Title
Replication Dynamics and Stochasticity in Single-Cell Poliovirus Infections

Permalink
https://escholarship.org/uc/item/1jh0q8g5

Author
Schulte, Michael

Publication Date
2014

Peer reviewed|Thesis/dissertation

Replication Dynamics and Stochasticity in Single-Cell Poliovirus
Infections

by

Michael Schulte

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication

This work is dedicated to Ella and Iris Schulte, who have always been a welcome distraction from science.

Acknowledgements

I thank my Ph.D. mentor, Raul Andino, for helpful guidance and advice.

I thank members of the thesis committee Joe DeRisi and Hao Li for helpful guidance.

I thank my collaborators Jeremy Draghi and Joshua Plotkin for a fruitful collaboration.

I thank members of the Andino laboratory for technical assistance, helpful discussions, and camaraderie, especially Dwight Barnes, Cecily Burrill, Kristin Butcher, Michelle Flenniken, Eileen Foy, Amethyst Gillis, Leonid Gitlin, Elizabeth Goldstein, Janet Gumbs, Tzachi Hagai, Mark Kunitomi, Adam Lauring, Gonzalo Moratio, Arabinda Nayak, Kazuhide Onoguchi, Isabel Ribeiro, Priya Shah, Adi Stern, Vanessa Strings, Michel Tassetto, Zachary Whitfield, Caroline Wright, Yinghong Xiao, and Ming Te Yeh.

I thank my classmates, especially Colin Dinsmore and Jacob Stewart-Ornstein, for both scientific discussions and non-scientific distractions.

I thank my friends, especially Caleb Stica, JP Cummins, and members of Zinc Finger, for encouragement and comaraderie.

I thank my very good friend, Isabel Ribeiro, for her indelible support.

Finally, I thank my family, especially my parents, for their support.

Acknowledgement of published material

Chapter 2 contains published material:

Schulte MB, Andino R (2014) Single-cell analysis uncovers extensive biological noise in poliovirus replication. J Virol. 88(11): 6205-6212.

Replication Dynamics and Stochasticity in Single-Cell Poliovirus Infections

Michael Schulte

RNA viruses have extremely high mutation rates and therefore exist as a large population of mutants. However, their replication often occurs from infections, which initiate with only a single infectious particle. We took a systems-level view and quantitative approaches to describe how individual poliovirus infections are affected by stochastic effects and to infer the mode of replication used by the virus within single-cell infections. Temporal, quantitative measurements of positive-sense genomes, negative-sense templates, virions, and infectious particles were the major data source. Stochastic mathematical modeling was used to bridge a gap between wet lab science and computational biology. We find that poliovirus is sensitive to both kinetic stochastic effects and spatial resource variabilities in individual infections. We also infer that poliovirus replicates with a geometric growth mode, with progeny resulting from a single infection being on average 5 genomic replication cycles away from the infecting parent. This replication mode not only allows the opportunity for significant amounts of intracellular selection but also creates the potential for expansive population structures.

Table of Contents

Page

List of Tables

List of Figures

Introduction

Poliovirus

Poliovirus (PV) is the prototypical picornavirus, a family named for its small RNA genomes. It is a non-enveloped RNA virus with a single, positive-sense genome of ~7.5 kilobases. Due to its medical importance as the causative agent of poliomyelitis and its ease of genetic manipulation, it has become one of the most well characterized viruses and has long served as a model system for studying RNA virus biology, pathogenesis, and evolution (1).

Error Catastrophe

Muller proposed that when mutation rates are high in asexual organisms, deleterious mutations would accumulate by an irreversible ratchet-like mechanism (2). Mutational meltdown is a quasispecies theory concept by which a small population accumulates deleterious mutations, which leads to loss of fitness and decline of the population size, which may lead to further accumulation of deleterious mutations and eventually, extinction (3-4). During mutational meltdown, the number of inviable viruses produced is too large relative to overall population size so that exceeds the number of viable virus progeny. This phenomenon is known as error catastrophe. In Chapter 1, we investigate a low-fidelity variant of poliovirus to determine if the theoretical concept of error catastrophe is applicable to the behavior of this variant.

1

Stochasticity

Viral infections often begin with a very small number of initiating particles. Accordingly, the outcome of an infection is likely to be affected by variability in the initial molecular interactions between virus and host. An important unanswered question is whether following viral entry into a host cell, a successful infectious cycle is a pre-determined outcome. In Chapter 2, we investigate the range of outcomes upon infection of single cells with either single (low multiplicity of infection, MOI) or multiple (high MOI) infectious viral particles.

Replication Mode

Life history theory posits that the sequence and timing of events in an organism's lifespan have been fine-tuned by evolution to maximize the production of viable offspring. In a virus, the life history strategy is largely manifested in its replication mode, which can be characterized by two extreme ends of a spectrum. In the "stamping machine" mode, all progeny are produced from the infecting virus via a single genomic replication cycle. In the "geometric replication" mode, multiple replication cycles within a single cellular infection create a complex tree of virion ancestry. RNA viruses have extremely high mutation rates, orders of magnitude greater than those of most DNA-based life forms, and these two modes of replication predict progeny populations with distinctly different frequencies and distributions of mutations. Importantly, the mutation distribution of a viral population has a significant impact on viral fitness, adaptability, and pathogenicity. In Chapter 3, we strive for a synthesis of life history theory with molecular

mechanisms and define the replication modalities shaping the structure and mutation distribution of a virus population in an intact single infected cell.

Thermodynamic Stability of Essential Proteins

According to the quasispecies theory, the limit on mutation rate is established by the relative fitness of sequence variants to that of the master (4). A critical point termed the "error threshold" exists where the mutation rate of a viral genome exceeds the replacement rate of the master sequence such that the production of the master sequence is outweighed by the production of an increasingly diverse mutant swarm. This "error catastrophe" scenario is often presented as the explanation for the limit on viral mutation rate. However, this divergent, evolutionary shift in sequence space can be distinct from a drop in absolute abundance as would occur during extinction. While high mutation rate is driving both processes, two very different outcomes occur based on environmental tolerance. By illuminating the mechanism by which mutations are tolerated, we can better understand how quasispecies diversity is shaped. In Chapter 4, we begin to examine a model stemming from the thermodynamic stability of essential proteins as a potential explanation of the mechanism for a limit on mutation rate.

References

1. Racaniello VR (2006) One hundred years of poliovirus pathogenesis. Virology 344(1):9-16.

2. Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 1: 2-9.

3. Lynch M, Burger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84(5):339-344.

4. Holland JJ, De La Torre JC, Steinhauer DA (1992) RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 176:1-20. Review.

Chapter 1:

The Effect of Reduced Polymerase

Fidelity on a Viral Population:

Implications for Error Catastrophe

and Pathogenicity

Introduction

Two opposing forces governing the dynamics of viral populations hinge on the fidelity of the viral polymerase. On one hand, survival stems from viral polymerases accurately replicating the viral genome and thereby producing viable progeny. On the other, adaptability stems from viral populations being made up of a heterogeneous mixture of genomes as a way of being prepared for different selective pressures. This heterogeneous mixture of genomes is easily produced by a polymerase that allows a few mutations with each replication cycle. In the past fifty years, our understanding of these forces has led to the conclusion that a delicate balance exists between too many and too few mutations.

Poliovirus is a well characterized model for RNA virus population dynamics and multiple polymerase mutants have been identified. Two recently isolated mutants, a high-fidelity polymerase and a low-fidelity polymerase, have opened a novel opportunity to examine this balance. G64S, the high-fidelity polymerase mutant, has been reported to introduce mutations at a rate of ~18% of WT and leans heavily toward accurate genome replication producing a population of restricted genetic diversity relative to WT (1). H273R, the low-fidelity polymerase mutant, has been observed to introduce mutations at a rate at least 2 times higher than WT, producing a more genetically diverse virus population (2). Interestingly, both mutants have been seen to display an attenuated phenotype in a mouse model of infection. While G64S can establish an infection and replicate in several tissues, its restricted diversity is thought to prevent or

6

hinder entry into the central nervous system (1). Preliminary studies of H273R, however, suggest that it may to be unable to establish a productive infection in most tissues (2).

Muller proposed that when mutation rates are high in asexual organisms, deleterious mutations would accumulate by an irreversible ratchet-like mechanism (3). Mutational meltdown is a quasispecies theory concept by which a small population accumulates deleterious mutations, which leads to loss of fitness and decline of the population size, which may lead to further accumulation of deleterious mutations and eventually, extinction (4-5). During mutational meltdown, the number of inviable viruses produced is too large relative to overall population size so that exceeds the number of viable virus progeny. One might thus expect that the high mutation rate of H273R is leading to error catastrophe, where mutations gradually accumulate in the population resulting in an overall reduction in fitness. Yet when H273R is grown in human tissue culture cells no drop in viral titer is seen over a dozen serial passages at a multiplicity of infection (MOI) of 0.1 (1). At this MOI, coinfection of a cell with multiple plaque-forming units (PFUs) is limited, reducing the possibility of complementation or recombination between two or more infectious genomes. Even so, because of the high particle/PFU ratio of RNA viruses (~200 to 1000 particles/PFU for poliovirus), it is probable that coinfection of multiple non-infectious genomes even at MOI around 0.1. In this case, either complementation or recombination between non-infectious genomes could be preventing the deleterious effect of the Muller's ratchet.

Here, we investigate the potential for the Muller's ratchet phenomenon to be

7

acting on H273R populations by passaging this variant at a very low MOI (0.001). We also further characterize the H237R variant population mutation frequency and determine this population's lethality and pathogenecity in a mouse model of poliovirus infection.
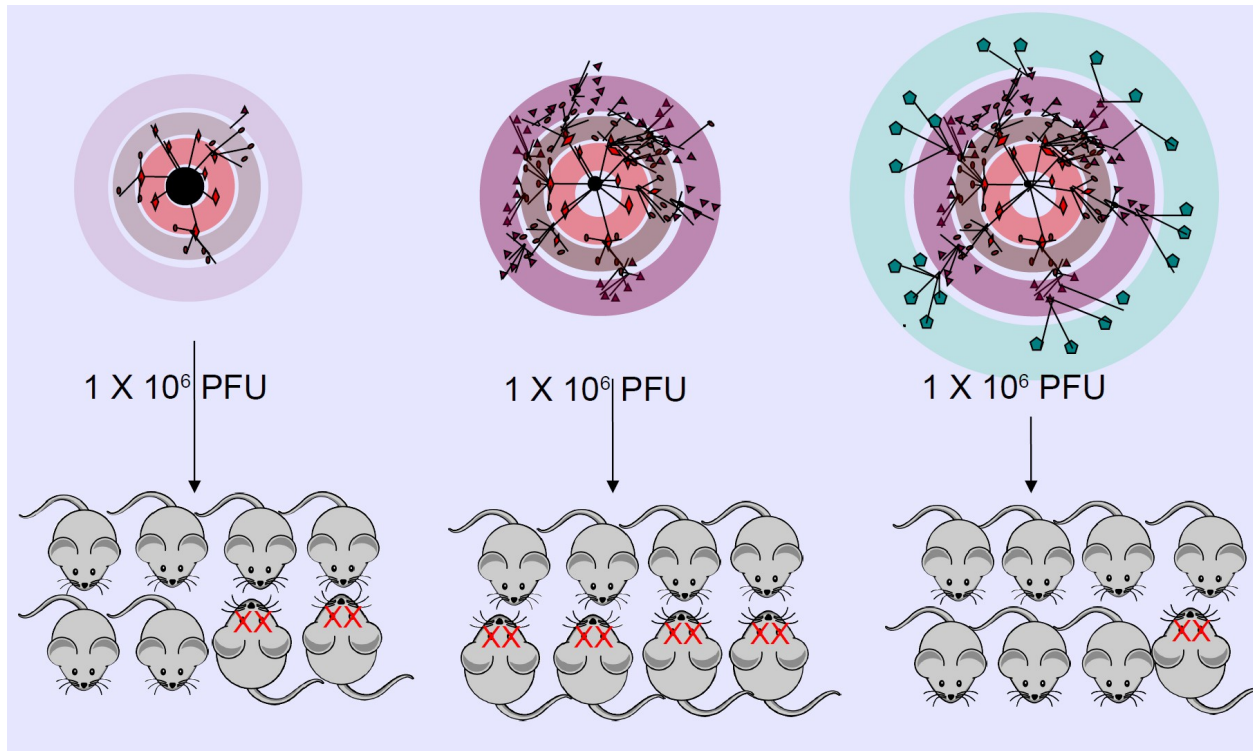
Figure 1:

Graphical representation of 3 poliovirus populations. On the left, G64S, the high-fidelity polymerase variant displays a larger proportion of the population as the center black dot, representing the WT sequence. It contains fewer mutants in the concentric rings which represent mutational steps away from the WT sequence. In the center, a WT population. On the right, H273R, the low-fidelity polymerase variant displays a larger, more diverse spread of mutants, with a greater number of mutational steps away from the WT sequence and the least amount of WT sequence. Also represented is the lethality in a mouse model of infection. One million infectious particles of each population produce differing lethality, with each polymerase variant being less lethal than WT.

Material and Methods

Cell Culture and Virus Generation

HeLa S3 cells were grown in tissue culture flask in DMEM/F12 50/50 medium
supplemented with 1x penicillin/streptomycin/glutamine and 10% newborn calf serum.
Cells were incubated at 37ºC and 5% CO_2. Poliovirus Mahoney typeI genomic RNA was
generated from *in vitro* transcription of prib(+)XpAlong. To generate virus, 20 μg of RNA
was electroporated into 4 X 10^6 HeLaS3 cells in a 4mm cuvette with the following
pulse: 300V, 24 Ω, 1000 μF.

Poliovirus RNA Transfection

HeLa S3 cells were collected, washed three times with PBS without salt, and
resuspended to a concentration of $5x10^6$ cells/ml. 800μl of cells were electroporated
with 20μg of RNA in a 0.4cm cuvette using an Electro Cell Manipulator 600 (BTX Inc.).
Cells were recovered in 16ml of medium and incubated at 37ºC in a 5% CO_2 incubator
until cytopathic effect.

Plaque assay

HeLa S3 cells were seeded in 6-well plates at a concentration of $1.5x10^6$ cells/well and
incubated overnight at 37ºC and 5% CO_2. Virus supernatant was diluted in a 1:10
dilution series in DMEM/F12 medium. Cells monolayers were washed once with PBS
then 250μl of virus dilution was added per well. To allow virus attachment, cells were

incubated with the virus inoculum for 30 mins at 37ºC and 5% CO_2. Each well was overlayed with 5ml of 1% agarose in 1X DMEM/F12 medium supplemented with 5% newborn calf serum. Plates were incubated for 2 days at 37ºC and 5% CO2. Agarose overlays were then removed and plates were stained crystal violet dye (0.1% crystal violet, 20% ethanol) to visualize the plaques which where then counted and a titer was calculated.

Passaging

HeLa S3 cells were seeded at a concentration of $5x10^5$ cells/well of a 12-well plate and incubated overnight at 37ºC and 5% CO_2. The next morning the cells were infected at the appropriate MOI. The viral inoculum was incubated with the cells for 30 mins at 37ºC and 5% CO_2 to allow virus attachment. Cells were then washed three times with PBS and 1ml of fresh media was added to the cells. Infections were incubated for 8 hours then frozen at -70C. Lysates were freeze-thawed 3X before titering.

Clonal isolation

After titers were determined by plaque assay, endpoint dilution was performed by diluting virus to ~10 PFU per 10 ml serum-free media. 100 μl of this viral dilution was add to 100 μl of cells in 96 well plates (10,000 cells/well in 10% serum media). Daily, wells were checked for CPE. Lysate from wells at CPE were transferred to Epp tubes, numbered, and frozen. After 7 days, after all infected wells were harvested, an online random number generator was used to select 24 clones in a chronologically unbiased

(and therefore likely fitness/mutation unbiased) manner. Samples were freeze-thawed 3X, then populations were amplified by infecting confluent wells of 24-well plates until CPE.

RT-PCR and Sequencing

RNA was extracted from the infected cells using Trizol Reagent. Trizol was added to the viral suspension at a 1:1 ratio then vortexed for 15 secs. 0.2 volumes of choloform was added to the reaction, mixed by hand then incubated at room temperature for 3 mins. Following centrifugation, the supernantant was precipitated by adding 0.5 volumes of isopropanol. The RNA pellet was washed once with 70% ethanol then resuspended in 20µl of RNase free H_2O.

Thermoscript RT-PCR system for First Strand cDNA Synthesis (Invitrogen) was used to synthesize cDNA from extracted RNAs. 1 µg of resuspended RNA was combined with 0.5mM dNTPs and 1µl dT primer and brought to a final volume of 13µl with ddH_2O. The sample was incubated for 5 mins at 65ºC then placed on ice for 1 minute. 4µl of 5X RT reaction buffer, 1µl RNASE out, 5nM DTT and 1µl Thermoscript was added to the reaction and brought to a total volume of 20µl with ddH_2O. The reaction was incubated at 50ºC for 1 hour. Then stored at -20ºC.

Using PV genome specific primers, the viral genome was amplified in a PCR reaction that contained 1x PCR buffer, 1.5mM $MgCl_2$, 2µl cDNA, 0.4µM of each primer, 200µM of

dNTP, 2.5U Taq DNA polymerase and water to 50µl. DNA was amplified using the following program:

Denaturing 95ºC - 2 min

Denaturing 95ºC 30sec

Annealing 56ºC 30sec   34 cycles

Extension 72ºC -1 min/1kb

Extension 72ºC 5 min

The amplified DNA fragments were purified using a PCR clean up column kit and sequenced by ELIM Biopharmaceuticals.

LD50 infections

6-7 week old cPVR mice were infected by IM route with 50µl diluted P3 stocks in each leg. Mice were monitored daily and paralysis and/or death was recorded.

Tissue tropism infections

4-5 week old cPVR mice were infected by IV route with 100µl diluted P3 stocks in the tail vein. 5 mice injected with each virus (WT or H273R) were harvested daily for 5 days. Spleen, kidney, muscle, spinal cord, and brain were collected, placed in 14 mL culture tubes on dry ice, then frozen at -70C. After thawing, tissues were homogenized in 2 mL

PBS. Homogenate was spun at 2,500 rpm for 10 min. Supernatant was then spun at

10,000 rpm for 2-3 min before resulting supernatant was titered.

Results

Low Multiplicity of Infection Passaging

In order to determine whether Muller's ratchet is operating on H273R populations in tissue culture, passaging of this population was done alongside WT and G64S at a MOIs of 0.1 and 0.001. Passages were titered by plaque assay between every passage, which is a more accurate but also more cumbersome approach than the more often used estimation of titer. After 15 serial passages of each population in duplicate at both MOIs, with titering between passages to keep multiplicity of infection consistent, no drop in titer was seen (Figure 2). "Blind" passaging, or passaging based on estimates of titer, of H273R populations was performed for 7 passages to investigate if artifacts of increased or decreased titer can arise from this less precise method. Indeed, after 6 blind passages, an increase of over a log was observed in 4 out of 4 replicates of H273R.
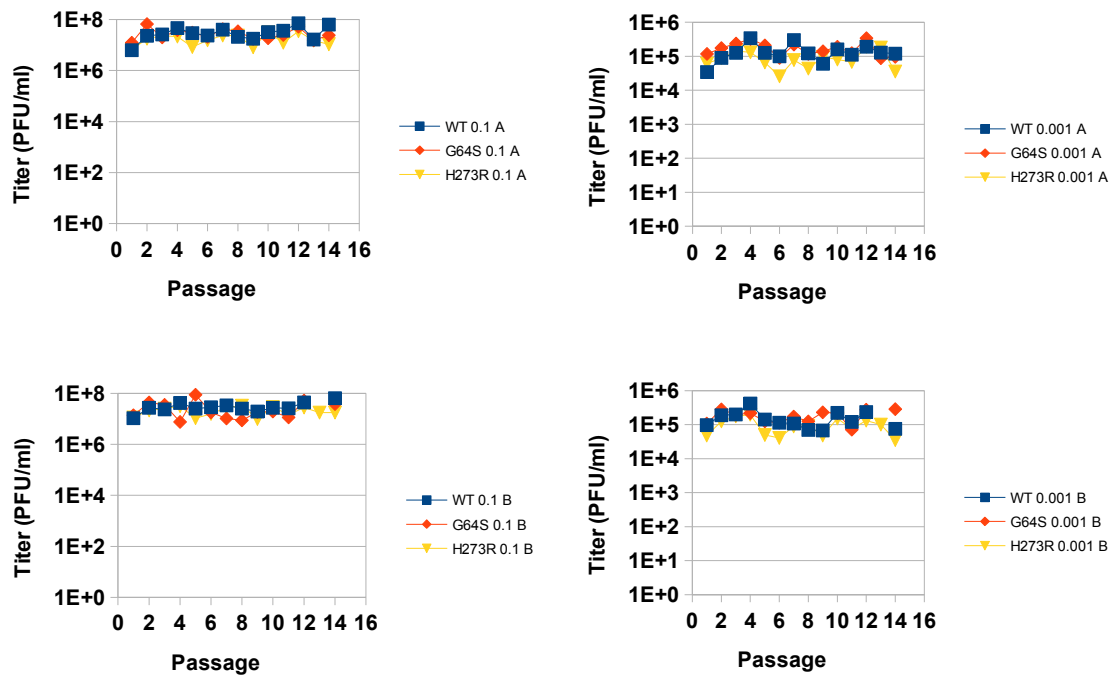
Figure 2:

Titering of well-controlled passages demonstrates no drop in titer of WT, G64S, and H273R after 15 passages.

Determination of H273R population mutation frequency

Previous results from the lab indicated a mutation frequency of nearly 2-fold higher than WT mutation frequency in an H273R population. As is standard when a viral stock is generated in the lab, it was characterized. In addition to the standard titering characterization, it was necessary to determine the mutation frequency of our H273R population. Concerns have been raised about biases in the determination of a stock's mutation frequency. In plaque-purifying to determine mutation frequency, it is possible and depending on the timing of plaque isolation, likely, that an experimentalist may accidentally bias his or her results. Plaques created by more-fit infectious particles are likely to form and grow faster. Therefore, the earliest visible plaques likely contain the most fit (presumably WT) founding genotypes. By picking plaques early and sequencing the population that comes form that plaque, an experimentalist can bias his or her results towards a more WT population. By picking plaques late and sequencing the population that comes form that plaque, an experimentalist can bias his or her results away from WT.

In order to alleviate this bias and create an unbiased approach at population mutation frequency analysis, I used an endpoint dilution approach to acquire samples expanded from single PFUs. By diluting stock virus to a very dilute solution and plating over cells in a 96-well plate, one can acquire plaque-like founding populations for mutation frequency analysis without the burden of precision in plaque picking or the bias of the timing of plaque picking. Wells that have come to cytopathic effect (CPE) are harvested and numbered, then frozen. After all infected wells reach CPE, samples are

17

randomized and a contingent is sequenced.

After sequencing the typical 24 amplified clones over the capsid region, sequencing analysis showed that both high-fidelity and low-fidelity mutants appear more WT-like than previously seen. Sequencing of over 50,000 bases revealed 10 mutations for WT, 5 mutations for G64S, and 13 mutations for H273R (Table 1). If the mutation frequency of this region is extrapolated over the entire genome, this equates to a mutation frequency of 1.3 mutations/genome for WT, 0.66 mutations/genome for G64S, and 1.75 mutations/genome for H273R (Table 1).

Population	Bases sequenced	Mutations	Mutations/genome*
WT	59095	10	1.3
G64S	56359	5	0.66
H273R	55412	13	1.75

Table 1:

Population mutation frequency analysis from 24 clones. Sequencing shows both the low-fidelity polymerase mutant population (G64S) and the high-fidelity polymerase mutant population (H273R) as more WT-like in mutation frequency than previously described. *Mutation frequency in the capsid region has been extrapolated to the entire genome.

Examination of H273R tissue tropism and lethality

Lethality and tissue tropism experiments were carried out in a mouse model of poliovirus infection with the above-described WT and H273R populations. Lethal dose, 50% (LD_{50}) studies of WT and H273R indicate that H273R is less lethal than WT in the mouse model of infection. It is, however, more lethal than previously described in the lab- nearly 10-fold more lethal (Figure 3). At non-lethal concentrations, H273R causes paralysis in both hind legs but the infection appears to be cleared and the mice regain the use of the previously-paralyzed limbs. Tissue tropism experiments indicate that my H273R population does appear to establish infection in several tissues, and is able to cross the blood-brain barrier and infect the brain (Figure 4).

