative contribution of free virus and synaptic transmission to HIV-1 spread [1]. Here, we expand this model to better quantify multiple components of infection. In order to study the dynamics of recombinant viruses, this model can be expanded to include multiple infection and infection with multiple strains of virus. Here, based on the experimental setup, we consider single mutant strain C (equipped with glowing cyan fluorescent protein), single mutant strain Y (equipped with glowing yellow fluorescent protein), double mutant strain G (recombinant strain that includes both mutations and glows green), and double mutant strain W (non-glowing recombinant strain that includes neither mutation). We denote $x_{i,j,k,l}(t)$ as the number of cells infected with i copies of strain C, j copies of strain Y, k copies of strain G, and l copies of strain W. Further, we denote $\mathcal{N} = i + j + k + l$ as the total number of viruses that the cell is infected with. As N is the maximum multiplicity of infection, $0 \leq \mathcal{N} \leq N$ must be satisfied.

Upon an infection event (either free virus or cell-to-cell), the infected cell produces a virus package that contains two copies of viral RNA. We assume that if recombination happens (with

probability $0 \leq \rho \leq \frac{1}{2}$) that a distinct virus strain is created if possible (which is equivalent to not making this assumption and allowing $0 \leq \rho \leq 1$). Figure 1 in the main text shows a possible recombination event between strains Y and C to create recombinant strain G. We denote $P(C|i, j, k, l)$ to be the probability for a cell infected with i copies of strain C, j copies of strain Y, k copies of strain G, and l copies of strain W to infect with a copy of strain C upon an infection event. This can happen in the absence of recombination if at least one of the viral copies is of the C strain, and then that strain is chosen for infection. If one viral copy of G and one viral copy of W are chosen, and recombination occurs, then a C or Y virus is created. We assume that viral copies are chosen randomly based on their density within the cell. Therefore, we have that

$$P(C|i, j, k, l) = \frac{1}{\mathcal{N}^2} \left(i(i + k + l) + (1 - \rho)ij + \rho kl \right). \quad (1)$$

Similarly, we have that

$$P(Y|i, j, k, l) = \frac{1}{\mathcal{N}^2} \left(j(j + k + l) + (1 - \rho)ij + \rho kl \right), \quad (2)$$

$$P(G|i, j, k, l) = \frac{1}{\mathcal{N}^2} \left(k(k + i + j) + (1 - \rho)kl + \rho ij \right), \quad (3)$$

$$P(W|i, j, k, l) = \frac{1}{\mathcal{N}^2} \left(l(l + i + j) + (1 - \rho)kl + \rho ij \right). \quad (4)$$

Therefore, we have that the total intensity to infect with strain α is given by

$$B_\alpha = \sum_{0 < \mathcal{N} \leq N} P(X|i, j, k, l) x_{i,j,k,l} \quad \text{for } \alpha = C, Y, G, W, \quad (5)$$

where again $\mathcal{N}$ is the total number of copies of virus within the cell, so here we are sum over all infected subpopulations.

We also define Z and X as the total number of cells and total number of infected cells respectively, so we have that

$$Z = \sum_{0 \leq \mathcal{N} \leq N} x_{i,j,k,l}, \quad (6)$$

$$X = \sum_{0 < \mathcal{N} \leq N} x_{i,j,k,l}. \quad (7)$$

To include synaptic transmission, we first define $S \geq 1$ to be the constant number of viral copies transferred per synaptic infection event. Therefore, for each synaptic infection event, we must choose S viral copies from the infected cell to pass on to the target cell. Assume that the infecting cell is infected with i copies of strain C, j copies of strain Y, k copies of strain G, and l copies of strain W. Synaptic transmission can result in infection by $\hat{i}$ copies of strain C, $\hat{j}$ copies of strain Y, $\hat{k}$ copies of strain G, and $\hat{l}$ copies of strain W, as long as $\hat{i} + \hat{j} + \hat{k} + \hat{l} = S$ is satisfied. The likelihood for each of these synaptic infection events is based on the probability to pick each strain for infection ($P(\alpha|i, j, k, l)$ for $\alpha = C, Y, W$, and G). For each transmission event, we multiply the probability to transmit the respective number of each strain $P(C|i, j, k, l)^{\hat{i}} P(Y|i, j, k, l)^{\hat{j}} P(G|i, j, k, l)^{\hat{k}} P(W|i, j, k, l)^{\hat{l}}$ by the number of ways to transmit that number of each strain, which is the multinomial coefficient $\binom{S}{\hat{i}, \hat{j}, \hat{k}, \hat{l}} = \frac{S!}{\hat{i}! \hat{j}! \hat{k}! \hat{l}!}$. Therefore, for $\hat{i} + \hat{j} + \hat{k} + \hat{l} = S$, we define the synaptic infection intensity as

$$G_{\hat{i}, \hat{j}, \hat{k}, \hat{l}} = \frac{S!}{\hat{i}! \hat{j}! \hat{k}! \hat{l}!} \sum_{0 < \mathcal{N} \leq N} P(C|i, j, k, l)^{\hat{i}} P(Y|i, j, k, l)^{\hat{j}} P(G|i, j, k, l)^{\hat{k}} P(W|i, j, k, l)^{\hat{l}} x_{i,j,k,l}. \quad (8)$$

Then the differential equation model is given by

$$\dot{x}_{0,0,0,0} = rx_{0,0,0,0} - \frac{\beta + \gamma}{Z} x_{0,0,0,0} X, \quad (9)$$

$$\begin{aligned} \dot{x}_{i,j,k,l} = & \frac{\beta}{Z} (x_{i-1,j,k,l} B_C + x_{i,j-1,k,l} B_Y + x_{i,j,k-1,l} B_G + x_{i,j,k,l-1} B_W - x_{i,j,k,l} X) \\ & + \frac{\gamma}{Z} \left(\sum_{\hat{i}+\hat{j}+\hat{k}+\hat{l}=S} x_{i-\hat{i},j-\hat{j},k-\hat{k},l-\hat{l}} G_{i,\hat{j},\hat{k},\hat{l}} - x_{i,j,k,l} X \right) - ax_{i,j,k,l}, \end{aligned} \quad (10)$$

with the convention that any negative index is 0 and with the appropriate adjustments for maximum multiplicity of infection N . Here r is the division rate of uninfected cells, a is the death rate of infected cells, β is the rate of free virus transmission, and γ is the rate of synaptic transmission. Parameter descriptions (for all models) are given in Table S1. Note that the less detailed model used in [1] can be obtained by summing up the equations for all infected subpopulations. Since these are ordinary differential equations, the only difference between free virus and synaptic transmission is that multiple copies of virus can be transmitted during synaptic infection, as both processes are non-spatial.

Latent infection. It turns out (see Section 2) that the basic model described above is not sufficient to describe experimental data. As the next level of complexity, we also include latent infection in the description. We assume that upon infection, a cell that is only infected with a single copy of virus may not actively produce new virus. Once this cell is infected by additional copies of virus, it activates and starts glowing according to its viral contents. We denote silent cells with a single copy of a C or Y virus as $w_{1,0,0,0}$ and $w_{0,1,0,0}$ and silent cells with a single copy of a G or W virus as $w_{0,0,1,0}$ and $w_{0,0,0,1}$ respectively. We assume that latently infected cells grow according to the same law as the uninfected cells. The probability for an uninfected cell to become infected and not actively produce virus (become latent) is denoted by ε . Therefore, in the presence of latently infected cells, the total number of cells is given by

$$Z = \sum_{0 \leq \mathcal{N} \leq N} x_{i,j,k,l} + w_{1,0,0,0} + w_{0,1,0,0} + w_{0,0,1,0} + w_{0,0,0,1}. \quad (11)$$

The model with latent infection is then given by

$$\dot{x}_{0,0,0,0} = rx_{0,0,0,0} - \frac{\beta + \gamma}{Z} x_{0,0,0,0} X, \quad (12)$$

$$\dot{x}_{1,0,0,0} = \frac{\beta}{Z} (x_{0,0,0,0} B_C (1 - \varepsilon) - x_{1,0,0,0} X) - \frac{\gamma}{Z} (x_{1,0,0,0} X) - ax_{1,0,0,0}, \quad (13)$$

$$\dot{w}_{1,0,0,0} = \frac{\beta}{Z} (x_{0,0,0,0} B_C \varepsilon - w_{1,0,0,0} X) - \frac{\gamma}{Z} (w_{1,0,0,0} X) + rw_{1,0,0,0}, \quad (14)$$

$$\dot{x}_{0,1,0,0} = \frac{\beta}{Z} (x_{0,0,0,0} B_Y (1 - \varepsilon) - x_{0,1,0,0} X) - \frac{\gamma}{Z} (x_{0,1,0,0} X) - ax_{0,1,0,0}, \quad (15)$$

$$\dot{w}_{0,1,0,0} = \frac{\beta}{Z} (x_{0,0,0,0} B_Y \varepsilon - w_{0,1,0,0} X) - \frac{\gamma}{Z} (x_{0,1,0,0} X) + rw_{0,1,0,0}, \quad (16)$$

...

$$\dot{x}_{2,0,0,0} = \frac{\beta}{Z} ((x_{1,0,0,0} + w_{1,0,0,0}) B_C - x_{2,0,0,0} X) - \frac{\gamma}{Z} (x_{2,0,0,0} X) - ax_{2,0,0,0}, \quad (17)$$

$$\dot{x}_{0,2,0,0} = \frac{\beta}{Z} ((x_{0,1,0,0} + w_{0,1,0,0}) B_Y - x_{0,2,0,0} X) - \frac{\gamma}{Z} (x_{0,2,0,0} X) - ax_{0,2,0,0}, \quad (18)$$

$$\dot{x}_{1,1,0,0} = \frac{\beta}{Z} ((x_{1,0,0,0} + w_{1,0,0,0}) B_Y + (x_{0,1,0,0} + w_{0,1,0,0}) B_C - x_{1,1,0,0} X) - \frac{\gamma}{Z} (x_{1,1,0,0} X) - ax_{1,1,0,0}, \quad (19)$$

...

$$\begin{aligned} \dot{x}_{i,j,k,l} &= \frac{\beta}{Z} (x_{i-1,j,k,l} B_C + x_{i,j-1,k,l} B_Y + x_{i,j,k-1,l} B_G + x_{i,j,k,l-1} B_W - x_{i,j,k,l} X) \\ &+ \frac{\gamma}{Z} \left(\sum_{\hat{i}+\hat{j}+\hat{k}+\hat{l}=S} x_{i-\hat{i},j-\hat{j},k-\hat{k},l-\hat{l}} G_{\hat{i},\hat{j},\hat{k},\hat{l}} - x_{i,j,k,l} X \right) - ax_{i,j,k,l}. \end{aligned} \quad (20)$$

again with the convention that any negative index is 0, with the appropriate adjustments for maximum multiplicity of infection N , and here also assuming $S \geq 3$ for simplicity of writing out the equations. For $S \leq 2$, subpopulations with $0 < \mathcal{N} \leq 2$ can also gain cells via infection by cell-to-cell transmission (with rate γ).

A phenomenological reinfection term. It turns out (see Section 2) that an additional layer of complexity is needed to describe the experimental data. In the experiments, spatial effects of infecting nearest neighbors are non-negligible. In order to incorporate these effects phenomenologically, we introduce “reinfection” terms in the equations. Namely, we assume that if a cell is infected with a double mutant virus, there is a rate at which it leaves that subpopulation and becomes infected with single mutants C and/or Y. Assuming that $k + l > 0$, we have that the model with the additional ξ terms is given by

$$\begin{aligned} \dot{x}_{i,j,k,l} &= \frac{\beta}{Z} (x_{i-1,j,k,l} B_C + x_{i,j-1,k,l} B_Y + x_{i,j,k-1,l} B_G + x_{i,j,k,l-1} B_W - x_{i,j,k,l} X) \\ &+ \frac{\gamma}{Z} \left(\sum_{\hat{i}+\hat{j}+\hat{k}+\hat{l}=S} x_{i-\hat{i},j-\hat{j},k-\hat{k},l-\hat{l}} G_{\hat{i},\hat{j},\hat{k},\hat{l}} - x_{i,j,k,l} X \right) \\ &+ \xi (x_{i-1,j,k,l}/2 + x_{i,j-1,k,l}/2 - x_{i,j,k,l}) - ax_{i,j,k,l}, \end{aligned} \quad (21)$$

with the appropriate adjustments for latent infection (cells leave the latently infected with G and latently infected with W populations with rate ξ and enter the appropriate coinfecting subpopulation).

2 Fitting procedures and best fit parameters

Based on previously estimated division and death rates for this cell type, we set $r = 0.924 \text{ days}^{-1}$ and $a = 0.2 \text{ days}^{-1}$ [1]. The death rate of infected cells is assumed lower for this in vitro cell

Notation	Description	Units	Status
r	division rate of uninfected cells	days ⁻¹	fixed $r = 0.924$
a	death rate of infected cells	days ⁻¹	fixed $a = 0.2$
β	rate of free virus transmission	days ⁻¹	fitted
γ	rate of synaptic cell-to-cell transmission	days ⁻¹	fitted
ξ	rate of recombinant reinfection	days ⁻¹	fitted
S	number of viruses transmitted per synaptic infection event	NA	fitted
N	maximum multiplicity of infection	NA	fixed $N = 20$
ε	latent infection probability	NA	fitted
ρ	recombination probability	NA	fitted
η	initial percentage of infected cells	NA	fitted
$P(\alpha i, j, k, l)$	probability of infection with strain α from $x_{i,j,k,l}$ cell	NA	NA
Z	total number of cells	NA	NA
X	total number of actively infected cells	NA	NA
f	shaking factor	NA	fixed $f = 1.33$
μ	mutation rate	NA	fixed $\mu = 3 \times 10^{-5}$

Table S1: Description of model parameters. Units are included if applicable.

line compared to estimates obtained from HIV-infected patients in vivo, and is consistent with previously published decline rates of HIV-infected cells in vitro [3]. We denote by $\mathcal{C}$ the percentage of cells infected with only strain C, by $\mathcal{YC}$ the percentage of cells infected with strain Y and strain C (but no other strains), and similarly with other strains, using calligraphic notation for the percentages. We denote $\mathcal{O}$ as the overall percentage of infected cells.

It is important to note that the double mutant recombinant strain W does not glow, and thus cannot be quantified experimentally. Therefore, we must make the appropriate adjustments to all glowing populations when comparing the experimental data to the mathematical model. For instance, experimental data for $\mathcal{C}$ correspond to mathematical model population $\mathcal{C} + \mathcal{CW}$ (as $\mathcal{CW}$ glows as if it were infected only with C) and experimental data for $\mathcal{O}$ correspond to mathematical population $\mathcal{O} - \mathcal{W}$. These adjustments are included in all steps below, where in terms of the mathematical model we redefine for instance $\mathcal{C} := \mathcal{C} + \mathcal{CW}$, and similarly for all other populations.

We further note that (especially for static experiments), the population of infected cells can reach values very close to 100%. For experiments in which this occurs, we ignore all data points where the overall percentage of infected cells starts to decline, as the dynamics are no longer in growth phase (because there are no uninfected cells left to infect).

Based on the experimental setup, we assume that initially there is an equal percentage of cells singly infected with single mutant strain C and singly infected with single mutant strain Y, and that all other cells are uninfected. Since an equal amount of C and Y are initially infected, C and Y are symmetric from a mathematical perspective (as are G and W). Therefore, for the purposes of finding the best fits, we take averages of symmetric populations and use the following six populations for each experiment: $\mathcal{O}$, $(\mathcal{C} + \mathcal{Y})/2$, $\mathcal{YC}$, $\mathcal{G}$, $(\mathcal{GC} + \mathcal{GY})/2$, and $\mathcal{GYC}$.

While the experimentally measured sizes of these populations differ by orders of magnitude, we nonetheless would still like all of them to contribute to determining the parameters. To this end, we normalize the data so that the final data point is always 1 (by simply multiplying the time series for each population by 1 divided by the final data point in that particular time series). Finally, since the populations grow over orders of magnitude during the time-course of the experiments, we

fit the natural log of the data (and thus ignore any 0 data points); this results in a more balanced fit where both early and late values matter.

All fits were completed using the *lsqcurvefit* function in Matlab, which calculates the parameters that minimize the sum of the squared difference between the experimental data and mathematical model over all experimental data points. Solutions of the differential equations were completed using *ode45*, which implements the Runge-Kutta method. We used random initial parameter guesses and repeated the procedure multiple times to ensure that global minima were obtained.

We began with the shaking experiment, since this simpler setting includes only free virus transmission. As the first step, we used the $\mathcal{O}$, $(\mathcal{C} + \mathcal{Y})/2$, and $\mathcal{YC}$ populations to fit the rate of free virus transmission, β , and the initial percent of infected cells, η , using the basic model (without latent infection) given by Equations (9-10). However, we found that it was necessary to include latent infection (model given by Equations (12-20)) because latent infection was needed for achieving a good fit for the coinfecting $\mathcal{YC}$ cells. As illustrated in Figure S1, it is not possible to get a good fit for the single mutant populations ($\mathcal{Y}$, $\mathcal{C}$, and $\mathcal{YC}$) without latent infection (here for simplicity we ignored the small recombinant populations, set $\rho = 0$). We can see that the $\mathcal{YC}$ population is underestimated, and the $(\mathcal{C} + \mathcal{Y})/2$ and $\mathcal{O}$ populations are overestimated. Therefore, the basic model without latent infection and reinfection given by Equations (9-10) is not used in any steps of the data fitting.

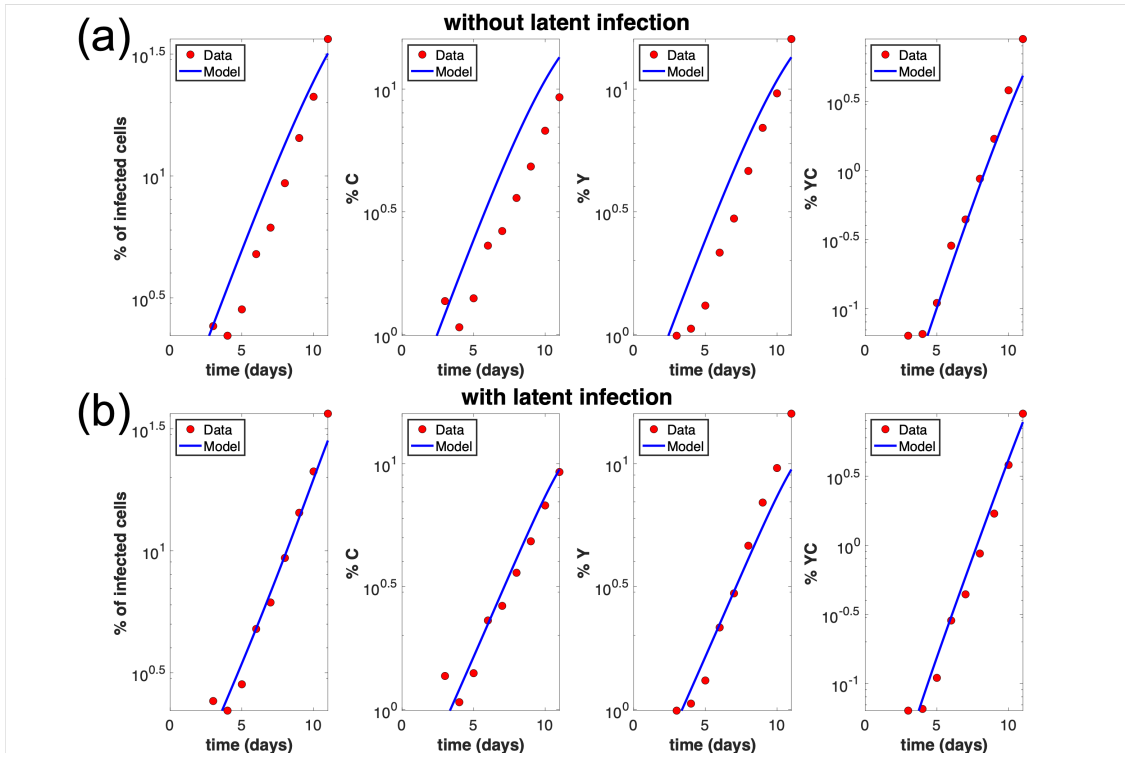


Figure S1: Comparison of best fits for shaking experiment A6 in the presence and absence of latent infection. Including latent infection results in a much better fit. The horizontal axis represents time (days) and the vertical axis represents the percentage of cells in the given category. (a) Best fit without latent infection. Here we use the mathematical model without latent infection given by Equations (9-10), and we see that the percent of cells coinfecting with both Y and C is an underestimate, and the percent of cells infected with only Y (and only C) is an overestimate. (b) Best fit with latent infection. Here we use the mathematical model with latent infection given by Equations (12-20). Parameters for this fit are given in Table S3. Here we get an F-statistic of $\mathcal{F} = 16.4$ and p-value of 0.007.

