Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Huisheng (Julie) Zhu Nov 6, 2021

Fitness Estimation for Viral Variants in the Context of Cellular Coinfection

by

Huisheng Zhu

Katia Koelle
Adviser

Department of Biology

Katia Koelle

Adviser

David Cutler

Committee Member

Anice Lowen

Committee Member

Seunghwa Rho

Committee Member

2021

Fitness Estimation for Viral Variants in the Context of Cellular Coinfection

By

Huisheng Zhu

Katia Koelle

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Biology

2021

Abstract

Fitness Estimation for Viral Variants in the Context of Cellular Coinfection
By Huisheng Zhu

Animal models are frequently used to characterize the within-host dynamics of emerging zoonotic viruses. More recent studies have also deep-sequenced longitudinal viral samples originating from experimental challenges to gain a better understanding of how these viruses may evolve in vivo and between transmission events. These studies have often identified nucleotide variants that can replicate more efficiently within hosts and also transmit more effectively between hosts. Quantifying the degree to which a mutation impacts viral fitness within a host can improve identification of variants that are of particular epidemiological concern and our ability to anticipate viral adaptation at the population level. While methods have been developed to quantify the fitness effects of mutations using observed changes in allele frequencies over the course of a host's infection, none of the existing methods account for the possibility of cellular coinfection. Here, we develop mathematical models to project variant allele frequency changes in the context of cellular coinfection and, further, integrate these models with statistical inference approaches to demonstrate how variant fitness can be estimated alongside cellular multiplicity of infection. We apply our approaches to empirical longitudinally sampled H5N1 sequence data from ferrets and SARS-CoV-2 sequence data from hamsters and ferrets. Our results indicate that previous studies may have significantly underestimated the within-host fitness advantage of viral variants. In addition, cellular coinfection could explain the leveling-off we observed in the advantageous variant's increase. These findings underscore the importance of considering the process of cellular coinfection when studying within-host viral evolutionary dynamics.

Fitness Estimation for Viral Variants in the Context of Cellular Coinfection

By

Huisheng Zhu

Katia Koelle

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Biology

2021

Acknowledgements

I am very grateful for the patient guidance that Dr. Koelle and Brent provided me with over the past two years. This project would not be possible without your help. We thank Dr. Thomas Friedrich and Dr. Katarina Braun at the University of Wisconsin for providing us with the measured allele frequencies of G778A that are plotted in Figure 2b of Wilker et al. We also thank Dr. Anice Lowen for helpful comments on determining what generation times to use in our model.

Table of Contents

Introduction

Zoonotic pathogens are often poorly adapted to their spillover hosts. Viral adaptation, however, can occur during epidemiological spread following spillover, resulting in increases in viral transmission potential as the pathogen establishes itself in the host population [1]. This has been observed most notably in influenza viruses that have successfully established in humans (e.g., [2, 3]). The pandemic coronavirus SARS-CoV-2 provides a more recent example, with variant lineages that are better adapted to human hosts (such as D614G [4]) emerging and replacing earlier viral lineages. Viral adaptations that improve transmission potential often arise from their effect on within-host replication dynamics. For example, mutations that enable viruses to replicate more efficiently within hosts (in particular, in transmission-relevant tissues) could enhance transmission potential, as could mutations that allow for a more effective evasion of the host immune response.

In vivo studies could in principle be used to identify mutations that improve viral fitness in a spillover host. For example, experiments using the ferret animal model identified a set of influenza A subtype H5N1 mutations that increase viral replication within the nasal turbinate of hosts (a transmission-relevant tissue) and also increase transmissibility [5, 6]. The fitness effects of mutations such as these have been estimated by interfacing quantitative models with data on how variants carrying these mutations change in frequency over the course of infection [7–9]. However, these approaches assume that fitness is an individual-level property of a variant. While this may be the case when cells are only singly infected, many viral infections involve significant levels of cellular coinfection. For example, due to incomplete viral genomes, influenza viruses heavily rely on complementation to produce viral progeny [10–12]. High levels of cellular coinfection in other viruses, such as HIV, is also likely, given the pervasiveness of recombinant genomes that are identified during viral sequencing [13, 14].

Cellular coinfection can impede the ability of high-fitness variants to rise to high frequencies within an infected host. This is because of the phenomenon of 'phenotypic hiding' [15, 16]. Phenotypic hiding comes about as a consequence of viral protein products being shared within coinfected

cells. Delivery of a viral genome carrying a highly beneficial mutation results in the production of a viral protein that can provide a replicative benefit to all of the viral genomes present in the coinfected cell. Similarly, a viral genome carrying a deleterious (and potentially even lethal) mutation can be rescued by protein products derived from coinfecting viral genomes. Cellular coinfection thus results in natural selection no longer acting on individual viral genomes, but instead on viral collectives. This effectively reduces the strength of selection, such that deleterious mutations are purged more slowly [17] and beneficial mutations are also fixed more slowly [18]. As a result, the extent of cellular coinfection impacts the dynamics of allele frequency changes in an infection and affects fitness inference.

Here, we first develop a set of mathematical models to project changes in the allele frequencies of viral variants within infected hosts. Our models specifically allow for cellular coinfection and the effect of phenotypic hiding on allele frequency changes. Using Bayesian inference approaches, we then demonstrate how these mathematical models can be interfaced with longitudinally sampled allele frequency data to jointly estimate the relative fitness of a variant and cellular multiplicity of infection levels. Finally, we apply our developed methods to estimate the fitness effect of adaptive mutations that was identified in an influenza H5N1 experimental challenge study performed using the ferret animal model and in SARS-CoV-2 experimental *in vivo* competition study performed using the hamster animal model. Our findings indicate that the fitness effect of this mutation is considerably higher than previously estimated and that cellular coinfection precipitously slowed down the rate of within-host influenza virus adaptation.

Materials and Methods

Deterministic within-host evolution model

Several studies to date have used longitudinal allele frequency data to estimate the relative fitness of a mutant allele over a wild-type allele within an infected host or from passage studies [9, 19, 20]. None of these models, however, account for the impact that cellular coinfection can have on variant allele frequency changes over time. To accommodate cellular coinfection, we first start with an evolutionary model that projects allele frequencies from one viral generation to the next in the absence of coinfection:

$$q^m\left(t_{g+1}\right) = \frac{q^m\left(t_g\right)e^{\sigma_m}}{q^m\left(t_g\right)e^{\sigma_m} + \left(1 - q^m\left(t_g\right)\right)e^{\sigma_w}} \tag{1}$$

where $q^m\left(t_g\right)$ is the frequency of the variant (mutant) allele in viral generation g, σ_m (with range $-\infty$ to ∞) is the selective advantage/disadvantage of the focal mutation, and e^{σ_m} (with range ≥ 0) is the relative fitness of the variant allele over the wild-type allele. The fitness of the wild-type allele (e^{σ_w}) is defined as 1. This model is a simplification of a model first presented in [9]. That model considers an arbitrary number of viral haplotypes and further incorporates *de novo* mutation in its projection of allele frequencies. Here, we ignore *de novo* mutation over the course of infection and limit our analysis to two viral haplotypes: a wild-type viral genotype and a variant genotype carrying a mutant allele at a single locus. We adopt these simplifications to focus attention on the effect of cellular coinfection in within-host evolution.

To extend this initial model to allow for the effect of cellular coinfection, we first assume that viral genomes enter cells independently of other viral genomes. Under this assumption, viral genomes are distributed across cells according to a Poisson distribution. Given a mean overall cellular multiplicity of infection (MOI) of M, the variant's mean MOI in viral generation t_g is simply given by $M_m = q^m\left(t_g\right)M$ and the wild-type virus's mean MOI is simply given by $M_w = \left(1 - q^m\left(t_g\right)\right)M$. The probability that a cell is infected with k variant viral genomes and l wild-type

viral genomes is then:

$$P(k,l) = \left(\frac{e^{-M_m}(M_m)^k}{k!}\right)\left(\frac{e^{-M_w}(M_w)^l}{l!}\right) \tag{2}$$

Under the assumption that viral protein products within cells have additive effects, the fitness of a viral genome present in a cell carrying k variant viral genomes and l wild-type viral genomes is given by:

$$F(k,l) = \frac{k}{k+l}e^{\sigma_m} + \frac{l}{k+l}e^{\sigma_w} \tag{3}$$

Note that this fitness does not depend on whether the focal genome is a variant viral genome or a wild-type viral genomes, since all viral genomes within a cell share their protein products and thus have the same fitness.

The realized mean fitness of a viral variant in the context of cellular coinfection is calculated by taking a fitness average of the viral variant across its cellular contexts:

$$\overline{e^{\sigma_m}} = \frac{\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} kP(k,l)F(k,l)}{MOI_m} \tag{4}$$

Similarly, and the realized mean fitness of the wild-type virus in the context of cellular coinfection is given by:

$$\overline{e^{\sigma_w}} = \frac{\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} lP(k,l)F(k,l)}{MOI_w} \tag{5}$$

Examination of these equations indicates that the realized mean fitness of the viral variant and of the wild-type virus approach e^{σ_m} and e^{σ_w}, respectively, as cellular MOI becomes small, as expected. As cellular MOI becomes large, $\overline{e^{\sigma_m}}$ and $\overline{e^{\sigma_w}}$ converge in their values, as expected.