H273R LD50

Legend:
- MS H273R 1e8
- MS H273R 1e7
- MS H273R 1e6
- WT 1e7
- MV H273R 2e7
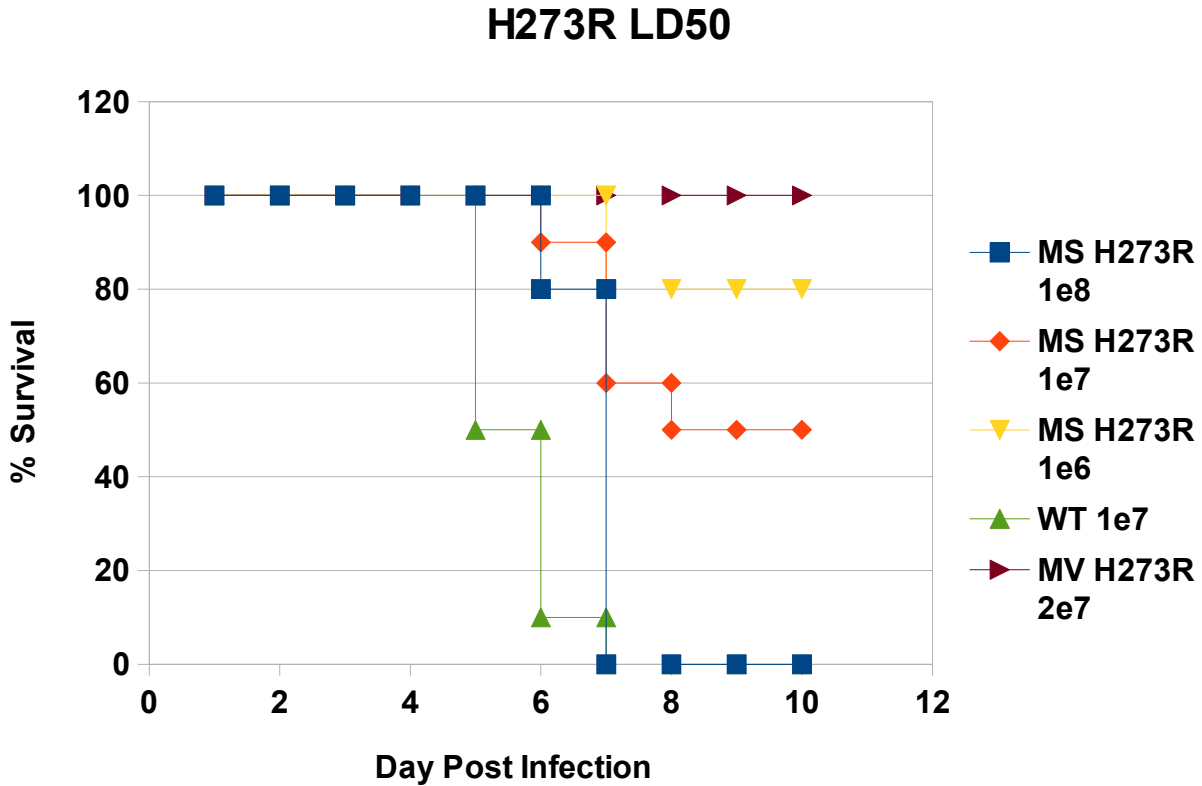
Figure 3:

Survival curve, or LD$_{50}$, of my (MS) H273R population alongside WT population and a previously made H273R population (MV H273R). The LD$_{50}$ of MS H273R population was 6.25×10^7 while the LD$_{50}$ of the MV H273R was $>4.0 \times 10^8$.

Tissue Tropism of H273R population
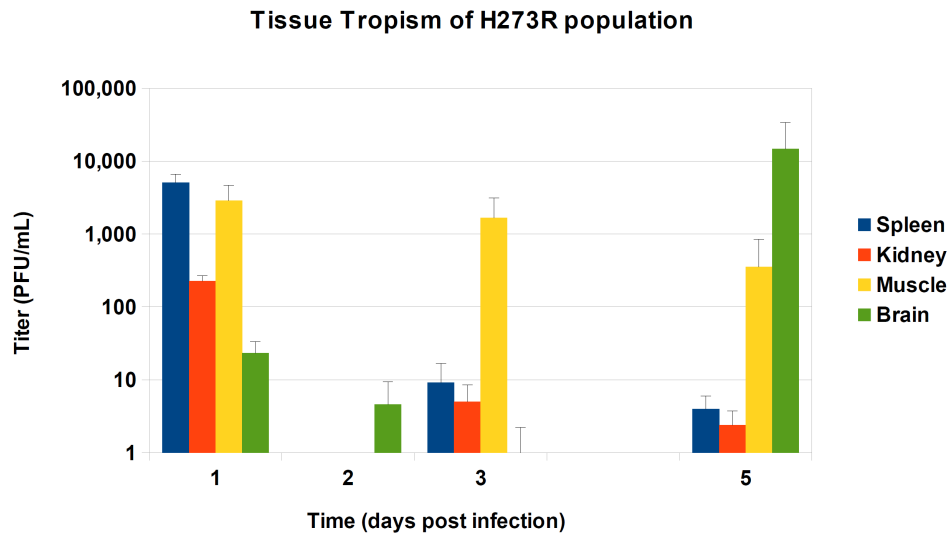
Figure 4:

Tissue tropism of H273R population. Infection was established in muscle and was present in the brain of 4 of 5 animals at 5 days post infection. This H273R population is clearly able to establish infection in a mouse model of infection and cross the blood-brain barrier to cause lethality. Note- only brain was titered on day 2. Also, samples from day 4 were not titered.

22

Discussion

In this work our objective was to describe the characteristics of a low-fidelity polymerase mutant population and investigate whether Muller's ratchet and subsequent mutational meltdown occurs in tissue culture and animal systems. Our experiments used sequencing, passaging, and infection in a mouse model to elucidate the nature of low-fidelity replication in population dynamics of poliovirus. Unfortunately, previous results were not reproducible and our H273R population appeared less low-fidelity and more WT-like than previously observed.

Error Catastrophe

Our tissue culture passaging experiments illustrated that tightly controlled passaging of H273R populations at either low or high MOI does not result in Muller's ratchet-like behavior or subsequent mutational meltdown. Previous results were likely artifactual as less well-controlled passaging (blind passaging) demonstrated the ability of small, accidental differences in multiplicity of infection to be compounded through subsequent passages creating the illusion of extinction.

However, recent results from the Cameron lab suggest that the H273R variant may exhibit error catastrophe when passaging occurs with a very small population. Clearly, the exact conditions under which populations are generated and passaged appear to yield different results.

Pathogenicity

Upon infection in an animal model, our H273R population also exhibited the ability to establish infection and spread to multiple tissues. It also displayed considerably more lethality than previously observed, with the ability to cross the blood-brain barrier to high titers by 5 days post infection. While these studies reinforce the conclusion that viral population diversity and disease pathogenesis are tightly linked, they fall short of producing a useful viral population for further study, as our H273R population displays more WT-like attributes than previously described.

References

1. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439:344-348.

2. Vignuzzi M, Andino lab notebook, 2004.

3. Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 1: 2-9.

4. Lynch M, Burger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84(5):339-344.

5. Holland JJ, De La Torre JC, Steinhauer DA (1992) RNA virus populations as quasispecies. *Curr Top Microbiol Immunol* 176:1-20. Review.

Chapter 2:

Single-Cell Analysis Uncovers

Extensive Biological Noise in

Poliovirus Replication

Introduction

When a virus infects a cell, it sets in motion a complex set of reactions. Some reactions, programmed by the viral genome, lead to virus replication and progeny production, while others, inherent to the host, act to restrict or limit viral replication. It is unclear how these contrasting forces shape the outcome of an infection. In principle, an infection is a seemingly deterministic series of processes – uncoating, translation, replication, encapsidation. However, infections often begin with so few molecules that the progress of any given infection may occur in a more stochastic manner than is often appreciated (1). Indeed, individual cells in a population infected with the same virus at the same multiplicity of infection have been observed to produce varied levels of viral progeny. The first rigorous observation of this variation during infection was made using single bacteriophage infections, where the large distribution in burst size (progeny per infected cell) could not be explained simply by the distribution in bacterial size (2). More recently, the effect of cell size on virus yield was also examined in a mammalian RNA virus (3). This study confirmed that while host cell size is a contributing factor to virus yield, it is insufficient to explain the variation in burst sizes. The source of variation remains unknown. We hypothesized that by removing cell-size-dependent variation, we should be able to uncover the extent of stochasticity in viral infection and define the contribution of other factors to the overall productivity of single-cell infections. Understanding this question may illuminate the dynamics of infection and pathogenesis and has implications to design therapeutic and preventive strategies.

Here we examined if non-deterministic, stochastic processes play a role in the outcome of viral infections. We determined the contribution of noise to RNA synthesis and infectious particle production in single-cell infections from cell-size selected populations. From each infected cell we accurately measured the generation of positive-strand RNA genomes; of negative-strand RNA templates, which are used as templates of replication for the positive-strand genome; and of infectious particles. Our measurements defined the variation in genome and viral progeny production across a cell population and allowed us to determine the correlation between the synthesis of viral RNA and infectious virus particle production in individual cells. Surprisingly, we do not observe tight correlations between the distributions of genomes and viral progeny, suggesting that stochastic effects have a significant impact on the outcome of infection. Furthermore, by comparing variation and stochasticity in cells infected at low and high multiplicity of infection, we observe that the sources of biological noise are different when cells are infected with multiple viral particles. While at low multiplicity of infection the kinetics of the early replication events is a significant source of variation, at high multiplicity of infection, access to cellular resources becomes a determining factor in the outcome of infection. Our findings have important implications for the evolution of viral strategies of transmission and pathogenesis and raise the question of how viruses balance the distinct dynamics in singly- and multiply-infected cells.

Materials and Methods

Cells and viruses

HeLaS3 cells (ATCC CCL-2.2) provided by R. Geller and J. Frydmann (Stanford University) were maintained in 50% DMEM/50% F-12 medium supplemented with 10% newborn calf serum, 100U/ml penicillin, 100U/ml streptomycin, and 2mM glutamine (Invitrogen). Poliovirus Mahoney typeI genomic RNA was generated from *in vitro* transcription of prib(+)XpAlong. To generate virus, 20 µg of RNA was electroporated into 4 X 10^6 HeLaS3 cells in a 4mm cuvette with the following pulse: 300V, 24 Ω, 1000 µF. The resulting virus was passaged at high multiplicity of infection (MOI ~1-10) three times then subjected to ultracentrifugation through a 30% sucrose cushion.

Infections

HeLasS3 cells in 12-well plates were rinsed 2X with PBS to remove unattached cells then infected in 100µL at an MOI of 10 or 0.1 for 20 minutes at 37ºC. The inoculum was removed and cells were washed 2X with PBS to remove any unattached virus. Cells were given 2% serum media for 7 hours, then frozen at -70C. All lysates were thawed once on ice, then refrozen. Lysates were homogenized at a final concentration of 0.06% NP-40, incubated on ice for 20 min, then vortexed and aliquoted.

For serum starvation experiments, cells were plated in serum media overnight, rinsed 2X with PBS, then given serum-free media for 48 hours. After infection with virus-containing PBS, cells were given serum-free media for 7 hours, then frozen at -70C.

For PIK93 experiments, after infection, cells were given 2% serum media containing 1.5 uM PIK93 in DMSO or DMSO (mock treated) for 7 hours, then frozen at -70C.

Isolation of individual infected cells

HeLaS3 cells in a 6-well plate (1.335 million cells per well) were rinsed 2X with PBS to remove unattached cells then infected in 100μL at an MOI of 10 or 0.1 for 20 minutes at 37ºC. The inoculum was removed and cells were washed 2X with PBS to remove any unattached virus. The cells were detached with 200 μL 0.05% trypsin with EDTA for 3 min at 37ºC. Trypsin was inactivated with 3 mL 10% serum media and cells were pelleted at 300Xg for 3 min at 4ºC. Cells were resuspended in cold 2% serum media and placed on ice. Cells were sorted via light scattering to 1 cell per well of a 96-well plate containing 200 μL 2% serum media per well using a FACSAriaIII flow cytometer (BD Biosciences) under BSL2 conditions. Tight gating of both forward and side scatter was performed to restrict the size heterogeneity of the population as well as eliminate cell debris and cell aggregates. Gating removed greater than 60% of the cell population, resulting in forward scatter pulse width variance (seen to correlate well with cell volume, 4) with a coefficient of variation of 0.06. The same gating settings were used to isolate cells from both infections. Previous results indicate that this trimming of the cell population enriches for cells in G1 and S phase, which not only represent the most common cell cycle states (85% of a mixed state population, 5), but also have been shown to produce viral titers close to the average of a mixed state population (3). Isolated individual cells in 96-well plates were incubated for 7 hours at 37ºC, then frozen

at -70ºC. Plates were thawed once on ice, then refrozen. Lysates were homogenized by adding 20 µL of 0.66% NP-40 per well. Plates were sealed with adhesive plate seals (Thermo Scientific) and vortexed on high for 20 seconds. Plates were incubated on ice for 30 minutes, vortexed again, then samples were aliquoted into fourths.

"Infection find" assay

After homogenization, 30 µL of cell lysate from MOI of 0.1 infected cells was added to 30,000 cells in 1.8 mL 5% serum media in a deep-well 96-well plate and incubated for 5 days at 37ºC. According to Poisson statistics, 30 µL lysate ensures a 99% chance of transferring at least 1 PFU to the fresh cells as long as the original lysate contained at least 34 PFU. Growth of HeLaS3 cells in wells which received lysate from uninfected cells acidified the media allowing for colorimetric distinction of lysates which did and did not contain infectious virus.

Determination of PFU

For single cell measurements: 50 µL of lysate was used in the first dilution of a $TCID_{50}$ assay, resulting in a final concentration of 0.001% NP-40, which we found did not inhibit growth of HeLaS3 cells. For population measurements: 166 µL of lysate was used in the first dilution of a $TCID_{50}$ assay. Calculations were based on the Reed-Muench method. PFU was determined to be 3.3 times the $TCID_{50}$. Measurements above the limit of detection but below the limit of quantification were included in distribution histograms (Fig. 2Aiii, 2Bii) but excluded from other calculations and figures.

RNA extraction, Reverse Transcription (RT), and qPCR

RNA was extracted via the PureLink RNA Micro Kit (Life Technologies) according to the manufacturers instructions. cDNA was synthesized from total RNA using SuperScript III Reverse Transcriptase (Life Technologies) and 125 nM strand-specific RT primer (+strand_RT: 5'-GGCCGTCATGGTGGCGAATAATGTGATGGATCCGGGGGGTAGCG-3'; -strand_RT: 5'-GGCCGTCATGGTGGCGAATAACATGGCAGCCCCGGAACAGG-3') in a 5 μL reaction. Separate RT reactions for positive and negative-strand RNAs were performed for each sample. RT products were treated with 0.5 units of Exonuclease I (Fermentas) to remove excess RT primer prior to qPCR. Strand-specific qPCR was based on a published protocol (5). cDNAs were analyzed by qPCR using 2× SYBR FAST Master Mix (Kapa Biosystems), 200 nM strand-specific qPCR primer (+strand_For: 5'-CATGGCAGCCCCGGAACAGG-3'; -strand_Rev: 5'-TGTGATGGATCCGGGGGGTAGCG-3'), and 200 nM Tag primer (5'-GGCCGTCATGGTGGCGAATAA-3') in a 10 μL reaction. A 10× dilution series of *in vitro* transcribed positive- and negative-strand RNA standards was run alongside experimental samples and used to construct a standard curve.

Bootstrapping for confidence intervals

Confidence intervals were acquired in R (R Core Team, Vienna, Austria, [http://www.R-project.org/]) using the bootstrapping package "boot" (R package version 1.3-9, Canty A, Ripley B). 1000 bootstrap replicates were performed for each statistic.

Sucrose gradients

HeLaS3 cells were infected for 2, 4, or 6 hours at an MOI of 10 in 15 cm dishes and simultaneously treated with 100ug/ml cycloheximide (CHX) for 2 minutes at 37ºC. Cells were washed with PBS+CHX and lysed with 0.5% NP-40 lysis buffer containing CHX on ice for 20 minutes. Cell debris was pelleted in a table-top centrifuge at 2000 rpm for 10 minutes at 4ºC, then nuclei were pelleted at 9000 rpm for 10 minutes at 4ºC. Cell lysates were loaded on a 10%-50% sucrose gradient containing CHX and ultracentrifuged at 35000 rpm for 3 hr. Gradients were analyzed using a Biocomp Gradient Station with a BioRad Econo UV Monitor.

Results

Estimating experimental noise

To determine the variance within infection of individual cells, we infected HeLaS3 cells with poliovirus Mahoney type 1 at two multiplicities of infection (MOI): 10 and 0.1. Individual cells from each infection were isolated using a cell sorter with tight forward and side scatter gatings to restrict cell size (4). Cell lysates were obtained and divided, allowing multiple measurements to be taken from each cell (Fig. 1A). We measured positive-strand viral RNA, negative-strand viral RNA, and plaque-forming units (PFUs) produced by each individual cell.
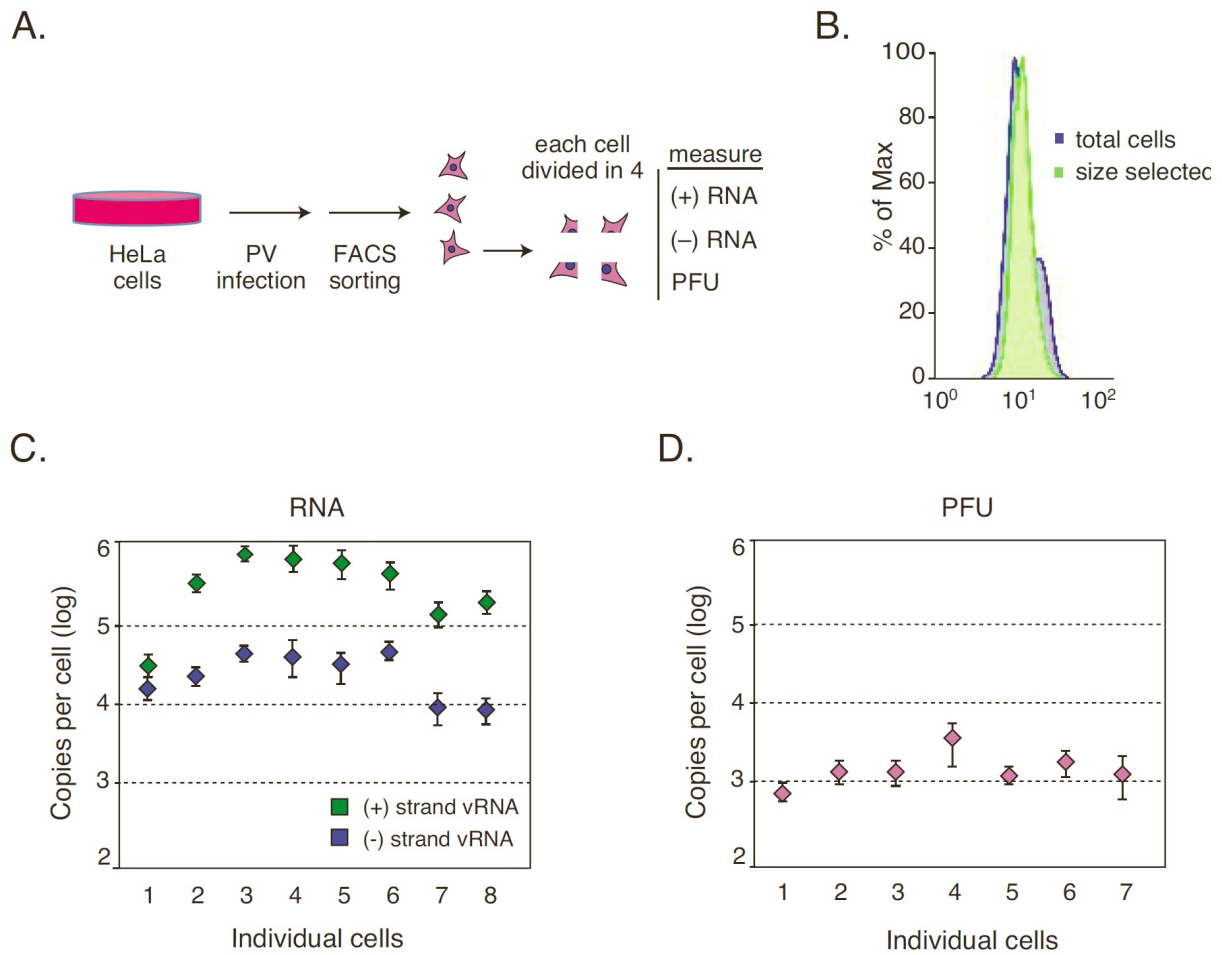
Figure 1:

Measurement error. (A) Schematic of single-cell isolation, division, and measurement.

(B) Chromosomal copy number analysis shows enrichment of G1/S phase cells. (C)

qRT-PCR measurements of positive-strand viral RNA and negative-strand viral RNA

from 8 quartered control cells had average CVs of 0.21 and 0.23, respectively. (D) PFU

measurements from 7 quartered control cells had an average CV of 0.31. Error bars

indicate standard deviation.

To estimate experimental error of our measurements, we divided single cells into four fractions, and we measured either viral RNA concentration or virus production in each of the four fractions separately. We utilized the coefficient of variation (CV), or relative standard deviation, for each set of measurements to normalize measurements with distinct mean values. Measurements of positive-strand RNA genomes and negative-strand RNA templates by qPCR from 8 control cells had average CVs of 0.21 and 0.23, respectively (Fig. 1B). Measurements of plaque-forming units (PFUs) from 7 control cells had a slightly higher average CV of 0.31 (Fig. 1C). The standard deviation measured in this experiment must result from experimental noise derived from either cellular lysate fractionation into quarters, extraction, and/or all downstream analysis steps. Based on these measurements, we assumed that variance significantly greater than this experimental error represents the biological noise characteristic of the dynamic biological processes under study.

Fluctuations in virus replication revealed by measurements in single infected cells

We first carried out an experiment in which cells were infected with a high multiplicity of infection. We isolated 106 cells infected at an MOI of 10. These cells had positive-strand RNA genome measurements ranging from 13,907 copies to 720,360 copies/cell (Fig. 2Ai). The mean positive-strand RNA genomes/cell was 230,620 copies with a CV of 0.62 (Fig. 2C). Measurements of negative-strand RNA templates ranged from 2,157 copies to 45,990 copies/cell (Fig. 2Aii). The mean negative-strand RNA

templates/cell was 16,865 copies with a CV of 0.61 (Fig. 2C). Thus, the distribution of

RNA accumulation from individual infected cells is highly disperse, with some cells

producing few copies of viral RNA and others extremely high concentrations. PFU

measurements from these same cells ranged from below our limit of detection of 269 to

4225 PFU/cell (Fig. 2Aiii). Within the range we can detect PFU, the mean PFU per cell

was 976 with a CV of 0.66 (Fig. 2C).

A.

i. — (+) Strand RNA copies per cell (10^5) vs Number of cells

iii. — PFU per cell (10^2) vs Number of cells

ii. — (−) Strand RNA copies per cell (10^3) vs Number of cells

iv. — (+) strand/(−) strand ratio vs Number of cells

MOI=10

B.

i. — (+) Strand RNA copies per cell (10^3) vs Number of cells

ii. — PFU per cell (10^2) vs Number of cells
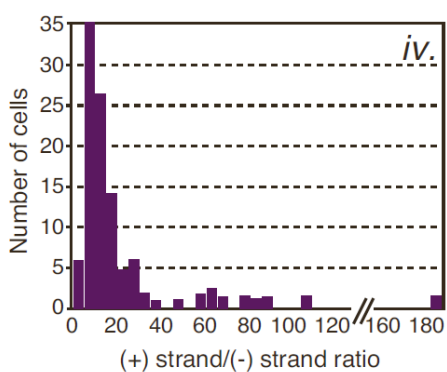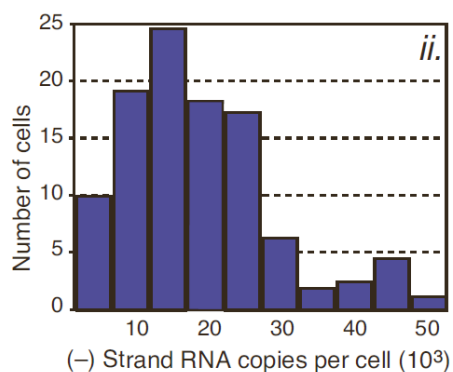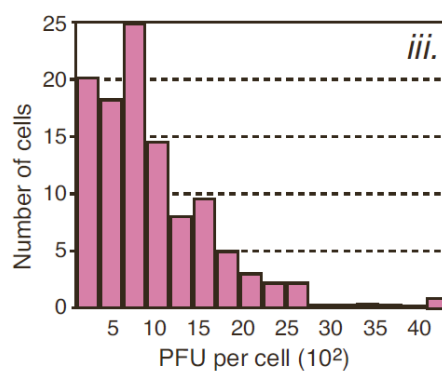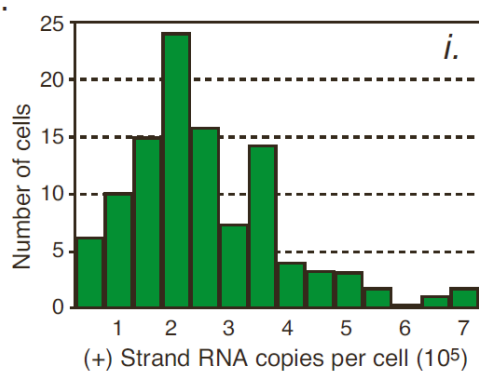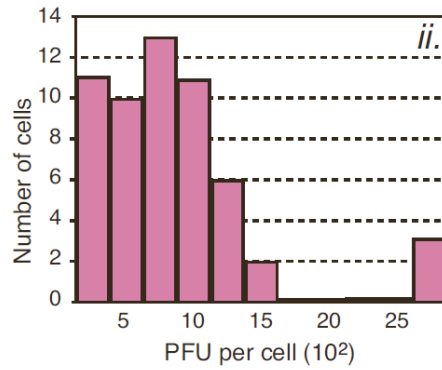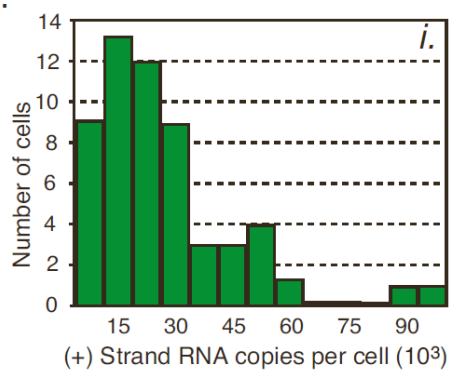
MOI=0.1

C.

