Distribution Agreement

Signature:

______________________________ ______________________________

Anand Bhardwaj Date

Effects of Life History and Genome Architecture on ssRNA Virus Evolution and Extinction

By

Anand Bhardwaj
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Population Biology, Ecology and Evolution

Leslie Real, Ph.D.
Advisor

David Cutler, Ph.D.
Committee Member

Jacobus de Roode, Ph.D.
Committee Member

Lance Waller, Ph.D.
Committee Member

Michael Zwick, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Effects of Life History and Genome Architecture on ssRNA Virus Evolution and Extinction

By

Anand Bhardwaj

B.S., Emory University, 2007

Advisor: Leslie Real, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Population Biology, Ecology and Evolution
2013

Abstract

Effects of Life History and Genome Architecture on ssRNA Virus Evolution and Extinction

By

Anand Bhardwaj

Single-stranded RNA viruses have evolved to survive extremely high mutation rates. The ubiquity and effect of ssRNA viral diseases makes an understanding of the theoretical and mechanical underpinnings of rapid viral evolution vital to our ability to control them. In this body of work, we explore some of the ways in which ssRNA viruses can uncouple the rate at which variation is generated (mutation rate) from the rate at which variation is observed (measured rate of molecular evolution).

A combination of replication strategies and genome architecture allow ssRNA viruses to evolve rapidly while avoiding many of the consequences of their error-prone replication process. However, this also means that ssRNA viruses exist at the very periphery of viable parameter space. Our models of viral evolution suggest that this can be exploited as a means of viral control, an idea that is reflected in the relatively new and experimental process of lethal mutagenesis.

We also highlight the general need for more molecular data and better estimates of viral replication parameters. Ironically, the latter is rare in scientific literature because of a lack of awareness of their impact on the rates of ssRNA virus evolution, and not because of any particular difficulty in obtaining them.

Effects of Life History and Genome Architecture on ssRNA Virus Evolution and Extinction

By

Anand Bhardwaj

B.S., Emory University, 2007

Advisor: Leslie Real, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Population Biology, Ecology and Evolution
2013

Contents

Chapter 1

On Painting a Better Portrait

It inevitably rests on any researcher to justify the very existence and relevance of her or his work. For doctoral researchers, after 5-odd years of toil, this request might seem unfair. However, in our times, when funding is sparse, and prioritization is a must, it is important to understand and to state the broader context in which to place our research. So what is the utility of this body of work? The short answer to that question is: it helps us paint a better portrait of the archetypical virus.

Why do we care about ssRNA *virus evolution?*

Evolution, as we understand it, is the result of the interaction of multiple processes - a mutation process generates prime variation; a filter process then eliminates a majority of this variation. We count up what little makes it through the filter, call it the measured rate of evolution and use it to infer something about the filter process itself. For most evolutionary processes, this filter is some combination of selection and drift. In the case of single stranded RNA (ssRNA) viruses, however, this filter process is largely driven by the unique demographic conditions and replication processes of ssRNA viruses. These viruses have evolved unique life history strategies that allow them to *get away with* a rapid but error-prone replication process. In a sense, ssRNA viruses have evolved a replication process that uncouples the rate at which variation is generated (mutation rate) from the rate at which variation is observed (the measured rate of molecular evolution).

ssRNA viruses are ubiquitous and are responsible for many of the most reconizable diseases known to man: Rabies, influenza, dengue, hepatitis C, measles, polio, and ebola are just a few of the more recognizable ssRNA viruses. The ubiquity and consequences of viral diseases, measured in terms of mortality and morbidity, make the understanding of viral evolution vital to our ability to control them. Many of our major concerns relating to ssRNA viruses stem from their evolution and rapid emergence [Lederberg, 1998, Sardanyes et al., 2009, Domingo et al., 2001].

In addition, from a purely intellectual standpoint, RNA virus mutation rates, which are orders of magnitude greater than comparable DNA-based viruses, allow RNA viral evolution to occur at ecological times scales, and allow us to ask and address questions about the interaction of ecological and evolutionary processes [Holmes, 2009, Pybus and Rambaut, 2009].

Many of our expectations of the patterns and rates of virus evolution comes from population genetic literature [Orr, 2000, Orr, 2003]. While these are elegant and appropriate for describing evolution in broad strokes, they fall short of describing the complexity of viral replication and demographics and the various strategies by which viruses dodge the deleterious effects of a high error rate.