Variant allele frequency changes in the context of cellular coinfection can then be projected using a modified version of Eqn. 1, where realized mean fitnesses replace individual-level viral fitnesses:

$$q^m\left(t_{g+1}\right) = \frac{q^m\left(t_g\right)\overline{e^{\sigma_m}}}{q^m\left(t_g\right)\overline{e^{\sigma_m}} + \left(1 - q^m\left(t_g\right)\right)\overline{e^{\sigma_w}}} \tag{6}$$

Simulated data

We simulated the models described above to ascertain the effect of cellular coinfection on variant allele frequency changes at various levels of coinfection. We also simulated mock datasets and used them to test the statistical inference methods described in detail below. We simulated one mock dataset using the deterministic within-host evolution model, with observed variant allele frequencies that include measurement noise (noise that is due to an inaccurate measuring process, rather than underlying noise in the viral dynamic process) . To implement measurement noise, we let the *observed* variant allele frequency in generation t_g, $q_o^m(t_g)$, be drawn from a beta distribution with shape parameter $\alpha = \nu q^m(t_g)$ and shape parameter $\beta = \nu(1 - q^m(t_g))$:

$$q_o^m(t_g) \sim Beta(\nu q^m(t_g), \nu(1 - q^m(t_g)))$$

where ν quantifies the degree of measurement noise. The parameter ν is constrained to be positive, with higher values corresponding to less measurement noise. We simulated a second mock dataset using the stochastic within-host model, similarly assuming beta-distributed measurement noise.

Empirical H5N1 data

As an application of the approaches developed here, we used longitudinal allele frequency data from an influenza A subtype H5N1 experimental challenge study in ferrets [21]. We specifically focused on inferring the relative fitness of a single nucleotide variant on the hemagglutinin gene segment (G788A) in the VN1203-HA(4)-CA04 virus. This variant was present in the viral inoculum stock at a frequency of 4.40% and increased in frequency over the course of infection in each of the four ferrets that were challenged with this inoculum. Although G788A allele frequencies were measured in [21] on days 1, 3, and 5 post-inoculation, we excluded the day 5 samples from

our analyses. This is because up to (and including) day 3, the viral population in each of the four ferrets exhibited low levels of genetic diversity, with G788A being the only variant present at substantial frequencies. By day 5, additional variants on the hemagglutinin gene segment had emerged, with some reaching high frequencies. Because there is genetic linkage between these later variants and G788A, the G788A frequency changes between days 3 and 5 are likely due in part to selection acting on these later variants. Because our model does not reconstruct viral haplotypes or consider epistatic interactions between loci, we thus decided to exclude day 5 from our analysis to be able to focus more specifically on estimating the fitness of G788A in the context of cellular coinfection.

Empirical SARS-CoV-2 data

We also applied our model to longitudinal allele frequency data from SARS-CoV-2 D614G *in vivo* competition experiments done in Syrian hamsters [22]. Six hamsters were inoculated with a 1:1 mixture of SARS-CoV-2^{D614} and SARS-CoV-2^{G614}, and nasal wash samples were taken daily from days 2 to 8, and day 12. The G variant quickly rose in frequencies at the beginning of the infection but lingered around 95% afterward until day 12. This observation led us to hypothesize that while there was little or no cellular coinfection at first, MOI increased as the infection proceeded and reduced the with-in host realized mean fitness of the G variant. Through a comparison of SARS-CoV-2^{D614} trajectories and cycle threshold (Ct) values, a Ct value of above 21 was chosen as an indicator of little or no coinfection happening. All of the six ferrets in [22] had a Ct value of above 21 until day 3. We first fit the existing deterministic model without coinfection to data from days 2 and 3 to infer e^{σ_G}, the relative fitness of SARS-CoV-2^{G614} to SARS-CoV-2^{D614}. Using the inferred e^{σ_G} value, we then fitted our deterministic model to data from day 4 to estimate MOI.

Statistical inference

The deterministic within-host model contains four parameters: the relative fitness of the variant virus (e^{σ_m}) over the wild-type virus, the mean cellular multiplicity of infection (M), the initial

frequency of the variant virus in a host ($q^m(t_0)$), and the magnitude of measurement noise (ν). When interfacing this model with longitudinal allele frequency data, we estimate the first three parameters but do not estimate ν. We do not estimate ν because it can be parameterized from allele frequency measurements from replicate samples. To estimate e^{σ_m}, M, and $q^m(t_0)$, we rely on Markov Chain Monte Carlo (MCMC) approaches.

Let $P(q_o^m(t_g))$ be the probability of observing a variant allele frequency of q_o^m in generation t_g. This probability is given by the beta probability density function, with shape parameters $\nu q_{sim}^m(t_g)$ and $\nu(1-q_{sim}^m(t_g))$, evaluated at $q_o^m(t_g)$, where $q_{sim}^m(t_g)$ is the model-simulated allele frequency in generation t_g. This simulated variant allele frequency depends on parameters e^{σ_m}, M, and $q^m(t_0)$, and for the stochastic model also N. For the deterministic model, the likelihood of the model is then given by:

$$\prod_g P(q_o^m(t_g)) \tag{7}$$

where g indexes the generation times of all the measured variant allele frequency data points. For the stochastic model, $P(q_o^m(t_g))$ is used to calculate the particle weights in the pMCMC algorithm.

Statistical inference code was implemented using Python 3.7.4 and is available from `https://github.com/koellelab/withinhost_fitnessInference`.

With-in host dynamics modeling for SARS-CoV-2 D614G (Ongoing work)

To examine how the with-in host dynamics of SARS-CoV-2^{D614} and SARS-CoV-2^{G614} differ and whether the difference can provide an explanation for the replacement of SARS-CoV-2^{D614} by SARS-CoV-2^{G614}, we used the single strain basic within-host dynamics model presented in [23], which is shown in Eqn. 8.

$$\begin{aligned}
\frac{dT}{dt} &= -\beta T V_i \\
\frac{dI}{dt} &= \beta T V_i - \delta I \\
\frac{dV_i}{dt} &= \alpha p I - c V_i \\
\frac{dV_n}{dt} &= (1-\alpha) p I - c V_n
\end{aligned} \tag{8}$$

where T represents target cells, I represents infected cells, V_i represents infectious viruses, and V_n represents non-infectious viruses. V_i infects T at a constant rate of β. Infected cells produce virions at rate p and die at rate δ. A portion of α of the produced virions become infectious, leaving the rest non-infectious. All of the viruses are cleared at rate c.

Assuming fast with-in host dynamics, V_i equilibrates very rapidly with respect to I within the first few days when there is exponential viral growth. Thus approximately,

$$\frac{dV_i}{dt} \approx 0 \quad V_i \approx \frac{\alpha p}{c} I \tag{9}$$

Plugging back into the second differential equation in the model, we get

$$\frac{dT}{dt} = (\frac{\beta T \alpha p}{c} - \delta) I = rI \tag{10}$$

where r, the intrinsic rate of increase, is defined as viral fitness during exponential growth phase. The ratio of r between SARS-CoV-2^{D614} and SARS-CoV-2^{G614} should thus equal e^{σ_G}. As the study in [23] reported parameter estimates for ferrets inoculated with viruses collected from Wuhan experimental and patient samples, which correspond to SARS-CoV-2^{D614}, we can estimate β^G assuming all the other parameters in the model are shared between SARS-CoV-2^{D614} and SARS-CoV-2^{G614}:

$$r^G = e^{\sigma_G} r^D \quad \beta^G = e^{\sigma_G} \beta^D - (1 - e^{\sigma_G}) \frac{\delta}{T \alpha p} \tag{11}$$

where T can be approximated by T_0 based on the fast dynamics assumption.

We could also extent the single strain model in Eqn. 8 to two strains to recapitulate the with-in host dynamics undergoing competition:

$$\begin{aligned}
\frac{dT}{dt} &= -\beta^D T V_i^D - \beta^G T V_i^G \\
\frac{dI^D}{dt} &= \beta^D T V_i^D - \delta I^D \\
\frac{dI^G}{dt} &= \beta^G T V_i^G - \delta I^G \\
\frac{dV_i^D}{dt} &= \alpha p I^D - c V_i^D \\
\frac{dV_i^G}{dt} &= \alpha p I^G - c V_i^G \\
\frac{dV_n^D}{dt} &= (1-\alpha) p I^D - c V_n^D \\
\frac{dV_n^G}{dt} &= (1-\alpha) p I^G - c V_n^G
\end{aligned} \tag{12}$$

where the initial V_i^D and V_i^G should be assigned in accordance with the inoculum mixture ratio.

Results

The extent of cellular coinfection impacts variant frequency dynamics

The within-host models developed above differ from previous models focused on within-host viral evolution by incorporating the possibility of cellular coinfection and its effects on variant frequency dynamics. Simulations of our deterministic model show, as expected, that a beneficial mutation does not increase in frequency as rapidly when cellular coinfection levels are high compared to when they are low (Fig.1A). Our simulations also show that a deleterious mutation does not decrease in frequency as rapidly when cellular coinfection levels are high compared to when they are low (Fig. 1B). Both of these effects are a direct consequence of phenotypic hiding that occurs in cells that are infected by more than one viral genome.

Our stochastic within-host evolution model recapitulates the general patterns observed in simulations of the deterministic model, with demographic stochasticity playing a more pronounced role at lower effective viral population sizes, as expected (Figs 1C and D).