	(+) Strands		(−) Strands		(+)/(−) ratios		PFU	
MOI	mean	variance*	mean	variance	mean	variance	mean	variance
10	230,620	0.62	16,865	0.61	20	1.3	976	0.66
0.1	22,870	0.82	−	−	−	−	860	0.66

Figure 2:

Distributions of products from single-cell infections. (A) RNA and infectious virus distributions from cells infected at an MOI of 10; n=106, (i) positive-strand RNA, (ii) negative-strand RNA, (iii) Infectious virus, (iv) positive-strand to negative-strand ratio. (B) RNA and infectious virus distributions from cells infected at an MOI of 0.1; n=56, (i) positive-strand RNA, (ii) Infectious virus. (C) Table summarizing means and variances from single-cell distributions. *variance expressed as CV. Positive-strand RNA from MOI of 10 infections had a CV of 0.62 (95% confidence interval: 0.54-0.72). positive-strand RNA from MOI of 0.1 infections had a CV of 0.82 (95% confidence interval: 0.67-1.01).

We report the first true average positive to negative strand ratio for a positive-strand virus in infected cells, with a ratio of 20 at 7 hours post infection. Previous studies have reported the ratio of average population measurements, which is statistically distinct, and arguably less informative, than the ratio measured within single-cells described here. Cells displayed a wide range of positive to negative strand ratios ranging from 2 to almost 200 resulting in a CV of 1.3. While we argue that the average ratio of 20 measured from within single-cells is a more accurate description of true strand ratio, we use the same data to compute a ratio of averages of 14 to compare to previous methods. Previously reported ratios of averages of 36 (7) and 30 (8) at 6 hours post infection along with observed decreases in the ratio between 4 - 6 hours post infection (7) suggest that our ratio of averages of 14 at 7 hours post infection is in line with previous studies using different methodologies.

We next determined whether initiating infection at a low multiplicity (MOI of 0.1) further increased the biological noise in virus replication. Given that the majority of cells are expected to be uninfected at an MOI of 0.1 (9), it was necessary to identify infected cells from the population of cells before measuring RNA and infectious virus production. Accordingly, a small portion of isolated cell lysate was removed and used to infect fresh uninfected cells to determine the presence of infectious virus (Methods). A total of 56 infected cells were then analyzed for RNA and virus production. These infected cell extracts contained between 1,580 and 91,897 copies/cell of the positive-strand RNA genome (Fig. 2Bi). This represents a mean positive-strand RNA genomes/cell of 22,870 copies with a CV of 0.82 (Fig. 2C). While negative-strand RNA templates were

detectable in these cells, the amount of copies accumulated was too low for confident

quantification. PFU measurements from these cells ranged from below our limit of

quantification to 2713 PFU/cell (Fig. 2Bii). The mean PFU/cell was 860 with a CV of

0.66 (Fig. 2C).

Biological noise is greater in infections initiated at low MOI

To determine the contribution of the number of viruses initiating infection to the

fluctuations observed in RNA and virus yield, we compared RNA and virus production

distributions from an MOI of 0.1, where infections likely began with only a single

infectious particle, with distributions from an MOI of 10, where infections began with a

Poisson distribution around 10 infectious particles. We find that infections that began at

an MOI of 10 produce, on average, 10 times more genomes than those initiated at an

MOI of 0.1 (i.e. infections initiated by a single genome, Fig. 2C). This is consistent with

the idea that RNA replication is independently initiated by individual genomes, so more

initiating genomes should in principle produce proportionally more viral RNA.

Interestingly, we observed a significant difference in the fluctuations of genome

productivity between the two infections, with a CV of 0.62 for an MOI of 10, and a

significantly larger CV of 0.82 for an MOI of 0.1 (Fig. 2C). We argue that this difference

in variance is likely to originate from the kinetic stochasticity of the early reactions that

lead to productive infection. Infections beginning with multiple viral genomes (MOI of 10)

average the effect on individual RNA replication reactions, are overall less susceptible

to stochastic kinetics, and therefore produce a number of genomes closer to the mean

41

than infections beginning with a single genome (MOI of 0.1).

The single infectious particles that initiate low MOI infections are clearly more sensitive to effects on RNA replication. Low MOI infections performed with sub-inhibitory concentrations of the RNA replication inhibitor PIK93 (10), produced significantly less genomes and PFU than high MOI infections performed under the same conditions (Fig. 3). Low MOI infections produced less than 75% as many genomes and less than 40% as many PFUs as high MOI infections relative to mock-treated cultures. Low MOI poliovirus infections have previously been reported to produce less genomes than high MOI infections in the presence of this inhibitor (10). We believe longer delays in the initiation of the initial reactions, or increased kinetic stochasticity, caused by RNA replication inhibition dampens the productivity capacity of low MOI infections.
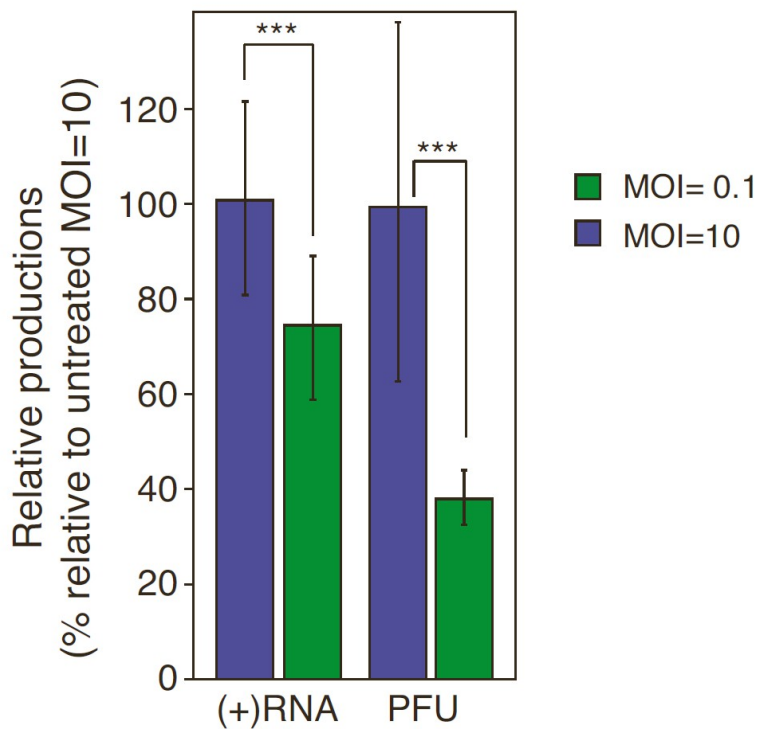
Figure 3:

Low MOI infections are more sensitive to RNA replication inhibition. In the presence of 1.5 uM PIK93, MOI of 0.1 infections produce less than 75% as many genomes and less than 40% as many PFUs as MOI of 10 infections compared to mock-treated cultures; n=6. ***Student's t-test p<0.05. Error bars indicate standard deviation.

Infectious particle production is independent of MOI and RNA production

Surprisingly, the effects of MOI and stochastic noise did not determine infectious

particle production. The mean and variance of the distributions of plaque forming units

between the two infections were nearly equivalent (Fig. 2C). This observation suggests

that infectious virus production is modulated by a limiting step in the infection that is not

largely affected by the MOI-dependent stochastic noise. Furthermore, these

observations were also confirmed at the population-level where, genome production is

proportional to MOI, but infectious virus production appeared to be independent of MOI

and restricted by factors that do not depend on the initial kinetics of infection (Fig. 4A,

4B). We speculate that infectious virus production is limited by a global cellular resource

required for virus particle formation. Consistent with this hypothesis, we observed a

slightly higher mean PFU/cell in cell populations which had not been size-restricted

(population PFU measurements, Fig. 4B). This global cellular resource limitation does

not appear to be affected by serum starvation (Fig. 4C). Polysomes profiling analysis on

sucrose gradients indicates that free 40S and 60S ribosomal subunits are available

throughout an infection, suggesting that ribosomal subunits are not limiting for virus

production (Fig. 4D).

A. Single-cell measurements

B. Population measurements

C.

D.

45
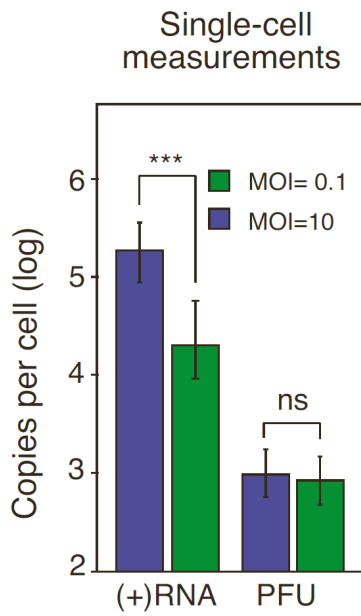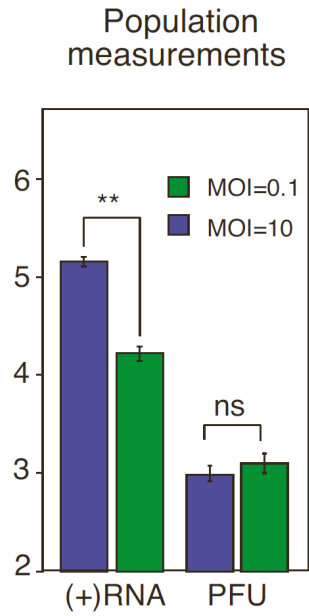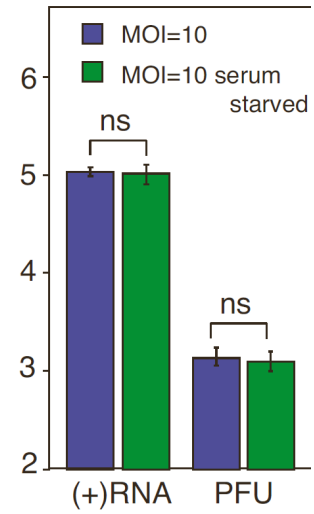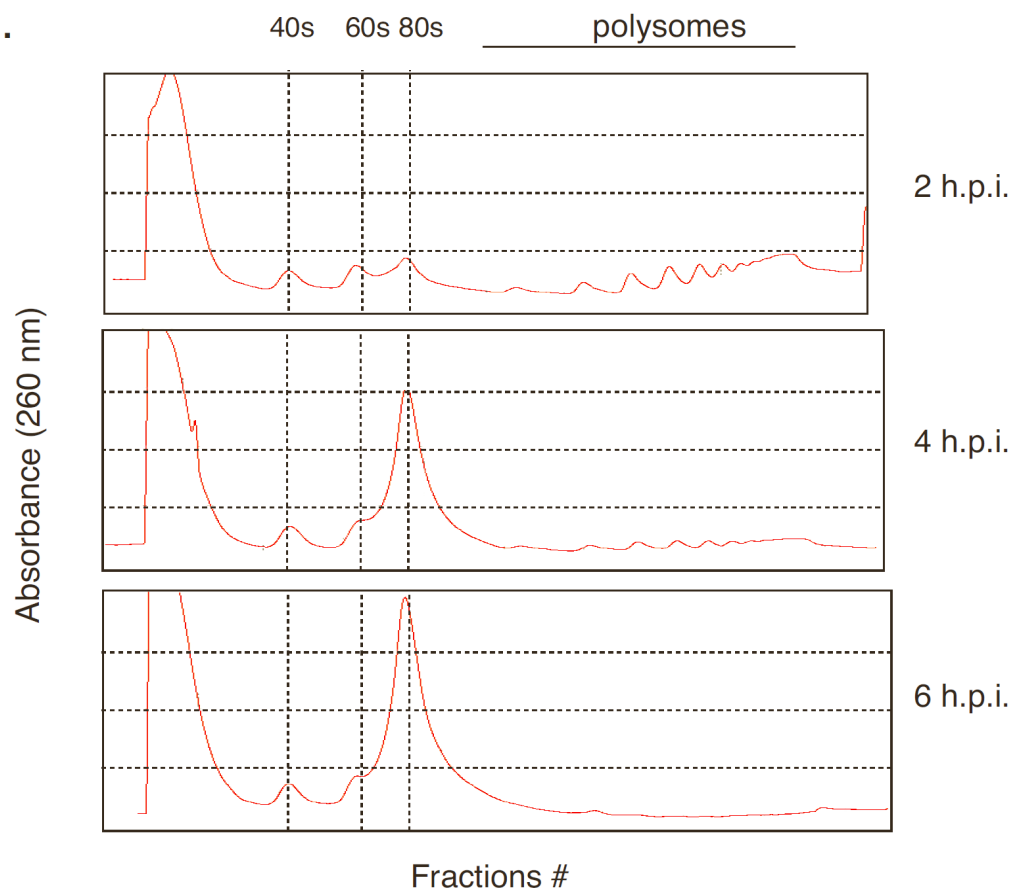
Figure 4:

Genome production scales with MOI, but PFU production does not. Comparison of single-cell measurements and population measurements: (A) Single-cell measurements, MOI=10: RNA n=106, PFU n=86; MOI=1: RNA n=56, PFU n=45. (B) Population measurements, n=6. (C) Serum starvation has no significant effect on RNA or PFU production of MOI of 10 infections; n=6. **Student's t-test P<0.001. ***Kolmogorov-Smirnov test P<0.001. ns, not significant. Error bars indicate standard deviation. (D) Ultraviolet absorbance at 260 nm of sucrose-gradient fractionated lysates from a time course of poliovirus infections show free 40S and 60S ribosomal subunits throughout infection.

Correlations between RNA and virus production at single-cell level

Next, we took advantage of our ability to take multiple measurements per individual cell to determine correlations between RNA synthesis and virus production. By comparing the quantity of multiple viral products from each cell, we were able to determine how biological noise affects early (RNA production) and late (particle production) steps in virus replication. At a low multiplicity of infection (MOI of 0.1), we observed a relatively good correlation between genomes and infectious particles produced per cell (nearly 30%, Fig. 5A). This suggests that although the initial reactions of low MOI infections are highly susceptible to the effects of kinetic stochastic noise, once these reactions have initiated, the infection progresses in a fairly deterministic manner. The variation in the initial steps of translation and RNA synthesis contribute significantly to the overall trajectory of the infection. This is in line with data from Fig. 3, which suggests that dampening the kinetics of these initial reactions lowers the average productive trajectory.

Given that infections starting with a higher MOI are less susceptible to this initial kinetic noise (Fig. 2C), we expected a better correlation between the distributions of viral RNA and infectious particle production for the infection starting at an MOI of 10. Surprisingly, we observed very little correlation between the distributions of RNA and infectious virus from the MOI of 10 infection- 4% between genomes and PFU, and 8% between genomes and negative-strand RNA templates (Fig. 5B, 5C). Because these correlations were obtained from measurements of individual infections, we can exclude the Poisson distribution of initial infecting genomes as a contributing factor to this

relationship. This lack of correlation reveals stochastic influences in multiply-infected

cells beyond the kinetic noise of the initial reactions and may be the result of local

fluctuations within individual cells of critical factors involved in virus replication.

A.

MOI = 0.1 $R^2 = 0.29$

B.

MOI = 10 $R^2 = 0.04$

C.

MOI = 10 $R^2 = 0.08$

D.

- MOI=0.1
- MOI=0.1 serum starved

49
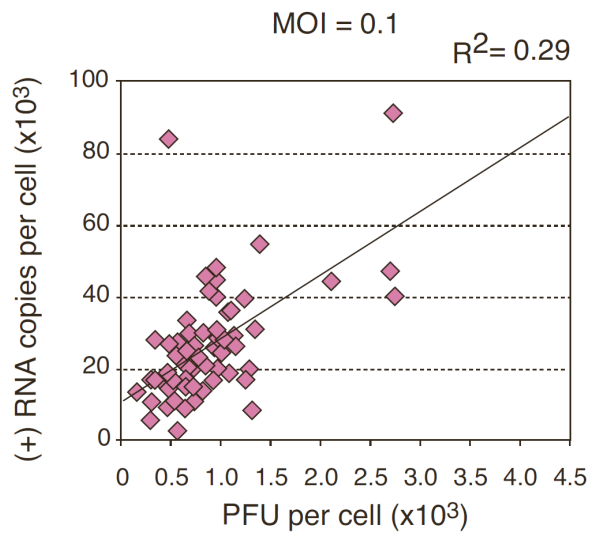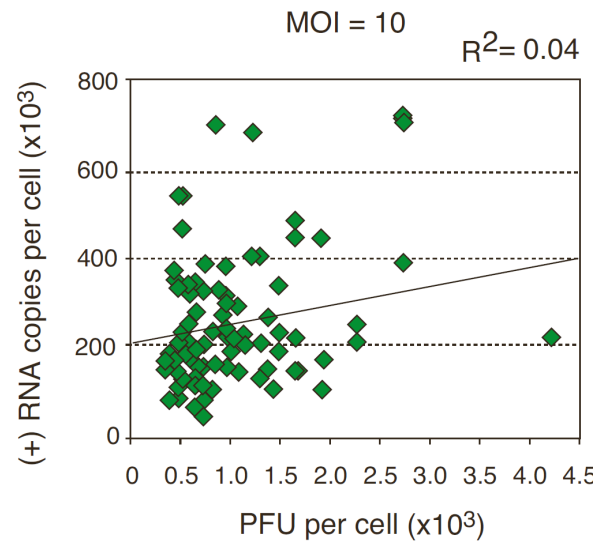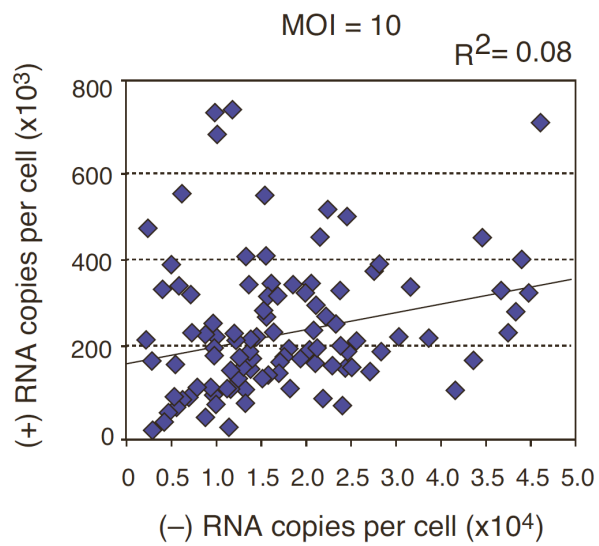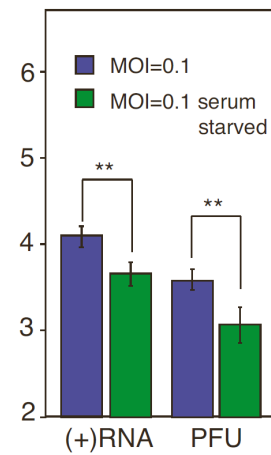
Figure 5:

Correlations between products of single-cell infections and effects of serum starvation on low MOI infections. (A) Correlation between positive-strand RNA and PFU from an MOI of 0.1 infection shows a significant portion of the noise in one product can be explained by the noise in the other (R^2=0.29, 95% confidence interval: 0.05-0.76); n=45. (B) Correlation between positive-strand RNA and PFU from an MOI of 10 infection shows a poor correlation (R^2=0.04, 95% confidence interval: 0.00-0.19); n=86. (D) Serum-starved MOI of 0.1 infections produce less than 40% as many genomes and less than 35% as many PFUs as serum-fed MOI of 0.1 infections; n=6 **Student's t-test p<0.005. Error bars indicate standard deviation.

Low MOI infections provide critical insights into this local resource stochasticity. When low MOI infections initiate in serum-starved cells, significantly less genomes and PFU are produced. Low MOI infections of serum-starved cells produced less than 40% as many genomes and less than 35% as many PFUs as serum-fed cultures (Fig. 5D). Serum starvation does not affect the global productive capacity of high MOI infections (Fig. 4C). Therefore, serum starvation must be creating local deficiencies in resources to which only low MOI infections are sensitive. We believe local variation in resources create hotspots and coldspots for initiating infection. In serum-starved cells, fewer hotspots exist, and single infectious particles have less of a chance of initiating at a hotspot and establishing a productive infection.

Discussion

The determinants of a successful viral infection are not well understood. Here we examine whether the outcomes of infection are deterministic or governed by stochastic events. We find considerable variability in the outcomes of an infection. The sources of noise appear different for viral infections beginning with only one or many infectious particles. Infections initiated with only a single infectious particle have the highest degree of variability in viral replication, independent of cell size. In infections initiated by multiple infectious particles we observed less variance in the distribution of RNA accumulation, suggesting that the initial steps of infection are particularly susceptible to stochastic kinetics. However, in multiply-infected cells, the lack of correlation between RNA produced and number of infectious particles produced suggests that when viral RNA replication initiates in multiple locations, the infection suffers from access to host resources. Interestingly, particle production also demonstrates a high degree of fluctuation (Fig 2C), indicating additional sources of variation exist at later stages of the replication cycle.

Our data highlights the stochastic nature of the interaction between the infecting virus and the host cell and indicate that biological noise significantly impacts the distributions of viral RNA and particle production. The source of noise may be variations in the biochemical reactions that underlie virus replication which stem from resource limitations or simply variations in the kinetics of the initiation of an infection. Infections that begin quickly would be expected to transition to exponential growth more rapidly

52

and be highly productive, while infections that falter or lag would, conversely, be less productive. In addition, cells have many mechanisms to block viral replication, whereas all viruses have mechanisms to evade them and hijack the cellular resources. Differences in the effectiveness of host innate immune responses at the earliest times of infection could have a significant impact on the initial kinetics of replication.

A major observation from this study was that despite the difference in production of genomes between singly-infected and multiply-infected cells, similar amounts of virus are produced. The observation of similar burst size distributions between singly-infected and multiply-infected cells was first made by Delbruck using phage (2). More recently, it was reported that virus yield of vesticular stomatitis virus, a mammalian RNA virus, does not correlate with multiplicity of infection (12), but it was unknown what limited virus production. Clearly, at least in the case of poliovirus, there are sufficient cellular resources for an order of magnitude increase in genome production with MOI but a global cellular resource limits virus particle production. This is likely due to differences in the nature of these processes. Viral genome replication is dependent on the polymerase, which is an enzyme and can be utilized repeatedly while virion production is dependent on capsid proteins, which are continuously consumed during virion synthesis. The synthetic requirements for viral capsid protein production exceed those for genome replication and the synthesis of other viral proteins for several additional reasons: 1) virions are comprised of a single genome and 60 copies of each of the viral capsid proteins- meaning that production of a virion requires only a single genome replication event but 60 translation events (12), 2) the rate of genome replication within

infected cells is estimated to be 5 times faster than the rate of translation (13), and 3) because of the functional compartmentalization of viral nonstructural proteins (14, 15), cytosolic concentrations of capsid proteins are likely much lower than the compartmentalized concentrations of the nonstructural proteins- meaning that replication occurs with higher, localized concentrations of viral nonstructural proteins while encapsidation occurs under more dilute, global concentrations of viral capsid proteins.

A surprising observation from this study was the importance of apparent local resource variability on the initiation of infection. Because infections initiated by multiple infectious particles are thought to proceed by generating multiple, compartmentalized replication foci that progress through the infection with limited crosstalk (14-16), local differences in the subcellular availability of resources for replication, such as ribosomes, energy, nucleotides, membranes, etc, likely create variation in the productive capacity of each focus. In multiply-infected cells, noise could arise from random fluctuations in the spatial organization of reactions as these foci are subject to more limited and variable access to cellular resources. As a result of this variability, asynchronies or inefficiencies could be introduced to the overall replication cycle that could skew the productivity of each infection thus generating "architectural" or spatial stochasticity. Similarly, local resource variation appears to create hotspots and coldspots of replication initiation, as evidenced by the differential effects of serum starvation on high and low MOI infections (Fig. 4C, 5D).

Similarities between the burst size distributions of singly-infected and multiply-

infected cells were originally interpreted to stem from a self-interference phenomenon whereas bacteria which were "simultaneously infected with several virus particles of the same kind" would "react as if only one of the virus particles was effective" (17). Our data suggests that, at least for poliovirus, self-interference at the level of genome synthesis does not occur. It is possible that a fundamental difference exists between bacteria and mammalian host cells, that, for example, restricts bacteriophages to only a single nucleation site for replication while allowing mammalian viruses many sites of replication. The observations that self-interference of genome synthesis does not occur and that the mammalian host cells restricts virion production creates an intriguing question: what advantage does the virus gain by allowing multiple infecting genomes to replicate? Is it simply a kinetic issue, with the benefit coming from the virus making the maximum number of virions before the cell mounts a response or apoptoses (18)? Or is there a more complex answer, with the benefit coming from multiple infecting genotypes creating a more diverse population of progeny to gain adaptive advantage (19)? Future studies should define the particular cellular resource limitation imposed on virus production and elucidate the evolutionary advantage of the mechanisms of replication within multiply-infected cells.