We can use the F-test for nested model to demonstrate the necessity of the additional level of

complexity by including latent infection (and thus an additional parameter ε). For the experiment shown in Figure S1, we get an F-statistic of $\mathcal{F} = 16.4$, where

$$\mathcal{F} = \frac{\frac{R_1 - R_2}{p_2 - p_1}}{\frac{R_2}{n - p_2}} \quad (22)$$

with n the number of data points, p_i the number of parameters, and R_i the sum of the squared errors in the respective models. Here the model with subscript 1 is the nested model without latent infection given by Equations (9-10) and the model with subscript 2 is the model with latent infection given by Equations (12-20). The null hypothesis is that the more complex model does not provide a significantly better fit than the reduced nested model. The F-statistic is a ratio of two quantities that are expected to be roughly equal under the null hypothesis, which produces an F-statistic of approximately 1. Here the F-statistic is 16.4 and p-value is 0.007, and so we reject the null hypothesis and use the more complex model with latent infection.

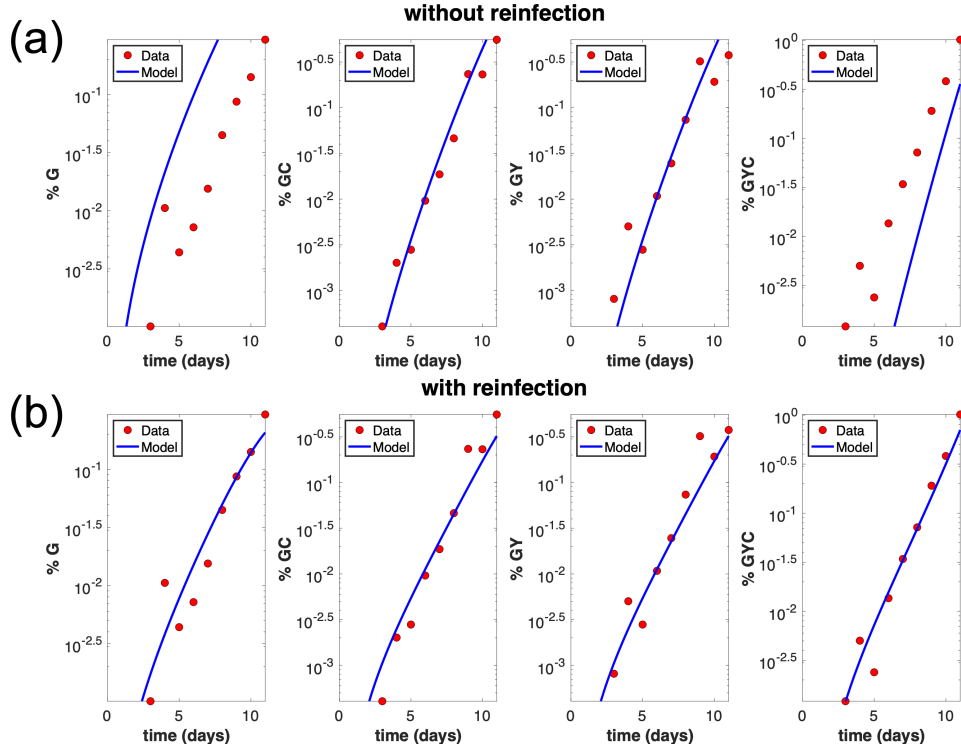


Figure S2: Comparison of best fits for shaking experiment A6 in the presence and absence of reinfection. Including reinfection results in a much better fit. (a) Best fit without reinfection, system (12-20): the percent of cells coinfecting with G, Y, and C is an underestimate, and the percent of cells infected with only G is an overestimate. (b) Best fit with reinfection, system (21). Parameters for this fit are given in Table S3. F-statistic is $\mathcal{F} = 39.3$ and p-value of 0.002.

The next step involved fitting the recombinant populations of $\mathcal{G}$, $(\mathcal{GC} + \mathcal{GY})/2$, and $\mathcal{GYC}$ to estimate the recombination probability, ρ . However, using the model with latent infection but without reinfection, which are given by Equations (12-20), resulted in severe underestimation of $\mathcal{GYC}$ cells, see Figure S2. This is because of the tendency of newly created recombinant G viruses to reinfect the cells that they are produced from (which must contain at minimum a copy of C and a copy of Y). This spatial effect results in higher experimentally observed fractions of $\mathcal{GYC}$. Therefore, we included phenomenological reinfection terms in the ODE model. If these terms are not included, the $\mathcal{GYC}$ population will be an underestimate, and the percent of cells infected with only $\mathcal{G}$ will be an overestimate, see Figure S2. This model includes a phenomenological reinfection term ξ for

recombinant strains G and W, is given in Equation (21). Again, we can use the F-test for nested model to demonstrate the necessity of the additional level of complexity by including reinfection (and thus an additional parameter ξ). For the experiment shown in Figure S2, we get an F-statistic of $\mathcal{F} = 39.3$, where the nested model without reinfection is given by Equations (12-20) and the full model with reinfection is given by Equation (21). This gives a p-value of 0.002, and so we reject the null hypothesis and use the more complex model with reinfection.

Below we describe two fitting procedures that we implemented, and which are based on the above modeling work. The first procedure is what we called a **simultaneous fitting method**, which provides an overall lower error. The second procedure is the **sequential fitting method**, which gives similar parameter estimates (with a slightly larger overall error), but is a much faster algorithm. Additional justification for parameter estimates is given in Section 6.

Simultaneous fitting method. Each experimental pair includes a shaking experiment (only free virus transmission) and a static experiment (both free virus and cell-to-cell transmission). This method uses a corresponding shaking and static experiment and fits them simultaneously for all parameters using the data for $(\mathcal{C} + \mathcal{Y})/2$, $\mathcal{Y}\mathcal{C}$, $\mathcal{G}$, $(\mathcal{G}\mathcal{C} + \mathcal{G}\mathcal{Y})/2$, and $\mathcal{G}\mathcal{Y}\mathcal{C}$. Therefore, we have ten data time series (five from each experiment) and fit them for eight parameters (β , ε , ξ , $\eta_{shaking}$, η_{static} , ρ , γ , and S) using the most complex model with latent infection and reinfection, which is given by Equation (21). For the shaking experiment, we set $\gamma = 0$ as there is only free virus transmission (with rate β). For the corresponding static experiment, we allow $\gamma > 0$ and simultaneously fit the rate of free virus transmission β/f , where f is the shaking factor. This is because shaking increases the efficiency of free virus infection on average by a factor of $f = 1.33$ [1].

This method eliminates the possibility of finding globally sub-optimal fits that exists with the sequential fitting procedure described below. However, this comes at a great computational cost, as each individual experiment (with four virus strains and maximum multiplicity of infection $N = 20$) involves $\binom{20+4}{4} = 10,626$ cell supopulations and differential equations. Fitting two experiments simultaneously for all parameters is thus computationally expensive, and can take several hours per fit.

There are eight pairs of experiments (eight shaking experiments and eight corresponding static experiments). The experiments are spilt into three subsets. Subset A contains three static experiments (A1,A2,A3) and three corresponding shaking experiments (A4,A5,A6). Here, A1 and A4 have the same initial conditions, A2 and A5 have the same initial conditions, and A3 and A6 have the same initial conditions. Similarly, subset B contains three static experiments (B1,B2,B3) and three corresponding shaking experiments (B4,B5,B6), and subset C contains two static experiments (C1,C2) and two corresponding shaking experiments (C3,C4). The best fit parameters using the simultaneous fitting procedure, along with the 95% parameter confidence intervals (based on an asymptotic normal distribution for the parameter estimate), are included in Table S2. Here for all experiments the shaking factor $f = 1.33$, maximum multiplicity of infection $N = 20$, $r = 0.924$, and $a = 0.2$.

Sequential fitting method. A less computationally intensive sequential fitting method is also possible. Rather than the straightforward fitting of all data for all parameters for a pair of corresponding experiments, each step of the hierarchical fitting procedure below describes which infected

experiment	β_{shaking}	ε	ξ	η_{shaking}	η_{static}	ρ	γ	S
A4/A1	2.56 ± 0.66	0.50 ± 0.15	1.99 ± 1.06	0.0022 ± 0.0010	0.00012 ± 0.000089	0.23 ± 0.17	1.23 ± 0.16	4
A5/A2	2.57 ± 0.53	0.48 ± 0.13	2.3 ± 0.96	0.0034 ± 0.0012	0.00061 ± 0.0030	0.18 ± 0.10	1.05 ± 0.11	3
A6/A3	2.35 ± 0.30	0.39 ± 0.10	3.06 ± 0.96	0.0054 ± 0.0013	0.00062 ± 0.00030	0.19 ± 0.073	1.19 ± 0.11	3
B4/B1	1.97 ± 0.51	0.39 ± 0.17	1.18 ± 0.67	0.0027 ± 0.0015	0.00047 ± 0.00029	0.39 ± 0.31	0.78 ± 0.10	3
B5/B2	2.19 ± 0.48	0.45 ± 0.15	1.76 ± 0.77	0.0054 ± 0.0023	0.0015 ± 0.00088	0.20 ± 0.13	0.74 ± 0.10	3
B6/B3	2.15 ± 0.33	0.43 ± 0.12	1.79 ± 0.67	0.0090 ± 0.0032	0.0027 ± 0.0013	0.20 ± 0.095	0.73 ± 0.088	3
C3/C1	1.83 ± 0.55	0.37 ± 0.22	0.49 ± 0.35	0.0045 ± 0.0026	0.00084 ± 0.00053	0.49 ± 0.39	0.80 ± 0.12	3
C4/C2	1.56 ± 0.42	0.23 ± 0.24	0.97 ± 0.67	0.0079 ± 0.0044	0.0014 ± 0.00094	0.50 ± 0.40	0.82 ± 0.13	3

Table S2: Best fit model parameters using the simultaneous fitting method (mathematical model described in Section 1 and fitting procedure described in Section 2) for each pair of experiments. For the shaking experiments, $\gamma = 0$. For the static experiments, $\beta_{\text{static}} = \beta_{\text{shaking}}/f$, where f is the shaking factor. Parameter descriptions and units can be found in Table S1.

cell populations were used to fit which parameters.

We start with the shaking experiment, which only contains free virus transmission. We begin with the $\mathcal{O}$, $(\mathcal{C} + \mathcal{Y})/2$, and $\mathcal{YC}$ populations (set $\rho = \xi = 0$), to fit the rate of free virus transmission, β , the probability of latent infection, ε , and the initial percentage of infected cells, η using the mathematical model with latent infection given by Equations (12-20). We then fix these three parameters (β , ε , and η) and use the remaining populations of $\mathcal{G}$, $(\mathcal{GC} + \mathcal{GY})/2$, and $\mathcal{GYC}$ to fit the recombination probability, ρ , and rate of recombinant reinfection, ξ , using the mathematical model with reinfection given by Equation (21). While ρ and ξ affect the non-recombinant populations, we assume this effect is negligible and so it is included in a separate step after the other parameters have been fixed.

For the corresponding static experiment, we fix the rate of free virus transmission to be β/f and further we fix ε and ξ as what was obtained in the shaking experiment. We then use only the $\mathcal{O}$ population to fit the rate of synaptic transmission, γ , and the initial percentage of infected cells, η . We then finally use the recombinant populations of $\mathcal{G}$, $(\mathcal{GC} + \mathcal{GY})/2$, and $\mathcal{GYC}$ to fit the number of viral copies transferred per synapse, S . We use these populations because changing S has the most significant effect on the smaller recombinant populations. Increasing S increases the number of coinfection events, which increases the percentage of cells infected with $\mathcal{GYC}$ and decreases the percentage of cells infected with $\mathcal{G}$. For all steps in fitting the static experiment, we use the full model with reinfection.

The best fit parameters for this method are included in Table S3 and Table S4. The 95% parameter confidence intervals, calculated at the relevant step of the fitting, are also included.

	β	ε	ξ	η	ρ	corresponding static experiment
experiment A4	2.23 $\pm$ 0.58	0.40 $\pm$ 0.16	1.44 $\pm$ 1.0079	0.0017 $\pm$ 0.00083	0.31 $\pm$ 0.14	experiment A1
experiment A5	1.94 $\pm$ 0.25	0.28 $\pm$ 0.10	2.90 $\pm$ 1.30	0.0037 $\pm$ 0.0010	0.35 $\pm$ 0.10	experiment A2
experiment A6	1.99 $\pm$ 0.15	0.28 $\pm$ 0.063	2.43 $\pm$ 1.055	0.0072 $\pm$ 0.0011	0.28 $\pm$ 0.077	experiment A3
experiment B4	1.63 $\pm$ 0.24	0.22 $\pm$ 0.12	1.27 $\pm$ 0.91	0.0026 $\pm$ 0.0010	0.36 $\pm$ 0.17	experiment B1
experiment B5	1.60 $\pm$ 0.20	0.21 $\pm$ 0.11	1.93 $\pm$ 1.23	0.0081 $\pm$ 0.0027	0.34 $\pm$ 0.14	experiment B2
experiment B6	1.58 $\pm$ 0.20	0.20 $\pm$ 0.12	1.90 $\pm$ 0.92	0.017 $\pm$ 0.0058	0.31 $\pm$ 0.096	experiment B3
experiment C3	1.56 $\pm$ 0.22	0.20 $\pm$ 0.12	0.77 $\pm$ 0.63	0.0034 $\pm$ 0.0011	0.49 $\pm$ 0.27	experiment C1
experiment C4	1.50 $\pm$ 0.28	0.16 $\pm$ 0.16	1.13 $\pm$ 0.72	0.0069 $\pm$ 0.0031	0.40 $\pm$ 0.17	experiment C2

Table S3: Best fit model parameters using the sequential fitting method for each shaking experiment. Parameter descriptions can be found in Table S1.

	β	ε	ξ	η	ρ	γ	S
experiment A1	2.23/ f	0.40	1.44	0.00021 $\pm$ 0.00010	0.31	1.12 $\pm$ 0.067	4
experiment A2	1.94/ f	0.28	2.90	0.00067 $\pm$ 0.00010	0.35	1.06 $\pm$ 0.037	3
experiment A3	1.99/ f	0.28	2.43	0.0018 $\pm$ 0.00041	0.28	1.01 $\pm$ 0.048	2
experiment B1	1.63/ f	0.22	1.27	0.00022 $\pm$ 0.00011	0.36	0.86 $\pm$ 0.064	3
experiment B2	1.60/ f	0.21	1.93	0.0013 $\pm$ 0.0010	0.34	0.78 $\pm$ 0.012	3
experiment B3	1.58/ f	0.20	1.90	0.0037 $\pm$ 0.0011	0.31	0.72 $\pm$ 0.048	3
experiment C1	1.56/ f	0.20	0.77	0.00036 $\pm$ 0.00020	0.49	0.88 $\pm$ 0.078	3
experiment C2	1.50/ f	0.16	1.13	0.00067 $\pm$ 0.00038	0.40	0.91 $\pm$ 0.091	3

Table S4: Best fit model parameters using the sequential fitting method for each corresponding static experiment. Parameter descriptions can be found in Table S1.

Since the simultaneous fitting procedure results in more globally optimal fits, we use its parameters (Table S2) for all figures and calculations in both the main text and Supplementary Materials. While the parameter estimates for the less optimal sequential fitting procedure described below are not identical, the overall error is similar, with the sequential fitting method generally resulting in less error in the shaking experiment and the simultaneous fitting method generally resulting in less error in the static experiment. The best fit curves also look similar, as seen in Figures S3 and S4 which shows the comparison of the two fitting methods along with the experimental data.

3 Relative contribution of free virus and synaptic transmission

Using the simultaneous fitting parameters found in the previous section (Table S2), we estimate the relative contribution of free virus transmission compared to synaptic transmission in driving infection. In contrast to our previous study [1], we now include latent infection (which is described further in Section 1). In order to compare relative contributions of the synaptic and free virus transmission pathways, we note that the rate of free virus transmission, β is on average 1.80 ± 0.31 fold higher than the rate of synaptic transmission (Table S5, column 5).

However, it is important to note that the presence of latent infection modifies the transmission process. In fact, the rate of productive infection (at least, relatively early in the infection process) is better characterized by $\beta(1 - \varepsilon)$. Taking this into account we obtain that the rate of free virus spread is on average 1.06 ± 0.16 fold higher than the rate of synaptic transmission (Table S5, column 6). This is very close to the estimated ratio of 1.1 in [1].