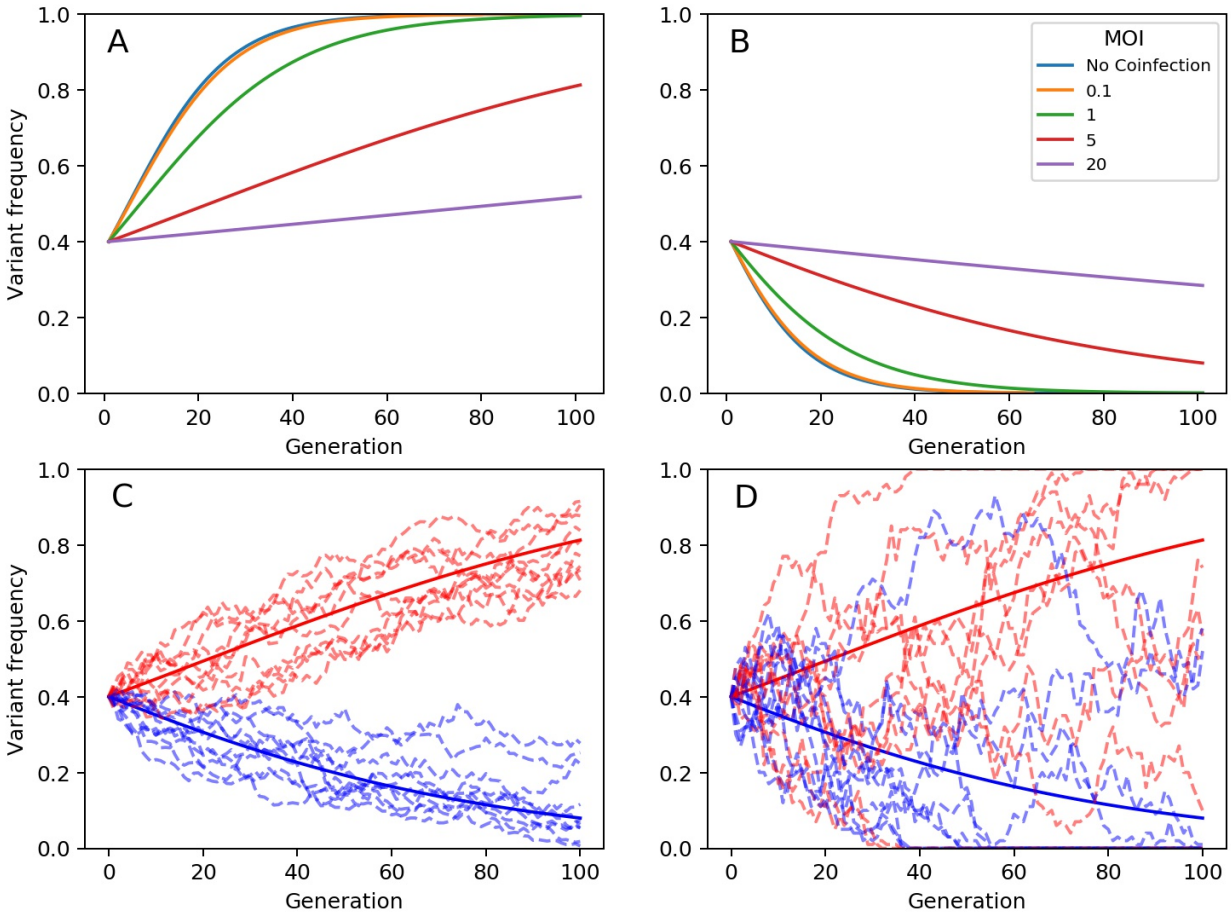


Figure 1: Model simulations showing changes in variant frequencies over viral generations. (A) Frequency changes of a beneficial mutation under the deterministic within-host model, parameterized at different mean cellular multiplicities of infection M. For all simulations shown, the variant's fitness is $e^{\sigma_m} = 1.1$ and its initial frequency is $q^m(t_0) = 0.4$. (B) Frequency changes of a deleterious mutation under the deterministic within-host model, parameterized at different mean cellular multiplicities of infection M. For all simulations shown, the variant's fitness is $e^{\sigma_m} = 0.9$ and its initial frequency is $q^m(t_0) = 0.4$. In (A) and (B), we consider MOI values of 0.1, 1, 5, and 20. Labeled as 'No coinfection', we also plot simulations of the model presented in Eqn. 1, which assumes that fitness is an individual-level property of a viral genome. (C) Frequency changes of a beneficial mutation (red; $e^{\sigma_m} = 1.1$) and of a deleterious mutation (blue; $e^{\sigma_m} = 0.9$) under the stochastic within-host model, parameterized with a mean cellular MOI of 5. Dashed lines show 10 stochastic realizations under each parameterization. Solid lines show simulations of the deterministic model under the same parameterization. Stochastic simulations used an effective viral population size of $N = 1000$. (D) Frequency changes of mutations, as in (C), only using an effective viral population size of $N = 100$.

Statistical estimation of variant fitness using the deterministic within-host model

Statistical inference with simulated data

We first aimed to determine if longitudinal allele frequency data could be used to infer variant fitness in the context of cellular coinfection under the assumption of deterministic within-host evolutionary dynamics. We therefore first generated a mock dataset by forward simulating the deterministic model and adding measurement noise (Fig. 2A). Prior to applying the MCMC methods described above to this mock dataset, we assessed the identifiability of the two parameters of greatest biological interest: variant fitness e^{σ_m} and mean cellular mulitiplicity of infection M. We did this by setting the magnitude of measurement noise ν and the initial mutant allele frequency $q^m(t_g = 0)$ to their true values and plotting the model likelihood over a range of MOIs and over a range of variant fitnesses. Our results indicate that there is a likelihood 'ridge' from low MOI-low fitness parameter combinations to high MOI-high fitness parameter combinations (Fig. 2B). The presence of this likelihood ridge is expected, given that higher variant fitness in the context of higher MOI compensates for the phenotypic hiding phenomenon that does not occur at lower MOI.

Given this likelihood ridge, it would be difficult to use MCMC to obtain posterior distributions of the model parameters without an informative prior on either variant fitness or MOI. We decided, for the sake of illustration, to adopt a prior on MOI. Specifically, we assumed a lognormal prior on MOI, with a mean of $log(2)$ and a standard deviation of 0.5. We ran the MCMC chain for 20,000 iterations (Figure S1). Posterior distributions for the initial frequency of the variant, MOI, and variant fitness are shown in Figures 2C-E. All true parameters fell within the 95% credible intervals of the estimated parameter values. In Figure 2A, we further plot 10 forward simulations, parameterized with draws from the posterior distributions, alongside the mock data. These results indicate that the deterministic within-host evolution model can be successfully interfaced with longitudinal variant allele frequency data to infer model parameters using MCMC.

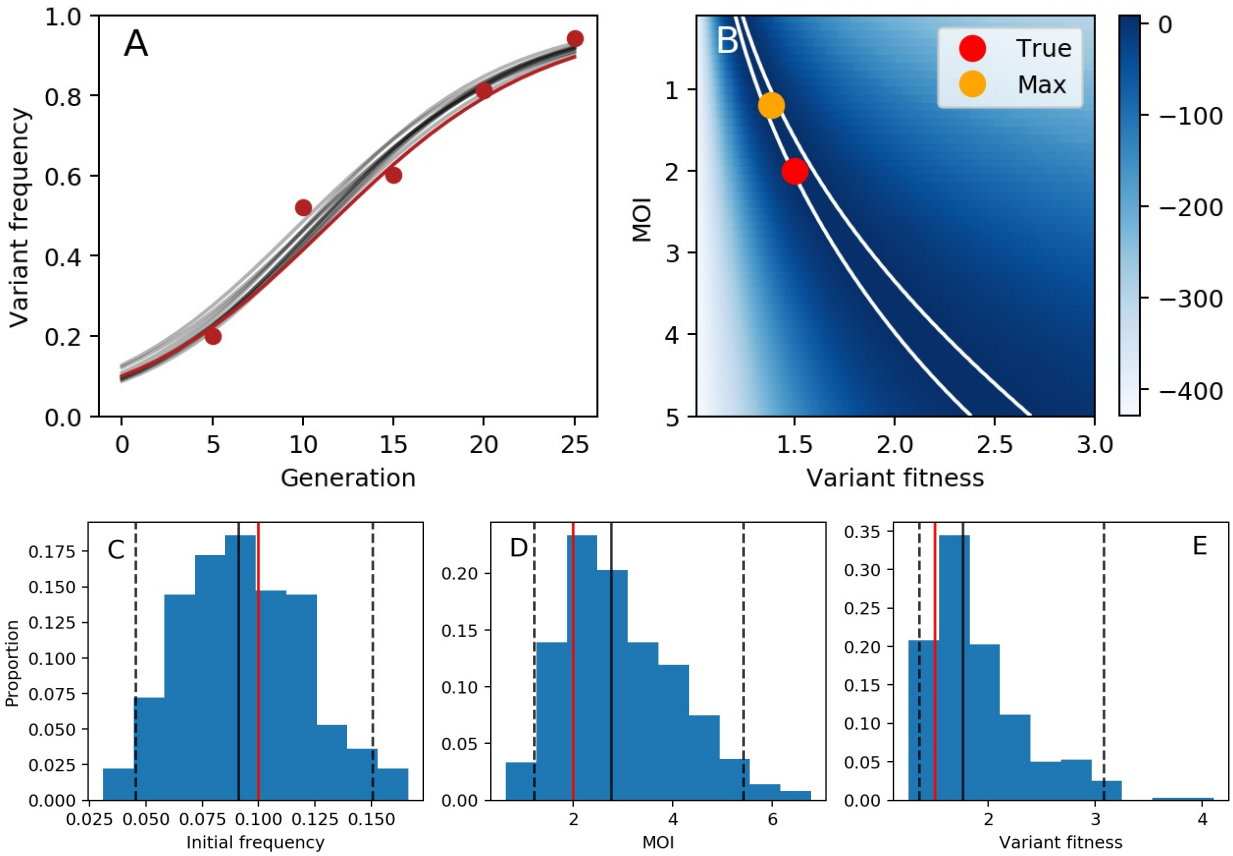


Figure 2: Variant fitness estimation under the assumption of deterministic evolutionary dynamics. (A) Mock data (red dots) generated from a forward simulation of the deterministic within-host evolution model with added measurement noise. The underlying deterministic dynamics are shown with a red line. The model is parameterized with variant fitness of $e^{\sigma_m} = 1.5$, a mean cellular MOI of $M = 2.0$, and an initial frequency of the variant of $q^m(t_0) = 0.10$. Measurement (observation) noise is set to $\nu = 100$. Grey lines show 10 model simulations, with parameters drawn from the MCMC posterior distributions. (B) Log-likelihood landscape, showing the log-likelihood of the model over a broad range of MOI and variant fitness values. When calculating these likelihoods, the initial frequency of the variant and the measurement noise were fixed at their true values. The red dot shows the true set of parameters used to simulate the mock data. The yellow dot shows the parameter combination yielding the highest log-likelihood. White boundary lines show the 95% confidence interval of parameter estimates. (C) Posterior distribution for the initial frequency of the variant. (D) Posterior distribution for the mean cellular multiplicity of infection M. (E) Posterior distribution for variant fitness. In (C)-(E), black solid lines show the median values of the posterior density, black dashed lines show the 95% credible intervals, and red solid lines show the true values.