Our observation of how random events play a perhaps larger than envisioned role in infection is relevant to the understanding of the dynamics of virus infection and its consequences for adaptation. For example, cells experiencing low-productivity infection may be more likely to evade host immune responses, which may in turn extend the overall infection cycle time and increase the chances for the virus to spread from host to

host (20). During infections, hosts are infected with a small number of viruses, which continuously face strong bottleneck events (21). In this scenario, variance in viral replication dynamics at the single-cell level can have a strong effect on the outcome of infection. Considerable attention has been given to the idea that organism survival and reproduction are subject to stochastic fluctuations and chance rather than fitness alone (22). Stochastic processes during replication lead to sampling errors over generations. These sampling errors, i.e. random genetic drift, can cause significant changes in the abundance of genetic variants. Such genetic drift can represent a significant hurdle for adaptation, and therefore can play a large role in determining the fate of an infection. Newly arising beneficial mutations may often be lost by chance and may need to occur many times before they succeed in reaching fixation (23), thus reducing the chance of effective adaptation to a new host or switch in the environment. Stochasticity manifested as genetic drift can also serve to isolate neutral mutations, opening new evolutionary avenues for virus adaptation (24). Future work will be necessary to extend these observations and elucidate the precise relationship between biological noise and virus adaptation.

References

1. Hensel SC, Rawlings JB, Yin J. 2009. Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. Bull Math Biol 71:1671-1692.

2. Delbruck M. 1945. The burst size distribution in the growth of bacterial viruses (bacteriophages). J Bacteriol 50:131-135.

3. Zhu Y, Yongky A, Yin J. 2009. Growth of an RNA virus in single cells reveals a broad fitness distribution. Virol 385:39-46.

4. Tzur A, Moore JK, Jorgensen P, Shapiro HM, Kirschner MW. 2011. Optimizing Optical Flow Cytometry for Cell Volume-Based Sorting and Analysis. PLoS ONE 6:e16053.

5. Hahn AT, Jones JT, Meyer T. 2009. Quantitative analysis of cell cycle phase durations and PC12 differentiation using fluorescent biosensors. Cell Cycle 8:1044-1052.

6. Plaskon NE, Adelman ZN, Myles KM. 2009. Accurate strand-specific quantification of viral RNA. PLoS One 4:e7468.

7. Novak JE, Kirkegaard K. 1991. Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. J Virol. 65(6): 3384-3387.

8. Andino R, Rieckhof GE, Baltimore D. 1990. A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. Cell 63:369-380.

9. Condit RC. 2001. pp 42-43 *In* Fields virology, Knipe DM, Howley PM (eds),

Lippincott Williams & Wilkins, Philadelphia, PA.

10. Arita M, Kojima H, Nagano T, Okabe T, Wakita T, Shimizu H. 2011. Phosphatidylinositol 4-kinase III beta is a target of enviroxime-like compounds for antipoliovirus activity. J Virol 85:2364-2372.

11. Timm A, Yin J. 2012. Kinetics of virus production from single cells. Virol 424: 11-17.

12. Hogle JM, Chow M, Dilman FJ. 1985. Three-dimensional structure of poliovirus at 2.9 A resolution. Science 229:1358-1365.

13. Regoes RR, Crotty S, Antia R, Tanaka MM. 2005. Optimal replication of poliovirus within cells. Am Nat 165:364-373.

14. Novak JE, Kirkegaard K. 1994. Coupling between genome translation and replication in an RNA virus. Genes Dev 8:1726-1737.

15. Egger D, Teterina N, Ehrenfeld E, Bienz K. 2000. Formation of the poliovirus replication complex requires coupled viral translation, vesicle production, and viral RNA synthesis. J Virol 74:6570-6580.

16. Baltimore D, Girard M, Darnell J. 1966. Aspects of the Synthesis of Poliovirus RNA and the Formation of Virus Particles. Virol 29: 179-189.

17. Delbruck M, Luria SE. 1942. Interference between bacterial viruses. Arch Biochem 1:111-141.

18. Agol VI, Belov GA, Bienz K, Egger D, Kolesnikova MS, Romanova LI, Sladkova LV, Tolskaya EA. 2000. Competing death programs in poliovirus-infected cells: commitment switch in the middle of the infectious cycle. J Virol 74:5534-5541.

19. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439:344-348.

20. Kutsch O, Benveniste EN, Shaw GM, Levy DN. 2002. Direct and quantitative single cell analysis of human immunodeficiency virus type 1 reactivation from latency. J Virol 76:8778-8786.

21. Escarmís, C, Lázaro E, Manrubia SC. 2006. Population bottlenecks in quasispecies dynamics. Curr Top Microbiol Immunol 299:141-170.

22. Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 104:8597-8604.

23. Burke MK. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. Proc Biol Sci 279:5029-5038.

24. Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature 263:353-355.

Chapter 3:

Experimentally Guided Mathematical Modeling Reveals RNA Virus Replication Principles Shaping the Mutation Distribution in Single Infected Cells

The work in this chapter is the product of an equal-contribution collaboration with

Jeremy Draghi.

Introduction

Replication modes

RNA viruses are excellent models for evolution. They replicate very quickly and have extremely high mutation rates (1). While this combination of traits creates the potential for rapid adaptation, it necessitates a life history strategy that balances the need for explosive, exponential growth with the requirement to maintain genomic integrity. The life history strategies of viruses are largely reflected in their intracellular mode of replication. Two classic replication modes have been described for single-stranded RNA viruses: the "stamping machine" mode (2) and the "geometric replication" mode (3). In the stamping machine mode (SM), templates made from the original infecting genomes are used for the production of all progeny genomes. In the geometric replication mode (GR), newly made progeny genomes are used to create further templates for additional rounds of replication within a single cellular infection cycle (Figure1). Progeny produced from stamping machine replication are a single generation away from the parental strand whereas progeny generated from geometric growth represent a distribution of generations from the parental strand, often resulting in a fractional mean number of generations (see Figure1). The iterative nature of GR creates branched genealogies that allow for expansive exploration of sequence space and results in a mutation distribution that is distinct from that of the SM (3). Recent studies with population genetics models (4) and RNA enzyme populations (5) have demonstrated how differences in the distribution of mutants can significantly impact the

adaptability of a population. Recent studies with poliovirus (PV) have also demonstrated

how mutational differences within a population can have dramatic effects on

pathogenecity (6) as well as fitness, virulence, and robustness (7).
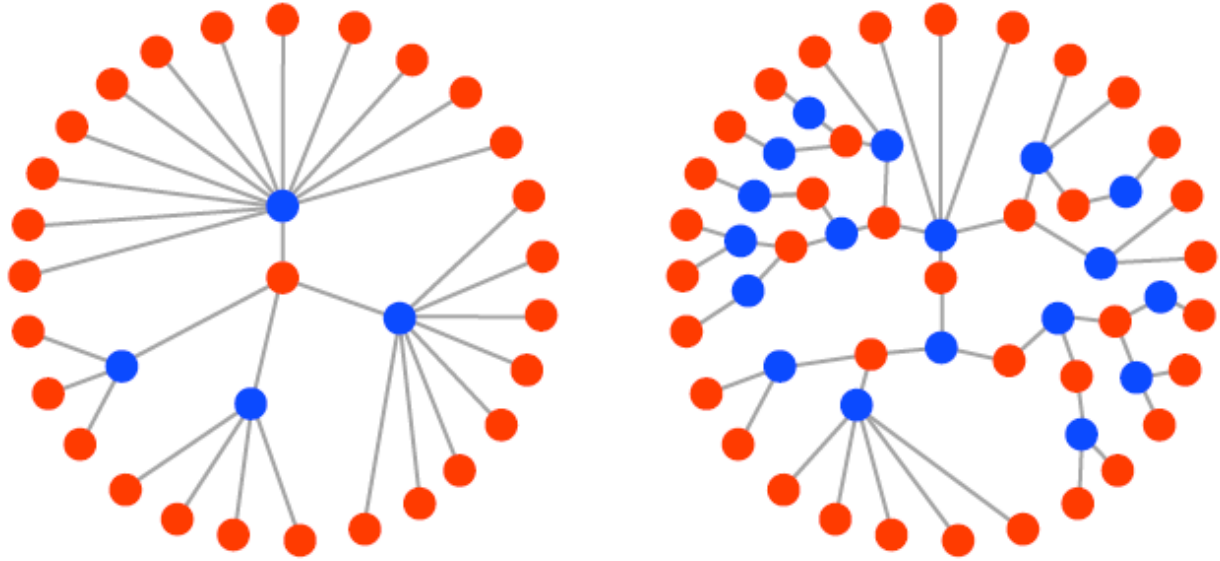
Figure 1:

Comparison of replication modes. Stamping machine (SM) progeny (left) are 1

generation from the initial infecting genome. Geometric replication (GR) progeny (right)

are a mean of 2.33 generations from the initial infecting genome. Red dots indicate

positive-sense strands. Blue dots indicate negative-sense templates.

Poliovirus intracellular dynamics

Poliovirus' simple genomic architecture and medical importance have made it one of the most extensively studied viruses (8). However, despite decades of mechanistic studies and recent revelations of the importance of population structures, the replication mode and resulting mutation distribution have yet to be determined. PV therefore proves an excellent candidate for the rigorous construction of a computational model of virus replication to predict population structure and mutation distribution. A major feature of PV intracellular dynamics is that the genome participates in multiple reactions: translation, replication, and encapsidation. Its 7.5 kb genome contains a single open reading frame which encodes 7 nonstructural proteins and 4 capsid proteins. Translation produces a single polyprotein which is cleaved into individual viral proteins such that each viral protein is produced at equimolar concentrations. Replication of the positive-sense genome by the virus-encoded RNA-dependent RNA polymerase produces a negative-sense strand which is used as a template for further genome synthesis. Evidence exists that suggests that the initial, infecting positive-sense genomes must be translated before they can replicate (9). The switch from translation to replication appears to be dependent on the concentration of a viral protein product, 3CD, which stimulates a transition from a linear, translating RNA to a circular RNA competent for replication (10-12). Encapsidation is thought to be a passive process resulting from protein-protein associations of capsid pentamers with the RNA replication machinery and protein-RNA association of capsid pentamers with viral RNA (13-15). Actively replicating genomes are preferentially encapsidated and packaging is biased to

exclude negative-sense strands, although the mechanism of this is not understood (14). Although multiple ribosomes can translate a genome at the same time and multiple viral polymerases can replicate a genome at the same time, the processes are mutually exclusive (10). Similarly, neither translation nor replication can occur after a genome is packaged into a virion.

Several studies have demonstrated that PV genomes are often localized to the cytosolic surfaces of the endoplasmic reticulum, Golgi bodies, lysosomes, or vesicles derived from these (16-20). Replication complexes are thought to form on these membranes in *cis*, resulting in a close association of translation products and positive-sense genomes (9, 21). Compartmentalization of replication complexes likely accounts for the observation that many functions of nonstructural proteins cannot be complemented in *trans* (9, 22). Only capsid proteins, 3CD, and 3D have been demonstrated to *trans*-complement (9, 14, 23). Taken together, these studies suggest that the essential transitions – from translation to replication, and from replication to encapsidation – are largely localized and influenced by the dynamics of the molecules in each compartment.

Modeling replication

In recent years, modeling approaches have begun to examine the trade-offs that come with having a genome that is a template for both replication and translation (24-28). These studies have raised mechanistic and evolutionary questions about the life cycle of single-stranded, positive-sense RNA viruses, but most have not produced models that can be directly compared to data. Several previous models are

deterministic in nature (24- 25, 28) and assume a well-mixed, spatially uniform cellular environment (24-28). Experimental evidence suggests that each of these assumptions is problematic and do not reflect the biological constraints and properties of viral replication.

The small number of the critical molecules that initiate an infection suggest that a stochastic model would more accurately describe early reactions and could make distinct predictions from previous deterministic approaches (29). Infections begin with relatively few virions, often only one, contacting a cell and continue with the translation of these few initial genomes. Random variation in the switch from translation to replication will be amplified by the subsequent exponential phase of the infection, and this amplification is likely to bias the mean dynamics of a set of infections. Indeed, recent single-cell studies demonstrated the significant impact of stochastic effects on poliovirus infections (30). Whether this random variation averages out, such that the mean dynamics of a population of infected cells can be well-approximated by a deterministic model, can only be determined by formulating and testing a stochastic model.

Here, we have developed a stochastic simulation model in which we compartmentalize reactions in an effort to accurately describe intracellular dynamics in both space and time. Additionally, rather than fixing each parameter on an estimated value, as most others have done, we use an Approximate Bayesian Computation approach to fit our parameters from temporal quantitative data. We find that by combining stochasticity and spatial structure, our model reflects and describes the

population dynamics and structure of the viral population during an infection cycle more accurately than previous models.

Fitting our model to RNA abundances over time, we find that poliovirus follows the geometric replication mode: multiple iterative cycles of genomic replication produce progeny virus. Posterior parameter fits indicate that progeny of a single cellular infection are approximately 5 generations away from the initial, infecting genomes. This replication mode produces populations with expansive, branched genealogies, creating the dramatic potential for the exploration of sequence space, as well as creating the potential for intracellular selection among related mutant genomes.

Model

Stochastic Simulation – Model Outline

We developed a stochastic simulation model that tracks discrete abundances of poliovirus molecular species within a cell and simulates individual reactions. Similar to Hensel et al. (31), we have modified the basic Gillespie algorithm (32) to balance accuracy and speed. This allows us to generate stochastic realizations of replication, translation, and other reactions unfolding in a single infected cell, based on a system of equations that describes each essential reaction in the poliovirus life cycle. Figure 2 depicts the events in poliovirus replications captured quantitatively by our model. Results from many replicate simulations are then averaged to predict the dynamics across a population of infected cells.
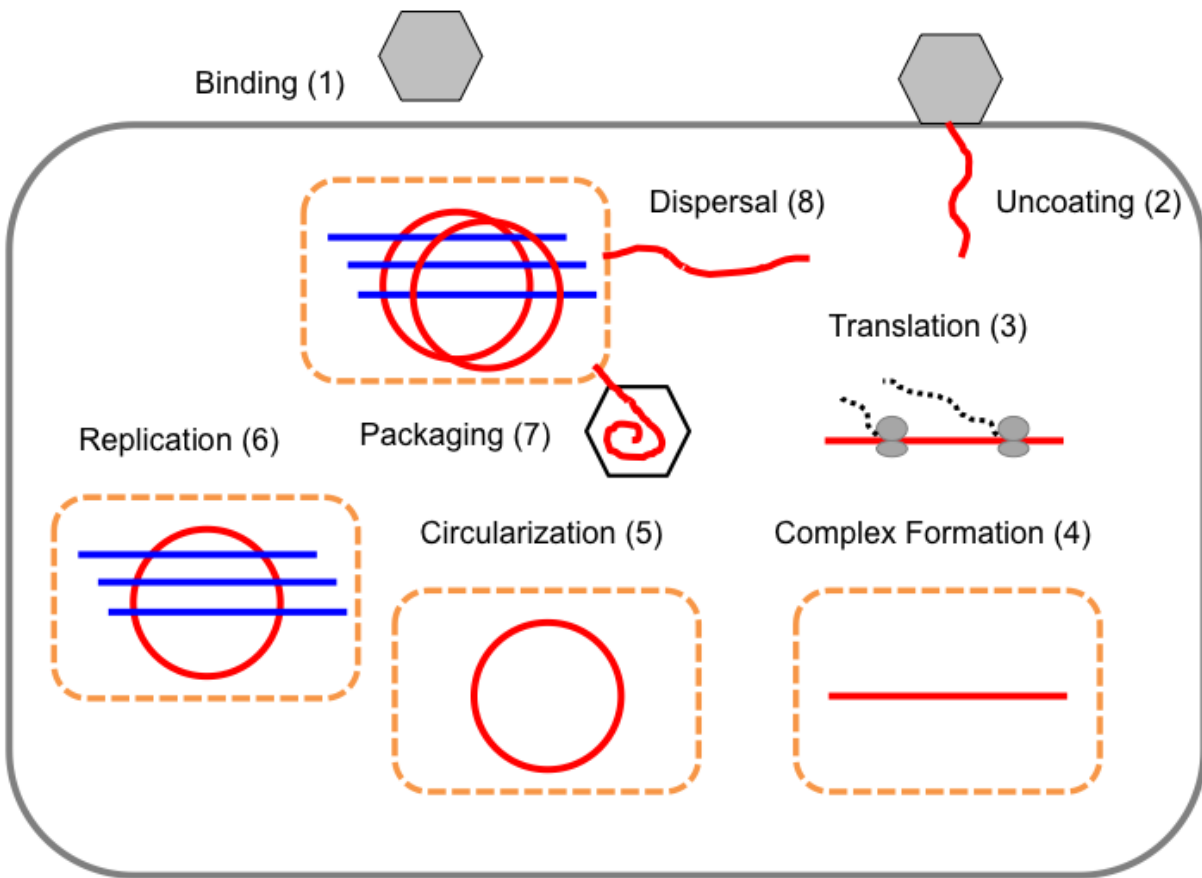
Figure 2:

The replication cycle of poliovirus as represented in our model. Numbered steps correspond to sections and equations in the Methods.

(1) Binding

We assume that the number of virions that bind to, and subsequently infect, a cell is Poisson distributed with a mean equal to the multiplicity of infection (MOI). This formulation assumes that bound virions do not interfere with the binding of additional virions during the period of infection. We denote these coated positive-sense RNA genomes at RNA$^+_{initial}$; their distribution is therefore:

$$RNA^+_{initial} \sim Poisson(MOI)$$ Eq. 1

(2) Uncoating

A quantitative description of uncoating was derived from the data presented in Brandenberg et al. (33); based on these data, we choose the two-parameter gamma distribution to model stochasticity in this process. To account for differences in experimental protocols, we excluded the t=0 measurement from Brandenberg et al.'s data and, taking the t=8 minute measurement as the starting point, fit the gamma distribution to the average cumulative measurements. Using the optim() function in R, we obtained an estimate of 0.678 for the shape parameter, and 0.02 for the rate parameter (n = 28, $R^2 \approx 0.92$). Each of the RNA$^+_{initial}$ molecules transitions to a translationally competent, linear-form positive-sense RNA, RNA$^+_{lin,}$ after a waiting time, t_{uncoat}, drawn from Eq. 2.

$$t_{uncoat} \sim Gamma(0.02, 0.678)$$ Eq. 2

(3) Translation

Translation is the first role of positive-sense genomes in a cell, and it continues

70

as the primary role throughout the infection. Because poliovirus translates a single polyprotein, we assume that all protein products are produced at equal rates based upon a single rate-constant of initiation. We also assume that poliovirus genomes, and not cellular factors, are rate-limiting, and neglect the delay between the initiation of translation, and the appearance of the protein products. With these assumptions, translation can be modeled as a first-order equation with a single parameter, c_{trans}, yielding a rate of translation:

$$r_{trans} = c_{trans} RNA_{lin}^{+}$$ Eq. 3

Here, and throughout the model, we consider these rates to describe Poisson processes, rather than changes in continuously valued quantities. We track three protein products of translation: the procapsid units, which we abbreviate CAP, and protein products 3A and 3CD. Based on evidence from complementation experiments, we assume that CAP units and 3A diffuse freely, while 3CD accumulates within complexes with translating genomes (9, 14, 22). Eq. 3 applies to translation of both complex-associated and free genomes. Global abundances of CAP and 3A are tracked, while abundances of 3CD are tracked individually for each replication complex. 3CD arising from the translation of free genomes is ignored.

(4) Replication Complex formation

We assume that two events must happen before a translating positive-sense sense strand can replicate: it must attach to a membrane, representing nucleation of a replication complex, and it must circularize through association with 3CD (10, 12). Once

71

a strand associates with a membrane, we consider that it has formed a complex, and

assume that all subsequent translation events will add to the local concentration of 3CD.

We model this first step by introducing a rate, $r_{compart}$, at which the RNA$^+_{lin}$ species forms

complexes. We also assume that the viral protein product 3A facilitates this complex

formation (34). Finally, we assume that complex formation is limited by the supply for

suitable membrane, which limits the number of possible complexes in a cell to com_{max}

(35). We therefore introduce a first-order reaction scaled by the number of existing

complexes, com, the maximum, com_{max}, and the concentration of the protein 3A:

$$r_{com} = c_{com}\left(1 - \frac{com}{com_{max}}\right) threeA \qquad \text{Eq. 4}$$

While other viral and cellular proteins are involved in complex formation, we

assume that their influence is adequately represented by tracking the concentration of

3A. We also represent the consumption of some number of 3A molecules in the

formation of each complex by a parameter c_{3A}. If insufficient 3A is available upon

complex formation, newly translated proteins are consumed by the existing complex

until c_{3A} have been allocated. Therefore, we are assuming that 3A binding is

cooperative, and that incomplete complexes have much higher affinity for 3A than does

the reaction to form a new complex.

(5) Circularization

We model circularization -- the transition of a positive-sense genome from a

linear, translating molecule to a circularized molecule competent for replication -- as a

72

first-order reaction driven by the concentration of the viral protein 3CD in each complex (indexed by *i*).

$$r^i_{circ} = c_{circ} threeCD_i \qquad \text{Eq. 5}$$

This formulation reflects experimental data supporting the direct role of 3CD in circularization (10, 12), and the low rate of rescue of 3CD-deficient strains by complementation in *trans* (9).

(6) Replication

We distinguish replication rates for positive and negative strand synthesis with separate rate constants c_{rep+} and c_{rep-}. We ignore polymerase concentrations and instead assume that both types of replication are first-order reactions modified by a common cellular resource limit. This limitation is parameterized by rep_{max}, the maximum number of replication events per cell permitted by some limited resource, and implemented with a running counter, *rep*, of synthesized RNAs. We also assume that per-capita replication does not differ between replication complexes, allowing us to write a mass-action equation for both replication reactions.

$$r_{rep+} = c_{rep+} RNA^+_{circ} \left(1 - \frac{rep}{rep_{max}} \right) \qquad \text{Eq. 6a}$$

$$r_{rep-} = c_{rep-} RNA^- \left(1 - \frac{rep}{rep_{max}} \right) \qquad \text{Eq. 6b}$$

Note that r_{rep+} measures the rate at which replication is initiated on positive-sense templates, producing negative-sense strands, and similarly, r_{rep-} measures the rate of

positive-strand production.

We allowed newly synthesized positive-sense genomes one of three fates: 1) associate with capsid protomers and become encapsidated 2) diffuse into the cytoplasm where they can translate and potentially create new, independent compartments or 3) remain in the complex in which they were generated and act as templates for further RNA replication. We assume that positive-sense genomes that remain in the complex are immediately competent for replication. We were unable to fit the sharp transition to exponential growth seen in our strand measurement data without allowing for this third option. Allowing newly synthesized positive-sense genomes to remain in the complex and act immediately as replication templates is consistent with previous reports indicating a coupling between translation and replication as we still require initial, infecting genomes to be translated before transitioning to replication (9). Negative-sense strands also stay in the complex in which they were produced and are immediately competent for replication.

We assume that only the positive-sense replication-competent form is packaged (22, but see 15), and that, following Nugent et al. (14), genomes can only be packaged as they are synthesized from a negative-sense strand. We therefore first determine whether the newly synthesized positive-sense strand is packaged; then, for unpackaged genomes, we calculate whether they remain in the replication complex.

(7) Packaging

We assume that the rate of initiation of packaging is proportional to the global

74

concentration of a virus-derived protein product, CAP, representing capsid protomers.

These protomers form pentamers, of which twelve are required for each capsid; each

packaging event therefore consumes sixty units of CAP. To account for the evidence

that deficiencies in capsid proteins can be complemented in *trans*, we allowed capid

proteins to diffuse freely throughout the cell. As with 3A molecules and complex

formation, if a packaging event begins when the global abundance of CAP is less than

sixty, then further packaging is halted until this deficit is filled.