A recurring theme of this dissertation is an appeal for more data and better estimates of parameters. In the following chapters, I hope to demonstrate the enormous value of having information on the mode of replication, template number, and fecundity of viral populations. High quality estimates of these parameters are rare - not because of any particular difficulty in obtaining them [Drake, 1993, Drake and Holland, 1999, Garcia-Villada and Drake, 2012], but because of a general lack of awareness of their impact on the rates of molecular evolution.

In the following chapters I look at different aspects of viral evolution, and explore the ways in which the dynamics of the viral replication process, viral demographic parameters, and genome architecture affect the measured rate of molecular evolution in single stranded RNA viruses.

In **Chapter 2**, I explore the upper limits of viral evolution through the lens of demographic extinction, and the phenomenon of lethal mutagenesis. This is a relatively new and experimental process of viral control, in which artificially increasing the mutation rate of a viral population can cause rapid extinction. This process exploits the long-standing idea that RNA viruses, by virtue of their highly error-prone replication process, are near some critical upper bound in mutation rates, more so than DNA-based organisms. What exactly constitutes this upper bound is unclear. The theoretical underpinnings behind this phenomenon are still poorly understood, and existing models have poor predictive ability [Bull et al., 2013]. I examine the limitations of existing models that attempt to explain lethal mutagenesis, and then proceed to present my own models, which I believe are more accurate because they account for critical complexities in the viral intracellular replication process that are ignored by existing models.

While viral extinction as a means of control is important in its own right, understanding the processes that determine the rate of viral mutation and evolution in nature are equally important. In **Chapter 3**, I dissect a long-standing assumption about viral evolution - that adaptive substitutions can only occur on viral genomes that are free of deleterious mutations. I look at the validity of this assumption with different types of viral genome architecture. I use a massive dataset of RNA sequences collected from patients of H3N2 Seasonal Influenza Type A, and Dengue Type I, II, and

III from around the world, and Bayesian phylogenetic techniques to estimate the rate of neutral, adaptive and overall evolution in these two viral systems, to elucidate how genome architecture affects the *effective adaptable population size* of viruses.

Chapter 4 is a brief note and rebuttal to a peer reviewed and published paper from 2012 which uses a demonstrably wrong mathematical model to assert the existence of an “error threshold” in nature based on an alleged (but fundamentally wrong) relationship between mutation rate and the rate of neutral evolution. My rebuttal is supported by long-standing theoretical results and by new evidence I present in Chapter 3.

In **Chapter 5** I develop and present a general mathematical framework to describe the expected rate of adaptive evolution in ssRNA viruses that can take into account and explore the effects of variation in replication dynamics and demographics. I use this mathematical framework to predict conditions under which rapid fixation of small effect adaptive mutations are possible, and how the measured rate of molecular evolution can be of limited informative value in the absence of replication parameters like mode of replication, template number and viral fecundity and population size.

Chapter 2

Demographic Extinction of ssRNA Viruses

Anand Bhardwaj, Leslie A. Real & David J. Cutler

In this study model I the effects of variation in the process of intracellular replication of single stranded RNA viruses on the rate of extinction of viral infection chains. I highlight the importance of mutations on opposite sense templates, which act as mutation-accumulation bottlenecks, amplifying the frequency of lethal mutations. I explore the drawbacks of ambiguous and biologically simplistic mathematical descriptions of the phenomenon of lethal mutagenesis, which cannot be used to make reliable predictions of the conditions required to drive populations to extinction.

2.1 Introduction

Viruses, and single stranded RNA viruses in particular, have extremely high mutation rates that are serveral orders of magnitude higher than those for humans and other DNA-based organisms [Sanjuan et al., 2010, Drake and Holland, 1999]. High mutation rates in tandem with - and perhaps as a consequence of - rapid replication allow viruses to quickly evolve and adapt to novel environments [Elena and Sanjuan, 2005]. In order to understand the limits to viral evolution in the wild at the population level and to effectively exploit this as a potential means of control at the host level, a better understanding is needed of the ways in which ecological characteristics of viruses - like fecundity and the mechanism of intracellular replication - affect the ability of viral populations to survive their high mutation rates.

Viruses have a life cycle that might help them mitigate some of the deleterious effects of mutation. Typically, a single stranded RNA viral particle enters a host cell, creates some number of opposite sense RNA templates, and produces a number of progeny. Some of these progeny can themselves act as the basis for secondary templates for use in a subsequent round of repliation. After some t_c rounds of replication, some target fecundity is reached and viral particles are released into the extracellular environment, either all at once by lysing the cell, or gradually, by budding off the cell membrane.