Statistical inference with experimental H5N1 challenge study

We now apply the same MCMC approaches to experimental data from an influenza A subtype H5N1 challenge study performed in ferrets. Figure 3A shows the frequencies of the G788A variant that was present in the inoculum stock at a frequency of 4.40% and increased in all four of the

experimentally infected ferrets. For the reasons provided above, we used only days 1 and 3 for estimation of variant fitness. We also used the measured stock frequency of 4.40% as the day 0 data point for all ferrets. While technically the stock frequency and the ferrets' day 0 data points constitute very different samples, we felt comfortable with this assumption because of the likely very large transmission bottleneck size between the inoculum and index ferrets. Although an estimate of this transmission bottleneck size is not reported on in [21], a study using barcoded virus found that three-quarters of viral barcodes present in the inoculum were transmitted to index (donor) ferrets in experimental challenges that used 10^4 plaque-forming units (p.f.u.) of virus inoculum [24], which is two orders of magnitude less virus than used in [21]. In the barcoded virus study, some of the barcodes that were transmitted had frequencies as low as 0.5% in the stock, indicating that the transmission bottleneck size was likely hundreds to thousands of virions. Under the assumption of random sampling of virions from the stock, this means that the frequency of G788A on day 0 of the ferrets was likely very close to 4.40%, with measurement noise significantly outweighing any noise stemming from the wide transmission bottleneck. Indeed, calculations involving the binomial distribution (for the transmission bottleneck) and the beta distribution (for measurement noise) indicate that measurement noise dominates transmission bottleneck process noise when the bottleneck size is larger than the measurement noise parameter ν. With a bottleneck size in the hundreds to thousands and the value of ν we use for this dataset (see below), measurement noise is much larger than transmission bottleneck size noise, thereby allowing us to make the assumption that the day 0 allele frequencies of G788A in the ferrets is equal to the stock frequency of G788A.

In fitting our model to these data, we first converted days post inoculation to viral generations by assuming an 8 hour influenza virus generation time based on [25]. Replicate samples for this experiment were not available, so we set the degree of measurement noise ν to 100, but consider the sensitivity of our results to this value (see below). We used an informative prior on the mean cellular MOI, specifically a lognormal prior with a mean of $log(4)$ and a standard deviation of 0.4. We used this prior based on studies that indicate that 3-4 virions are generally required to yield progeny virus from an infected cell [11]. However, we note that a wide range of estimates

exist in the literature on the extent of viral complementation required for successful influenza virus progeny production, with findings indicating that this depends on the host cell type and on the viral strain considered [10, 12]. We ran the MCMC chain for 20,000 iterations (Figure S2). Posterior distributions for mean cellular MOI and variant fitness are shown in Figures 3B and C, respectively. The joint density plot of MOI and variant fitness (Fig. 3D) indicates that there is a positive correlation between these two parameters, consistent with our findings on simulated data (Fig. 2B). Posterior distributions for the initial frequencies of the variant in each ferret are shown in Figure S3.

The results shown in Fig. 3B indicate that cellular MOI is relatively high, although the informative prior used played a large role in shaping this parameter's posterior distribution. Our estimate of variant fitness (relative to wild-type fitness) lies between 2.11 and 7.91, with a median value of 3.15. This stands in stark contrast to a previous fitness estimate for this variant of approximately $e^{0.35} = 1.42$ [9]. However, this previous estimate was based on a model that did not consider cellular coinfection. With high levels of coinfection thought to occur in within-host influenza virus infections [11] and our inference of relatively high cellular MOI (Fig. 3B), higher fitness was inferred for G788A to be able to account for its observed rapid rise in the context of phenotypic hiding. Indeed, the joint density plot shown in Fig. 3D indicates that if we had constrained MOIs to be lower(closer in line with the estimates from [10]), our variant fitness estimates would have been considerably closer to those previously inferred for G788A.

Our inferred fitness estimate of $\sim 2-8$ for G788A may initially seem unreasonably large. However, several studies that have estimated variant fitness using *in vitro* experiments have arrived at estimates of similar magnitude. For example, a recent *in vitro* study of dengue virus evolution performed at low MOI found that, of the beneficial mutations that were identified, some had relative fitness effects exceeding 2 [26]. An *in vitro* study focused on HIV similarly found that beneficial mutations could have pronounced effects on viral fitness, with the largest estimated relative fitness of a single mutation being 6.6 [27]. These studies show that the fitness effects of viral mutations can be quite high, particularly when under strong selection pressure. While our relative fitness

estimate of $\sim 2-8$ for G788A falls in the range of other estimates present in the viral literature, there are also studies that have inferred lower fitness values for beneficial mutations. For example, the highest relative fitness value estimated for an influenza B mutation that conferred resistance to a neuraminidase inhibitor was 1.8 [28].

The results presented in Figure 3 assume measurement noise ν of 100 and a viral generation time of 8 hours. To ascertain the effects of these assumptions on our results, we first re-estimated MOI, variant fitness, and initial variant frequencies under the assumption of both higher (ν = 25) and lower (ν = 400) levels of measurement noise (Figures S4 and S5). With higher levels of measurement noise, 95% credible interval ranges for MOI and variant fitness were both wider than when measurement noise was set to ν = 100. In contrast, with lower measurement noise, 95% credible interval ranges for MOI and variant fitness were both considerably more narrow than when measurement noise was set to ν = 100, with variant fitness estimates falling in the range of 2.25-3.6. At both higher and lower levels of measurement noise, median estimates for MOI and variant fitness were not considerably impacted. We also considered the sensitivity of our results to the viral generation time assumed (Figures S6 and S7). With a shorter generation time of 6 hours, the posterior distribution for MOI remained similar to one inferred using a viral generation time of 8 hours. However, variant fitness estimates were lower, with the 95% credible interval range of 1.66 - 3.47 and a median value of 2.36. With a longer generation time of 12 hours, the posterior distribution for MOI again remained similar to one inferred using a viral generation time of 8 hours. Variant fitness estimates using a 12 hour viral generation time were considerably higher, however, with the 95% credible interval range of 2.88 - 9.31 and a median value of 4.58. These results underscore the importance of accurately parameterizing the viral generation time when performing variant fitness estimation.

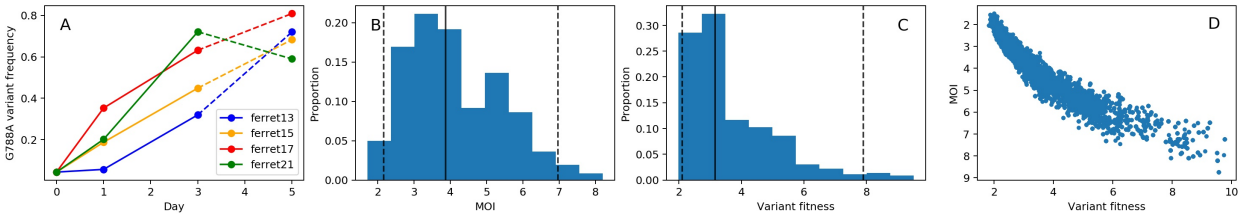


Figure 3: Fitness estimation for variant G788A, assuming deterministic within-host dynamics. (A) Measured G778 allele frequencies over the course of infection for 4 experimentally infected ferrets. Days 0 (stock frequency), 1, and 3 are used in the estimation of variant fitness. (B) Posterior distribution for the mean cellular multiplicity of infection. (C) Posterior distribution for variant fitness. In (B) and (C), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals. (D) Joint density plot for MOI and variant fitness.

In Figure 4, we show 10 forward simulations of the deterministic model, parameterized using draws from the posterior distributions. These indicate that the model, simulated using parameter estimates inferred from MCMC, reproduces observed G788A allele frequency patterns on days 0, 1, and 3 (the days included in the statistical analyses). The model, however, significantly over-predicts G788A frequencies on day 5 in ferret 15 and ferret 21 (Figs 4B and D). It is interesting to note that in both ferrets 15 and 21, one additional variant (G738A) rose to high frequencies between days 3 and 5. Previous work has inferred a large relative fitness value for this variant ($e^{0.9} = 2.5$) as well as (slightly negative) epistatic interactions between it and G788A [9]. Haplotype reconstruction indicates that the 'A' allele at site 738 arose in the genetic background of the 'G' allele at site 788 [9, 21]. With the 'A' allele at site 738 conferring a large fitness advantage, and its genetic linkage to the 'G' allele at site 788, we would anticipate that this mutation would slow or even reverse the rise of variant G788A between days 3 and 5 in these ferrets due to this process of clonal interference. Indeed, our model projections significantly overestimate the frequency of G788A on day 5 in both of these ferrets, indicating that selection efficiently acted on G738A, impeding the projected increase in the frequency of G788A between days 3 and 5. It is also interesting to examine the dynamics of additional variants in ferrets 13 and 17, where the model predicts G788A frequencies relatively well on day 5, although this data point was not used during model fitting. Ferret 13 had one other variant arising between day 3 and day 5 (variant G496T). A previous study using these data inferred a large relative fitness value for this variant ($e^{0.7} = 2.0$)

[9]. Our model simulations, however, projected the allele frequency of G788A on day 5 well in the absence of considering this variant. As such, we would predict that this G496T variant had lower relative fitness than previously estimated. Ferret 17 also had one other variant rising to high frequencies between day 3 and day 5 (variant C736A). It is unclear whether previous work inferred this mutation to be strongly beneficial or strongly deleterious, since A736C (rather than C736A) was the mutation identified as being under positive selection. Regardless, our model slightly over-projects the frequency of G788A on day 5, such that we expect C736A to have contributed to some extent to allele frequency changes of G788A through linkage effects.