Using the approximation that each available CAP molecule independently

contributes to the probability of packaging, we derive the probability that a newly

synthesized positive-sense strand is packaged to be:

$$p_{pack} = 1 - e^{-c_{pack} CAP} \qquad\qquad \text{Eq. 7}$$

(8) + Strand Dispersal

The probability of a newly synthesized positive-sense strand to remain within its

replication complex, assuming it was not packaged, is given by the parameter c_{stay}. The

total probability is therefore:

$$p_{stay} = (1 - p_{pack}) c_{stay} \qquad\qquad \text{Eq. 8}$$

(9) Replication Phenotype

Our primary goal is to infer the number of replication cycles between the infecting

and the progeny virions. Defining a complete replication cycle to include both copying to

a negative-sense strand, then back to a positive-sense strand, we label the mean of this

value μ_{rc}. The principal purpose of this mean is to link the mutation rate of replication to

the mean mutation frequency in the progeny population, so the appropriate measure is to average over virions, not infected cells. For *k* replicate simulations, let n_i represent the number of progeny produced by each replicate *i*, and g_{ij} represent the number of replication cycles in the ancestry of each virion *j* in replicate *i*; then we calculate μ_{rc} as follows.

$$\mu_{rc} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} g_{ij}}{\sum_{i=1}^{k} n_i} \qquad \text{Eq. 9}$$

Stochastic Simulation – Computational Approach

Gillespie (32) argued that discrete stochastic simulation is an accurate method for predicting the dynamics of chemical reactions, and is preferred when there are small numbers of some species of reactants. This approach requires exponentially distributed random numbers to quantify the periods between reactions. These random number draws contribute important stochasticity when rates are relatively low; however, when the overall reaction rate is very high, these draws are computationally expensive and contribute little variance. We therefore implemented a dual approach: when the rate is below a threshold, we draw exponential times; when it is above a threshold, we use the inverse of the rate – the expected time – as our interval between reactions. Figure 3 shows that this approximation delivers accurate results for the best (lowest error) inferred parameter set.
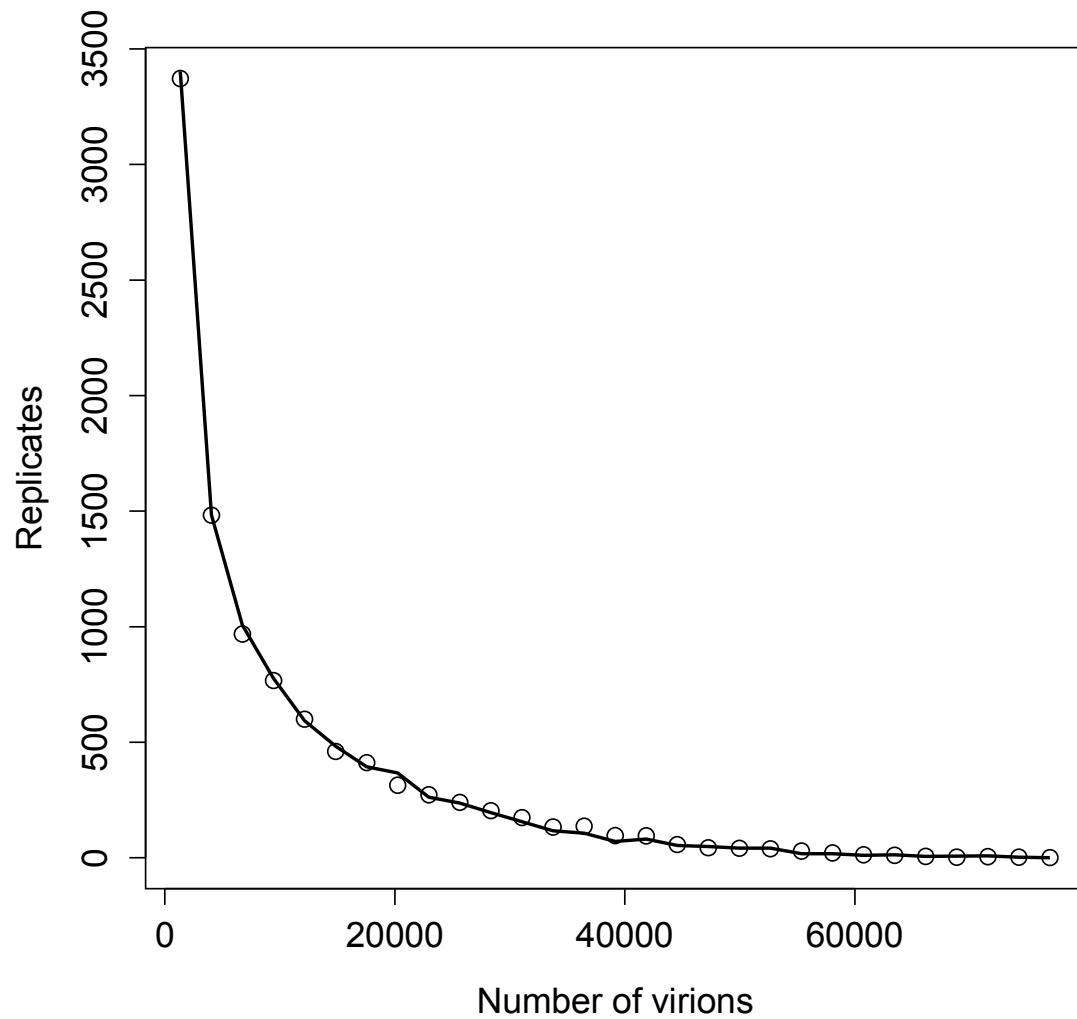
Figure 3:

Distribution of virions in 10,000 replicates for simulations with (points) and without (line) a deterministic threshold for waiting times (see Methods). Parameter values are the 'best' set given in Table 1.

Parameter inference

We used temporal, quantitative RT-PCR data of both positive-sense genomes and negative-sense strands to estimate the free parameters in our model. We chose to use measurements of positive- and negative-sense RNA at multiple time points for three multiplicities of infection (1, 10, and 100), as well as measurements of virion numbers at multiple times for MOI 10; this amounted to twenty-seven measured means. This relatively high number of data dimensions, combined with the computationally intensive and highly stochastic nature of our simulations, made a traditional maximum likelihood approach impractical. Instead, we turned to Approximate Bayesian Computation, using as our summary statistic the sum of the squared deviations of the average simulated RNA concentrations (and average fraction of virions for MOI 10) from their corresponding empirical means. After determining that the simplest rejection algorithm was impractically slow for our problem, we chose to implement a relatively new method called Sequential Monte Carlo (36-41).

Sequential Monte Carlo (SMC) consists of several rounds of parameter selection which form successively better approximations of the posterior distribution. In each round x, a population of size n_x parameters sets is generated iteratively by choosing a parameter set from the preceding round x-1, perturbing its values, then accepting or discarding the new parameter set based on the distance of its measured summary statistic from the summary statistic representing the data. Parameter sets with distances less than ε_x are accepted; a diminishing series of thresholds, $\varepsilon_1 > \varepsilon_2 > \varepsilon_3$, etc., progressively focuses the search on those parameter values that best match the data.

78

In round 1, parameter sets are drawn from the prior distributions; this first round is therefore identical to the basic rejection algorithm, but with a fairly large ε_1 to reduce computation time.

The advantage of searching for better parameter sets near previously identified good values is a much higher frequency of acceptance, and therefore much less computational time. However, the parameters accepted in later rounds are then biased toward common values in the previous rounds. The SMC algorithm removes this bias by weighting the selection of parameter sets against those that are most similar to their parent round, and toward those that resemble the prior distribution. Let $K_\sigma(\theta_a, \theta_b)$ represent the probability of perturbing θ_a into θ_b with a Gaussian kernel of standard deviation σ, w_{xi} represent the weight of parameter set i in round x, θ_{xij} represent the value of parameter j in set i of round x, and $\pi(\theta)$ represent the prior probability of θ. Then, for round two and later, we calculate these importance weights as in Eq. 10 (adapted from 37).

$$w_{x,i} = \frac{\sum_{j=1}^{10} \pi(\theta_{x,i,j})}{\sum_{k=1}^{n_{x-1}} w_{x-1,k} \sum_{j=1}^{10} K_{\sigma_{x-1,j}}(\theta_{x-1,k,j}, \theta_{x,i,j})} \qquad \text{Eq. 10}$$

Beaumont (38) suggests that the Gaussian perturbations applied to each proposed parameter set should be scaled with regard to the variance in that parameter in the previous round. In practice, we identified a trade-off based on the scaling of these perturbations; smaller perturbations lead to increased acceptance rates but more positively skewed importance weights; because the weighting of each accepted parameter set is normalized relative to the highest observed importance weight, this

strong skew effectively dilutes the inferential power of the analysis. We found that using the standard deviation of each parameter as the standard deviation for the perturbation balanced this trade-off adequately for our model: the weights of the 1135 sampled parameter sets had an entropy of 10.01 bits, compared to a maximal value of 10.15 bits. This high entropy confirms that any remaining skew in the importance weights does not severely diminish the effective sample size.

Implementing this method requires several additional choices: the shapes of prior distributions, the number of rounds, the values of ε, and the number of replicates, n, to perform for each evaluation of the model. This last decision turned out to be crucial; inferring based on the mean of a larger number of replicates ($n \geq 1000$) tended to select parameter sets with highly variable behavior. Reducing n led to a higher rate of parameter set rejection but more biologically plausible dynamics. We therefore chose to accept parameter sets that passed a given ε for multiple, sequential sets of n replicates. For round 1, one thousand parameters sets were accepted based on five sets of $n=20$ replicates at MOI = 10 only with $\varepsilon = 12$. For round 2, 1135 parameters sets were accepted based on five sets of $n=20$ replicates at MOI = 1 with $\varepsilon = 16$, five sets of $n=20$ replicates at MOI = 10 with $\varepsilon = 12$, and five sets of $n=20$ replicates at MOI = 100 with $\varepsilon = 7$. Thresholds for round 2 were calibrated to achieve an acceptance rate of about 1 in 10,000.

To aid in visual exploration of the data, we chose five representative parameter sets as follows. From an initial batch of 513 parameter sets, we chose the fifty sets with the overall lowest error. From these, we sampled sets of five at random and calculated

80

the summed pairwise distance in parameter space of those five sets from each other.

To adjust for the different scales and uncertainties in each parameter, the contribution of

each parameter to the distance measure was divided by its standard deviation over the

whole set of 513 values. These summed distances provided a metric of the parameter

diversity captured in a choice of five parameter sets; we examined one thousand

randomly drawn sets of five and chose the set with the highest summed distance.

These five parameter sets are shown in Table 1.

	c_{trans}	c_{com}	c_{circ}	c_{rep+}	c_{rep-}	c_{pack}	com_{max}	rep_{max}	c_{3A}	c_{stay}
Best	-2.106	-6.676	-6.452	-4.164	0.804	-2.193	4.630	13.205	2.642	0.073
Set 1	-1.475	-6.304	-4.79	-6.397	1.047	-5.207	6.963	13.296	3.236	0.693
Set 2	-2.321	-5.111	-5.848	-5.603	1.92	-5.742	7.07	13.912	1.922	0.082
Set 3	-2.765	-6.157	-5.3	-3.6	1.112	-2.109	6.254	15.203	3.247	0.021
Set 4	-1.503	-5.716	-7.349	-5.369	1.91	-5.213	7.604	14.002	2.554	0.082
Set 5	-1.958	-5.857	-4.178	-6.062	0.79	-9.195	4.997	12.707	2.688	0.499
Low	-2.027	-5.510	-6.636	-5.241	2.191	-5.71	7.233	14.963	2.222	0.033
High	-2.163	-5.826	-4.639	-5.203	0.934	-9.257	7.042	14.864	3.765	0.188

Table 1:

'Best' parameter set used in Figure 3, and representative parameter sets used in Figure

4.

Methods and Materials

Cells and viruses

HeLaS3 cells (ATCC CCL-2.2) provided by R. Geller and J. Frydman (Stanford University) were maintained in 50% DMEM/50% F-12 medium supplemented with 10% newborn calf serum, 100U/ml penicillin, 100U/ml streptomycin, and 2mM glutamine (Invitrogen). Poliovirus Mahoney typeI genomic RNA was generated from *in vitro* transcription of prib(+)XpAlong. To generate virus, 20 µg of RNA was electroporated into 4 X 10^6 HeLaS3 cells in a 4mm cuvette with the following pulse: 300V, 24 Ω, 1000 µF. The resulting virus was passaged at high multiplicity of infection (MOI ~1-10) three times then subjected to ultracentrifugation through a 30% sucrose cushion.

Infections

Four wells of HeLaS3 cells in 12-well plates were washed, trypsinized, and counted twice each on a hemocytometer then averaged to determine cell count. To synchronize infections, plates were placed on ice, cells were washed with cold serum-free media and infected at MOIs 1, 10, and 100. Plates were incubated at 4ºC for 30 minutes with rocking every 10 minutes to adhere virus. After removal of the inoculum, cells were washed 2X with warm serum-free media. Cells were then incubated at 37ºC in 2% serum media until harvest. To harvest, plates were frozen at -70ºC.

RNA extraction, Reverse Transcription (RT), and qPCR

Plates were thawed on ice and refrozen at -70°C 3X. RNA was extracted via the PureLink RNA Micro Kit (Life Technologies) according to the manufacturer's instructions. cDNA was synthesized from total RNA using SuperScript III Reverse Transcriptase (Life Technologies) and 125 nM strand-specific RT primer (+strand_RT: 5'-GGCCGTCATGGTGGCGAATAATGTGATGGATCCGGGGGTAGCG-3'; -strand_RT: 5'-GGCCGTCATGGTGGCGAATAACATGGCAGCCCCGGAACAGG-3') in a 5 μL reaction. Separate RT reactions for positive and negative-strand RNAs were performed for each sample. RT products were treated with 0.5 units of Exonuclease I (Fermentas) to remove excess RT primer prior to qPCR. Strand-specific qPCR was based on a published protocol (42). cDNAs were analyzed by qPCR using 2× SYBR FAST Master Mix (Kapa Biosystems), 200 nM strand-specific qPCR primer (+strand_For: 5'-CATGGCAGCCCCGGAACAGG-3'; -strand_Rev: 5'-TGTGATGGATCCGGGGGTAGCG-3'), and 200 nM Tag primer (5'-GGCCGTCATGGTGGCGAATAA-3') in a 10 μL reaction. A 10× dilution series of *in vitro* transcribed positive- and negative-strand RNA standards was run alongside experimental samples and used to construct a standard curve.

Virion immunoprecipitation

Lysates from MOI 10 infections were homogenized with a final concentration of 0.06% NP-40. Immunoprecipitation was performed using Protein A-coated Dynabeads and anti-poliovirus antibody according to a published protocol (42).

84

Sucrose gradients

HeLaS3 cells were infected for 1, 2, 3, 4, and 5 hours at an MOI of 10 in 15 cm dishes then simultaneously treated with 100ug/ml cycloheximide (CHX) for 2 minutes at 37ºC. Cells were washed with PBS+CHX and lysed with 0.5% NP-40 lysis buffer containing CHX on ice for 20 minutes. Cell debris was pelleted in a table-top centrifuge at 2000 rpm for 10 minutes at 4ºC, then nuclei were pelleted at 9000 rpm for 10 minutes at 4ºC. Cell lysates were loaded on a 10%-50% sucrose gradient containing CHX and ultracentrifuged at 35000 rpm for 3 hr. Fractions were collected on a Biocomp Gradient Station with a BioRad Econo UV Monitor. Fractions were pooled based on the spectrophotographic trace into 2 fractions (ribonucleoprotein and monosome/polysome fractions), RNA was extracted and subjected to qRT-PCR.

Results

Temporal quantitative data

To obtain estimates of our unknown parameters we fit our model to experimental measurements of poliovirus RNA during an infection. We infected HeLaS3 cells with poliovirus type I Mahoney at three multiplicities of infection: 1, 10, and 100. Samples were taken in biological triplicate every hour. Strand-specific qRT-PCR was performed to quantify positive-sense and negative-sense poliovirus RNA against in vitro transcribed standard RNAs of each sense (42). Along with cell counts, this allowed for temporal measurements of the average positive-sense and negative-sense poliovirus RNA copies per cell. Negative-sense RNA was not detectable until 2 hours post infection for MOIs 10 and 100, and 3 hours post infection for MOI 1. Positive-sense RNA was clearly quantifiable for all timepoints at the MOI 10 and 100 but did not rise above background levels until 3 hours post infection for MOI 1. Figure 4 illustrates this data alongside parameters from the second round of SMC.

Figure 4:

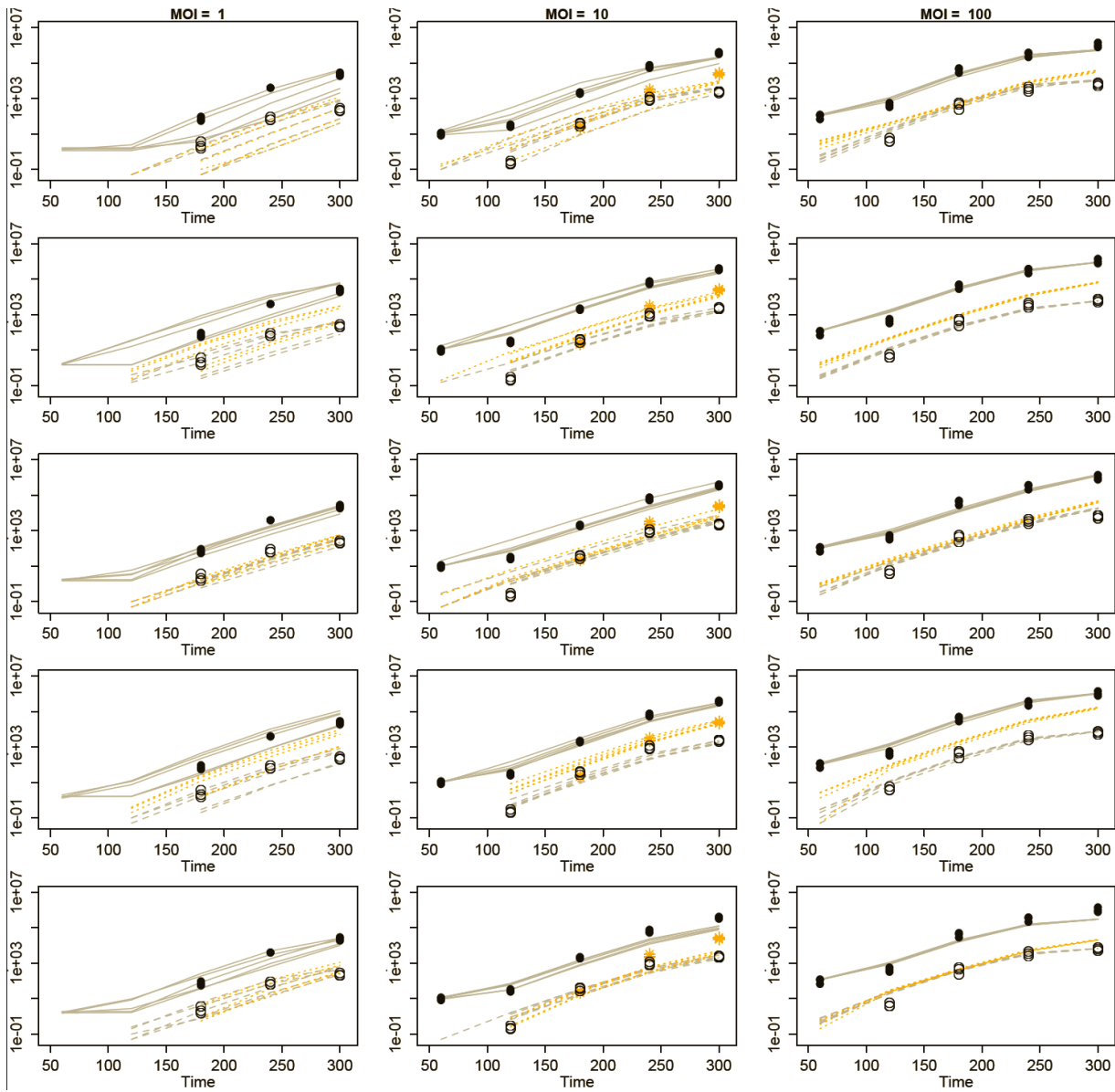Projected mean abundances of positive-sense RNA (solid lines vs filled circles), negative-sense RNA (dashed lines vs hollow circles) and virions (orange dotted lines vs stars). Each row represents a different example parameter set (see Results); each line is the mean of twenty individual-cell simulations, and the means of five sets of twenty replicates are plotted in each panel. Parameter values are given in Table 1.

To further refine our parameter estimates, we quantified the percentage of genomes encapsidated over a time course of infection at an MOI of 10. Using a newly developed virion immunoprecitipation assay (42), we observed de novo virion assembly between 2 hours and 3 hours post-infection. Along with total positive-sense RNA measurements from this time course, we obtained a percentage of genomes encapsidated in quadruplicate at 3 hours, 4 hours, and 5 hours post infection.

As an additional test of our best parameter set, we experimentally determined the percentage of genomes associated with translation machinery throughout a time course of infection. We fractionated infected cell lysates and quantified positive-sense RNAs in polysome fractions relative to total positive-sense RNA copies. This data renders a percentage of genomes associated with translation machinery and provides an additional set of data to evaluate the parameters sets produced by SMC. When measured at an MOI of 10 at 1, 2, 3, 4, and 5 hours post infection, the majority of positive-sense RNAs appeared to be associating with translation machinery, consistently averaging near 85%.

Inference of Replication Parameters

We performed inference of the ten free parameters in our model using a sequential Monte Carlo variant of Approximation Bayesian Computation. This algorithm produces progressively more accurate estimates of each parameter over several rounds; Figure 5 illustrates that, for most parameters, round one restricts the credible range of each parameter in comparison to the flat prior and round two leads to further

focusing. The data appear to be uninformative for at least one parameter, c_{pack}; a second parameter, com_{max}, appears to be poorly constrained by the comparison to MOI = 10 in round one, but somewhat constrained by the broader measurement against all three MOIs in round two. Round two also appears to significantly move the mode of two other parameters, c_{com} and c_{3A}.
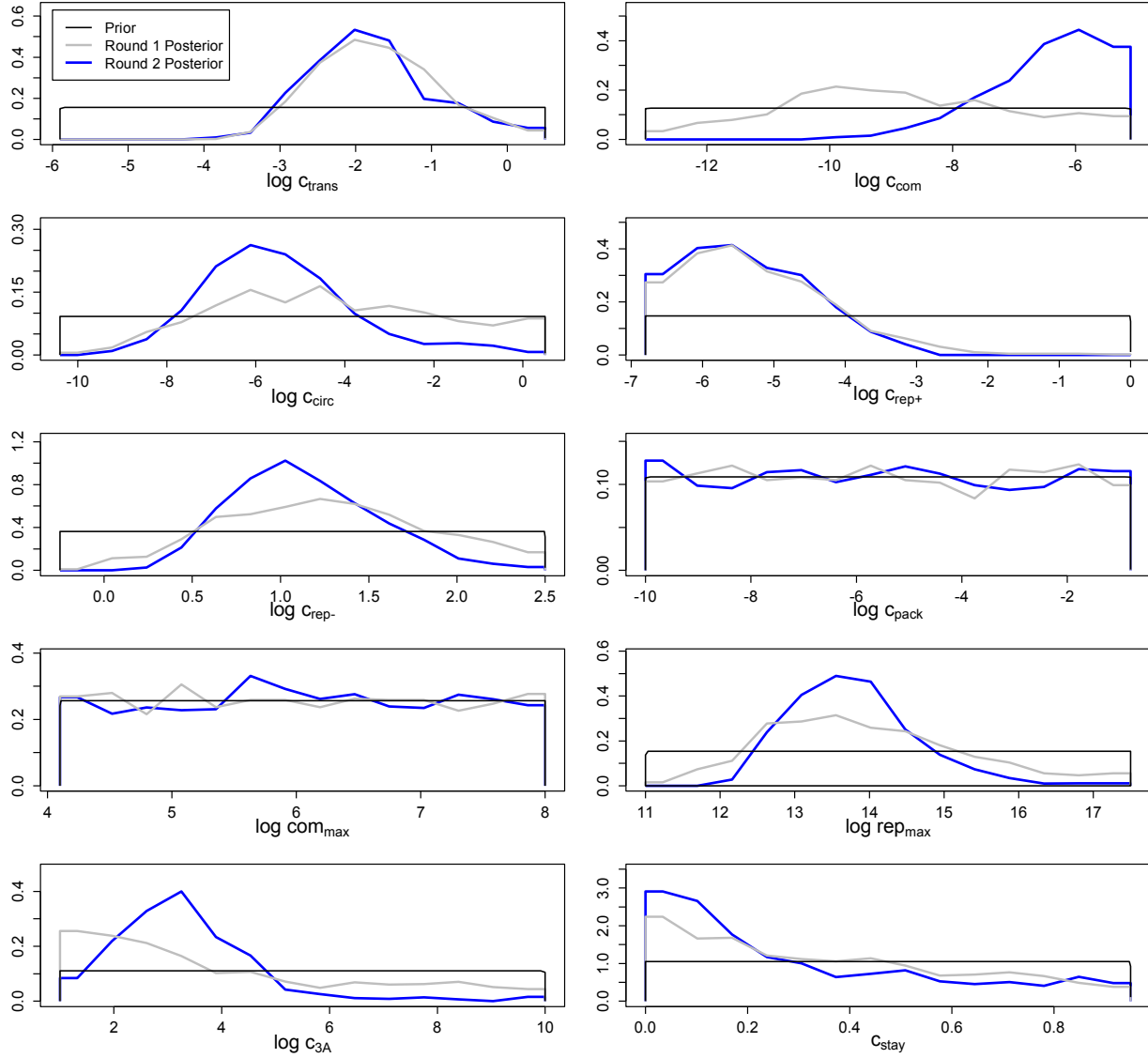
Figure 5:

Prior and posterior distributions after each of two rounds of inference by Approximate Bayesian Computation. Empirical posteriors for rounds one and two are based on 1000 and 1135 points, respectively, grouped into fifteen bins.