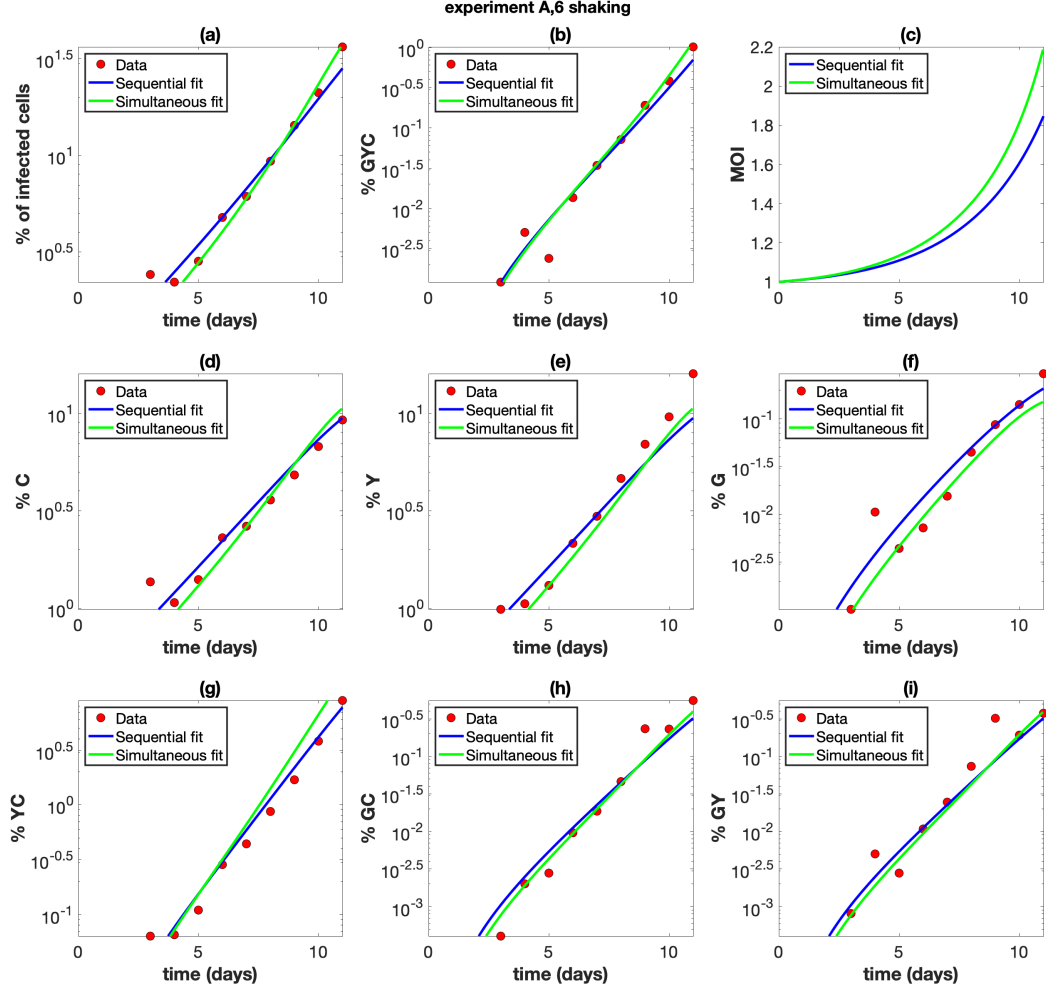


Figure S3: Comparison of the sequential fitting method (blue lines) with the simultaneous fitting method (green lines) for shaking experiment A6. The experimental data are also shown (red dots). The parameters for each fit are given in Tables S3 and S2. (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y.

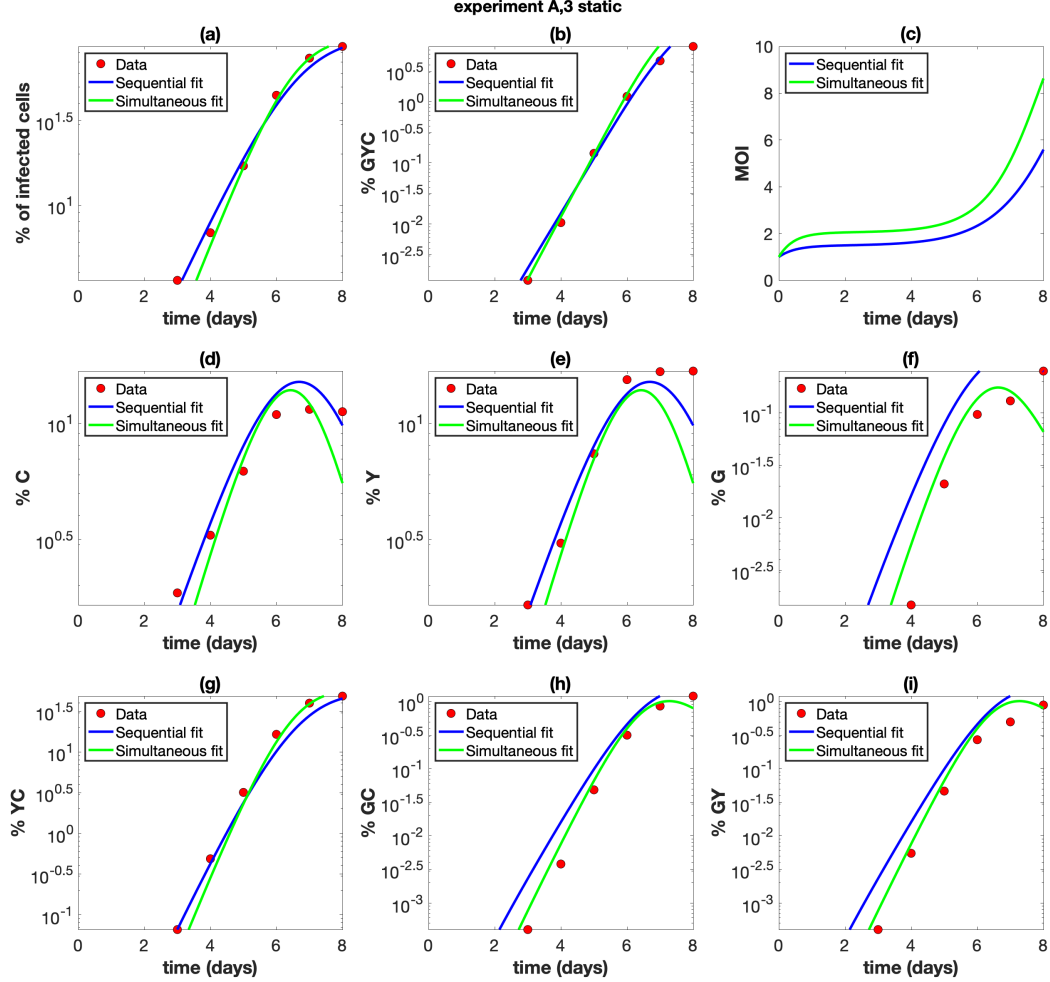


Figure S4: Comparison of the sequential fitting method (blue lines) with the simultaneous fitting method (green lines) for static experiment A₃. The experimental data are also shown (red dots). The parameters for each fit are given in Tables S4 and S2. (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y.

experiment	β_{shaking}	β_{static}	γ	$\beta_{\text{static}}/\gamma$	$\beta_{\text{static}}(1 - \varepsilon)/\gamma$
A4/A1	2.56	$2.56/f$	1.23	1.56	0.78
A5/A2	2.57	$2.57/f$	1.05	1.84	0.96
A6/A3	2.35	$2.35/f$	1.19	1.48	0.91
B4/B1	1.97	$1.97/f$	0.78	1.90	1.16
B5/B2	2.19	$2.19/f$	0.74	2.23	1.22
B6/B3	2.15	$2.15/f$	0.73	2.21	1.26
C3/C1	1.83	$1.83/f$	0.80	1.72	1.08
C4/C2	1.56	$1.56/f$	0.82	1.43	1.10

Table S5: Comparing the contribution of free virus versus synaptic transmission for best fit model parameters using the simultaneous fitting method (Table S2) for each pair of experiments. We have $\beta_{\text{static}} = \beta_{\text{shaking}}/f$, where f is the shaking factor. Parameter descriptions can be found in Table S1.

4 Role of synaptic transmission in evolution

The rate of generation of cells infected with at least one copy of the double mutant strain is a measure of the evolutionary potential. Using the best fit experiment parameters, we can explore the role of the different transmission pathways (free virus versus synaptic cell-to-cell versus both pathways) on the spread of different strains using the mathematical models described in Section 1. In this way, we are able to simulate beyond the many constraints of the physical experiments.

In the main text, we describe evolutionary simulations that correspond to non-expanding populations of susceptible cells. Here we focus on populations that grow exponentially, which corresponds to the experimental conditions. Two types of simulations are explored: (1) We start with a small equal amount of cells singly infected with each of the single mutant strains. (2) We start with a small amount of cells singly infected with the wild type strain (which does not include either of the glowing mutations), and include the process of mutation. In both types of simulations we assume a large initial number of cells (populations of 10^9 are used in the figures). We run the simulation until the number of infected cells is a given fraction of the beginning number of cells, such as 40%. In order to understand the contribution of the different transmission pathways, we investigate the number of double mutant recombinant infected cells under three different conditions: i) with both free virus and synaptic transmission. Here we use the same parameters as in the static experiments. ii) with only free virus transmission. Here we use the same parameters as in the static experiments, but set $\gamma = 0$. iii) only synaptic transmission. Here we use the same parameters as in the static experiments, but set $\beta = \epsilon = \xi = 0$.

In setup (1) above, in the absence of mutations, the double mutant has to be created by recombination between the single mutant strains. This can only occur in cells that are coinfecting with both single mutant strains. The double mutant strain can then spread either by more recombination in single mutant coinfecting cells, or through productive infection of cells infected with the double mutant. As shown in Figure S5(a), a combination of free and synaptic transmission results in the highest number and percentage of cells infected with the double mutant. This is because synaptic transmission promotes cotransmission of both single mutants from coinfecting cells, and because $S = 3$, results in more opportunities for the double mutant to be created via recombination in coinfecting cells. Free virus transmission, however, promotes the spread of the double mutant once it is generated because it can result in cells that are only infected with the double mutant (where it spreads better). Furthermore, including free virus transmission in addition to synaptic transmission helps the two single mutant strains to “meet” in the same cell to begin with, as with only one of the transmission pathways the number of infected cells increases, but the multiplicity of infection is low and the percentage of infected cells actually decreases (see Section 5). Therefore, a combination of both free virus and synaptic transmission is better than either one on their own, even though they

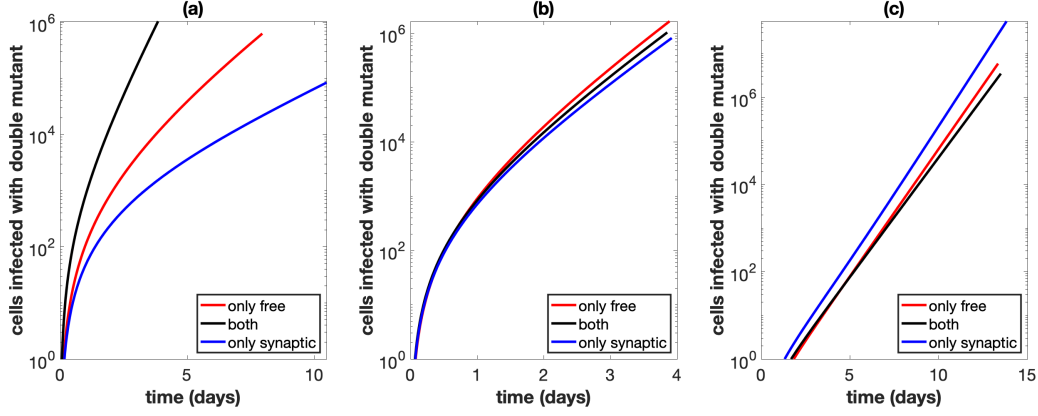


Figure S5: Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount. Parameters are the best fit parameters from static experiment B2, which are listed in Table S2. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

are both non-spatial in the context of the ODE system, and either one on their own results in a lower rate of infection which allows more time to get to the stopping threshold of 4×10^8 infected cells. For Figure S5(b) and (c), as in main text Figure 4(b) and (c), we assume that the overall growth rate is the same across the three cases. This leads to similar dynamics as discussed in the main text.

Next, we investigate the more realistic scenario (2), where we start with a small number of cells singly infected with the wild type, and include all forward and back mutations between the four strains, where for each infection event there is a μ chance to mutate at each of the two point locations. For instance, a wild type virus does not mutate with rate $(1 - \mu)^2$, mutates into a single mutant strain with rate $\mu(1 - \mu)$, and mutates into the double mutant strain with rate μ^2 , where $\mu = 3 \times 10^{-5}$ [4]. Here, we also assume that the self-reinfection terms apply only to the recombinant strain, since we start with the wild type and so most wild type is not generated by recombination in single mutant coinfecting cells. In this setup, the single mutants have to be first created by mutation, and then the double mutant can be created either by mutation or recombination between cells coinfecting with both single mutants (as direct mutation from the wild type to the double mutant is extremely rare). By turning recombination off, we find that most double mutants are created by mutation, which is promoted by synaptic infection, as for each infection event there is $S = 3$ chances for mutation to create the double mutant. As shown in Figure S6(b), when the infected cell population reaches 4×10^8 infected cells, synaptic transmission alone results in higher numbers of the double mutant compared to free virus transmission or a combination of both. The reason for this is two-fold: i) synaptic transmission is better to generate mutations because each transmission event results in $S = 3$ chances for mutation to occur, while free virus only results in 1, and ii) synaptic transmission only has the lowest rate of infection, which allows more time for mutants to be generated. For Figure S6(c) and (d), as in main text Figure 5(c) and (d), we assume that the overall growth rate is the same across the three cases, which again leads to similar dynamics as discussed in the main text.

Overall, it is clear that synaptic transmission plays a vital role in recombinant spread and infection evolution, because of multiple factors:

- Synaptic transmission can repeatedly cotransmit multiple mutant strains once they are to-

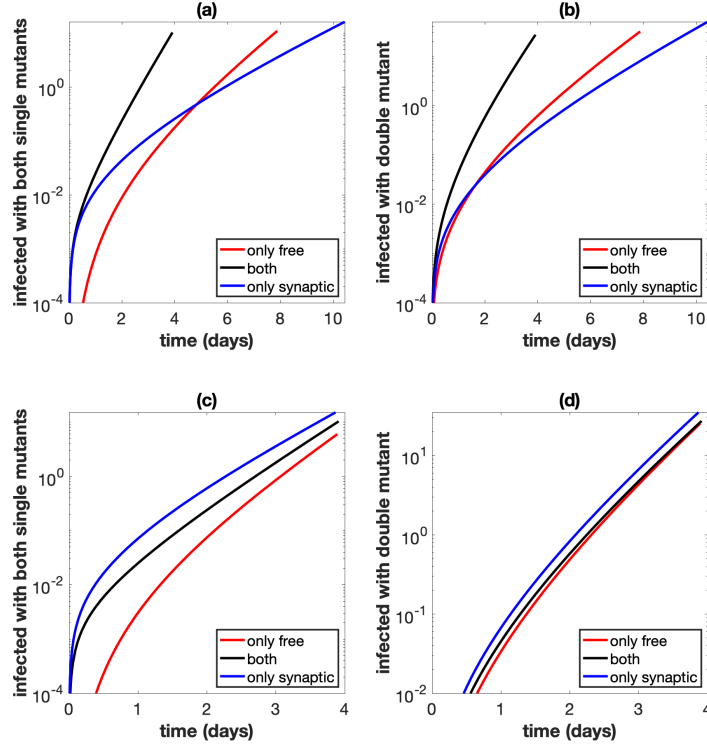


Figure S6: Evolutionary mutant dynamics. Same as Figure S5, but the simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

gether in the same cell. This creates an environment where interactions between different strains can occur with cells (such as recombination).

- Synaptic transmission is beneficial to generate mutant strains, which could have evolutionary advantages. This is because more viral copies are transmitted per infection event, which gives rise to more chances for mutation.
- Similarly, transferring more viral copies per infection event gives rise to more chances for recombination, if the infecting cell is infected with distinct strains that allow for recombination.

It is also worthwhile to note that since cells have to form synapses between each other in order to initiate cell-to-cell transmission, there are likely spatial constraints to synaptic transmission (as we have found with free virus transmission through the experimental data fitting). However, all of the benefits of synaptic transmission for evolution listed above are still true and relevant in the context of spatial cell-to-cell transmission. Furthermore, the ability of synaptic transmission to repeatedly cotransmit multiple mutant strains once they are together in the same cell is amplified in the context of spatial transmission, as spatial clumps of cells all infected with mutants strains can occur [2].

Finally, Figure S7 shows an example of a shaking experiment (A6) and the corresponding static experiment (A3). In panel (e), we see that the static experiment (where both transmission pathways operate) results in higher levels of the recombinant G strain, supporting the idea that synaptic transmission plays an important role in recombinant evolution by promoting recombinant generation through repeated cotransmission of single-mutant viruses and also by increasing the overall rate of virus growth. However, in the experimental setup it is difficult to modulate the rate of infection and isolate cell-to-cell transmission (by turning off free virus transmission) and so we use the mathematical model parameterized by the experimental data in order to further study recombinant dynamics.

5 Further theoretical considerations

Disregarding latent infection, we have that the basic reproductive ratio R_0 for the mathematical model is $\frac{\beta+\gamma}{a}$. If $R_0 > 1$ the number of infected cells will grow, and if $R_0 < 1$, then the number of infected cells will decline. However, to compare to the experiments, we are also concerned with the dynamics of the percentage of infected cells. We define the “reproductive ratio” for the percentage of infected cells, $R_{\%}$, as $R_{\%} = \frac{\beta+\gamma}{a+r}$. That is if $R_{\%} > 1$, the percentage of infected cells will increase, eventually reach 100%, and then all infected cells will eventually die. If $R_{\%} < 1$, the percentage of infected cells will decline and approach 0%.

Since the reproductive rate of uninfected cells is non-negative ($r \geq 0$), we have that $R_{\%} \leq R_0$. The infection dynamics can then be split into three cases:

- $R_{\%} \leq R_0 < 1$. Here, the infected cells will die out, and the number of uninfected cells will continue to grow toward infinity.
- $R_{\%} < 1 < R_0$. Here, the number of infected cells will increase, but the percentage of infected cells will decrease, as the number of uninfected cells will increase faster than the number of infected cells. So the number of uninfected and infected cells will approach infinity, and the percentage of infected cells will approach 0%.
- $1 < R_{\%} \leq R_0$. Here, the percentage of infected cells will increase, eventually reaching 100%, and then all cells will die.

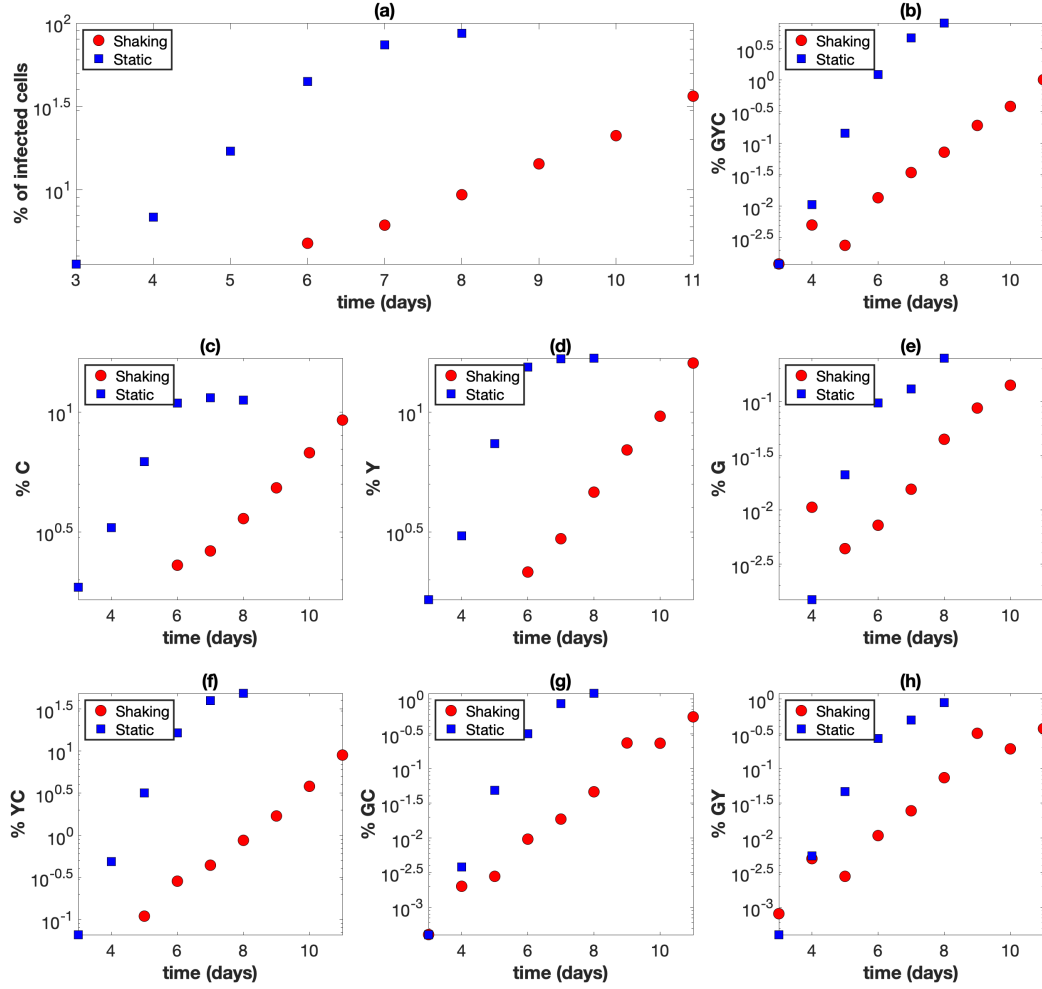


Figure S7: Shaking experiment A6 (red circles) and corresponding static experiment A3 (blue squares). The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The percentage of cells infected with at least one copy of C. (d) The percentage of cells infected with at least one copy of Y. (e) The percentage of cells infected with at least one copy of G. (f) The percentage of cells infected with at least one copy of Y and C. (g) The percentage of cells infected with at least one copy of G and C. (h) The percentage of cells infected with at least one copy of G and Y.