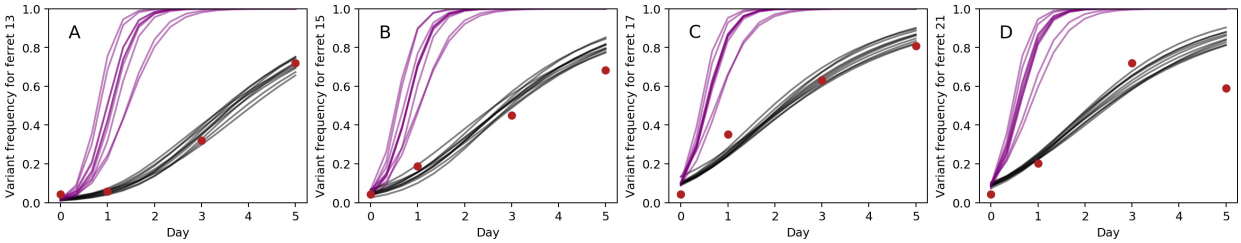


Figure 4: Deterministic model simulations (grey lines) and observed data points (red dots) are shown for (A) ferret 13, (B) ferret 15, (C) ferret 17, and (D) ferret 21. Only days 0, 1, and 3 were used in model fitting. Parameters for the model simulations were drawn from the posterior distributions of the parameters. Purple lines show model simulations under the same parameterizations of variant fitness and initial variant frequencies as the grey lines, but simulated in the absence of cellular coinfection. These no-coinfection projections were simulated using Eqn. 1.

In Figure 4, we further plot model simulations that assume no cellular coinfection. Specifically, we simulate Equation 1 where the dynamics are driven by the variant's individual-level fitness e^{σ_m} rather than by $\overline{e^{\sigma_m}}$. The frequency of G788A rises considerably faster in these simulations compared to those that incorporate cellular coinfection. This indicates that the speed of within-host viral adaptation is severely reduced by cellular coinfection.

Statistical inference with SARS-CoV-2 competition experiment (Ongoing work)

Here we set the degree of measurement noise ν to 45 based on our analysis of the spread of data in *in vitro* technical replicates done in the same study [22]. As shown in Figure 5A and B, after fitting the deterministic model without coinfection in Eqn. 1 to data from days 2 and 3, the log likelihood peaks when fitness, or e^{σ_G}, is equal to 1.33 with a 95% confidence interval of $[1.29, 1.38]$. We then fit our model to data from day 4 to 12 using the inferred fitness value of 1.33. As expected, in Figure 5C we observed an increase in log likelihood when we expanded the limit of MOI. However, the marginal increase is diminishing as MOI increases. We thus picked a MOI of 5, which was around the turning point, to simulate the trend in Figure 5D from days 4 to 12.

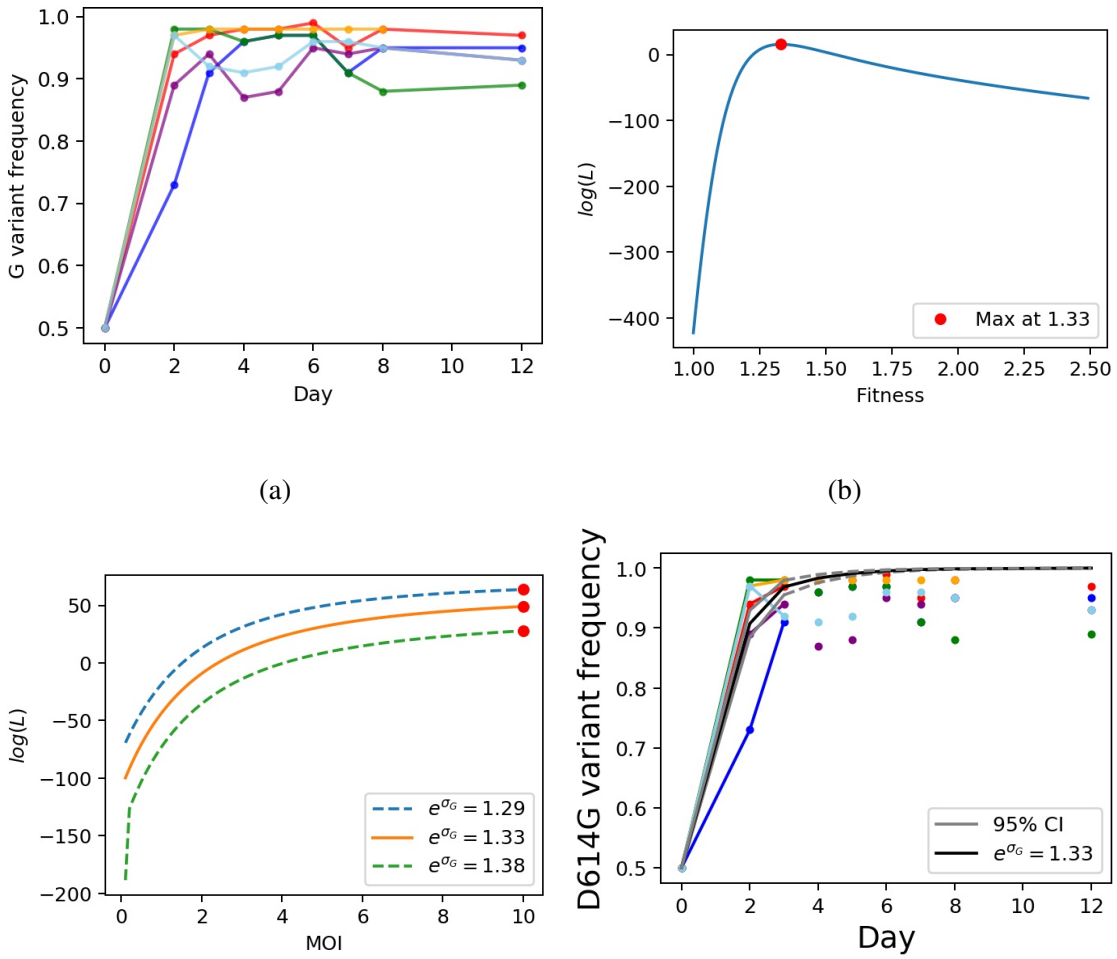


Figure 5: Fitness and MOI estimation for D614G, assuming deterministic with-in host dynamics. (A) Measured D614G allele frequencies over the course of infection from 6 experimentally infected hamsters. (B) The log likelihood curve for variant fitness when fitting day 2 and 3 data to the deterministic model without coinfection. (C) The log likelihood curve for cellular multiplicity of coinfection when fitting day 4 to 12 data to our deterministic model considering coinfection. (D) The black line shows that simulated trend with a fitness of 1.46 and an MOI of 5 between day 4 and 12. In (B,C), dashes lines show the 95% credible intervals on variant fitness inference.

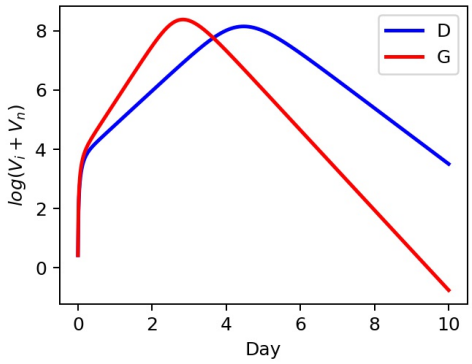
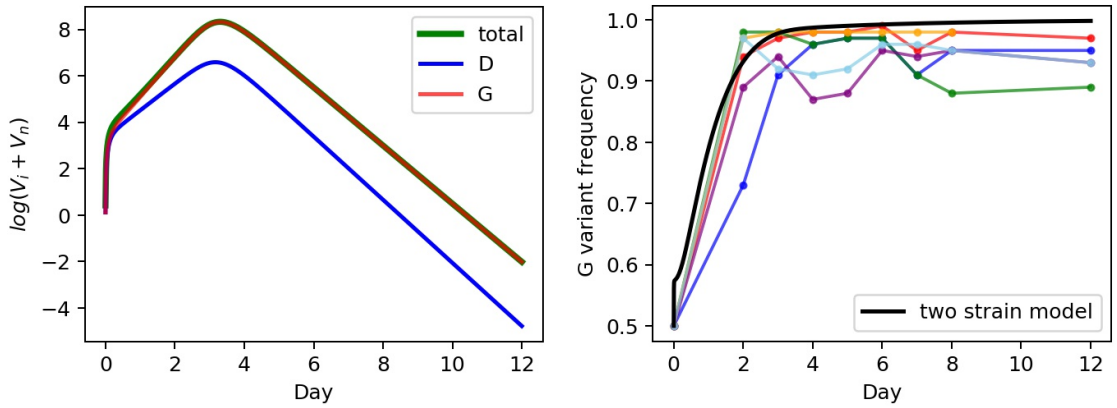


Figure 6: Trajectories of log total viral genome copies are shown by single strain modeling using Eqn. 8. The blue line represents the dynamics of SARS-CoV-2^{D614} using the first set of parameters (F13-E-1) in Table 1 of [23]. The red line the dynamics of SARS-CoV-2^{D614} under the exact same setting except for β^G calculated in Eqn. 11.

SARS-CoV-2 D614G With-in Host Dynamics (Ongoing work)

Through single strain modeling using β^D and calculated β^G, we observed in Figure 6 that the SARS-CoV-2^{D614} and SARS-CoV-2^{G614} had similar with-in host dynamics except that it took a shorter period of time for SARS-CoV-2^{G614} to reach the peak viral load.