Figure 5 indicates that ABC inference informed the values of nine of our ten parameters, but these marginal parameter distributions alone do not capture correlations between parameter values. Figure 6 shows evidence of significant correlations, and Figure 7 shows that parameter sets drawn from the marginal distributions in Figure 5 (i.e., uncorrelated parameter values) do a poor job of matching the data. While not unexpected, these significant correlations require that we either attempt to fully characterize the multidimensional structure of these parameter distributions, or work directly with the sampled parameter sets arising from our inference process. Given our limited capacity to sample the posterior, we opted for the latter choice.

	c_{trans}	c_{com}	c_{circ}	c_{rep+}	c_{rep-}	c_{pack}	com_{max}	rep_{max}	c_{3A}	c_{stay}
c_{stay}	+++	+		---	--			--	-	
c_{3A}	+	---	+	+						-
rep_{max}	-		+		++					--
com_{max}										
c_{pack}										
c_{rep-}	-		+					++		--
c_{rep+}	--	-	-						+	---
c_{circ}	---	-		-	+			+	+	
c_{com}	-		-	-					---	+
c_{trans}		-	---	--	-			-	+	+++

Figure 6:

Correlations between parameters in the round two posterior. Spearman rank correlations with magnitudes below 0.1 were ignored; those with magnitudes between 0.1 and 0.25 are noted as "+" or "-", those with magnitudes between 0.25 and 0.5 with "++" or "--", and those with magnitudes above 0.5 with "+++" or "---".
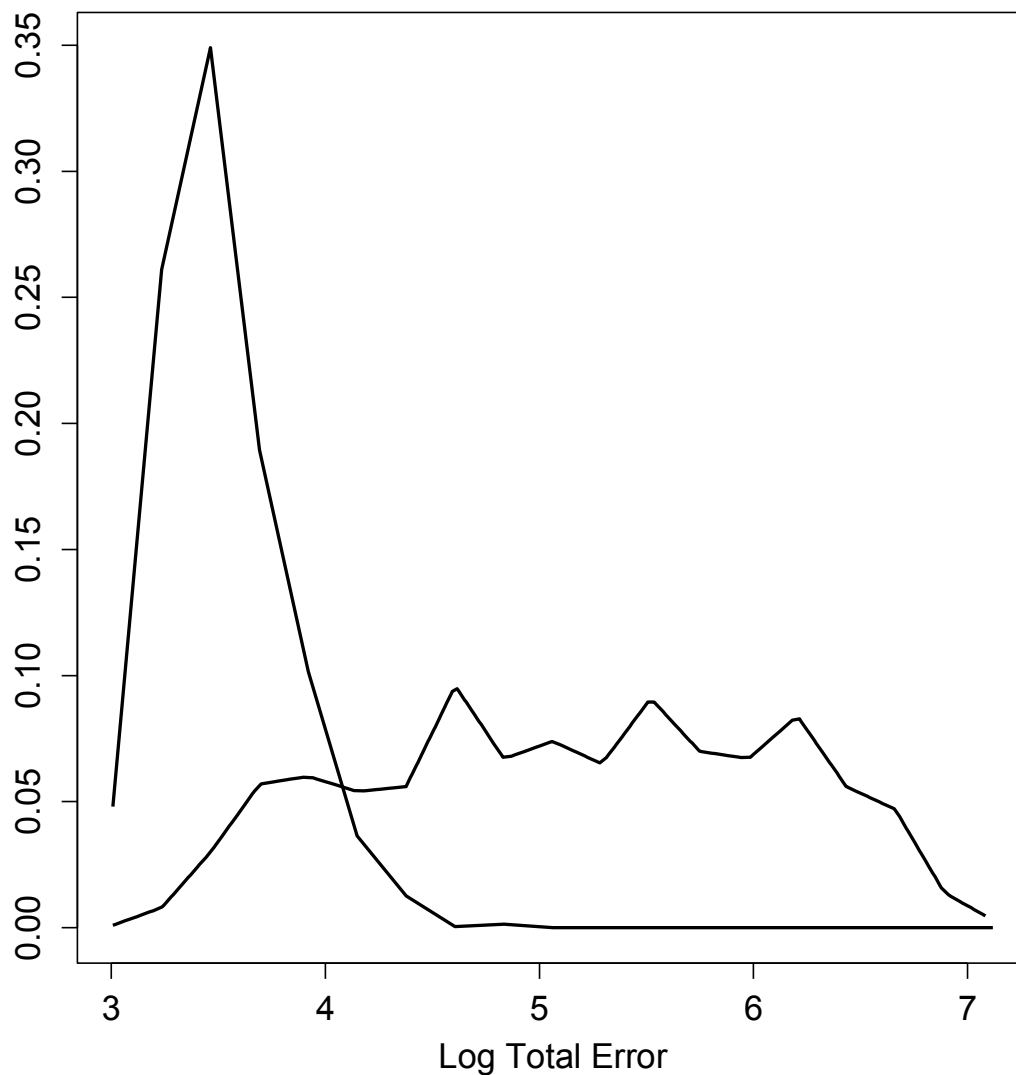
Figure 7:

Log of total error for inferred, weighted parameter sets in round two (solid) versus 1000

sets assembled from parameter values drawn independently from the weighted

posterior (dotted).

Each parameter in the posterior is supported over a significant range of possibilities. This remaining uncertainty reflects two factors: the data may be insufficient to determine each parameter, and the inference process may not have fully exploited the inferential power of the data. We took several approaches to quantify the sufficiency of the data versus the effectiveness of the inference process. First, we measured the mean error of parameter sets when compared to the data for each multiplicity of infection independently; we asked if performance at one MOI predicted performance at the other two. If so, the dimensionality of our data would be effectively lower than we had initially assumed. Surprisingly, pairwise correlations between mean error at one MOI and another were very weak: Spearman's rho is 0.031 for MOIs 1 & 10, -0.092 for MOIs 1 & 100, and 0.129 for MOIs 10 & 100.

Figure 4 shows that the inferred parameter sets generally capture the information in the RNA and virion data, although some parameter sets deviate consistently from the data for some values. Variability among replicate sets of twenty single-cell simulations are substantial, correlated across a time series, and greatest for the smallest MOI. Further inferential effort could improve either the accuracy of the mean predicted dynamics, or the precision of replicate simulation dynamics, though Figure 4 suggests that such improvements could only be modest.

To further investigate if the parameter values identified by ABC minimize the error in predicted RNA and virion dynamics, we explored the sensitivity of mean error to variation in each parameter for a single parameter set ('Best' in Table 1). These simulations again used sets of 20 single-cell replicates to calculate mean error, but did

so for 500 trials at each MOI to produce a high-resolution estimate of true deviation from the data. These results are plotted as 1/(1 + mean error) to show an intuitive goodness-of-fit measure, where high values indicate similarity to the data. Figure 8 shows that the 'best' parameter set is at or near a local maximum for goodness-of-fit for eight of ten parameters; the effects the remaining two parameters, c_{pack} and com_{max}, appear to be minimal for this parameter set. These results suggest that convergence of the posterior distributions is linked to the sensitivity of the model to each parameter, which supports the effectiveness of the ABC algorithm.
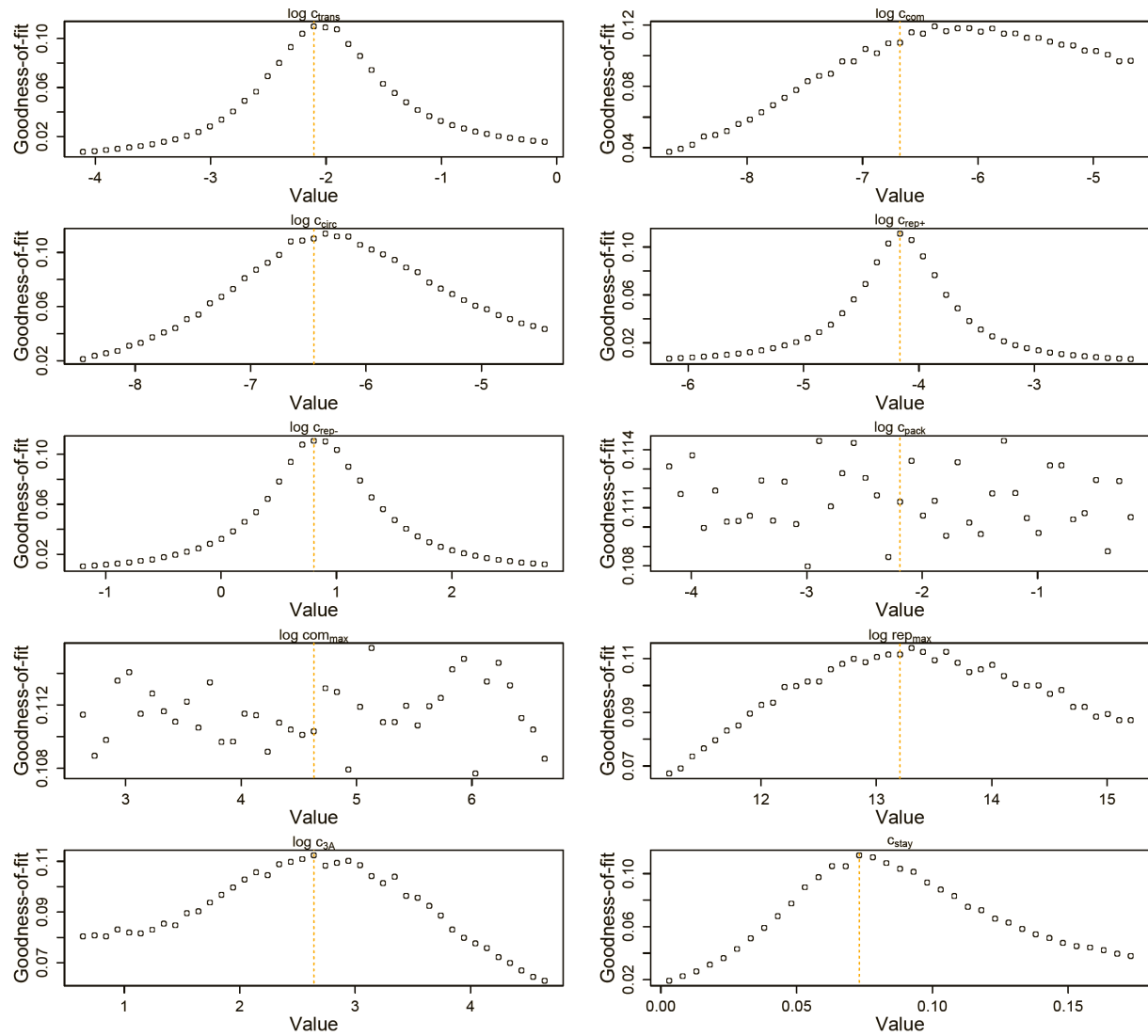
Figure 8:

Goodness-of-fit (1 / (1+ mean error)) of highly replicated simulations for MOI = 10 and the 'best' inferred parameter set. Each parameter value was varied independently for 1000 sets of 20 single-cell replicates for each MOI. Orange lines represent the base value of each parameter.

Predicted Replication Dynamics

Figure 9 shows the inferred posterior distribution of μ_{rc}, the mean number of replication cycles for a packaged virion based on two rounds of inference with measured RNA and virion abundances. This distribution is plotted for MOI=10; the predicted values at MOI = 1 and MOI = 100 are very similar and highly correlated (weighted means: MOI=1, 4.96; MOI=10, 5.06; MOI = 100, 4.85; Spearman's rho (unweighted): MOI 1 & 10, 0.92; MOI 1 & 100, 0.85; MOI 10 & 100, 0.96). While this distribution does show substantial variance, it is strongly inconsistent with a "stamping machine" mode of replication, which would have a μ_{rc} near one.
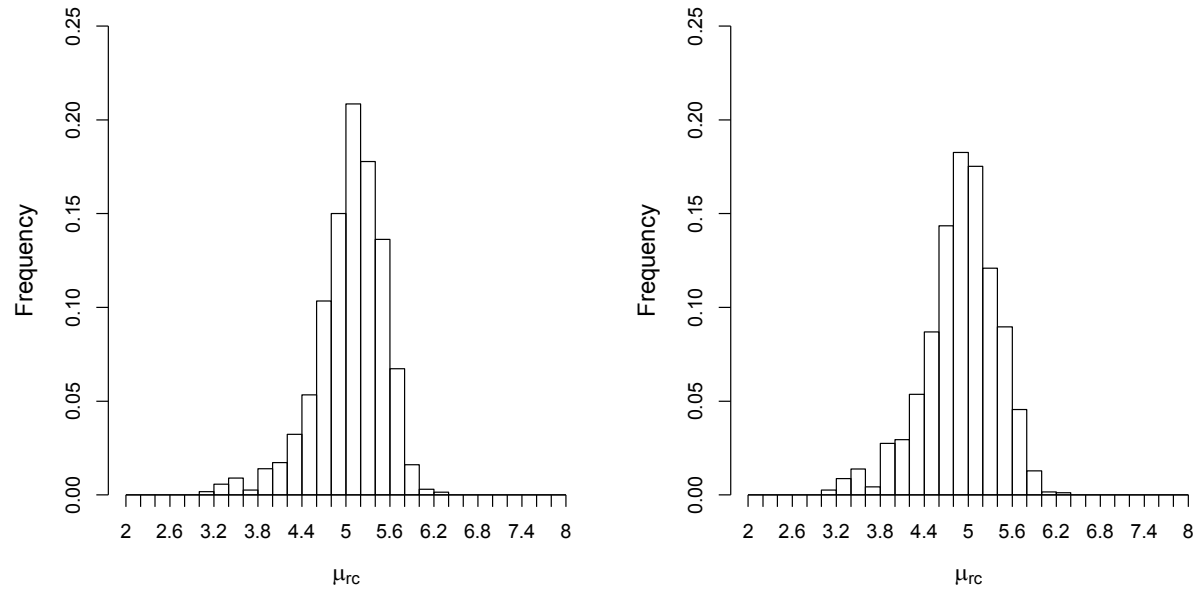
Figure 9:

Left: Posterior distribution of the mean number of generations of replications (μ_{rc}). Right:

Distribution reweighted by the fit of predicted fractions of translating positive-sense RNA

to empirical measurements.

To explore the robustness of this inference, we compared the predicted dynamics of the model to an additional type of data: the fraction of positive-sense RNA molecules translating at each time point. Many of the inferred parameter sets are consistent with the measured values, but a substantial fraction is clearly inconsistent (Figure 10). The summed squared error of the translating fractions is also correlated with μ_{rc} (Figure 11). To estimate how these new data inform our prediction of μ_{rc}, we calculated a weighting factor based on the relative rank of the summed squared error of translating fractions, such that the parameter set with the best fit was assigned a weight of 1, the next a weight of 1134/1135, etc. Reweighting the distribution of μ_{rc} by this additional factor produced the distribution shown in Figure 9 right; the mean μ_{rc} shifts from 5.06 to 4.9.
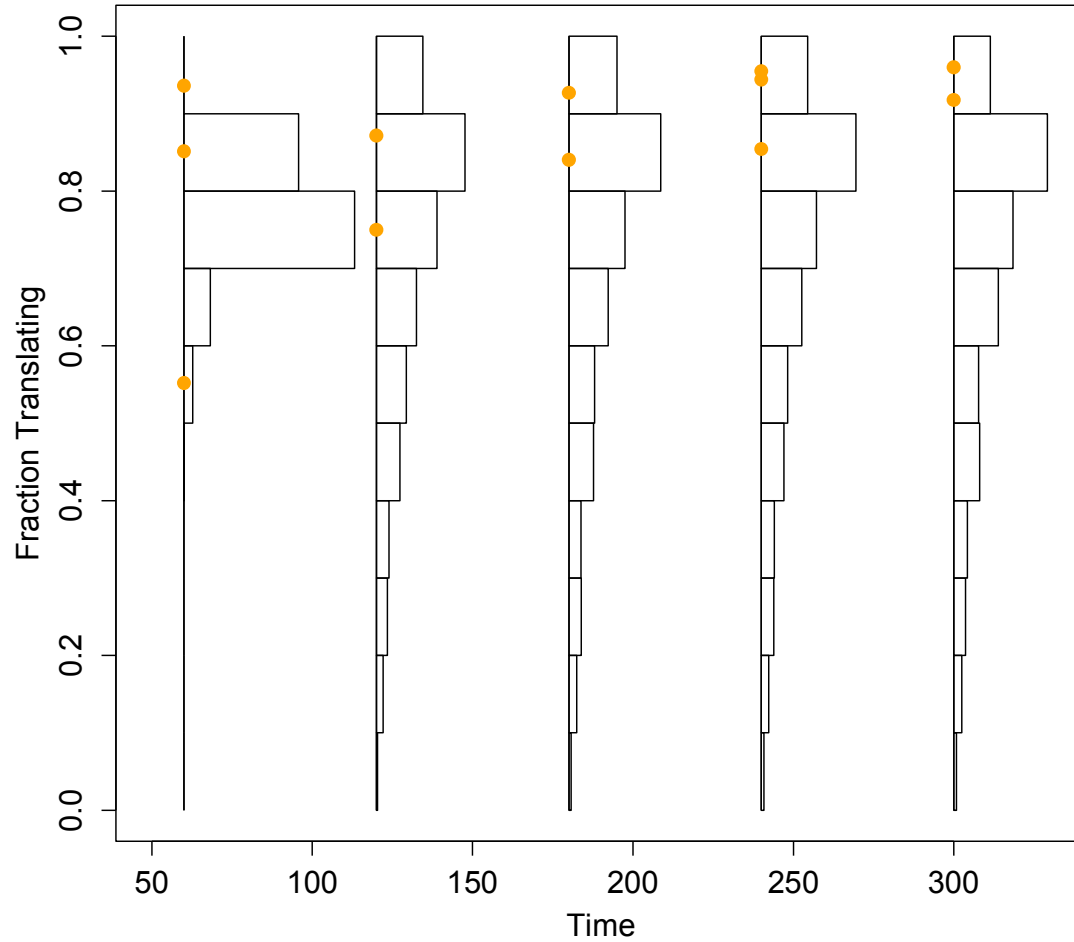
Figure 10:

Histograms of the projected fraction of positive-sense RNA undergoing translation for the mean simulated dynamics of each parameter set, compared to empirical measurements (orange dots). To better visualize the diversity of the predictions made from the inferred parameter sets, these histograms are not corrected by the importance weights.
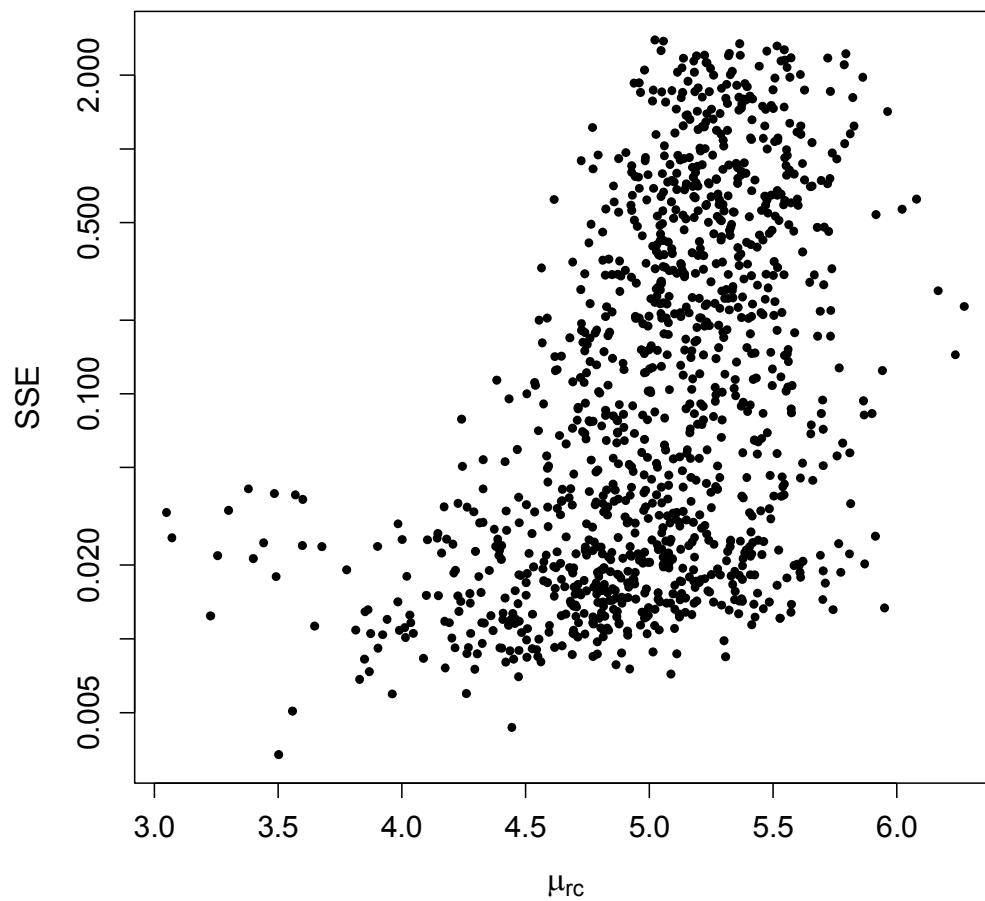
Figure 11:

Summed squared error (SSE) of fraction of translating positive-sense RNA for all 1135

parameter sets, plotted against μ_{rc} at MOI = 10.

Predicting the Distribution of Mutations

We simulated mutation and selection during infections to understand how replication dynamics shape the distribution of mutation frequencies among virions. To illustrate how mutant frequencies depended on μ_{rc}, we chose two parameter sets with values of μ_{rc} at the low and high end of the range supported by the posteriors in Figure 5, and included the 'best' parameter set as a representative of the more common values of μ_{rc}. Mutation frequencies for these parameter sets ('best,' 'low,' and 'high' – see Table 1) are plotted in Figure 12 for a range of replication deficits that produce a selective pressure. The per-capita probability that a mutant genome is chosen as a template for replication, relative to an unmutated genome, is given by one minus the deficit.
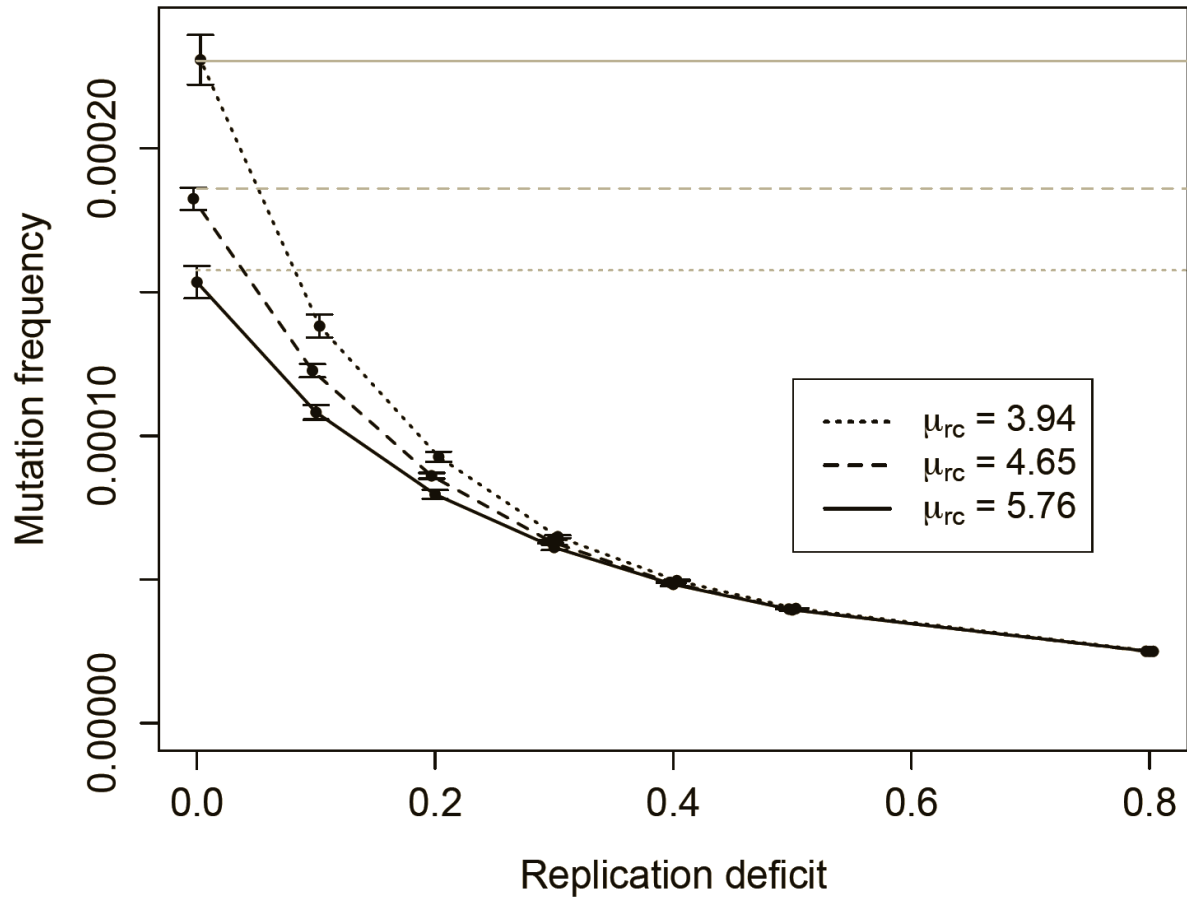
Figure 12:

Mean mutation frequencies for three parameter sets ('low,' μ_{rc} = 3.94; 'best,' μ_{rc} = 4.65; 'high,' μ_{rc} = 5.76). Mutation rate is 2 x 10^{-5} per replication event; 'replication deficit' reflects the reduced probability of a mutant template to replicate, relative to an unmutated strand. Grey lines indicate the expected mean for each parameter set with no selection (deficit of zero); the black line shows the mutation rate in one replication step, and therefore the expected frequency when mutants cannot replicate. Bars indicate 95% confidence intervals.