The precise number of rounds of viral replication t_c within the cell can vary significantly [Chao et al., 2002, Duffy et al., 2002, Garcia-Villada and Drake, 2012]. At one extreme, all progeny genomes produced per generation may originate from a single original template as a result of a single round of replication - the stamping machine mode of viral replication where mutations accumulate linearly. At the other extreme is binary replication, where the number of viral particles doubles after each round of replication, with templates for further replication being produced from early copies. Under this mode of replication, mutations accumulate geometrically. The precise mode of replication for specific viruses is unknown, but it presumably varies between these two extremes.

Here, and for the rest of this study, I define a generation as a single cell infection cycle, during which a virus infects a cell, replicates within the cell by highjacking its molecular machinery, ultimately releasing daughter viruses into the extracellular environment. Recent work suggests that explicitly modeling the within-host replication process reveals significant deviations in expected evolutionary outcomes, when compared to simpler models of evolutionary escape and emergence [Loverdo et al., 2012].

In this study, I examine the effects of variation in the processes of intracellular viral replication [Duffy et al., 2002, Holmes, 2009] on deleterious mutation accumulation and consequently the persistence of viral infection chains over the short term [Sardanyes et al., 2009].

2.1.1 Lethal Mutagenesis

RNA viruses have evolved and evidently tolerate extremely high mutation rates [Holland et al., 1982, Drake and Holland, 1999]. On the other hand, artificially increasing the mutation rate of riboviral populations 3-4 fold through treatment with chemicals like ribavirin has been shown to cause a catastrophic drop in population fitness, and population extinction by lethal mutagenesis within a few generations [Holland et al., 1990, Crotty et al., 2000, Crotty et al., 2001, Domingo et al., 2005]. Consequently, the consensus that RNA virus mutation rates are near some threshold for tolerance is a long-held one, beginning with, but not limited to Manfred Eigen's formulation of an evolutionary extinction threshold [Eigen, 1971, Eigen, 2002, Beibricher and Eigen, 2005] beyond which the error rate is too high for information contained in nucleotide molecules to be successfully passed on from generation to generation

However, the current understanding of lethal mutagenesis is defined by a demographic extinction threshold. Extinction occurs when the mutation rate of a virus is high engough to prevent successful population replacement, when a virus infecting a single cell can no longer produce enough viable progeny to go on to successfully infect one or more other cells [Bull et al., 2007].

Some deleterious mutations are of large enough effect to consistently prevent the perpetuation of riboviral cell infection chains. If these mutations can be classified as lethal, then lethal mutagenesis could be said to occur when the rate of mutations per generation is high enough to bring the expected number of lethal mutation-free viral particles released per generation below 1. Assuming that the number of lethal mutations occuring per genome in a single generation is Poisson distributed with mean U_cX_L, where U_c is the mean number of mutations per genome per cell infection cycle and X_L is the proportion of mutations that are lethal in effect, the fraction of viral progeny with no lethal mutations or the lethal mutation zero class p_0 can be described as: $p_0 = e^{-U_cX_L}$. The number of viral progeny per generation with no lethal mutations is $e^{-U_cX_L} \cdot N_c$, where N_c is the total viral fecundity per infected cell. This allows an extinction criterion to be set. For lethal mutagenesis to occur, the expected number of lethal mutation-free viral progeny produced per cell must be less than one [Bull et al., 2007]. That is,

$$e^{-U_cX_L} \cdot N_c < 1 \tag{2.1}$$

For extinction under this condition, the number of lethal mutations/genome/cell U_cX_L must be greater than $Ln(N_c)$.

The inequality above assumes that all viable viral particles released on burst survive and go on to successfully infect a secondary cell and even a single viable viral particle is capable of perpetuating cell infetion chains. This mathematical formulation also suggests that the critical mutation rate required to achieve lethal mutagenesis is dependant on the total fecundity of the cell. Given a fitness landscape, viruses with higher fecundity can therefore be expected to have higher critical mutation rates than less fecund viruses.

2.1.2 Limitations of a Formulation based on the Zero Class

The above formulation for the conditions required for achieving lethal mutagenesis, however, is flawed in many ways. The two main drawbacks are that of ambiguity and a lack of biological realism.

The conditions for lethal mutagenesis are satisfied when the mutation rate per generation is theoretically high enough to bring the expected size of the lethal mutant zero class p_0 below 1. However, this zero class is a function of both the mutation rate and the mean fecundity of the virus in the absence of mutation. Therefore, an extinction condition based on the size of the zero class can be satisfied by applying an extremely high lethal mutation rate to a high fecundity virus, or with a moderate lethal mutation rate for a low fecundity virus. The case I would like to make in this study is that these two situations are not equivalent.