We can also modify the model in order to introduce an infection steady state. This allows us to test model predictions with theoretically determined steady states from the differential equations. To do this, we modify the assumptions on the uninfected cells. Instead of assuming that uninfected cells reproduce with rate r , we assume that there is a constant influx of uninfected cells with rate λ , and that uninfected cells die with rate d . If we let x denote the number of uninfected cells, then in the model we replace the rx term with $\lambda - dx$.

Let y denote the total number of infected cells. The model (grouping all infected cells together) is then

$$\dot{x} = \lambda - \frac{(\beta + \gamma)xy}{x + y} - dx \quad (23)$$

$$\dot{y} = \frac{(\beta + \gamma)xy}{x + y} - ay \quad (24)$$

The infection free steady state is given by

$$x = \frac{\lambda}{d}, \quad y = 0, \quad (25)$$

and the infection steady state is given by

$$x = \frac{\lambda}{\beta + \gamma - a + d}, \quad y = \frac{\lambda(\beta + \gamma - a)}{a(\beta + \gamma - a + d)}. \quad (26)$$

We then verify that our simulations for each of the different transmission strategies results in the correct steady state. An example of this can be seen in Figure S8.

6 Further justification of parameter estimates

We have performed further analysis to justify our parameter estimates and ensure the reporting of global minima. Because of the high dimension (number of parameters being fitted) it is not computationally feasible to calculate the error for a fine mesh of points within reasonable parameter ranges and pick the parameters that correspond to the lowest error. Therefore, we employed the following strategy.

We used multiple fitting algorithms and multiple random initial parameter guesses to estimate best fit parameters. We implemented the Levenberg-Marquardt Algorithm, which uses an approximate Gauss-Newton direction to approximate the minimum. Depending on the range of the initial parameter guesses, this method returns the same best fit parameters about 85% of the time, and either gets stuck at a local minimum or does not converge the other 15% of the time. For example, Figure S9 shows a plot of random initial parameter guesses for Levenberg-Marquardt Algorithm and where they converge to for two parameters (β, ε) for experiment A6 (here we use the sequential fitting method because of the lower computational cost). The random initial guesses that converge are denoted in blue and the guesses that either do not converge or get stuck at a local minimum (with higher error) are denoted in green. The best fit parameters are denoted by the large red dot. If we zoom in on the best fit parameters (red dot), then a higher percentage of initial parameter guesses within that range will converge to it, and if we zoom out then a lower percentage of initial parameter guesses within that range will converge to the best fit parameters.

Further, we have also implemented the Trust-Region-Reflective Least Squares Algorithm, which is a robust global convergence algorithm based on trust regions (Figure S10). For this method, for 10^3 random initial guesses within a reasonable range (i.e. $\beta \in [0, 10]$) the algorithm converged to

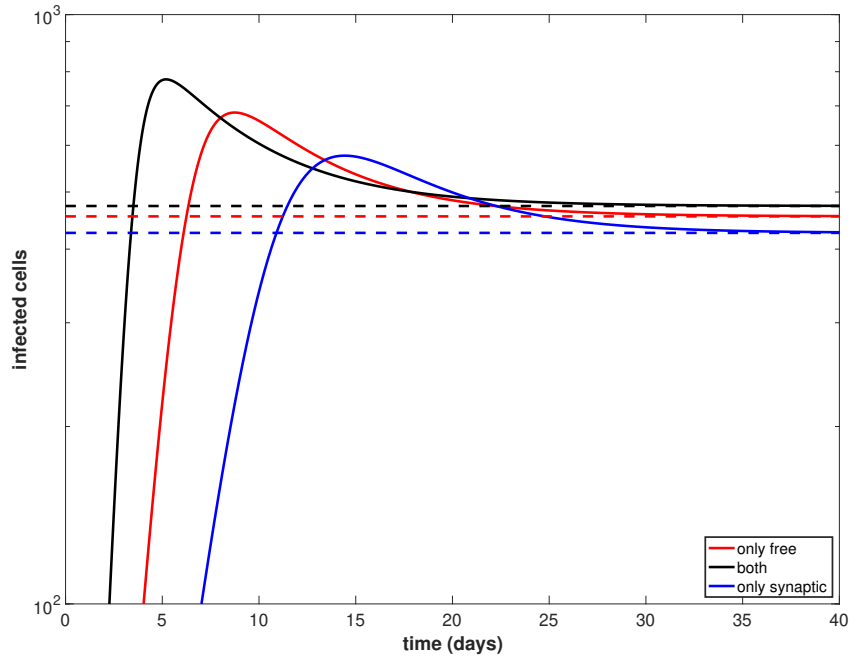


Figure S8: An example of convergence to the theoretical steady state when rx is replaced by $\lambda - dx$ in the uninfected cell equation. Parameters are as in Figures S5 and S6 (best fit parameters from static experiment B2, which are listed in Table S2), except we turn latent infection off and thus set $\varepsilon = 0$. Further, we arbitrarily set $\lambda = 100$ and $d = 0.1$. For each case, the solid line is the simulation of the model and the dashed line is the theoretical prediction of the steady state. The black lines represent both free virus and synaptic transmission, the red lines represent only free virus transmission, and the blue lines represent only synaptic transmission.

the same global minimum 100% of the time (again using the sequential fitting method because of computational cost). If we expand the range large enough (i.e. $\beta \in [0, 100]$), then it is possible for the algorithm to report different parameter estimates and get stuck at a local minimum with much higher error. This, along with reasonable confidence intervals and bands, helps to justify our parameter estimates.

Finally, we can visualize the error landscape with a two-dimension projection by fixing all but two parameters and generating a heatmap. Figures S11 and S12 show the log of the error for the simultaneous fitting of corresponding static experiment B2 and shaking experiment B5 and a range of values for specific parameter pairs (β, ε) and (ρ, ξ) while all other parameters are fixed. Here, we can calculate the error for a fine mesh of points and find the global minimum, which matches the results from the fitting algorithms. While this approach greatly reduces the dimension of the problem and is not a proof that global minima are obtained by the fitting algorithms, it provides further evidence for our parameter estimates.

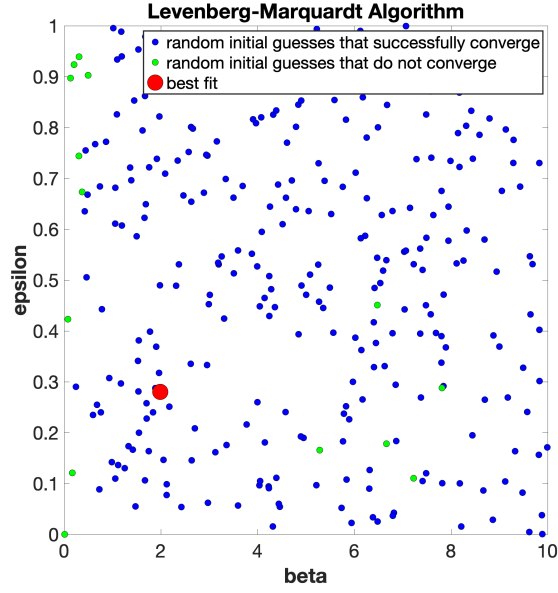


Figure S9: Plot of 300 random initial parameter guesses for the Levenberg-Marquardt Algorithm and where they converge to for two parameters (β, ϵ) for experiment A6 (sequential fitting procedure, see Table S3). The random initial guesses that converge are denoted in blue. The random initial guesses that either do not converge or get stuck at a local minimum are denoted in green. The best fit parameters are denoted by the large red dot.

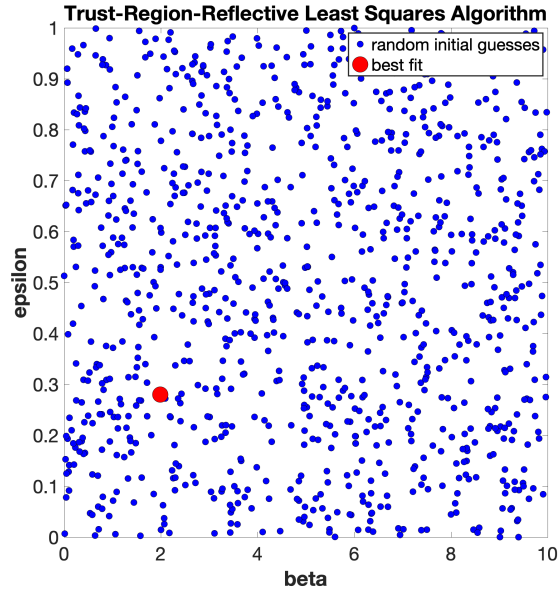


Figure S10: Plot of 10^3 random initial parameter guesses for the Trust-Region-Reflective Least Squares Algorithm and where they converge to for two parameters (β, ϵ) for experiment A6 (sequential fitting procedure, see Table S3). The random initial guesses that converge are denoted in blue. The best fit parameters are denoted by the large red dot. For the given range $(\beta \in [0, 10])$ all initial guesses converge to the best fit parameters, however it is possible to extend the range for β such that this is no longer the case.

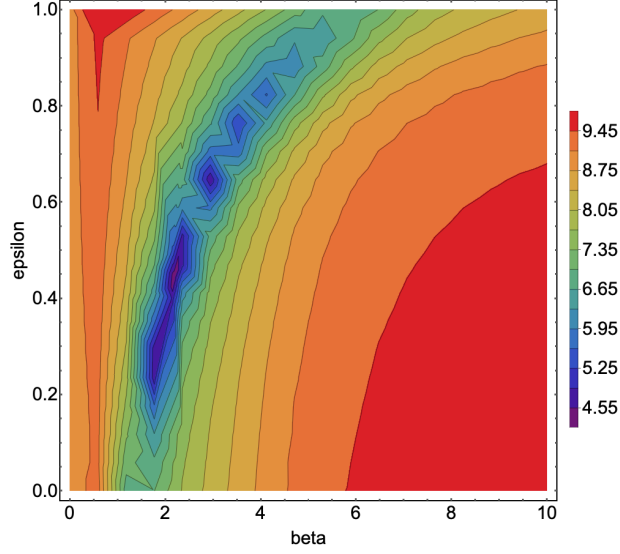


Figure S11: Heat plot of log error for the simultaneous fitting of corresponding static experiment B2 and shaking experiment B5 when all parameters other than β and ϵ are fixed. The horizontal axis represents β (the rate of free virus transmission) and the vertical axis represents ϵ (latent infection probability). The global minimum (which matches what is found by the fitting procedure, see Table S2) is at about (2.19,0.45).

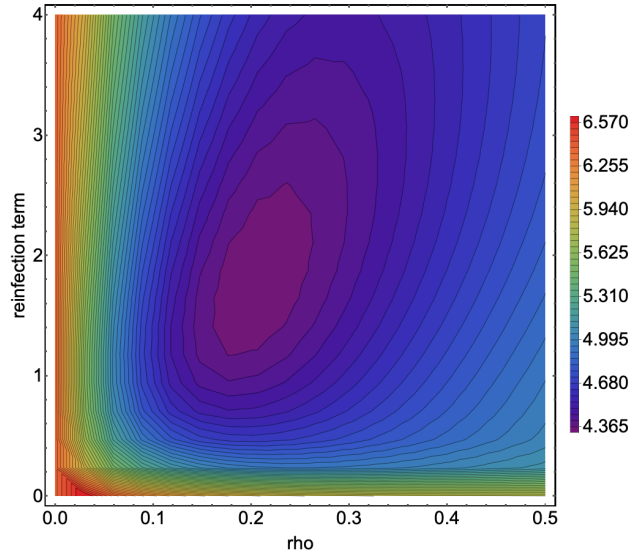


Figure S12: Heat plot of log error for the simultaneous fitting of corresponding static experiment B2 and shaking experiment B5 when all parameters other than ρ and ξ are fixed. The horizontal axis represents ρ (probability of recombination) and the vertical axis represents ξ (refection term). The global minimum (which matches what is found by the fitting procedure, see Table S2) is at about (0.20,1.76).

7 Best fit parameters and plots

The fits (using the simultaneous fitting procedure) are given in Figures S13-S28. We present each experiment with pointwise (non-simultaneous) functional confidence bands (also obtained using the simultaneous fitting procedure). As the plots are on a log scale, negative lower bounds to the confidence interval are not shown. Each of the eight pairs of experiments is presented with the shaking experiment first, followed by the corresponding static experiment.

8 Comparing the effect of viral transmission pathways on recombinant dynamics for all best fit parameter values

In this section, as in the main text, we examine the effect of the viral transmission pathways on recombinant dynamics. Therefore, we include figures that are similar to Figures 4 and 5 in the main text, for best fit parameters obtained for each of the eight pairs of experiments. Figures S29-S36 are comparable to Figure 4 in the main text, and Figure S37-S44 are comparable to Figure 5 in the main text. The parameters are obtained by the simultaneous fitting method.

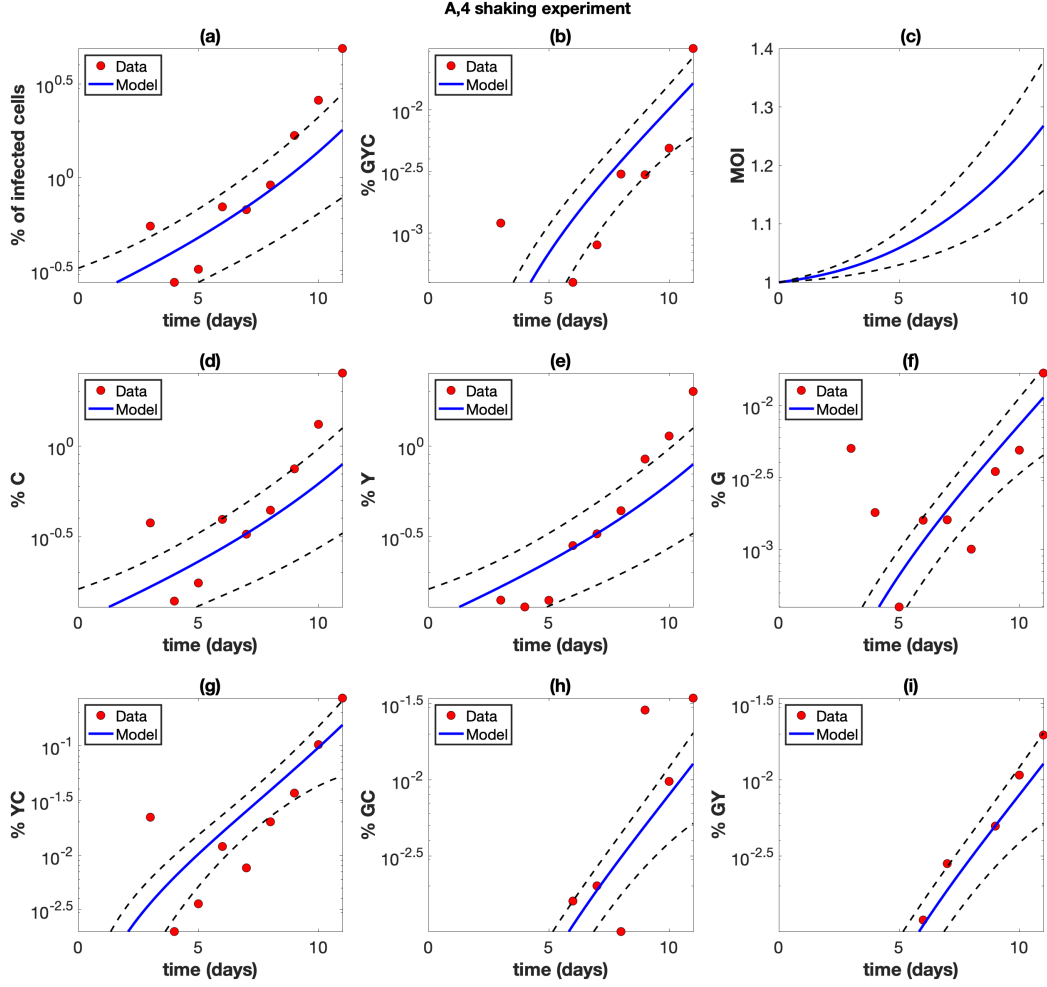


Figure S13: Shaking experiment A4. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

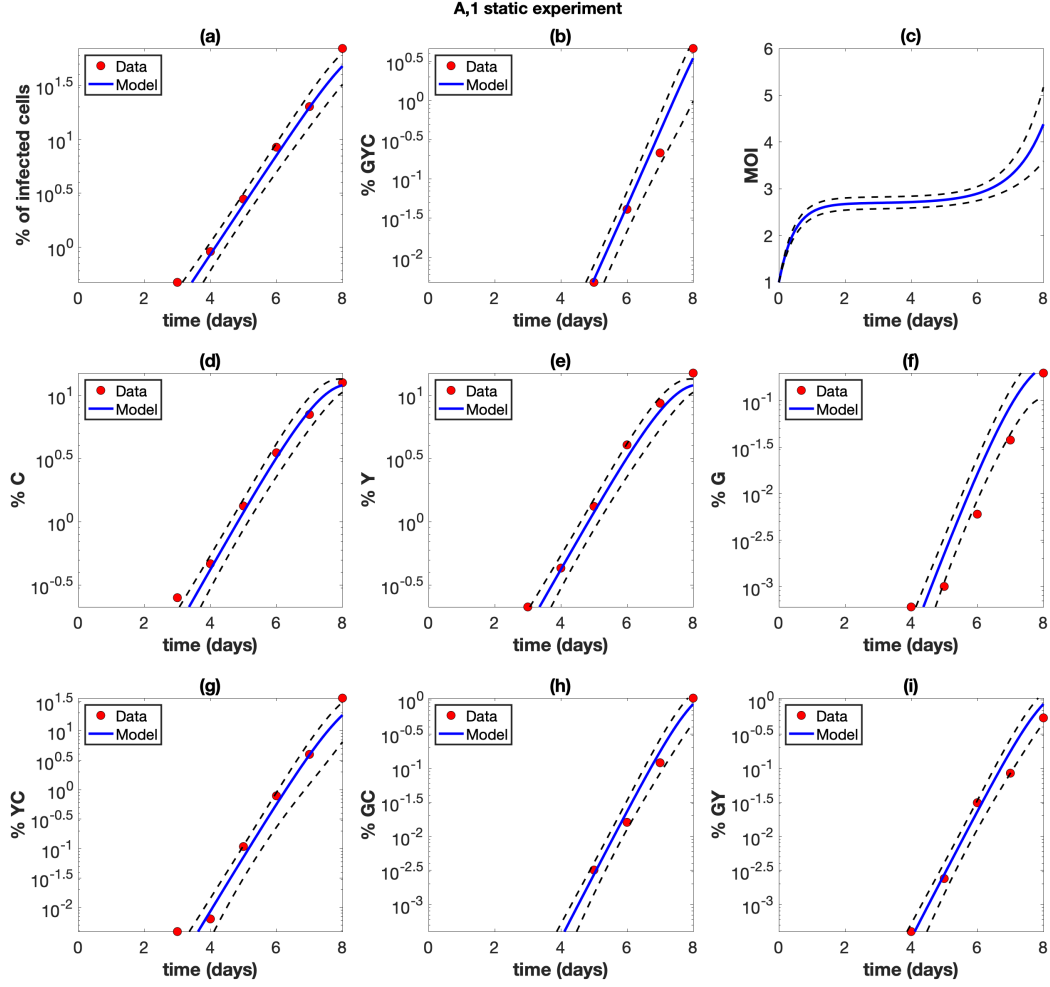


Figure S14: Static experiment A1. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