Several distinct features of mutation in this model are evident from Figure 12. Mutation frequency does not decrease linearly as intracellular selection approaches its maximal value; the curve results from the fact that mutant genomes that are immediately packaged are not subject to selection, while the contribution of rare, early mutants to average mutation frequency may be reduced by multiple rounds of intracellular selection. Knowing μ_{rc} and the mutation rate allows us to predict the fate of neutral or very unfit mutations, but the frequency of mutations of intermediate fitness may require simulation of the model to predict. A third feature is the sizable confidence intervals relative to the number of infections sampled (ten million for each point). This high variability reflects the large contribution of very rare mutations that arise early in an infection and can contribute thousands of mutant virions, especially when selection is weak.

The effect of minor alleles in the overall mutant distribution can be seen as a departure from a Poisson process. To remove a potential confounding variation in burst size, we compare the distribution of mutations from infections within 10% of the median burst size, and calculate a Poisson expectation for a median-sized burst with the same expected frequency. For the 'best' parameter set, median infections produced many more bursts with no copies of a given mutation (79.4% vs. 22.5% for the Poisson), but also many more bursts with five or more copies of the mutant (8% vs. 1.82% for the Poisson (n = 51,365).

The distribution of the number of generations between progeny virions and initial infecting genomes is displayed for three parameter sets ('best,' 'low,' and 'high' – see

Table 1) in Figure 13. Only a very small percentage of progeny are produced via a

single genomic replication cycle. Although all three parameter sets have means close to

5 generations, the distributions show a portion of the progeny virions representing up to

10 genomic replication cycles between parent and progeny within a single cellular

infection. The breadth of these distributions suggests that μ_{rc} alone is not adequate to

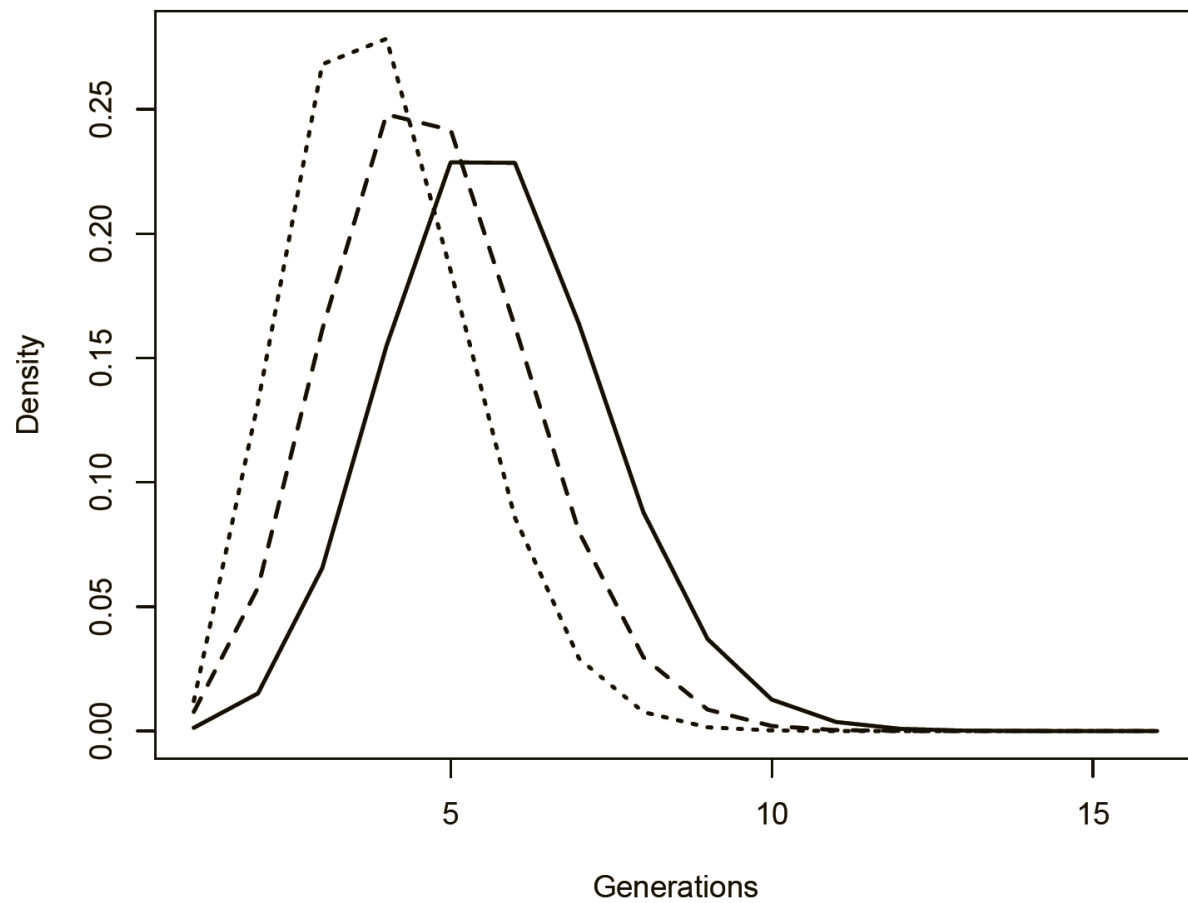completely predict the mutation structure of viral progeny.

Figure 13:

Distributions of μ_{rc} of progeny from single cell infections for three parameter sets ('low,' $\mu_{rc} = 3.94$; 'best,' $\mu_{rc} = 4.65$; 'high,' $\mu_{rc} = 5.76$).

107

Discussion

Virus infections are normally depicted as deterministic processes that follow a stereotypical path from infection to progeny production and death of the infected cell. However, experimental data show that some infected cells produce few progeny while others produce large populations of progeny (30). These observations suggest the common assumption that stochasticity is an important factor shaping the outcome of infection. By combining accurate experimental measurements and a stochastic model of viral replication, we have obtained a realistic model of how the molecular events driving the life cycle of the virus govern the outcome of infection in each cell. The intracellular replication mode of a virus strongly influences the frequency and distribution of mutations among progeny, which contributes significantly to the behavior of a population (6-7). Due to the complex nature of intracellular dynamics, assessing the mode of replication of viruses has long been a difficult task (but see 43). Here, we build on decades of mechanistic studies and recent modeling efforts to construct an stochastic computational model coupled with new Bayesian inference methods. This is combined with accurate temporal and quantitative data to produce a detailed picture of viral infection. We find positive- and negative-sense RNA measurements over multiple MOIs along with quantitative data on virion packaging to be sufficiently informative to infer that poliovirus replication features several layers of intermediate replication, in contrast to the oft-assumed 'stamping machine' model. The implications of this replication mode are as follows: 1) per-replication error rates in vivo are considerably

lower than measured rates from full-replication-cycle in vivo studies, 2) with the potential

accumulation of different mutations, many steps through sequence space are possible

in a single cellular infection, and 3) there exists a significant potential for intracellular

selection and competition among related virions, even in infections initiated by only a

single genome.

Accurate estimates of viral mutation rates are essential for studying viral

evolution and have crucial practical applications in drug and vaccine design. While

estimates of mutation rates exist for nearly two dozen viruses, estimates of replication

modes exist for only a few (44). Calculating per-replication event mutation rates from

observed mutant frequencies is not possible without knowledge of replication mode.

Thus, estimates of poliovirus per-replication event mutation rates can range over 10-fold

depending on the assumed replication mode (1, 44). Our replication mode inference

allows us to link estimates of per-replication event mutation rates to published mutant

frequencies. The most extensive poliovirus mutant frequency data set estimated an

average mutant frequency of $2\text{X}10^{-4}$ (45). Using our inferred value of approximately 5

complete replication cycles (defined by a cycle of 2 copy events: positive to negative to

positive), we calculate a per-replication event mutation rate of $2\text{X}10^{-4} / 5 * 2 = 2\text{X}10^{-5}$,

which is in agreement with the average estimates of poliovirus in vivo mutation rates

calculated from lethal mutation frequencies (45). Our finding of 5 complete intracellular

replication cycles is also in line with previous inferences of replication mode using the

Luria-Delbruck fluctuation test null-class method (44). However, our results highlight

some limitations for inferring mutation rates from frequencies: intracellular selection may

strongly affect mutation frequencies, and the strong stochastic nature of virus replication appears to deeply modulate minor alleles distribution, which in turn will result in imprecise estimates of the expected frequency. In particular, assuming that mutation frequency can be modeled as a Poisson process will lead to inappropriate confidence in measured frequencies. As a consequence, multiple empirical mutation frequencies measurements will be required to obtain a more precise determination of true mutation frequencies.

The branched genealogy inferred in our study implies the potential for significant amounts of intracellular complementation, selection, and competition between mutant genomes, even in infections initiated by a single genome (6, 9, 46). Figure 4(a) demonstrates the extent to which the frequency of a mutation can be skewed by negative selection during the course of an infection. On the other hand, a mutational event that occurred early in replication and conveyed an intracellular replication advantage could potentially give rise to hundreds or thousands of descendant virions in a single generation. If the mutation distribution data in Figure 4(b) were displayed as a tree (as in Figure 1), it would contain over 7000 terminal nodes, too many to resolve in a figure. Hence, the apparent potential for mutant interactions is vast. These results suggest that the evolutionary fate of mutations may depend strongly on their intracellular competitive ability, even when multiplicities of infection are low. Additionally, studies that rely on bottlenecks to reduce selection in viral mutation studies (e.g., 47) may be allowing more selection than they expect. Future population dynamics studies should consider the implications of the intracellular expansion of mutant phenotypes.

An significant benefit of computational modeling is that the information learned in the empirical process of the development of a model can yield important insights in the biologic process study. For example, our initial attempts to fit temporal strand measurement data were unable to match the sharp transition to exponential growth seen in the data. Only after removing the requirement for positive-sense genomes to be translated before becoming replication-competent was our model flexible enough to rapidly create templates for exponential replication. While Novak and Kirkegaard (9) demonstrate a requirement of the initial, infecting genomes to be translated before replication can occur, their data does not implicate that all genomes produced at any time during infection must be translated before replicating. Our study suggests that newly synthesized positive-sense genomes may or may not disperse to nucleate new replication complexes within a single cellular infection allows us to model intracellular dynamics in a novel way, permitting a portion of newly made positive-sense strands to immediately act as templates for replication without the requirement of translation. Of note, this observation also complicates efforts to use lethal mutant frequencies as a proxy for per-replication event error rates because potentially lethally-mutated genomes have the possibility to act as templates of replication without the proofreading step of translation.

Our model succeeds in describing many experimentally observed features of viral replication and is an excellent staging point for future and more accurate models of viral replication and evolution. With the realistic benefits of stochasticity, compartmentalized reactions, and parameters inferred from quantitative, temporal data,

it acts as a baseline intracellular viral replication algorithm. More quantitative data, including data on the formation and number of replication compartments, would further inform the model. Potential additions of intracellular selection, complementation, and recombination parameters would allow population evolution studies to explore intracellular dynamics with more precision than previous approaches. The ultimate goal is to generate a comprehensive model incorporating mechanistic replication dynamics learned from virology with selection and complementation dynamics learned from population genetics. This tool could be very powerful for informing future therapeutic and preventative strategies.

References

1. Drake JW (1993) Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci U.S.A. 90(9):4171-4175.

2. Stern, G. S. 1963. Molecular biology of bacterial viruses. W. H. Freeman and Company, San Francisco, Calif.

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

4. Draghi JA, Parsons TL, Wagner GP, Plotkin JB. 2010. Mutational robustness can facilitate adaptation. Nature 263:353-355.

5. Hayden EJ, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature 474(7349):92-95.

6. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439:344-348.

7. Lauring AS, Acevedo A, Cooper SB, Andino R (2012) Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. Cell Host Microbe 12(5):623-632.

8. Racaniello VR (2006) One hundred years of poliovirus pathogenesis. Virology 344(1):9-16.

9. Novak JE, Kirkegaard K (1994) Coupling between genome translation and

replication in an RNA virus. Genes Dev 8(14):1726-37.

10.   Gamarnik AV, Andino R (1998) Switch from translation to RNA replication in a positive-stranded RNA virus. Genes Dev 12(15):2293-304.

11.   Garmarnik AV, Andino R (2000) Interactions of viral protein 3CD and poly(rC) binding protein with the 5' untranslated region of the poliovirus genome. J Virol 74(5):2219-2226.

12.   Herold J, Andino R (2001) Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Mol Cell 7(3):581-91.

13.   Pfister T, Pasamontes L, Troxler M, Egger D, Bienz K (1992) Immunocytochemical localization of capsid-related particles in subcellular fractions of poliovirus-infected cells. Virology 188(2):676-84.

14.   Nugent CI, Johnson KL, Sarnow P, Kirkegaard K (1999) Functional coupling between replication and packaging of poliovirus replicon RNA. J Virol 73(1):427-35.

15.   Liu Y, Wang C, Mueller S, Paul AV, Wimmer E, Jiang P (2010) Direct interaction between two viral proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus morphogenesis. PLoS Pathog 6(8):e1001066.

16.   Schlegel A, Giddings TH Jr, Ladinsky MS, Kirkegaard K (1996) Cellular origin and ultrastructure of membranes induced during poliovirus infection. J Virol 70(10):6576-88.

17.   Bolten R, Egger D, Gosert R, Schaub G, Landmann L, Bienz K (1998) Intracellular localization of poliovirus plus- and minus-strand RNA visualized by

strand-specific fluorescent In situ hybridization. J Virol 72(11):8578-85.

18. Cui ZQ, Zhang ZP, Zhang XE, Wen JK, Zhou YF, Xie WH (2005) Visualizing the dynamic behavior of poliovirus plus-strand RNA in living host cells. Nucleic Acids Res 33(10):3245-52.

19. Egger D, Bienz K (2005) Intracellular location and translocation of silent and active poliovirus replication complexes. J Gen Virol 86(Pt 3):707-18.

20. den Boon JA, Diaz A, Ahlquist P (2010) Cytoplasmic viral replication complexes. Cell Host Microbe 8(1):77-85.

21. Egger D, Teterina N, Ehrenfeld E, Bienz K (2000) Formation of the poliovirus replication complex requires coupled viral translation, vesicle production, and viral RNA synthesis. J Virol 74(14):6570-80.

22. Ansardi DC, Porter DC, Anderson MJ, Morrow CD (1996) Poliovirus assembly and encapsidation of genomic RNA. Adv Virus Res 46:1-68.

23. Oh HS, Pathak HB, Goodfellow IG, Arnold JJ, Cameron CE (2009) Insight into poliovirus genome replication and encapsidation obtained from studies of 3B-3C cleavage site mutants. J Virol 83(18):9370-9387.

24. Krakauer DC, Komarova NL (2003) Levels of selection in positive-strand virus dynamics. J Evol Biol 16(1):64-73.

25. Regoes RR, Crotty S, Antia R, Tanaka MM (2005) Optimal replication of poliovirus within cells. Am Nat 165(3):364-73.

26. Sardanyés, J, RV Solé, SF Elena (2009) Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. J Virol

83(23):12579-12589.

27.   Thébaud G, Chadoeuf J, Morelli MJ, McCauley JW, Haydon DT (2010) The relationship between mutation frequency and replication strategy in positive-sense single-stranded RNA viruses. Proc Biol Sci 277(1682):809-817.

28.   Martinez, F, J Sardanyés, SF Elena, J-A Daròs (2011). Dynamics of a plant RNA virus intracellular accumulation: Stamping machine vs. geometric replication. Genetics 188(3):637-646.

29.   Srivastavawz, R, L You, J Summers and J Yin (2002). Stochastic vs. deterministic modeling of intracellular viral kinetics.

30.   Schulte MB, Andino R (2014) Single-cell analysis uncovers extensive biological noise in poliovirus replication. J Virol. 88(11): 6205-6212.

31.   Hensel SC, Rawlings JB, Yin J (2009) Stochastic kinetic modeling of vesicular stomatitis virus intracellular growth. Bull Math Biol 71(7):1671-92.

32.   Gillespie DT (1976) General method for numerically simulating stochastic time evolution of coupled chemical-reactions. J Comp Phys 22(4): 403-434.

33.   Brandenberg, B, LY Lee, M Lakadamyali, MJ Rust, X Zhuang, JM Hogle (2007) Imaging poliovirus entry in live cells. PLos Biology 5(7):1543-1555.

34.   Nai-Yun Hsu, Olha Ilnytska, Georgiy Belov, Marianita Santiana, Ying-Han Chen, Peter M. Takvorian, Cyrilla Pau, Hilde van der Schaar, Neerja Kaushik-Basu, Tamas Balla, Craig E. Cameron, Ellie Ehrenfeld, Frank J.M. van Kuppeveld, Nihal Altan-Bonnet (2010). Viral Reorganization of the Secretory Pathway Generates Distinct Organelles for RNA Replication. Cell 141(5): 799-811.

35. Guinea R, Carrasco L (1990) Phospholipid biosynthesis and poliovirus genome replication, two coupled phenomena. EMBO J 9(6): 2011-2016.

36. Sisson SA, Fan Y, and Tanaka MM (2009) Sequential Monte Carlo without likelihoods. Proc Natl Acad Sci USA 106(39)1688-1689.

37. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP(2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6(31):187-202.

38. Beaumont, MA (2010) Approximate Bayesian Computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics 41: 379-406.

39. Csilléry, K, Blum, MGB, Gaggiotti, OE, François, O (2010) Approximate Bayesian Computation (ABC) in practice. TREE 25: 410-418.

40. Lopes JS, Beaumont MA (2010) ABC: a useful Bayesian tool for the analysis of population data. Infect Genet Evol 10(6):826-833.

41. Toni T, Stumpf MP (2010) Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics 26(1):104-110.

42. Burrill CP, Strings VR, Schulte MB, Andino R. (2013) Poliovirus: generation and characterization of mutants. *Curr Protoc Microbiol* Chapter 15: Unit 15H.2.

43. Chao, L., C. U. Rang, and L. E. Wong. 2002. The distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage f6. *J. of Virology* 76:3276–3281.

44. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84(19): 9733-9748.

45. Acevedo A, Brodsky L, Andino R (2014) Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505: 686-690.

46. Turner, P. E., & Chao, L. (1999). Prisoner's dilemma in an RNA virus. *Nature*, *398*(6726), 441-443.

47. de la Penta, M., Elena, S. F. and Moya, A. (2000) Effect of Deleterious Mutation-Accumulation on the Fitness of RNA Bacteriophage M4. Evolution, 54: 686–691.

Chapter 4:

Investigating the Fundamental Basis for a Limit to Mutation Rate: Is the Thermodynamic Stability of Essential Proteins the Speed Limit to Evolution?

Introduction

According to the quasispecies theory, the limit on mutation rate is established by the relative fitness of sequence variants to that of the master. A critical point termed the "error threshold" exists where the mutation rate of a viral genome exceeds the replacement rate of the master sequence such that the production of the master sequence is outweighed by the production of an increasingly diverse mutant swarm. This "error catastrophe" scenario is often presented as the explanation for the limit on viral mutation rate. However, this divergent, evolutionary shift in sequence space can be distinct from a drop in absolute abundance as would occur during extinction. While high mutation rate is driving both processes, two very different outcomes occur based on environmental tolerance. By illuminating the mechanism by which mutations are tolerated, we can better understand how quasispecies diversity is shaped. Here, we begin to examine a model stemming from the thermodynamic stability of essential proteins as a potential explanation of the mechanism for a limit on mutation rate.

Zeldovich et al (1), put forth an evolutionary model in which the fitness of an organism can be directly inferred from its genomic sequence with a strict physiological assumption of the genotype-phenotype relationship. Essentially, the model states that mutations can be tolerated in an essential gene such that they do not disrupt the thermodynamic stability of its protein to the degree that renders it unstable and therefore nonfunctional. Using a distribution of protein stabilities, the model establishes a speed limit of molecular evolution of 6 mutations per essential portion of a genome

per replication (1). This number is interestingly close to the mutation frequency of 5.3 mutations per genome observed for wildtype poliovirus in tissue culture (2).

Here, we first take a simple, descriptive approach to observe whether this theory can hold water in our system by attempting to illustrate the differing sensitivities to both reduced and elevated temperatures of poliovirus populations with differing mutation frequencies. We then begin a larger, more systematic examination of the sensitivities of poliovirus populations to targeted mutagenesis of poliovirus proteins. Based on thermodynamic predictions from a novel prediction algorithm incorporating crystallography data of poliovirus proteins, this targeted mutagenesis is expected to disrupt protein stability to varying, well-delineated degrees. Deep sequencing of these populations is expected to reveal the fitnesses of each targeted mutation, allowing an inference of how specific, targeted disruptions affect stability and to what degree.

Materials and Methods

Cell Culture and Virus Generation

HeLa S3 cells were grown in tissue culture flask in DMEM/F12 50/50 medium supplemented with 1x penicillin/streptomycin/glutamine and 10% newborn calf serum. Cells were incubated at 37ºC and 5% CO_2. Poliovirus Mahoney typeI genomic RNA was generated from *in vitro* transcription of prib(+)XpAlong. To generate virus, 20 µg of RNA was electroporated into 4 X 10^6 HeLaS3 cells in a 4mm cuvette with the following pulse: 300V, 24 Ω, 1000 µF.

Poliovirus RNA Transfection

HeLa S3 cells were collected, washed three times with PBS without salt, and resuspended to a concentration of $5x10^6$ cells/ml. 800µl of cells were electroporated with 20µg of RNA in a 0.4cm cuvette using an Electro Cell Manipulator 600 (BTX Inc.). Cells were recovered in 16ml of medium and incubated at 37ºC in a 5% CO_2 incubator until cytopathic effect.

Plaque assay

HeLa S3 cells were seeded in 6-well plates at a concentration of $1.5x10^6$ cells/well and incubated overnight at 37ºC and 5% CO_2. Virus supernatant was diluted in a 1:10 dilution series in DMEM/F12 medium. Cells monolayers were washed once with PBS then 250µl of virus dilution was added per well. To allow virus attachment, cells were

incubated with the virus inoculum for 30 mins at 32ºC, 37ºC, or 39.5ºC and 5% CO_2.

Each well was overlayed with 5ml of 1% agarose in 1X DMEM/F12 medium

supplemented with 5% newborn calf serum. Plates were incubated with 5% CO2 for 2

days at 37ºC and 39.5ºC, and 3 or 4 days at 32ºC. Agarose overlays were then

removed and plates were stained crystal violet dye (0.1% crystal violet, 20% ethanol) to

visualize the plaques which where then counted and a titer was calculated.

Passaging

HeLa S3 cells were seeded at a concentration of $5x10^5$ cells/well of a 12-well plate and

incubated overnight at 37ºC and 5% CO_2. The next morning the cells were infected at

the appropriate MOI. The viral inoculum was incubated with the cells for 30 mins at

37ºC and 5% CO_2 to allow virus attachment. Cells were then washed three times with

PBS and 1ml of fresh media was added to the cells. Infections were incubated for 8

hours then frozen at -70C. Lysates were freeze-thawed 3X before titering.

Cloning

19 oligos for each of 24 positions were ordered in 500 pmole quantity. Oligos had a melt

temp of 62ºC-64ºC. Oligos were 29 bases in length, with the mutant codon flanked by 9

WT bases on each side. Overlap cloning using touch-down PCR was used to introduce

oligos in the pXpAlongrib+WT plasmid. Agarose gel purification was performed after

both the first and second rounds of overlap PCR. Round 2 "inserts" were digested with

SfoI and EcoRI and cleaned over simple PCR columns. pXpAlongrib+WT was digested

with SfoI and EcoRI, run on a 1.5% low-melt agarose gel in the cold room for 3-4 hours at 100 volts. Ligations were performed with 28 ng vector, 14 ng insert, 1 ul of T4 DNA ligase and 1 ul 10X ligation buffer at 16ºC overnight. Ligations were transformed in SURE2 cells via their protocol. All colonies were scraped together into 8 mLs of LB+carbampicillin and grown for an additional 4 hours before mini-prepped.

Results

Cursory, descriptive examination of the stability of poliovirus populations with varying degrees of mutations

To fully experimentally test this model would require analyzing the protein stabilities of an organism with a modular mutation rate over a range of temperatures. Luckily, several poliovirus variants exist which contain fidelity-altering mutations in the viral polymerase (3). In addition, because all of the PV proteins are essential, instability of any protein will result in a non-productive infection. By passaging the high-fidelity polymerase mutant and low-fidelity polymerase mutant alongside WT PV at different temperatures, one can get a comparative readout of fitness between the populations generated with different mutation rates by looking at overall production of plaque-forming-units (PFUs) at each temperature. Because of the limited range of temperatures (~32°C-39.5°C) that permit growth of tissue culture cells, one could use mutagenic nucleotides, such as ribavirin or 5-fluorouracil, to further exacerbate viral mutation rate (4).

As a cursory illustration of potential differing sensitivities to temperature (as a proxy for thermodynamic stability) of poliovirus populations generated with RNA-dependent RNA polymerase mutants with varying mutation rates (therefore having varying mutation frequencies), we simply titered viral stocks at various temperatures. Specifically, a WT poliovirus stock population, along with a G64S (the high-fidelity variant) stock population and a H273R (the low fidelity variant) stock population were

diluted to roughly 150 PFU/mL, then plaque assays were performed and incubated at 32ºC, 37ºC, or 39.5ºC. If this experimental setup was sensitive enough to provide insights into the model, one would expect to see an increase in viral titers at lower temperatures as mutant genomes, which would be inviable at higher temperatures are now viable due to stability of their mutated proteins. Likewise, one would expect that the low fidelity variant, which at 37ºC consistently displays a slightly lower titer than WT, would produce an even lower titer at higher temperatures. If an increase in titer was seen at lower temperatures due to a presumed switch in the viability of mutated genomes, one could determine the diversity of viable virions produced at each temperature via sequencing. If the model is correct, one would expect to see an increase in mutation frequency of PFUs with a decrease in temperature, or in other words, a more diverse population of PFUs at lower temperatures.