The second major source of ambiguity, as well as the primary lack of biological realism in the zero class formulation, comes from the measure of mutation rate used in the zero class formulation of the demographic error threshold. The mutation rate U_c referred to up to this point is a measure of the mean number of errors on the viral progeny of a cell infection cycle, compared to the genome of the virus that initiated the cell infection. This is a relatively common measure of mutation rate, as it can be measured empirically [Sanjuan et al., 2010].

This mutation rate per cell infection cycle or generation is distinct from some true biological mutation rate U, which is a measure of the number of errors accumulated per genome during a single replication event, of which there could be several within a single generation. U_c is dependent on the mechanism of viral replication within a cell. Assuming the absence of selection during intracellular viral replication, the mean number of mutations per genome U_t in the viral population within a cell at any time t is a function of the number of mutations per genome per replication U and the t rounds of replication the population has gone through. Therefore, $U_t = U2t$ [Drake and Holland, 1999] and the number of mutations per genome at the end of a cell infection cycle $U_c = U2t_c$, where t_c is the number of rounds of replication required to produce the viral progeny. The number of mutations per genome scales with twice the number of rounds of replication because each round of replication involves the formation of an opposite sense template, which can accumulate errors as well [Domingo et al., 2001].

The dynamics of viral intracellular replication suggests that while a mean value of U_c can be achieved either by a high biological mutation rate U and a low value of t_c or by a low value of U and a high value of t_c, the distribution of mutations per genome among the viral progeny is likely to be different in these two cases, as mutations accumulated during early replication events are likely to be passed on to all subsequent genomes further down along the intracellular replicative lineage of the virus.

2.2 Methods

2.2.1 Modeling Viral Intracellular Replication

In order to look at the impact of different replication strategies on the ability of a virus to be driven to extinction, I explicitly modeled the process of intracellular replication as a modified Walton-Gaston branching process [Feller, 1968]. This framework allows us to estimate the extinction probability of individual lineages within a cell in some pre-determined number of replication events, as well as the impact of fecundity and the number of opposite sense templates generated per replication event.

I assume here that the rate of mutation during the production of the opposite sense template is the same at the rate of mutation during the production of a daughter progeny from that template. This is a reasonable assumption in the case of single stranded RNA viruses where these steps are carried out by similar RNA-dependant RNA polymerase enzymes [Ahlquist, 2002]. This assumption would be biologically inaccurate in the case of retroviruses because of the inclusion of intermediate DNA steps during the replication process which involve additional enzymes with different rates of error.

Assuming discrete rounds of replication, if viral fecundity N_c can be described by an equation for exponential growth where $N_c = x \cdot g^{t_c}$, and x describes the number of opposite sense initial templates copied from a viral genome to produce its progeny, then mutations accumulate linearly when t_c is 1, and more geometrically as t_c approaches $Log_2(N_c)$. This assumes that a single viral particle initiates a cell infection.

Assuming that the number of direct offspring of any viral genome in a single round of replication is geometrically distributed of the form $\{qp^k\}$ with mean $p/q = g$, then we can also assume that the number of viable, lethal mutation free progeny produced from a single individual per replication event is $g \cdot e^{-2.U.t_c.X_L}$. This allows us to use the probability generating function for having s direct descendants after t rounds of a branching process with geometrically distributed offspring [Feller, 1968]:

$$P_t(s) = q \cdot \frac{(p^t - q^t - (p^{t-1} - q^{t-1}) \cdot p \cdot s)}{(p^{t+1} - q^{t+1} - (p^t - q^t) \cdot p \cdot s)} \tag{2.2}$$

This allows us to get an expression for the probability of an individual having no viable decendants after t_c rounds of replication. We can then combine that with the probability of a lethal initial template to get an expression for the probabilty of a viral infection chain going extinct in a single generation $P(E_1)$, as follows:

$$P(E_1) = (e^{-U \cdot X_L} \cdot \frac{(q \cdot (p^{t_c} - q^{t_c}))}{(p^{t_c+1} - q^{t_c+1})} + (1 - e^{-U \cdot X_L}) \cdot 1)/x \tag{2.3}$$

The above expression is the product of the probability of having a lethal template and the conditional probability of extinction given a lethal or non-lethal template. As in the original formulation based on the zero class, I assume that even a single viable virus produced at the end of a generation is sufficient to prevent extinction. An added advantage of this assumption is that it allows us to easily extend the above expression to get the probability of extinction in some n generations, by replacing t_c with $t_n = n.t_c$.