Using Eqn. 12, we were able to model the *in vivo* competition between the two strains, SARS-CoV-2^{D614} and SARS-CoV-2^{G614}. As shown in Figure 7A, the trajectory of the G variant almost superposed the trajectories of all of the viruses, indicating that the G variant predominates the viral population since the beginning of the infection. The black line showing the calculated frequency of G variant copies in 7B recapitulated the allele frequency dynamics of the hamster *in vivo* competition experiment data in [22] that we used.



(a) (b)

Figure 7: Two strain modeling for *in vivo* competition experiments using Eqn. 12. (A) Trajectories of log total viral genome copies for SARS-CoV-2^{D614} (red), and their sum (green). The first set of parameters (F13-E-1) in Table 1 of [23] were used. V_{i0} was equally divided into V_{i0}^{D} and V_{i0}^{G} to match the 1:1 inoculum mixture in [22]. (B) The black line shows the recovered SARS-CoV-2^{G614} variant frequencies as $(V_i+V_n)^{G}$ divided by $(V_i+V_n)^{total}$. The rest of the colored lines are the same data as in Figure 5A.

Discussion

Here, we have developed mathematical within-host models that can take into consideration cellular coinfection when projecting changes in viral allele frequencies over the course of an infection. We have further described and demonstrated how these evolutionary models can be statistically interfaced with viral sequence data to jointly estimate variant fitness relative to the wild-type allele along with the mean cellular multiplicity of infection. Our results indicate that ignoring the possibility of cellular coinfection can result in significant underestimation of a variant's selective advantage. This is important because a variant with a much higher selective advantage, once established monomorphically within a host, is expected to have a more precipitous impact on within-host viral dynamics than a variant with a smaller selective advantage. We might, for example, expect a variant with a higher selective advantage to result in higher peak viral loads and potentially longer durations of infection. This would impact both symptom development as well as onward transmission potential.

Our models, like all models, make some simplifying assumptions. First, we assume low viral diversity, with diversity comprising just one locus and two alleles (a wild-type and a variant allele). We have chosen to model evolution at a single locus to highlight the important contribution that cellular coinfection may play in the within-host evolution of viral pathogens. Our application to the G788A mutation in the H5N1 experimental challenge study in ferrets satisfied this assumption between days 0 through 3. Because other sites became polymorphic in each of the four studied ferrets by day 5, we excluded this time point from our statistical analyses. To consider the effect of cellular coinfection within a system with higher levels of genetic diversity, and the possibility of new variants arising over the course of infection, the models developed here should be extended using approaches developed already in [9]. These approaches include inference of viral haplotypes and the incorporation of *de novo* mutations into the presented model structures. With these additions, full genetic linkage between loci can be considered, and epistatic interactions between loci can also be inferred. Our models, as presented here, however, could still be applied to higher diversity viral systems if recombination occurred freely between loci, as may be the case between influenza gene segments or some viruses with high recombination rates.

A second assumption present in the current formulation of our models is that viral fitness is additive: if a coinfected cell harbors both variant and wild-type viral genomes, then the fitness of each viral genome is not only assumed to be equal, but also equal to the arithmetic mean fitness of the involved genomes. This may be a good assumption if the focal mutation impacts, for example, polymerase activity, with the viral polymerase protein being used for the replication of all viral genomes. However, it may also be the case that a mutation has a disproportionate effect on intracellular viral fitness. Future work should therefore examine the impact of a mutation's 'dominance' [29] on *in vivo* viral evolution.

A third assumption is one that is somewhat less transparent in the structure of our models, namely that we assume that there is no intracellular viral competition for host cell machinery. This assumption is reflected in the calculation of a variant allele's mean fitness ($\overline{e^{\sigma_m}}$). A single viral genome's fitness in a cell depends on the genotypes of the other genomes present in the cell, but

not on the cellular multiplicity of infection directly. If a variant genome is in a cell alone or with a large number of other variant genomes, for example, its fitness will be the same. However, if host cell machinery is limiting, one would expect the per genome fitness – which can be interpreted here as *per capita* viral yield or reproductive success – to be lower in highly coinfected cells. Indeed, empirical studies with influenza virus indicates that there is a saturating relationship between viral input and viral output from a cell [30]: at low cellular MOI, doubling the viral input yields a doubling of viral output, such that viral competition is not readily apparent; at high MOI, however, doubling the viral input does not appreciably change the overall viral output, indicative of limiting host cell machinery. Future work should therefore also examine the impact of intracellular viral competition on within-host viral evolution and extend models such as the ones we presented here to account for intracellular viral competition.

Finally, our model assumes that the mean cellular multiplicity of infection (MOI) is fixed across viral generations and that virion entry into cells is governed by a Poisson process. In terms of the former assumption, it is conceivable that MOI might change over the course of an infection. For example, at the beginning of a viral infection, MOI may be low because a very small viral population is initiating infection in a large environment of host cells. As viruses replicate within their host, viral population sizes increase and the number of target cells decreases. This may result in more individual-level selection at the beginning of the infection (due to low MOI), followed by a greater degree of phenotypic hiding later on in the infection (due to higher MOI). To accommodate these changes in MOI, the structures of the within-host models presented here would not need to be significantly altered; MOI could simply be made into a time-varying parameter. For simplicity, we here instead decided to assume that MOI is fixed over the course of infection, in part because of the lack of empirical data to inform MOI at multiple time points over the course of an infection. A further argument against incorporating dynamic changes in MOI is that spatially-structured within-host viral dynamics, such as those characterized for influenza [31], may result in cellular MOIs that are more uniform over time than expected from a spatially unstructured setting. In terms of the latter assumption (Poisson-distributed virions), there are a number of reasons why this assumption

may not be met. Virions could aggregate, such that virion entry into cells is not an independent process. Cells could also be heterogeneous with respect to their susceptibility to infection, for example due to their cell cycle state or due to antiviral states triggered by interferon. Both of these factors would result in virions being overdispersed across cells, rather than Poisson-distributed. While considering different assumptions of how virions are distributed across cells is beyond the scope of this study, future work should address the effect of viral overdispersion on variant fitness estimation.

Despite these limiting assumptions, a general takeaway from the evolutionary models presented here is that cellular coinfection will slow down the rate of viral adaptation within hosts when adaptation occurs through selection acting on single point mutations (or insertions/deletions) as we have considered here. (A caveat here is that cellular coinfection could accelerate viral adaptation if it heavily relies on genetic exchange, that is, recombination or reassortment.) Slower rates of viral adaptation is good news from the perspective of the host population, as this will also slow down viral adaptation at the population-level. This finding has clear implications for emerging zoonotic viruses that are adapting to a new host population. Analogously, cellular coinfection will result in less effective purging of deleterious mutations. By making natural selection a weaker evolutionary force, cellular coinfection may thus be one reason why stochastic processes appear to dominate within-host viral dynamics and why selection does not seem to act efficiently over the course of an acute infection for viruses such as seasonal influenza [32, 33]. There are other factors, however, that may also limit the ability for positive selection to act efficiently within hosts. For example, the temporal asynchrony between the timing of the immune response and when virus diversification occurs may explain why antigenic immune escape variants do not readily arise in individuals with some pre-existing immunity [34]. A second takeaway is that variants whose fitness levels (relative to wild-type) have been quantified using models that do not include cellular coinfection may have significantly underestimated variant fitness. Underestimation of variant fitness may underestimate the effect of a mutation on viral replication dynamics once those dynamics involve only the variant virus. Our results – that the fitness effect of certain mutations can be large – speak to the adaptive

potential of these viruses to new or changing host populations, even if adaptation may occur more slowly than might be expected.

References

(1) Antia, R.; Regoes, R. R.; Koella, J. C.; Bergstrom, C. T. *Nature* **2003**, *426*, 658–661.

(2) Matrosovich, M.; Tuzikov, A.; Bovin, N.; Gambaryan, A.; Klimov, A.; Castrucci, M. R.; Donatelli, I.; Kawaoka, Y. *Journal of Virology* **2000**, *74*, 8502–8512.

(3) Su, Y. C.; Bahl, J.; Joseph, U.; Butt, K. M.; Peck, H. A.; Koay, E. S.; Oon, L. L.; Barr, I. G.; Vijaykrishna, D.; Smith, G. J. *Nature Communications* **2015**, *6*, 1–13.

(4) Volz, E.; Hill, V.; McCrone, J. T.; Price, A.; Jorgensen, D.; O'Toole, Á.; Southgate, J.; Johnson, R.; Jackson, B.; Nascimento, F. F., et al. *Cell* **2021**, *184*, 64–75.

(5) Herfst, S.; Schrauwen, E. J.; Linster, M.; Chutinimitkul, S.; de Wit, E.; Munster, V. J.; Sorrell, E. M.; Bestebroer, T. M.; Burke, D. F.; Smith, D. J., et al. *Science* **2012**, *336*, 1534–1541.

(6) Imai, M.; Watanabe, T.; Hatta, M.; Das, S. C.; Ozawa, M.; Shinya, K.; Zhong, G.; Hanson, A.; Katsura, H.; Watanabe, S., et al. *Nature* **2012**, *486*, 420–428.

(7) Holland, J. J.; De La Torre, J. C.; Clarke, D.; Duarte, E. *Journal of Virology* **1991**, *65*, 2960–2967.

(8) Ganusov, V. V.; Goonetilleke, N.; Liu, M. K.; Ferrari, G.; Shaw, G. M.; McMichael, A. J.; Borrow, P.; Korber, B. T.; Perelson, A. S. *Journal of Virology* **2011**, *85*, 10518–10528.

(9) Illingworth, C. J. *Molecular Biology and Evolution* **2015**, *32*, 3012–3026.

(10) Brooke, C. B.; Ince, W. L.; Wrammert, J.; Ahmed, R.; Wilson, P. C.; Bennink, J. R.; Yewdell, J. W. *Journal of Virology* **2013**, *87*, 3155–3162.