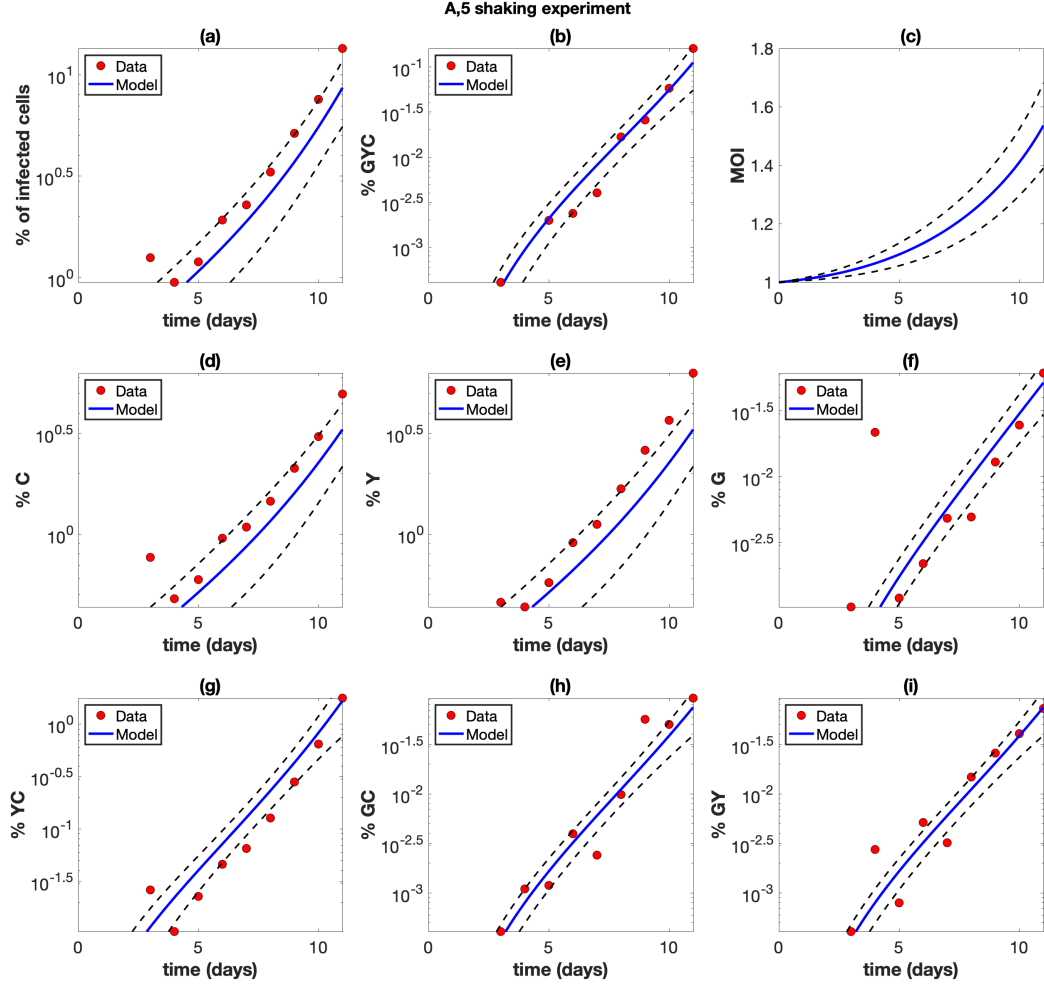


Figure S15: Shaking experiment A5. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

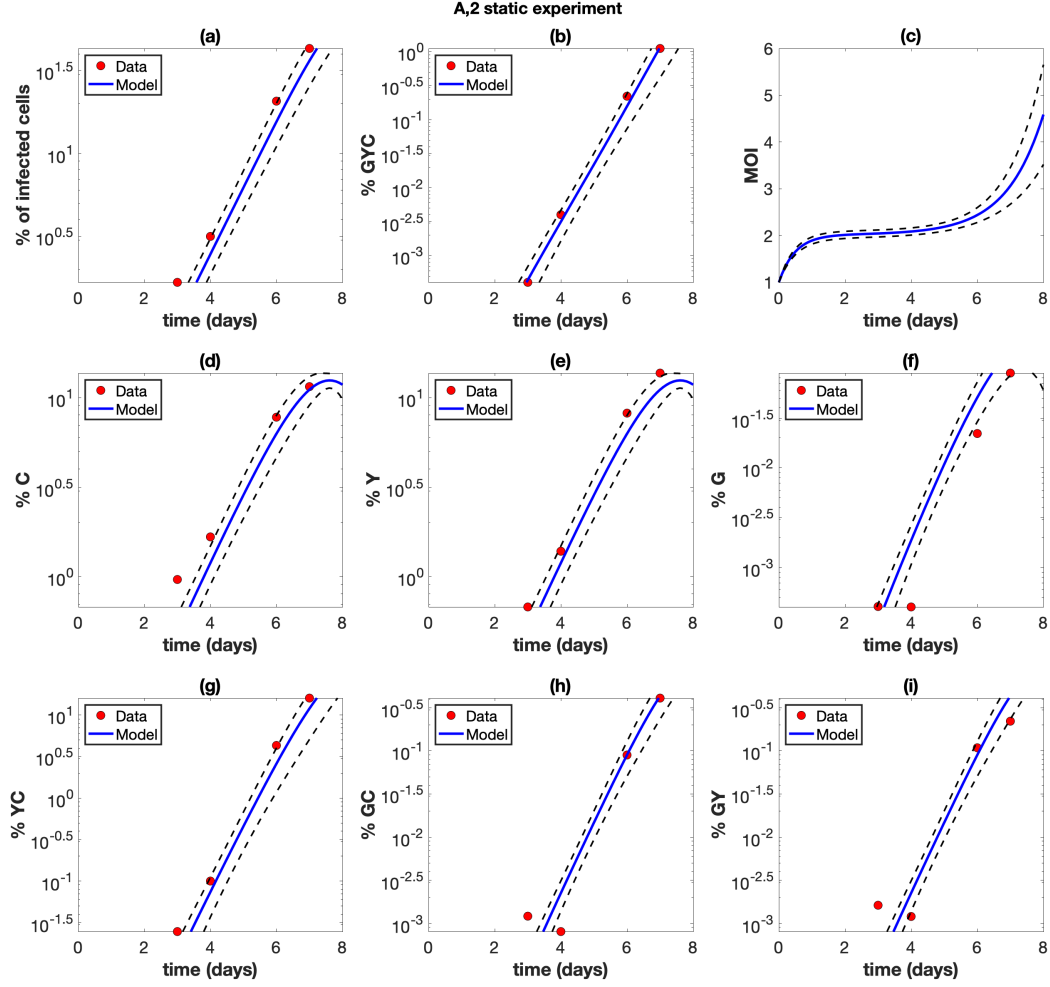


Figure S16: Static experiment A2. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

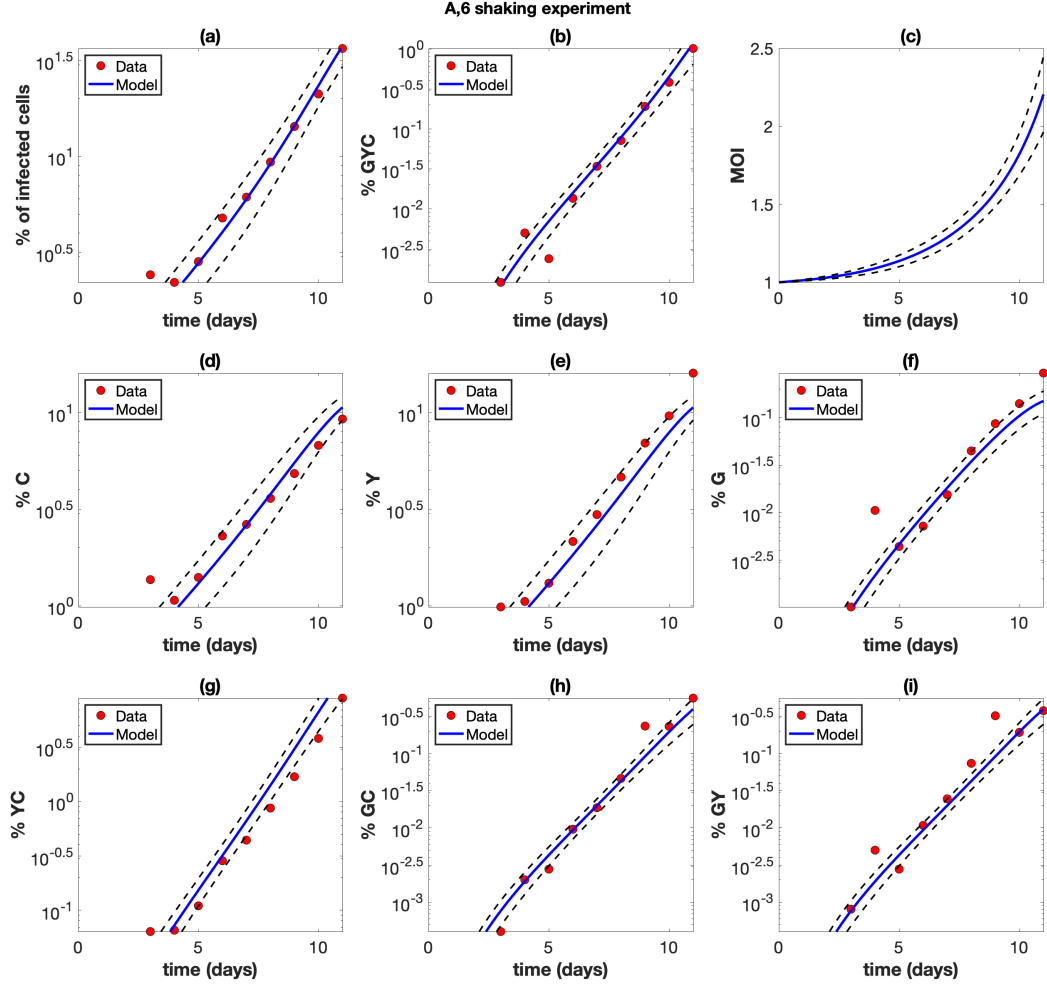


Figure S17: Shaking experiment A6. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

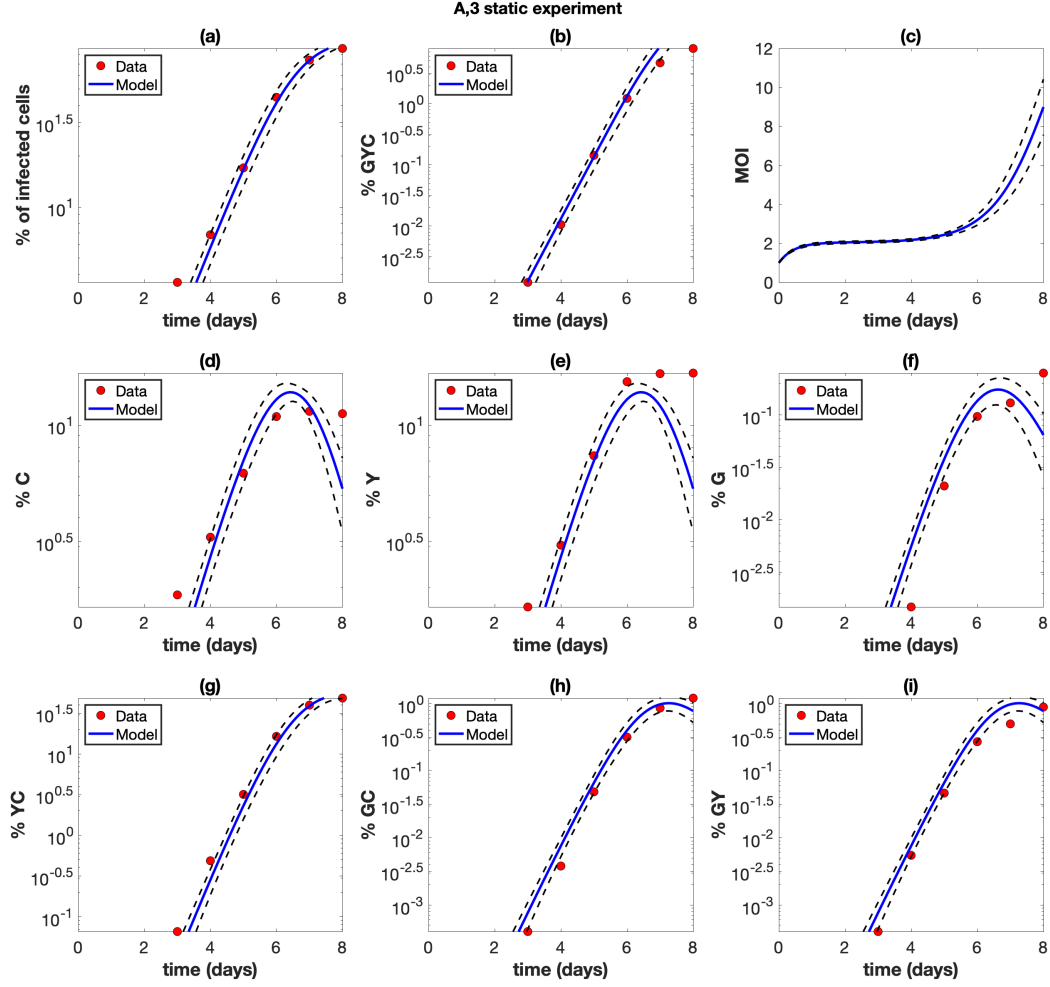


Figure S18: Static experiment A3. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

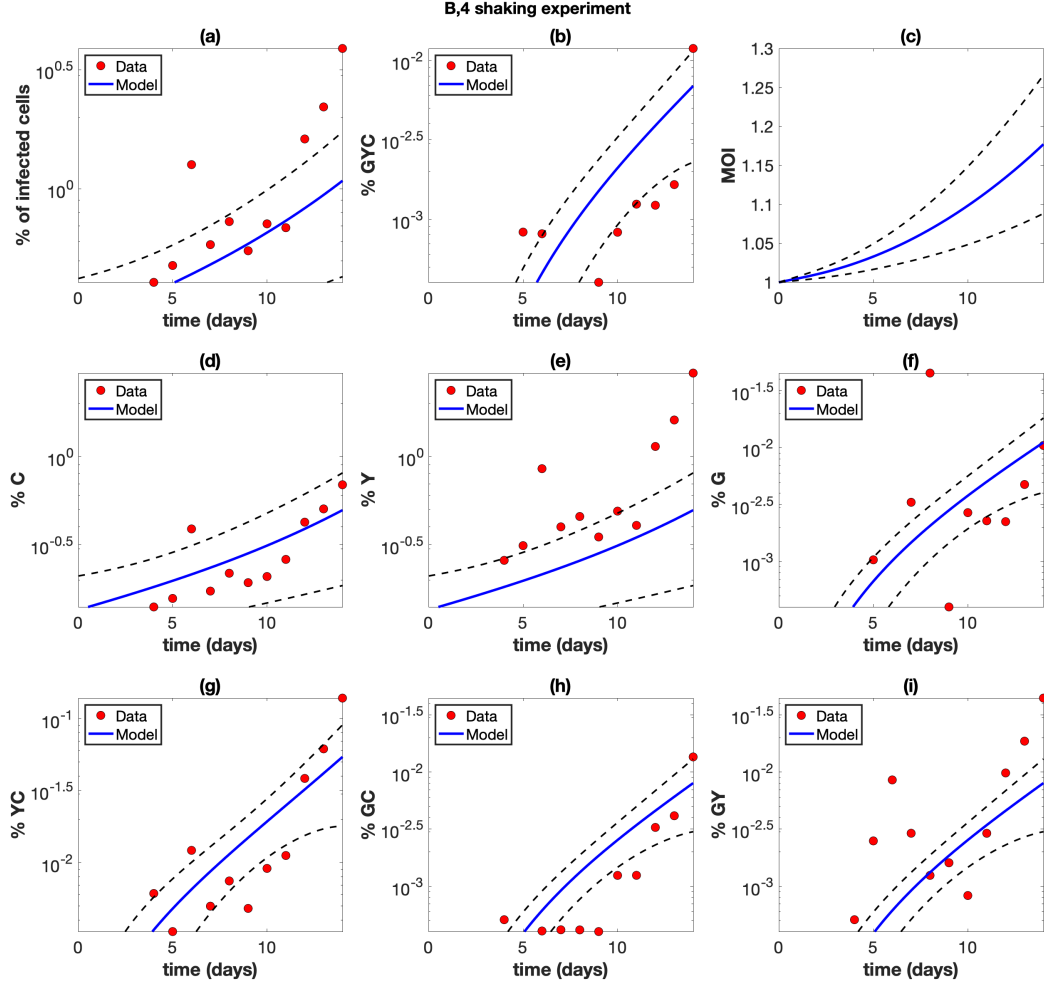


Figure S19: Shaking experiment B4. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