We observe that all populations are less viable at non-optimal temperatures (32ºC, 39.5ºC). Interestingly, we observe that the H273R population, which has a higher mutation frequency (see Chapter 1) shows a lower-than-WT proportion of its virions as viable PFUs when the temperature is elevated (Figure 1). The H273R population also shows a higher-than-WT proportion of its virions as viable PFUs when the temperature is lowered (Figure 1). While this data appears to support the idea that the H273R population harbors a larger portion of mutants which are sensitive to elevated temperature and also harbors a larger portion of inviable mutants that can be resuscitated by lowering temperature (which may be a proxy for increasing stability), the G64S (the high-fidelity variant) population shows a similar trend. In the most simplistic

126

view, the model predicts that a high-fidelity variant population (which is known to have a lower mutation frequency, see Chapter 1) should display a larger-than-WT portion of its virions as viable PFUs at elevated temperatures, because it contains fewer mutants that are close to the stability threshold, and a smaller-than-WT portion of its virions as viable PFUs at reduced temperatures, because it contains fewer inviable mutants that can be resuscitated by the presumed increased stability of reduced temperature.
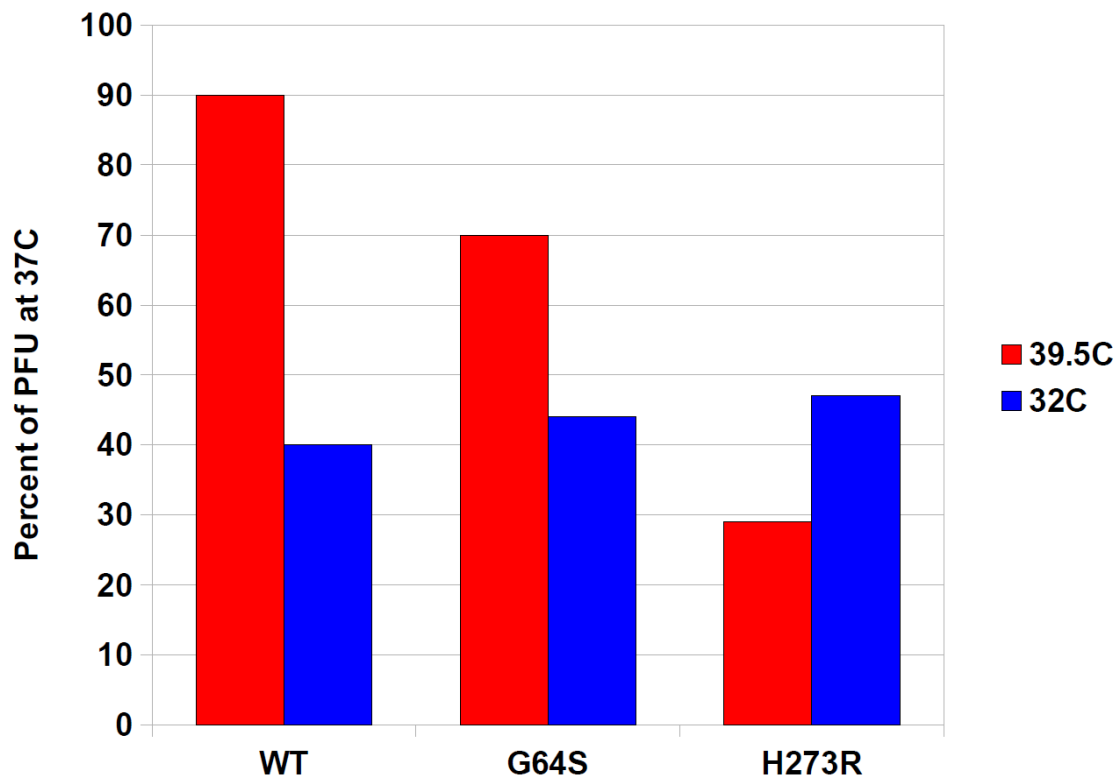
Figure 1:

Temperature sensitivities of poliovirus RNA-dependent RNA polymerase mutant populations. All populations shows reduced viability at non-optimal temperatures. Compared to WT, both G64S (the high-fidelity variant) and H273R (the low-fidelity variant) show reduced viability at elevated temperature and a higher level of viability at lowered temperature.

This experiment was repeated with the same viral stocks with the only change being allowing the 32ºC incubation to occur for 4 days rather than 3 days. While this extended incubation appeared to allow smaller plaque to develop and become visible, it did not affect the trends discovered in the first experiment (Figure 2). Both RNA-dependent RNA polymerase variant populations show fewer viable virions than WT at elevated temperatures and a higher portion of viable virions than WT at reduced temperatures.
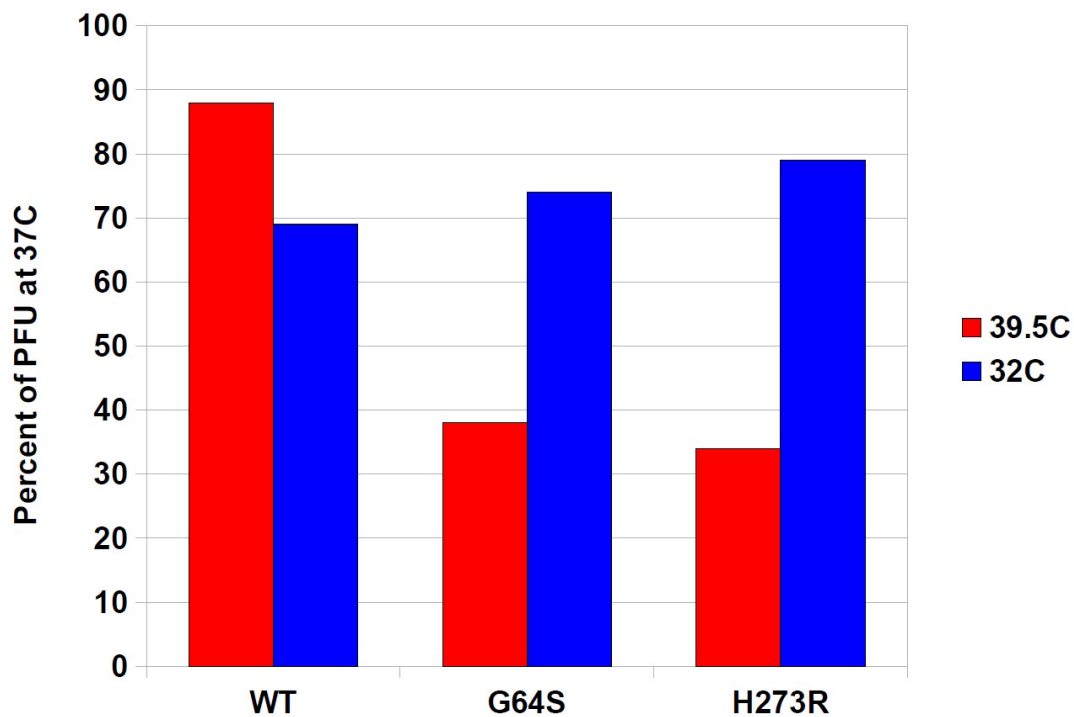
Figure 2:

Temperature sensitivities of poliovirus RNA-dependent RNA polymerase mutant populations (repeated). All populations shows reduced viability at non-optimal temperatures. Compared to WT, both G64S (the high-fidelity variant) and H273R (the low-fidelity variant) show reduced viability at elevated temperature and a higher level of viability at lowered temperature. While a longer incubation of 32ºC plaque assays allowed more plaque to develop and become visible, the trends displayed in Figure 1 are recapitulated.

Directed, experimental approach to perturb protein stabilities of polioviral populations

Next, we took a more directed, targeted approach at manipulating the thermodynamic stability of essential poliovirus proteins. Using thermodynamic predictions, specific resides of poliovirus proteins were targeted for mutagenesis. For selected residues, the WT amino acid was replaced with every non-WT amino acid to create a library of mutants representing every possible amino acid at a given position. These mutant populations were passaged several times and their relative proportions in the population were examined. Deep sequencing of these passaged populations is expected to give a fitness readout for each individual amino acid. This fitness readout will be used as a proxy for the stability of each mutant residue, although further in vitro measurements should be used to more concretely determine the nature of such fitness differences.

First, 24 residues of the poliovirus 3D protein were selected based on the expected severity and range of effects of mutations to each of the 19 non-WT amino acids at that position. Residues were chosen to in an attempt to balance all possibilities: residues where many mutants are expected to debilitate stability, residues where roughly half of the mutants are expected to debilitate stability, and residues where very few mutants are expected to debilitate stability. Also, residues in catalytic sites, those known to be part of an interaction interface, and those nearest to the surface were excluded, as lowered fitness of mutating these residues will likely have alternative, confounding effects and are therefore less likely to report directly on protein stability.

Thermodynamic predictions were made by Tzachi Hagai. Table 1 describes the 24

chosen residues. Figure 3 shows the physical location of each of the 24 selected

residues in the poliovirus 3D protein structure.

library#	PV genomic position	WT codon	WT residue
1	6062	GAA	Gly
2	6212	GAG	Glu
3	6290	TAT	Tyr
4	6305	CTA	Leu
5	6413	AAA	Lys
6	6512	ATT	Ile
7	6584	AAA	Lys
8	6635	TTT	Phe
9	6701	GCA	Ala
10	6731	CTA	Leu
11	6749	AAA	Lys
12	6767	AGA	Arg
13	6788	CTA	Leu
14	6806	CTG	Leu
15	6818	AAA	Lys
16	6881	ATG	Met
17	6926	AAG	Lys
18	6953	ATG	Met
19	7046	ATG	Met
20	7148	ATT	Ile
21	7157	GTA	Val
22	7211	AAC	Asn
23	7280	TTC	Phe
24	7295	AGG	Arg

Table 1:

Table noting each of the 24 selected residues, their shorthand library number, genomic

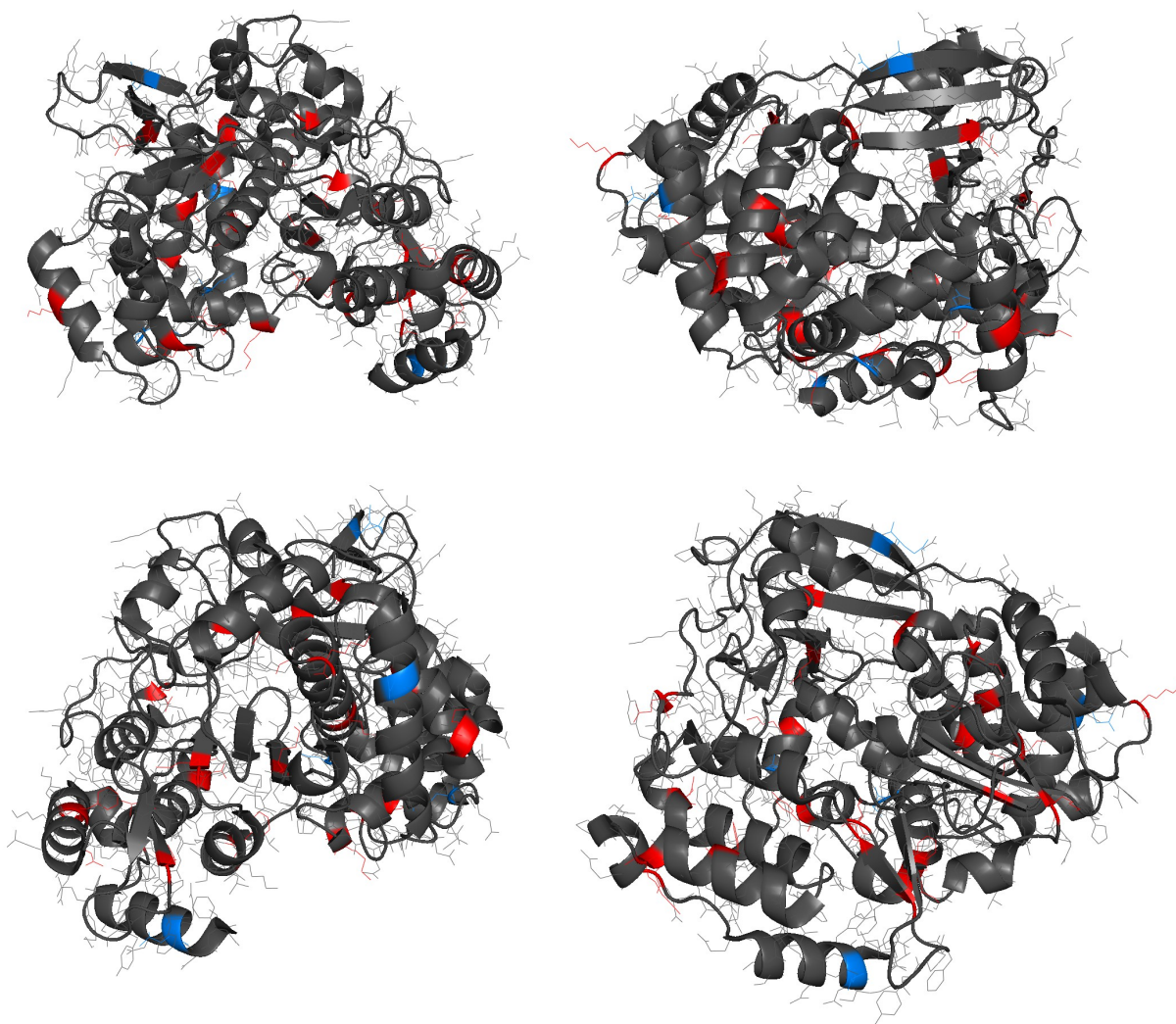position, and WT codon and residue.

133

Figure 3:

Four views of the poliovirus 3D protein structure with the 24 selected residues highlighted in red. Residues were selected in an attempt to directly perturb stability without affecting binding of other proteins or disrupting catalytic reactions.

Second, specific oligos introducing each of the 19 non-WT amino acids to each of the 24 positions were designed and ordered. Codons were chosen to increase the distance in sequence space between WT and mutant by mutating as many of the three bases in the mutant residue as possible. These 19 non-WT oligos were mixed in equimolar ratios and cloned into the standard WT poliovirus plasmid backbone as a library using an overlap cloning method. To determine the purity of the mutant population and to ensure sufficient removal of the WT starting material, individual clones were sequenced. WT carry-over from the cloning procedure was determined to be roughly 5%, which was deemed sufficiently low to more forward with. Once linearized, these plasmid libraries were used to construct a population of in vitro transcribed RNA, varying 20 ways over a single codon (19 mutants plus WT background).

Third, library in vitro transcribed RNA was transfected into HeLaS3 cells to produce a viral population containing all (or likely all) viable mutants representing all 19 non-WT amino acids at a particular residue in the 3D protein. Ideally, a low MOI would be used to generate the initial viral population. However, transfection does not provide the opportunity to control MOI. It is likely that much of the initial viral population produced from the transfection occurred in infections which were initiated by several mutants, allowing for some complementation to occur. Figure 4 illustrates the titers of each library population produced from the P0 transfection, shown relative to a WT transfection. If all mutants were as stable as WT, titers would be expected to be 100% of WT. If all mutants were instable to the point of being dead, titers would be expected to be produced only by background levels of WT.

Figure 4:

Passage 0 titers of the 24 selected residue library populations transfections relative to WT transfection. If all mutants were as stable as WT, titers would be expected to be 100% of WT. If all mutants were instable to the point of being dead, titers would be expected to be produced only by background levels of WT (roughly 5%).
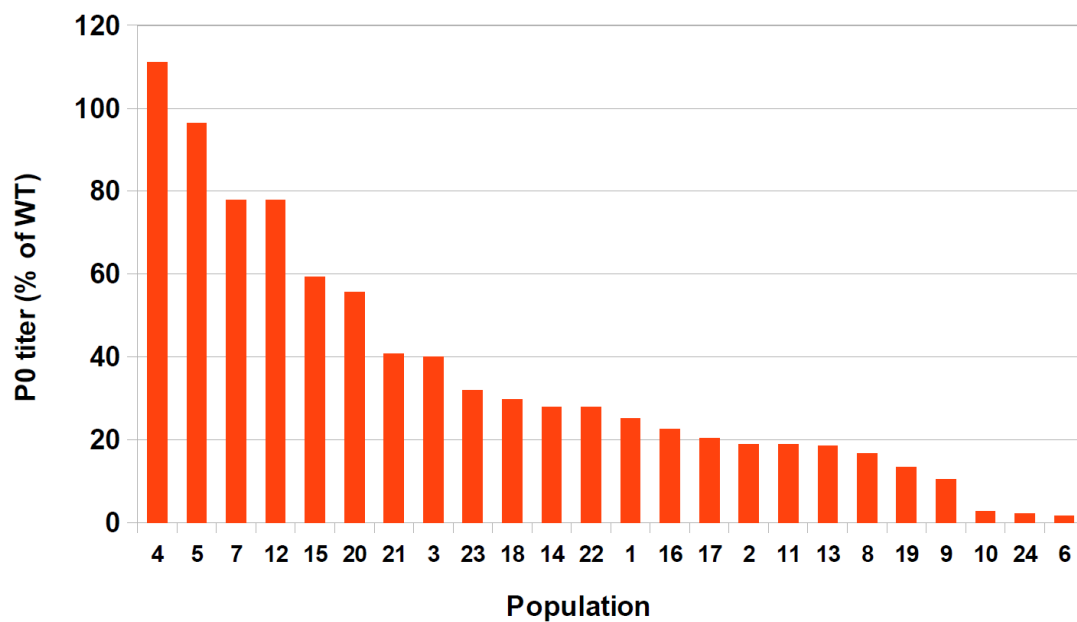
Populations were furthered passaged at low MOI (0.1) and titered. Figure 5 shows the overall increase in titer over passages 0-3. Most of the populations appear to drop or increase only mildly in titer between passage 0-1, suggesting that many of the mutants in these populations are unproductive and likely being selected against.
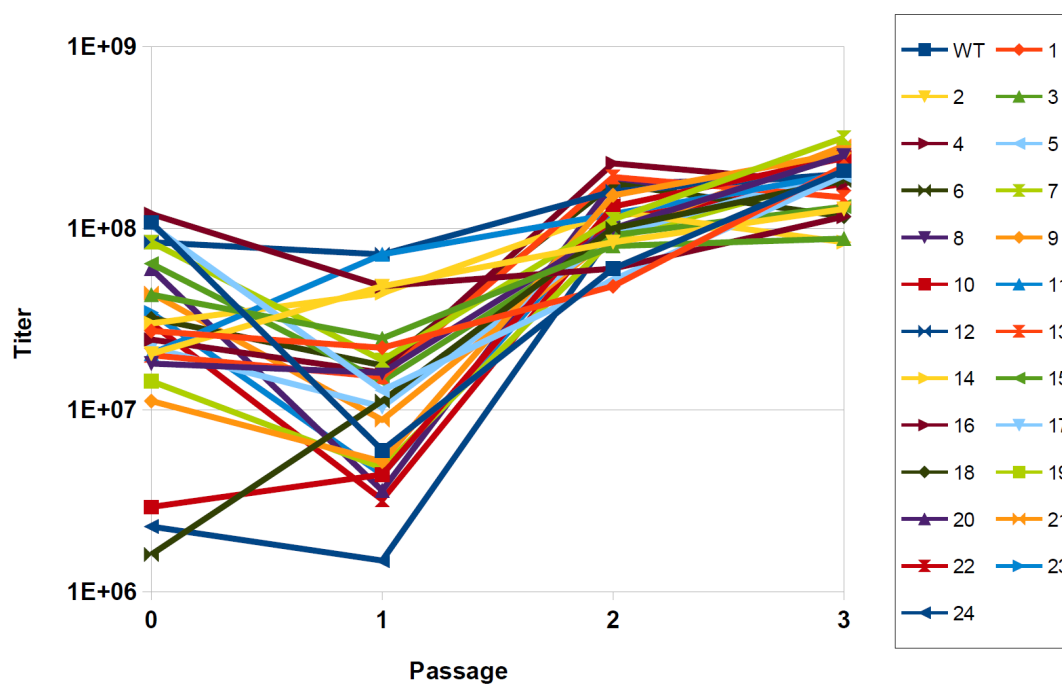
Figure 5:

Titers of the 24 selected residue library populations over 3 passages, alongside WT.
Most titers first appear to drop or increase mildly, suggesting a significant portion of the
population is being selected against in the first passage. Titers increase as expected
with the following passages.

Discussion

Despite their simplicity, the initial descriptive experiments in this chapter do well to suggest that the model of protein stability being a speed limit to molecular evolution can hold some weight. More-heavily mutated viral populations saw less of a drop in titer from a non-optimal lowered temperature, presumably from a resuscitation of mutants that, for thermodynamic stability reasons, would not have been viable under normal growth conditions (Figures 2, 3).

Further work is need to complete the directed protein perturbation experiments initiated in this project. The initial glimpses of information seem promising. P0 titers span a wide range of titers relative to WT, suggesting some residues are very insensitive to mutation while other are very sensitive to perturbation (Figure 4). Also, serial passaging of many of the mutant library populations showed a decrease or only a modest increase in titer in the first low MOI passage, suggesting that a considerable portion of the population may have been selected against (Figure 5). However, as an important point, all mutations have only been inferred to affect stability. Deep sequencing of these populations will be necessary to shed further light on the relative increase or decrease of each mutant in each population. Only the collection of correlations between fitness inferences drawn from deep sequencing data and the thermodynamic stability predictions of each particular mutant at each residue can act as a substantial test to the model of the limit to mutation rate.

References

1. Zeldovich KB, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *PNAS* 104(41): 16152-16157.

2. Vignuzzi M, Wendt E, Andino R (2008) Engineering attenuated virus vaccines by controlling replication fidelity. *Nat Med* 14(2): 154-161.

3. Arnold JJ, Vignuzzi M, Stone JK, Andino R, Cameron CE (2005) Remote site control of an active site fidelity checkpoint in a viral RNA-dependent RNA polymerase. *J Biol Chem* 280(27): 25706-25716.

4. Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *PNAS* 98(12): 6895-6900.

Conclusion

Error Catastrophe

In Chapter 1, our objective was to describe the characteristics of a low-fidelity polymerase mutant population and investigate whether Muller's ratchet and subsequent mutational meltdown occurs in tissue culture and animal systems. Our experiments used sequencing, passaging, and infection in a mouse model to elucidate the nature of low-fidelity replication in population dynamics of poliovirus. Unfortunately, previous results were not reproducible and our H273R population appeared less low-fidelity and more WT-like than previously observed.

Our tissue culture passaging experiments illustrated that tightly controlled passaging of H273R populations at either low or hight MOI does not result in Muller's ratchet-like behavior or subsequent mutational meltdown. Previous results were likely artifactual as less well-controlled passaging (blind passaging) demonstrated the ability of small, accidental differences in multiplicity of infection to be compounded through subsequent passages creating the illusion of extinction.

Stochasticity

In Chapter 2, we sought to uncover the sources of variation in single cell viral infections. We followed isolated individual cells infected with poliovirus at low or high MOI and measured viral genomic replication and infectious viral progeny in each cell. We report the first true positive to negative strand ratio of a positive-strand virus in

infected cells, with a ratio of 20 at 7 hours post infection. We find that while virus genome production is higher in cells infected at a high multiplicity, the production of infectious particles is largely independent of the number of viruses infecting each cell. Strikingly, by correlating RNA and particle production within individual infections, we find a significant contribution of stochastic noise to the outcome of infection. At low MOI, stochastic influences appear as kinetic effects which are most critical at the initial steps in infection. At high MOI, stochastic influences appear to dictate the viruses ability to harness cellular resources. We conclude that biological noise is a critical determinant of the overall productivity of viral infections. The distinct nature of stochasticity in the outcome of infection by low and high numbers of viral particles may have important implications for our understanding of the determinants of successful viral infections.

Replication Mode

In Chapter 3, we sought to combine years of knowledge of the molecular mechanism of poliovirus replication with new computational methods to infer the replication mode of the virus. We measured RNA dynamics through the poliovirus infection cycle, and applied these data to infer the parameters of a stochastic mathematical model. We find that poliovirus differs significantly from the stamping machine model. Our results suggests that for each infection cycle in a human cell, the average progeny virus is approximately 5 genomic replication cycles away from the infecting virus, providing the potential for an expansive mutant population to be generated from a single cellular infection event. Multiple genomic replication cycles

142

within a single cellular infection cycle create the potential for significant intracellular selection, especially in infections initiated by multiple genomes.

Thermodynamic Stability of Essential Proteins

Despite their simplicity, the initial descriptive experiments in Chapter 4 do well to suggest that the model of protein stability being a speed limit to molecular evolution can hold some weight. However, further work is need to complete the directed protein perturbation experiments initiated in this project. While the initial glimpses of information seem promising, all mutations have only been inferred to affect stability. Deep sequencing of these populations will be necessary to shed further light on the relative increase or decrease of each mutant in each population. Only the collection of correlations between fitness inferences drawn from deep sequencing data and the thermodynamic stability predictions of each particular mutant at each residue can act as a substantial test to the model of the limit to mutation rate.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author
Signature_____*Michael Schulte*_____ Date_ 09-01-14 _____