(11) Jacobs, N. T.; Onuoha, N. O.; Antia, A.; Steel, J.; Antia, R.; Lowen, A. C. *Nature Communications* **2019**, *10*, 1–17.

(12) Phipps, K. L.; Ganti, K.; Jacobs, N. T.; Lee, C.-Y.; Carnaccini, S.; White, M. C.; Manandhar, M.; Pickett, B. E.; Tan, G. S.; Ferreri, L. M., et al. *Nature Microbiology* **2020**, *5*, 1158–1169.

(13) Shriner, D.; Rodrigo, A. G.; Nickle, D. C.; Mullins, J. I. *Genetics* **2004**, *167*, 1573–1583.

(14) Neher, R. A.; Leitner, T. *PLoS Computational Biology* **2010**, *6*, e1000660.

(15) Wilke, C. O.; Novella, I. S. *BMC Microbiology* **2003**, *3*, 11.

(16) Zavada, J. *Archives of Virology* **1976**, *50*, 1–15.

(17) Froissart, R.; Wilke, C. O.; Montville, R.; Remold, S. K.; Chao, L.; Turner, P. E. *Genetics* **2004**, *168*, 9–19.

(18) Wodarz, D.; Levy, D. N.; Komarova, N. L. *Evolution Letters* **2019**, *3*, 104–115.

(19) Acevedo, A.; Brodsky, L.; Andino, R. *Nature* **2014**, *505*, 686–690.

(20) Foll, M.; Shim, H.; Jensen, J. D. *Molecular Ecology Resources* **2015**, *15*, 87–98.

(21) Wilker, P. R.; Dinis, J. M.; Starrett, G.; Imai, M.; Hatta, M.; Nelson, C. W.; O'Connor, D. H.; Hughes, A. L.; Neumann, G.; Kawaoka, Y., et al. *Nature Communications* **2013**, *4*, 1–11.

(22) Zhou, B.; Thao, T. T. N.; Hoffmann, D.; Taddeo, A.; Ebert, N.; Labroussaa, F.; Pohlmann, A.; King, J.; Steiner, S.; Kelly, J. N., et al. *Nature* **2021**, *592*, 122–127.

(23) Vaidya, N. K.; Bloomquist, A.; Perelson, A. S. *Viruses* **2021**, *13*, 1635.

(24) Varble, A.; Albrecht, R. A.; Backes, S.; Crumiller, M.; Bouvier, N. M.; Sachs, D.; Garcıa-Sastre, A., et al. *Cell host & microbe* **2014**, *16*, 691–700.

(25) Dou, D.; Hernández-Neuta, I.; Wang, H.; Östbye, H.; Qian, X.; Thiele, S.; Resa-Infante, P.; Kouassi, N. M.; Sender, V.; Hentrich, K., et al. *Cell reports* **2017**, *20*, 251–263.

(26) Dolan, P. T.; Taguwa, S.; Rangel, M. A.; Acevedo, A.; Hagai, T.; Andino, R.; Frydman, J. *Elife* **2021**, *10*, e61921.

(27) Da Silva, J.; Coetzer, M.; Nedellec, R.; Pastore, C.; Mosier, D. E. *Genetics* **2010**, *185*, 293–303.

(28) Burnham, A. J.; Armstrong, J.; Lowen, A. C.; Webster, R. G.; Govorkova, E. A. *Journal of Virology* **2015**, *89*, 4575–4587.

(29) Bushman, M.; Antia, R. *Journal of the Royal Society Interface* **2019**, *16*, 20190165.

(30) Martin, B. E.; Harris, J. D.; Sun, J.; Koelle, K.; Brooke, C. B. *PLoS Pathogens* **2020**, *16*, e1008974.

(31) Gallagher, M. E.; Brooke, C. B.; Ke, R.; Koelle, K. *Viruses* **2018**, *10*, 627.

(32) Dinis, J. M.; Florek, N. W.; Fatola, O. O.; Moncla, L. H.; Mutschler, J. P.; Charlier, O. K.; Meece, J. K.; Belongia, E. A.; Friedrich, T. C. *Journal of Virology* **2016**, *90*, 3355–3365.

(33) McCrone, J. T.; Woods, R. J.; Martin, E. T.; Malosh, R. E.; Monto, A. S.; Lauring, A. S. *Elife* **2018**, *7*, e35962.

(34) Morris, D. H.; Petrova, V. N.; Rossine, F. W.; Parker, E.; Grenfell, B. T.; Neher, R. A.; Levin, S. A.; Russell, C. A. *Elife* **2020**, *9*, e62105.

Supplementary materials

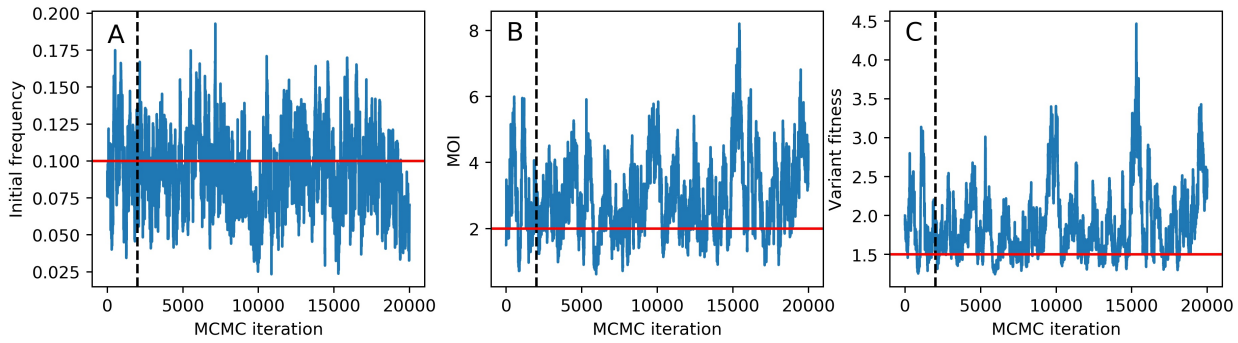


Figure 1: MCMC trace plots for parameters estimated by interfacing the deterministic within-host model with the simulated data. (A) Trace plot for initial frequency of the variant. (B) Trace plot for mean cellular multiplicity of infection. (C) Trace plot for variant fitness. 20,000 MCMC iterations were run. Following the removal of the first 2,000 MCMC iterations as burn-in, the MCMC chain was sampled every 50 iterations.

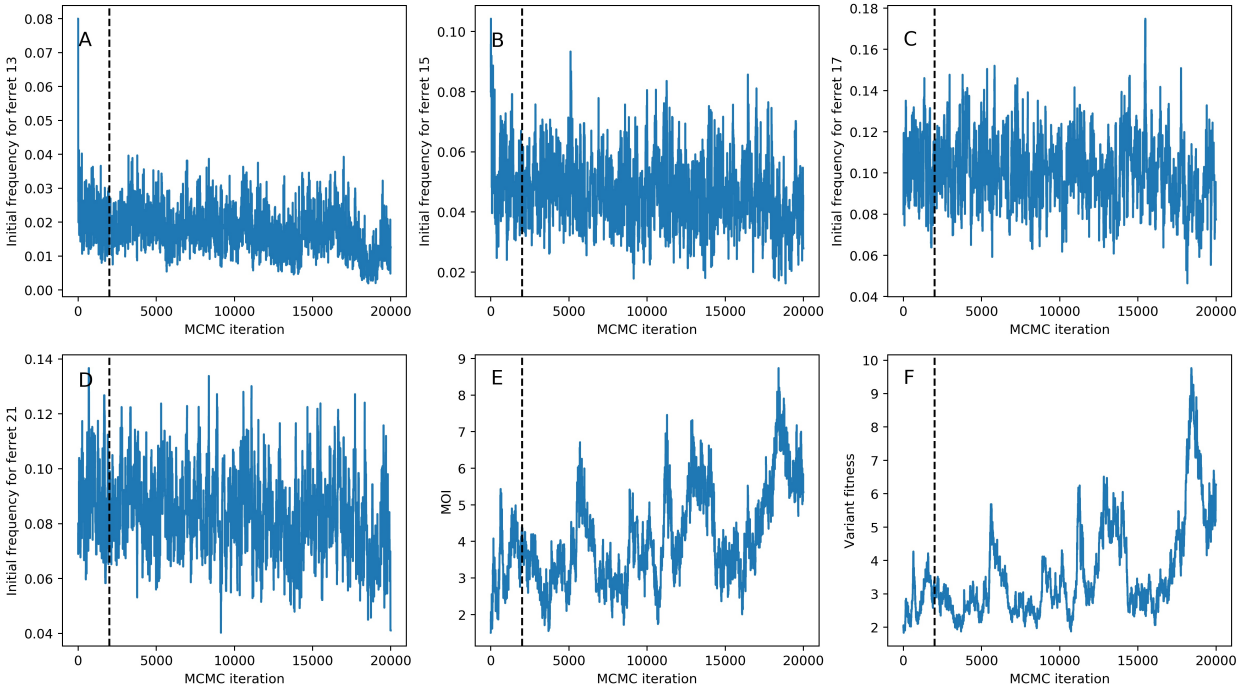


Figure 2: MCMC trace plots for parameters estimated by interfacing the deterministic within-host model with the influenza H5N1 experimental challenge study data. (A) Trace plot for initial frequency of the G788A variant in ferret 13. (B) Trace plot for initial frequency of the G788A variant in ferret 15. (C) Trace plot for initial frequency of the G788A variant in ferret 17. (D) Trace plot for initial frequency of the G788A variant in ferret 21. (E) Trace plot for mean cellular multiplicity of infection. (F) Trace plot for G788A variant fitness. 20,000 MCMC iterations were run. Following the removal of the first 2,000 MCMC iterations as burn-in, the MCMC chain was sampled every 50 iterations.