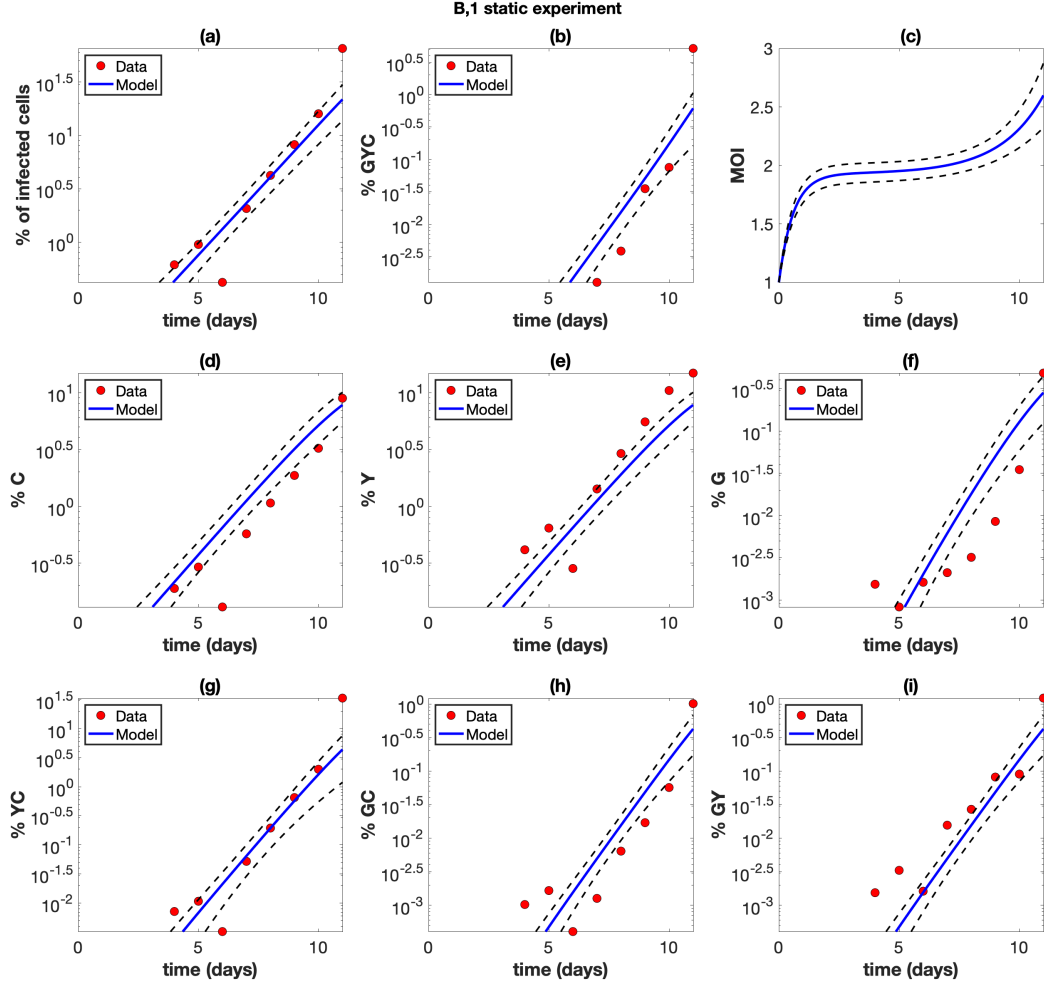


Figure S20: Static experiment B1. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

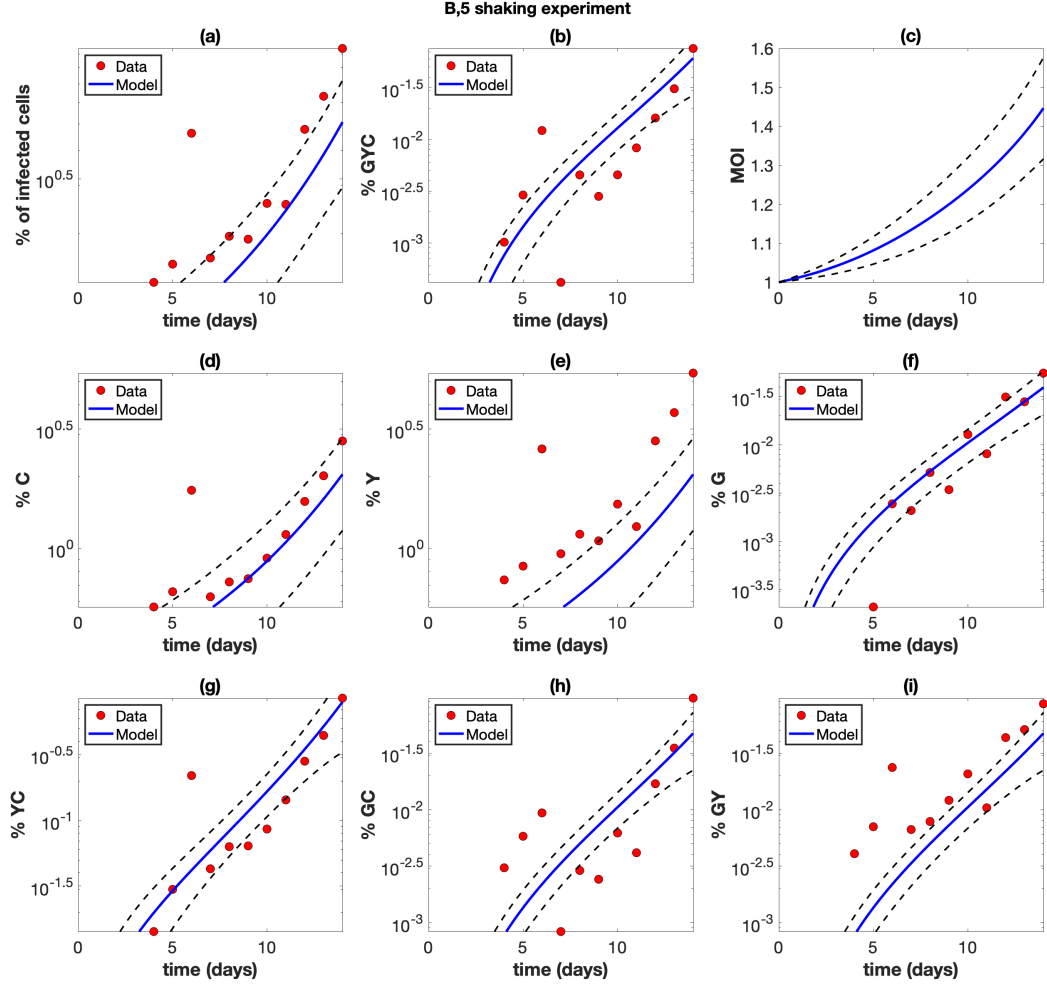


Figure S21: Shaking experiment B5. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

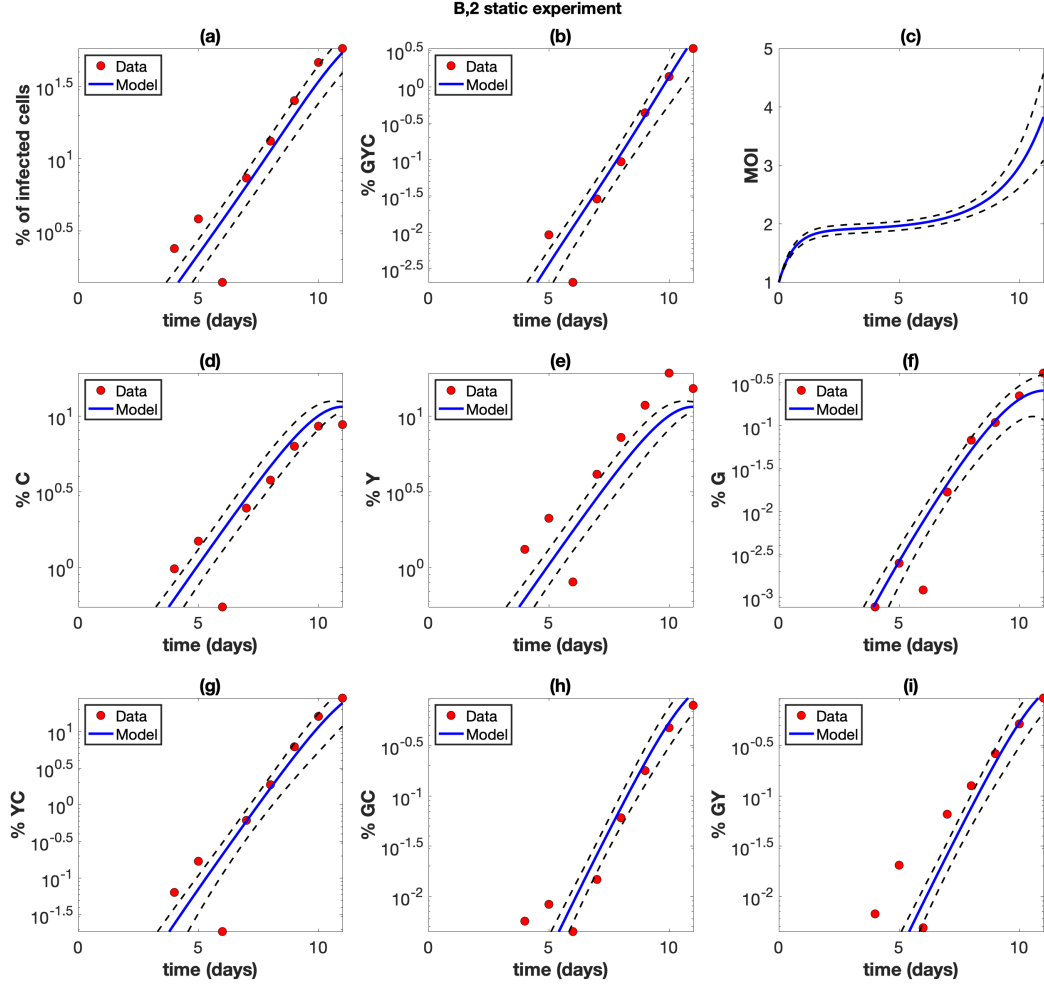


Figure S22: Static experiment B2. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

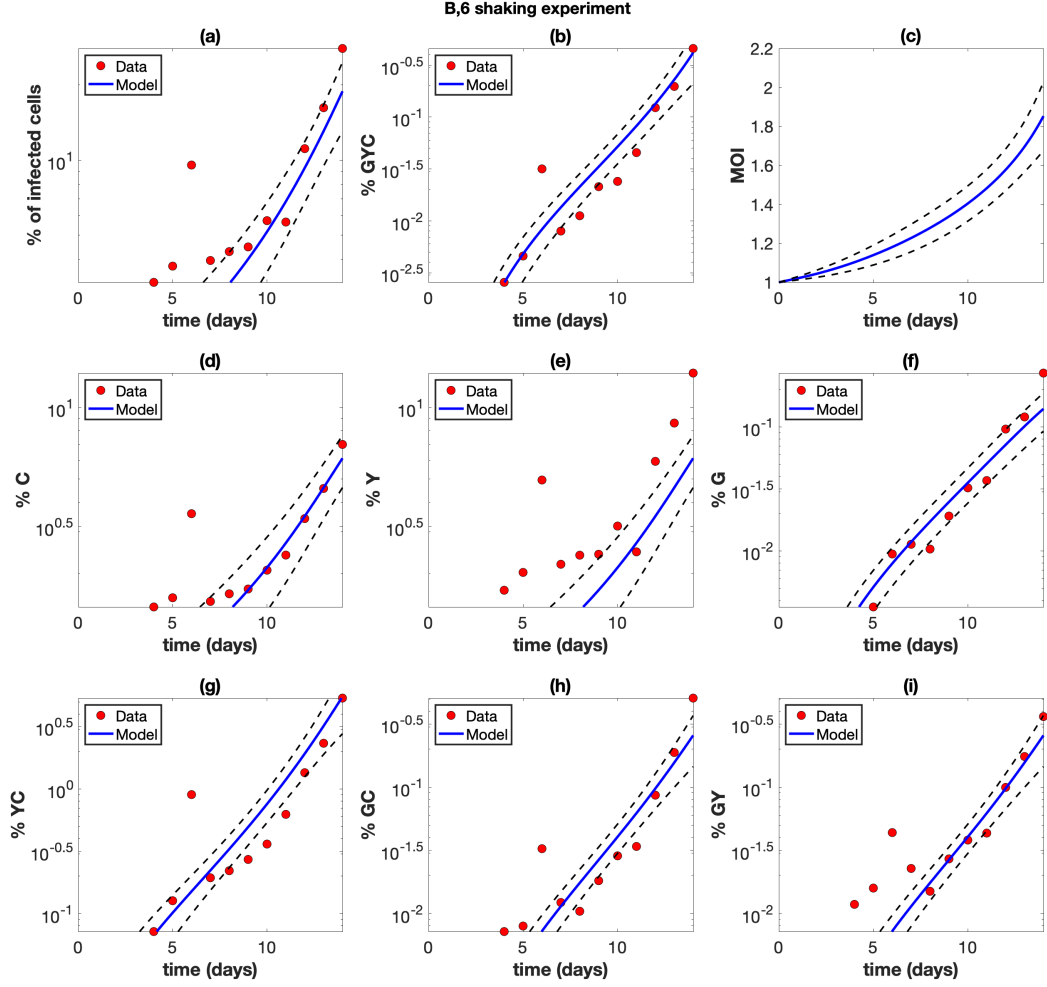


Figure S23: Shaking experiment B6. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

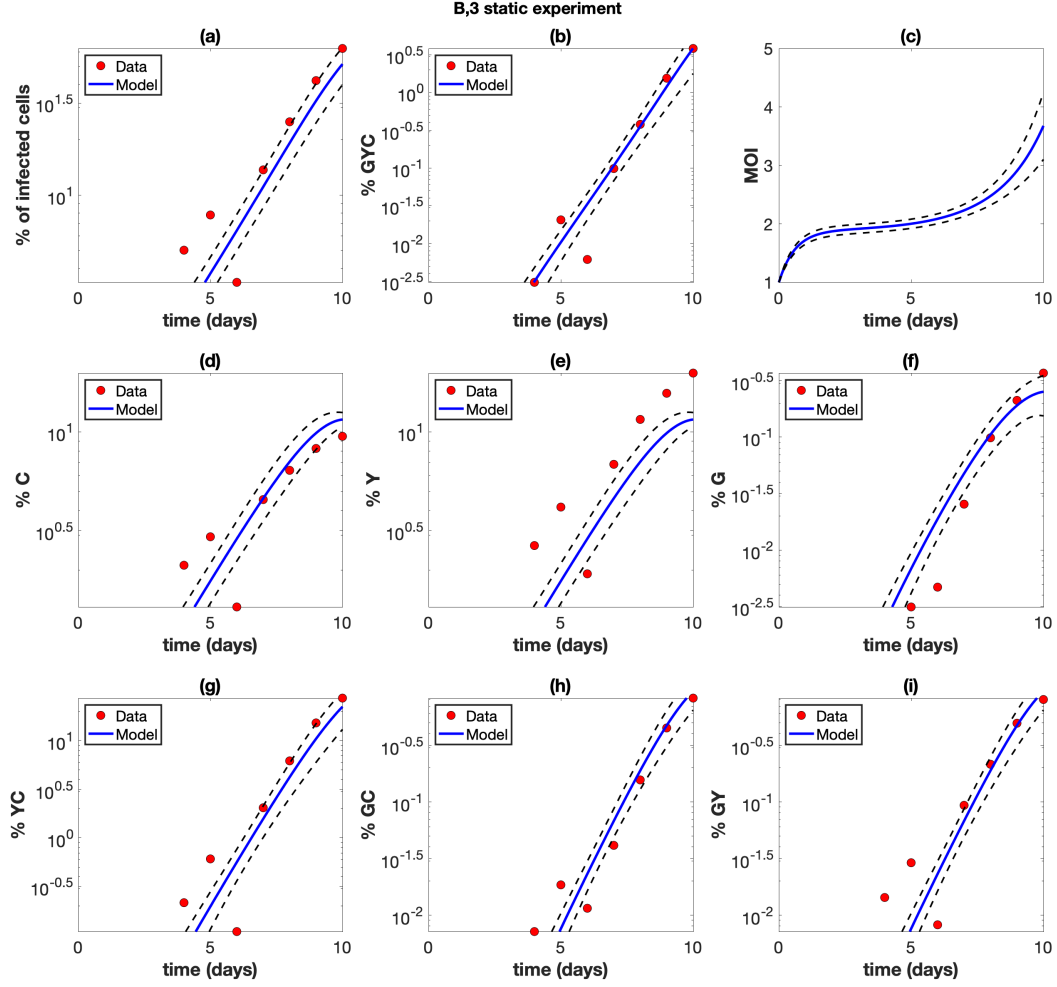


Figure S24: Static experiment B3. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

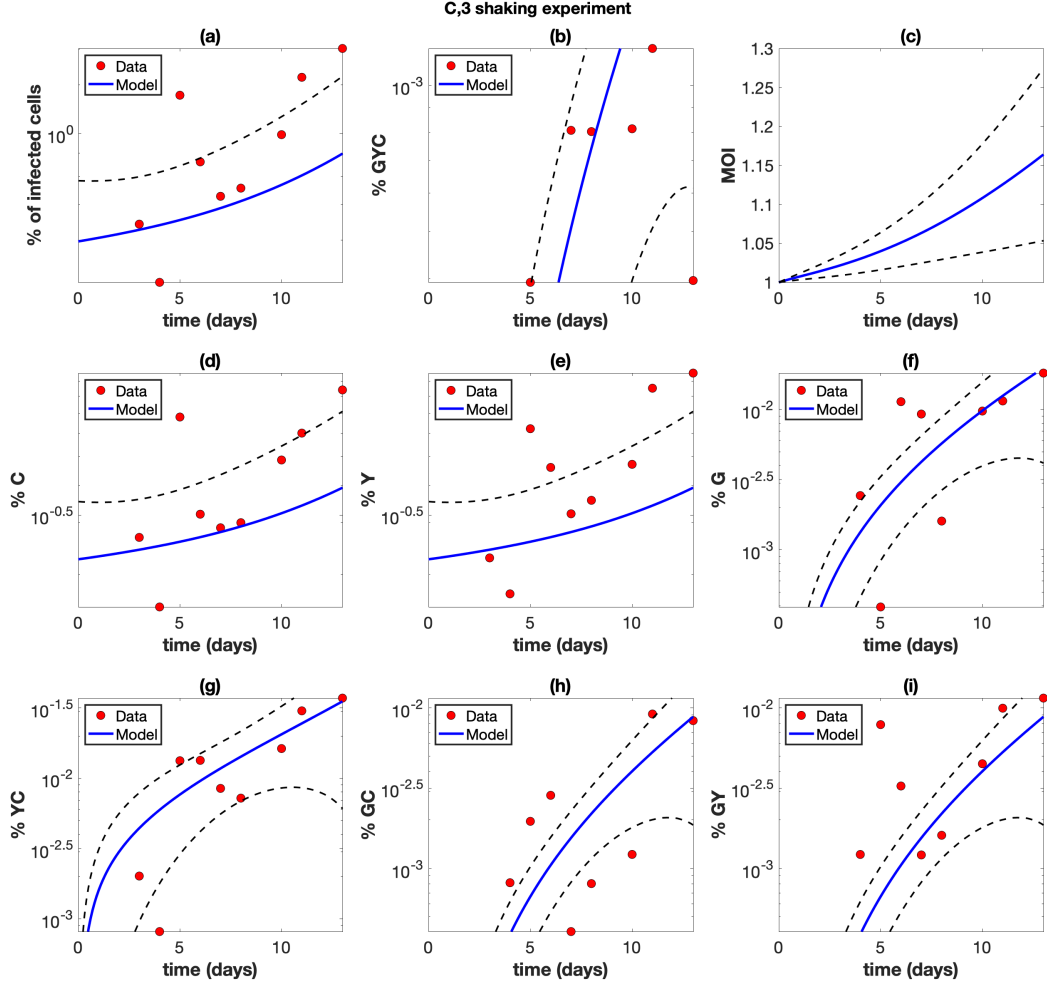


Figure S25: Shaking experiment C3. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

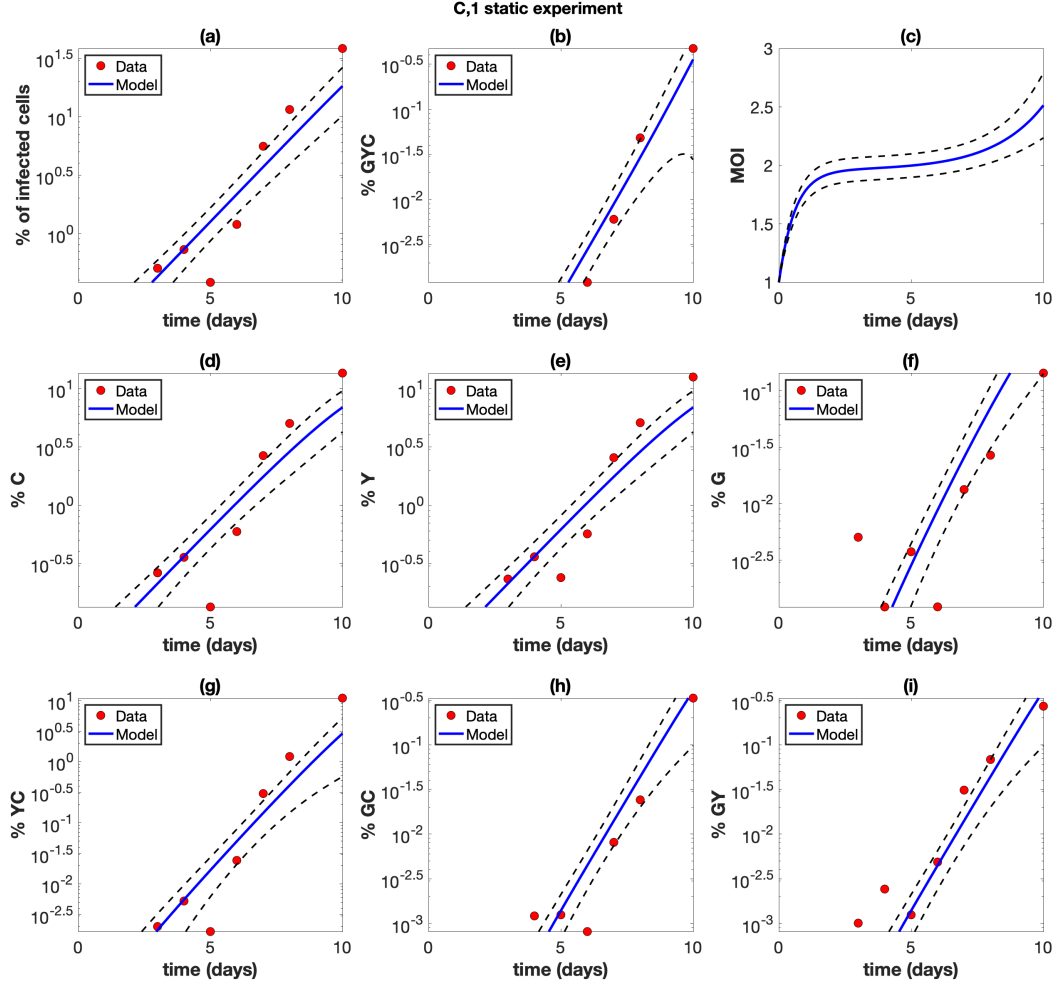


Figure S26: Static experiment C1. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

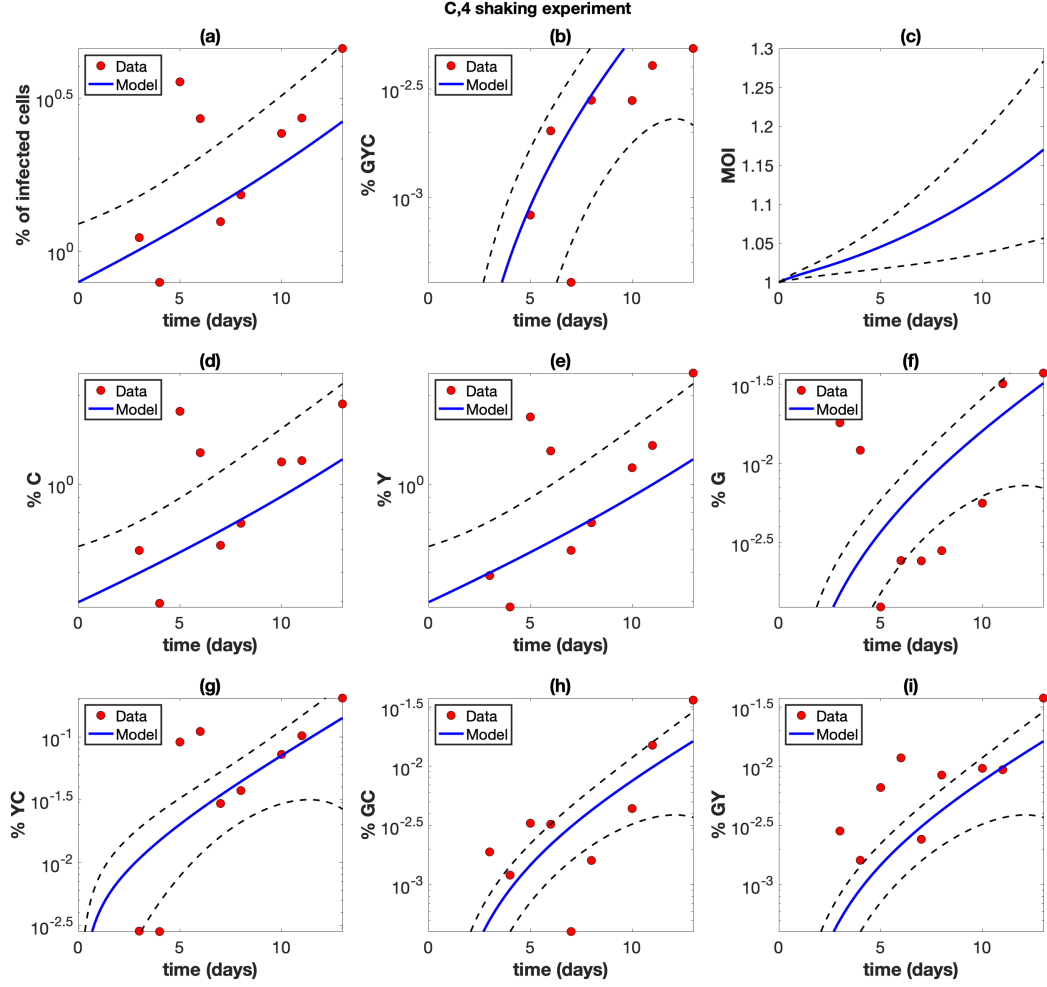


Figure S27: Shaking experiment C4. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

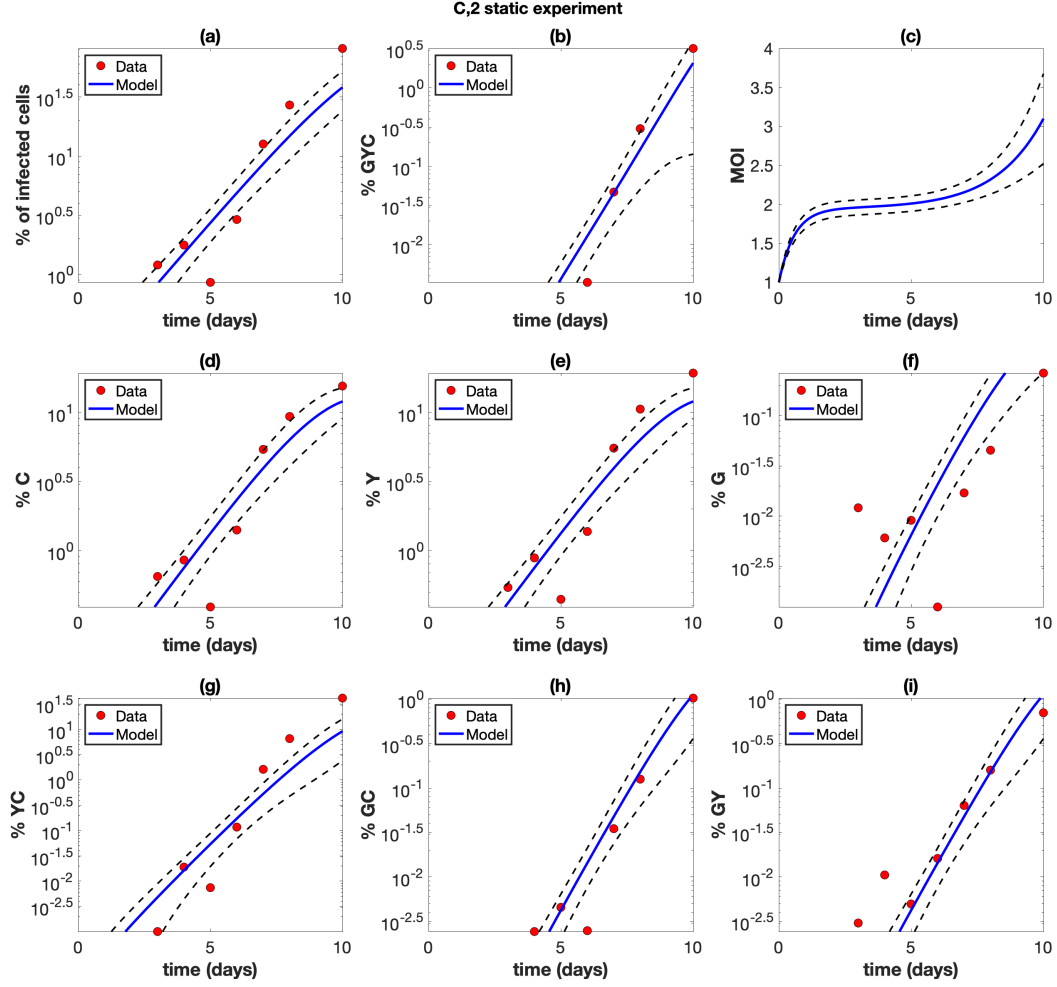


Figure S28: Static experiment C2. The experimental data (red circles) are presented with best fit curves from the model (blue lines). The mathematical model is described in Section 1 and the fitting procedure is described in Section 2. Best fit parameters are included in Table S2. The horizontal axis represents time (days). (a) The overall percentage of infected cells. (b) The percentage of cells infected with at least one copy of G, Y, and C. (c) The average multiplicity of infection (MOI) over all infected cells. (d) The percentage of cells infected with at least one copy of C. (e) The percentage of cells infected with at least one copy of Y. (f) The percentage of cells infected with at least one copy of G. (g) The percentage of cells infected with at least one copy of Y and C. (h) The percentage of cells infected with at least one copy of G and C. (i) The percentage of cells infected with at least one copy of G and Y. The dashed black lines represent the pointwise 95% prediction confidence bounds.

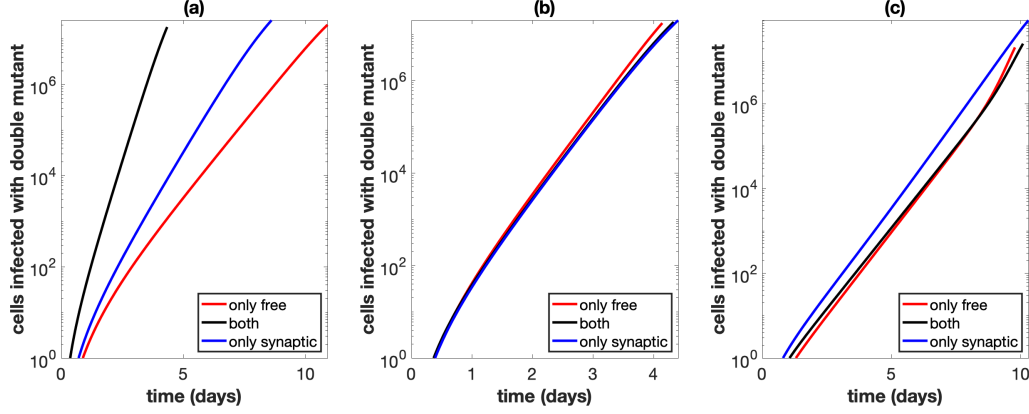


Figure S29: Similar to Figure 4 in the main text, using best fit parameters from experiment A1. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

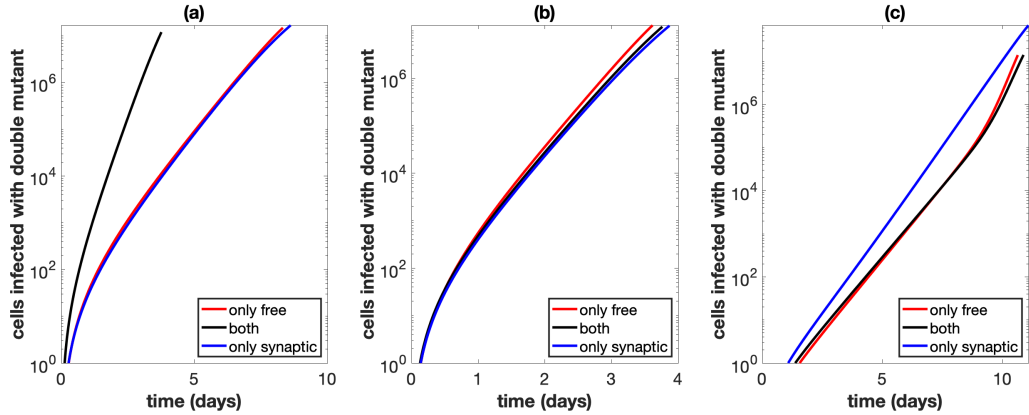


Figure S30: Similar to Figure 4 in the main text, using best fit parameters from experiment A2. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

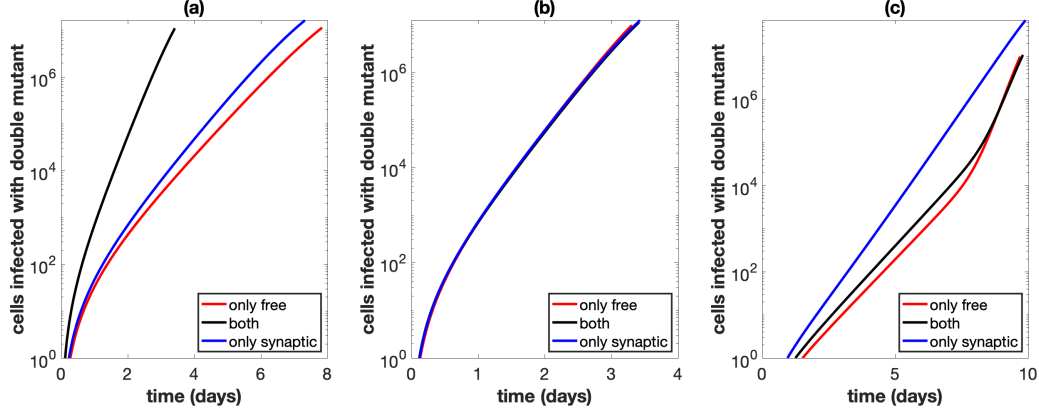


Figure S31: Similar to Figure 4 in the main text, using best fit parameters from experiment A3. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

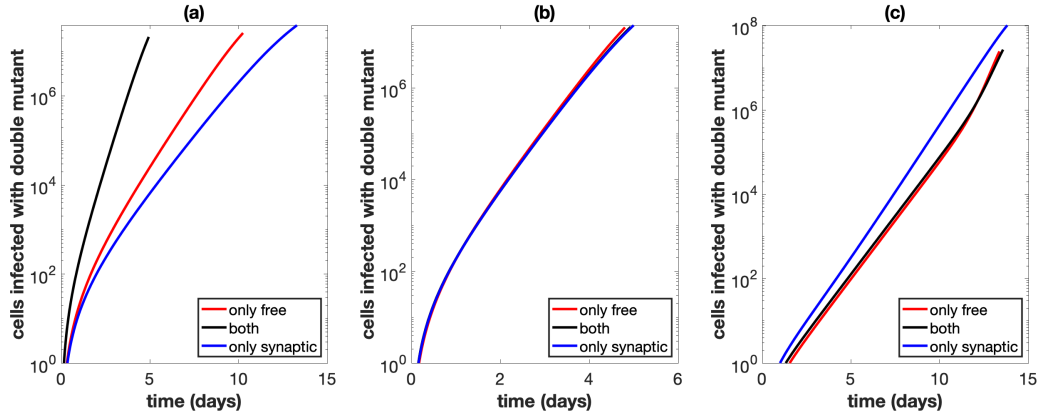


Figure S32: Similar to Figure 4 in the main text, using best fit parameters from experiment B1. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

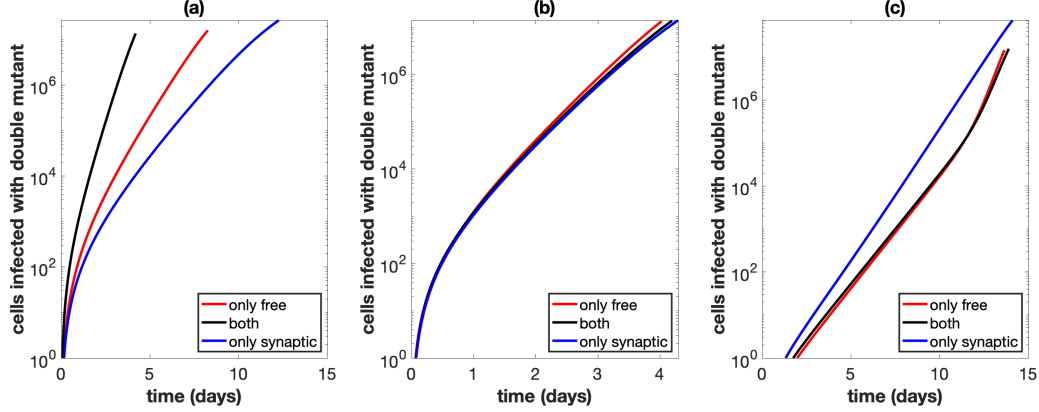


Figure S33: Similar to Figure 4 in the main text, using best fit parameters from experiment B2. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

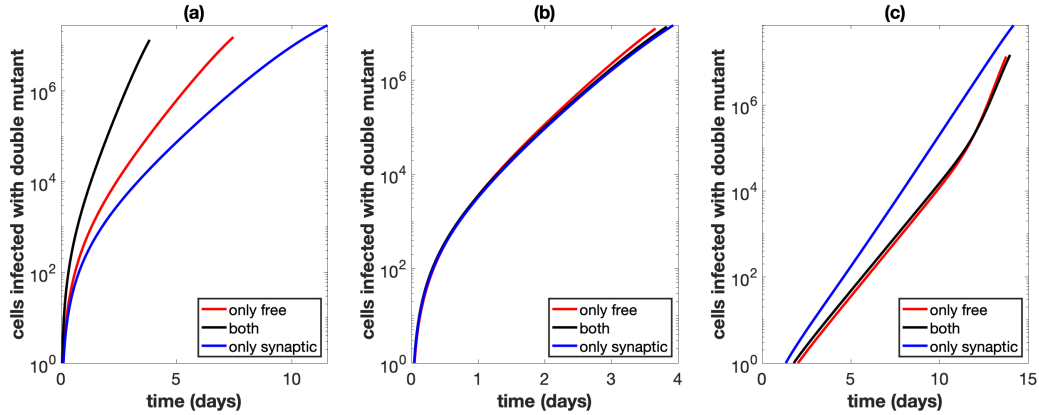


Figure S34: Similar to Figure 4 in the main text, using best fit parameters from experiment B3. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