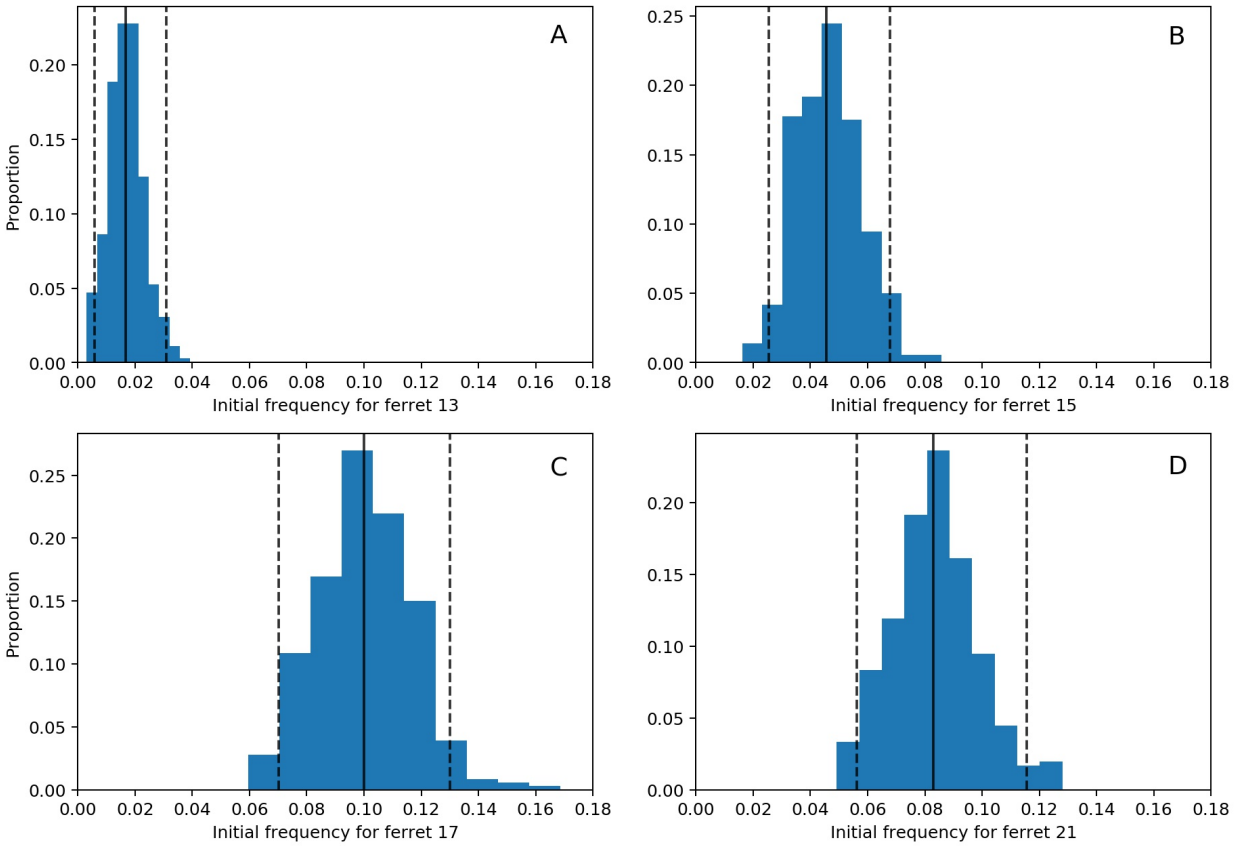


Figure 3: Posterior distributions of initial G788A frequencies for (A) ferret 13, (B) ferret 15, (C) ferret 17, and (D) ferret 21, from fitting the deterministic within-host model. In (A)-(D), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals.

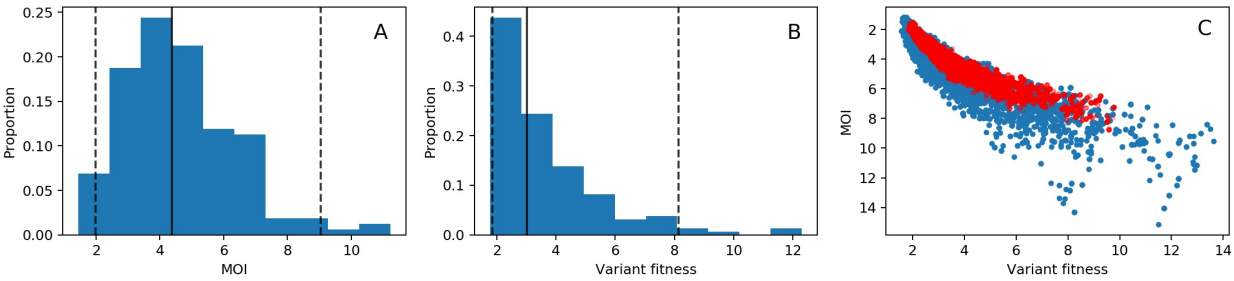


Figure 4: Parameter estimation for variant G788A, assuming deterministic within-host dynamics and measurement noise of $\nu = 25$. (A) Posterior distribution for the mean cellular multiplicity of infection. (B) Posterior distribution for variant fitness. In (A) and (B), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals. (C) Joint density plot for MOI and variant fitness (blue). For comparison, we have superimposed the joint density plot shown in Figure 3D (red).

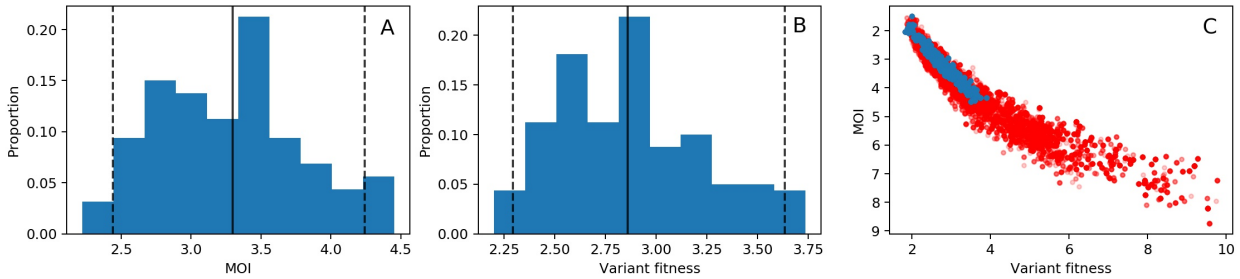


Figure 5: Parameter estimation for variant G788A, assuming deterministic within-host dynamics and measurement noise of $\nu = 400$. (A) Posterior distribution for the mean cellular multiplicity of infection. (B) Posterior distribution for variant fitness. In (A) and (B), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals. (C) Joint density plot for MOI and variant fitness (blue). For comparison, we have superimposed the joint density plot shown in Figure 3D (red).

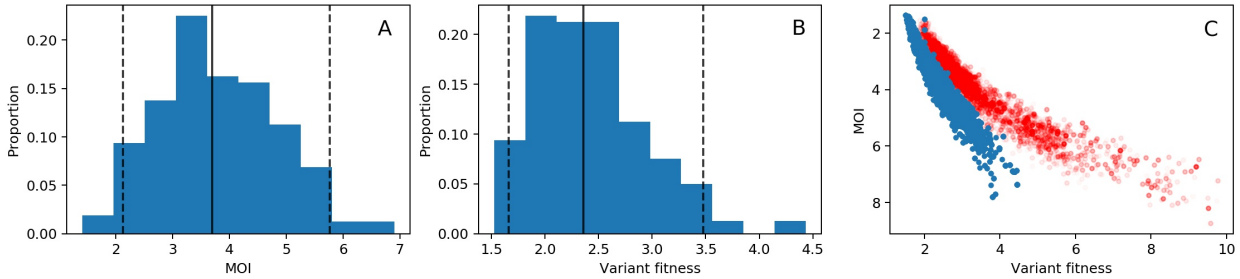


Figure 6: Parameter estimation for variant G788A, assuming deterministic within-host dynamics and a viral generation time of 6 hours. (A) Posterior distribution for the mean cellular multiplicity of infection. (B) Posterior distribution for variant fitness. In (A) and (B), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals. (C) Joint density plot for MOI and variant fitness (blue). For comparison, we have superimposed the joint density plot shown in Figure 3D (red).

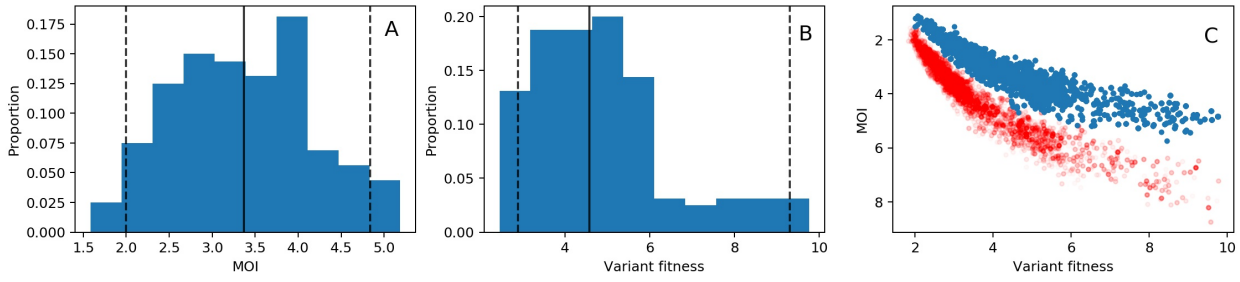


Figure 7: Parameter estimation for variant G788A, assuming deterministic within-host dynamics and a viral generation time of 12 hours. (A) Posterior distribution for the mean cellular multiplicity of infection. (B) Posterior distribution for variant fitness. In (A) and (B), black solid lines show the median values of the posterior densities and black dashed lines show the 95% credible intervals. (C) Joint density plot for MOI and variant fitness (blue). For comparison, we have superimposed the joint density plot shown in Figure 3D (red).