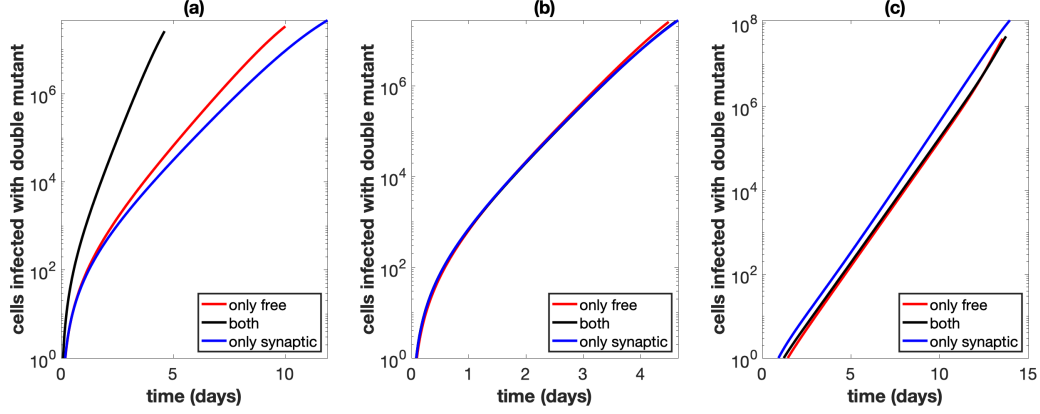


Figure S35: Similar to Figure 4 in the main text, using best fit parameters from experiment C1. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

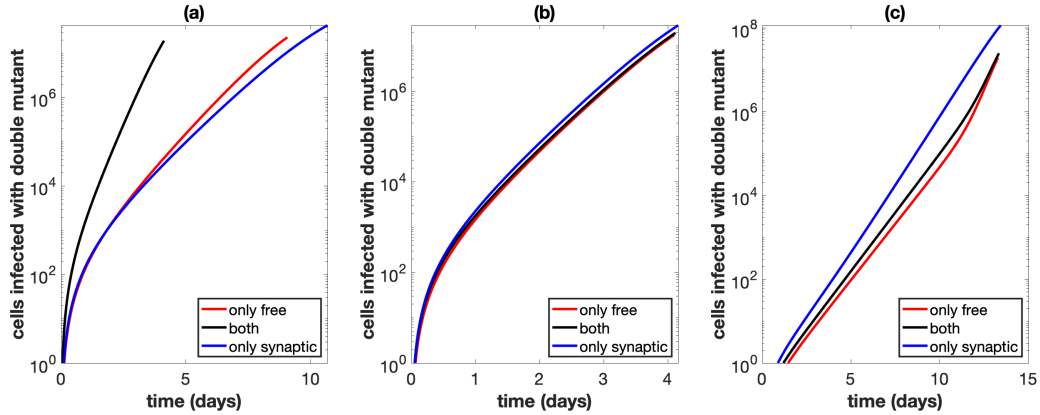


Figure S36: Similar to Figure 4 in the main text, using best fit parameters from experiment C2. Best fit parameters are included in Table S2. Total number of cells infected with at least one active copy of double mutant virus plotted against time. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Parameters are exactly as in the experiment. (b) The overall growth rate is the same across the three cases. (c) The overall growth rate is the same across the three cases and the simulation starts with only a single infected cell, which is coinfecting with both single mutant strains.

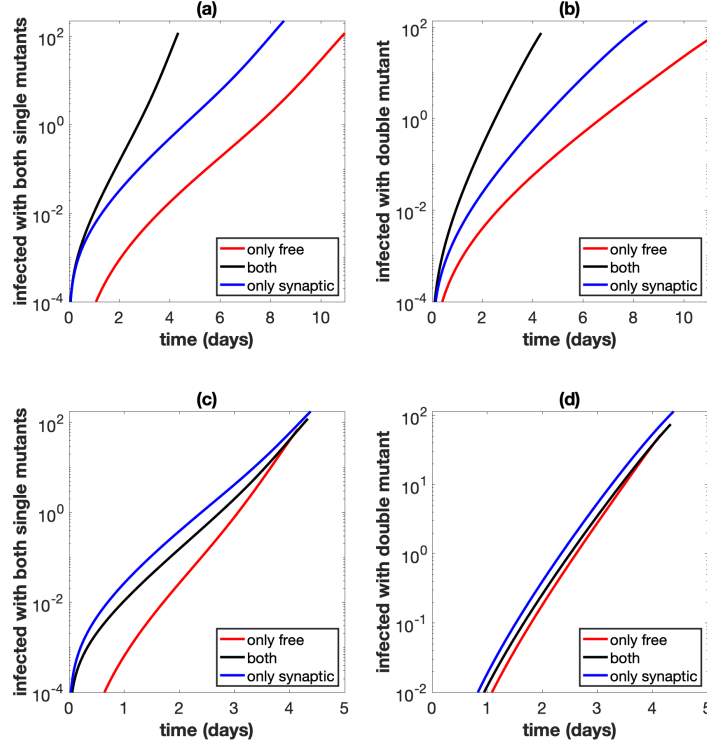


Figure S37: Similar to Figure 5 in the main text, using best fit parameters from experiment A1. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

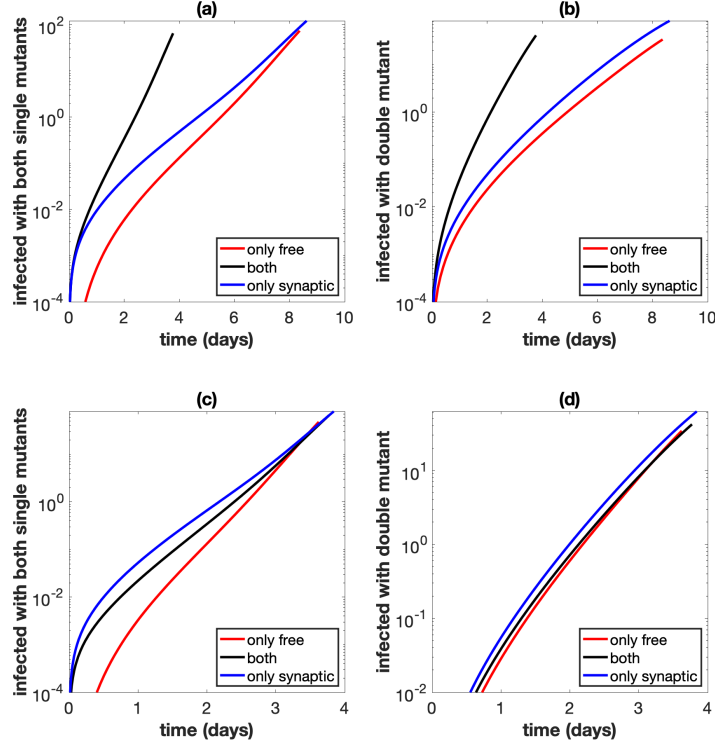


Figure S38: Similar to Figure 5 in the main text, using best fit parameters from experiment A2. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

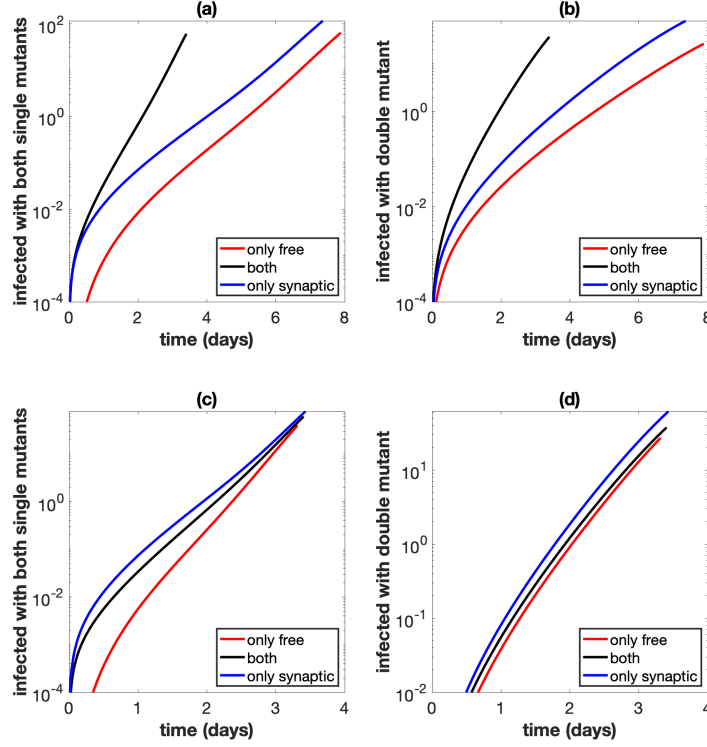


Figure S39: Similar to Figure 5 in the main text, using best fit parameters from experiment A3. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

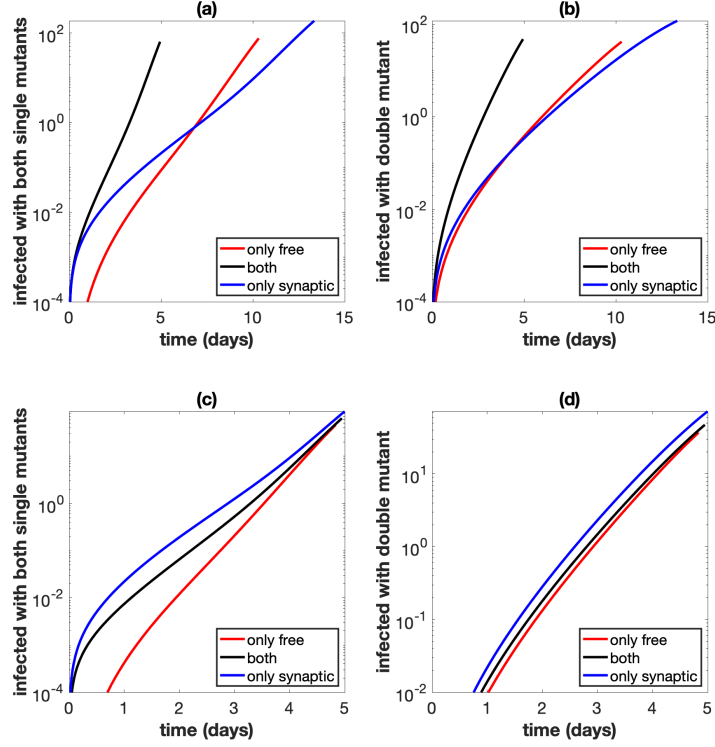


Figure S40: Similar to Figure 5 in the main text, using best fit parameters from experiment B1. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

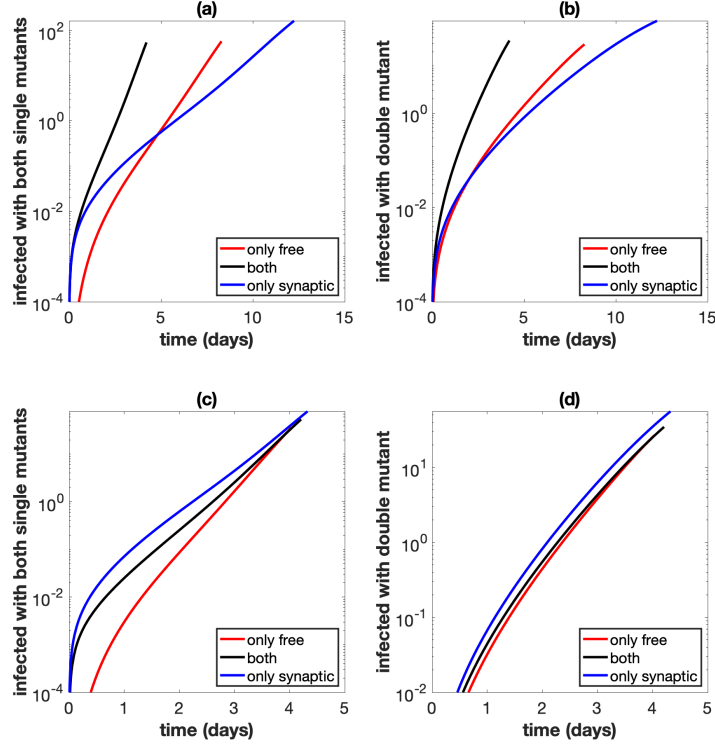


Figure S41: Similar to Figure 5 in the main text, using best fit parameters from experiment B2. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

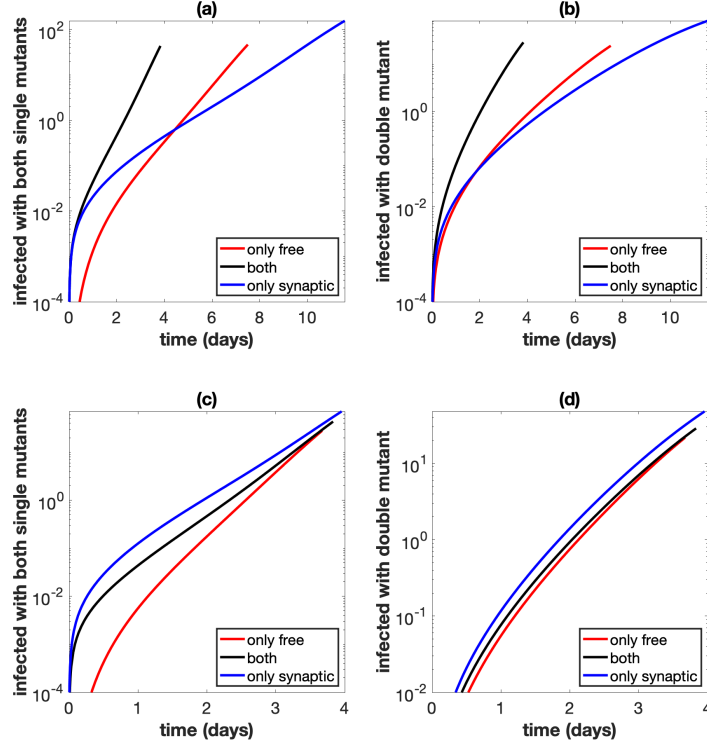


Figure S42: Similar to Figure 5 in the main text, using best fit parameters from experiment B3. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

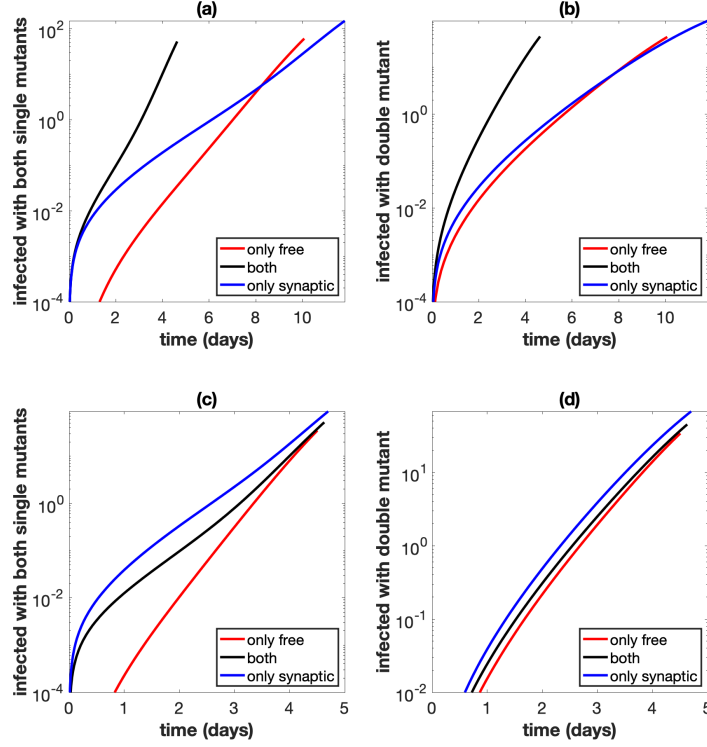


Figure S43: Similar to Figure 5 in the main text, using best fit parameters from experiment C1. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

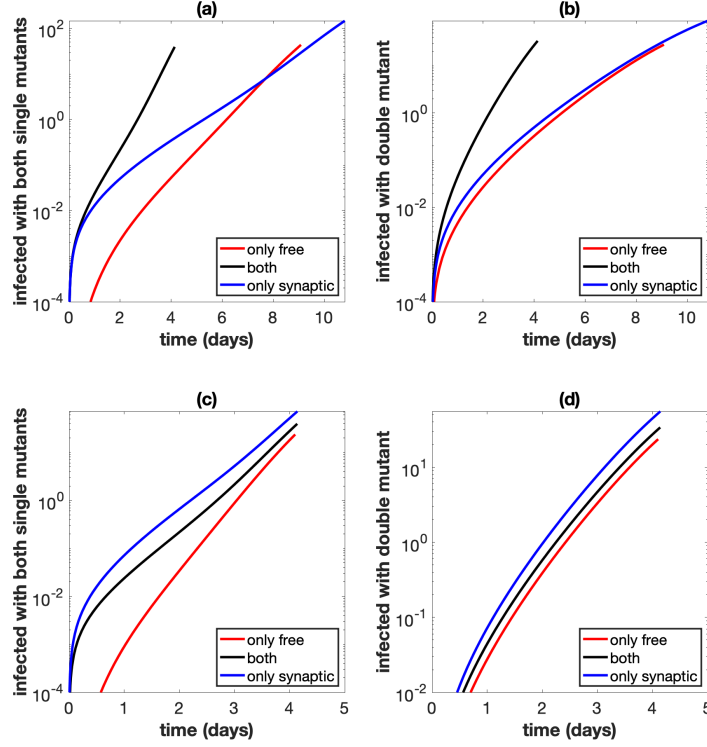


Figure S44: Similar to Figure 5 in the main text, using best fit parameters from experiment C2. Best fit parameters are included in Table S2. Simulations start with a small equal amount of cells singly infected with the wild type, and mutations are included. We assume we have 10^9 initial cells, and run the simulation until the number of infected cells reaches 40% of this initial amount, while also assuming that uninfected cells do not divide. The combination of both transmission pathways is represented by the black lines. For the only free virus transmission case (red lines), synaptic transmission is turned off. For the only synaptic transmission case (blue lines), free virus transmission and reinfection are turned off. (a) Total number of cells infected with at least one copy of both single mutant strains. Parameters are exactly as in the experiment. (b) Total number of cells infected with at least one active copy of double mutant virus. Parameters are exactly as in the experiment. (c) Same as panel (a), but the overall growth rate is the same across the three cases. (d) Same as panel (b), but the overall growth rate is the same across the three cases.

References

- [1] Natalia L Komarova, Daniela Anghelina, Igor Voznesensky, Benjamin Trinité, David N Levy, and Dominik Wodarz. Relative contribution of free-virus and synaptic transmission to the spread of hiv-1 through target cell populations. *Biology letters*, 9(1):20121049, 2013.
- [2] Jesse Kreger, Natalia L Komarova, and Dominik Wodarz. Effect of synaptic cell-to-cell transmission and recombination on the evolution of double mutants in HIV. *Journal of the Royal Society Interface*, 17(164):104–115, 3 2020.
- [3] David N Levy, Grace M Aldrovandi, Olaf Kutsch, and George M Shaw. Dynamics of hiv-1 recombination in its natural target cells. *Proceedings of the National Academy of Sciences*, 101(12):4204–4209, 2004.
- [4] Louis M Mansky and Howard M Temin. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *Journal of virology*, 69(8):5087–5094, 1995.