2.2.2 Simulation

I used a mechanistic stochastic simulation to test my predictions. All simulations were written and implemented in R [R Development Core Team, 2013], and were based on Poisson processes for growth and mutation. In each simulation a single individual was allowed to replicate and accumulate mutations until the population reached a target size. Growth was modeled stochastically: the number of secondary particles n produced at time t by a single template i was sampled from a Poisson distribution with mean $g = e^r$. The number of mutations $m_{i(t)}$ accumulated by an individual i at replication t was sampled from a Poisson distribution with mean U.

For simulations with non-lethal deleterious mutations, the fitness effect w of each mutation was sampled from a previously published distribution of fitness effects of random point mutations in Vesicular Stomatitis Virus VSV [Sanjuan et al., 2004, Sanjuan, 2012]. Compensatory or back mutations were not allowed and fitness effects were combined additively to get total fitness effect E of all mutations accumulated by individual i at replication t ($E_{i(t)} = \sum_{j=1}^{m_{i(t)}} w_j$). To account for errors that are passed on to all offspring produced in a single round of replication when templates themselves are mutated, an additional template error step was incorporated into the stochastic algorithm. I assumed that there is no selection on viral particles during the initial rounds of intracellular replication, but that selection is imposed on the viruses by the intercellular environment [Drake and Holland, 1999]. The number of mutations accumulated by each individual and the additive fitness effects of each mutation were recorded for each individual in the growing population and at the end of the replication. All simulations were repeated 1000 times each unique combination of parameters.

2.3 Results

Overall, I found large variation in the probability of extinction explained by differences in the mode of replication, number of templates and mean fecundity of the virus. This variation is entirely unaccounted for in the simple extinction condition described in Equation 2.1.

2.3.1 Variation in $P(E_1)$ under extinction conditions

The probability of extinction in a single cell infection cycle varied with the value of mean fecundity at mutations rates that satisfity the extinction condition for lethal mutagenesis (Figure 2.1a). This variation in extinction probabilities was even more drastic with an increase in the number of initial templates for replication. The probability of extinction in a single cell infection cycle at critcal mutation rates that satisfy

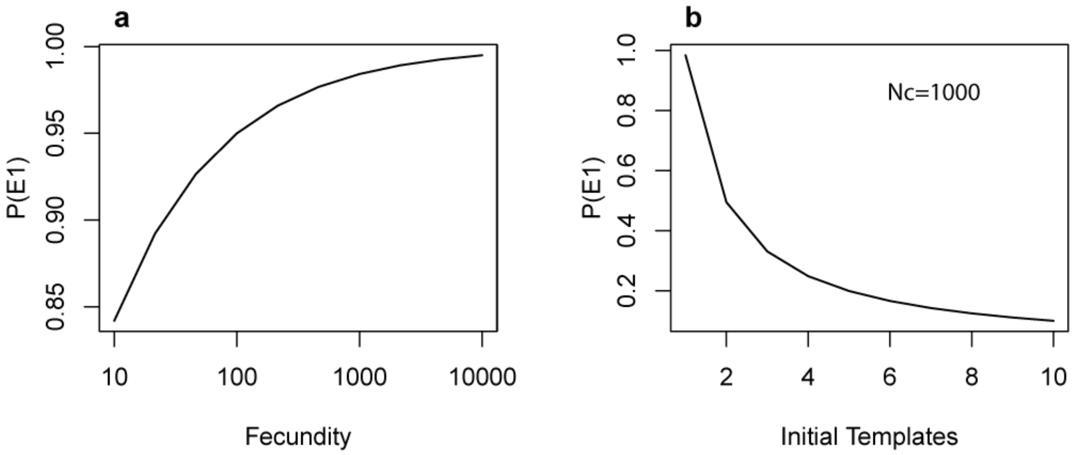

Figure 2.1: Variation in the probability of exctinction at critical mutation rates that satisfy the zero class extinction condition described in Equation 2.1. a) Preventing replacement for a high-fecundity virus is associated with a greater probability of extinction than preventing replacement in a low-fecundity virus; b) The probability of extinction in a single generation varies radically with the number of initial templates. With many templates, viral populations may easily survive critically high mutation rates.

the zero class extinction condition described in Equation 2.1 drops rapidly with an increase in the number of initial templates (Figure 2.1b).

2.3.2 Variation in $P(E_1)$ with replication parameters

Under conditions of linear replication, I found that the probability of extinction in a single cell infection cycle increases with mutation rate per generation. There doesn't appear to be much of an effect of fecundity on extinction curves, as the probability of exctintion given a particular mutation rate does not vary significantly for any value of mean fecundity on the order of 100 or higher (Figure 2.2a).

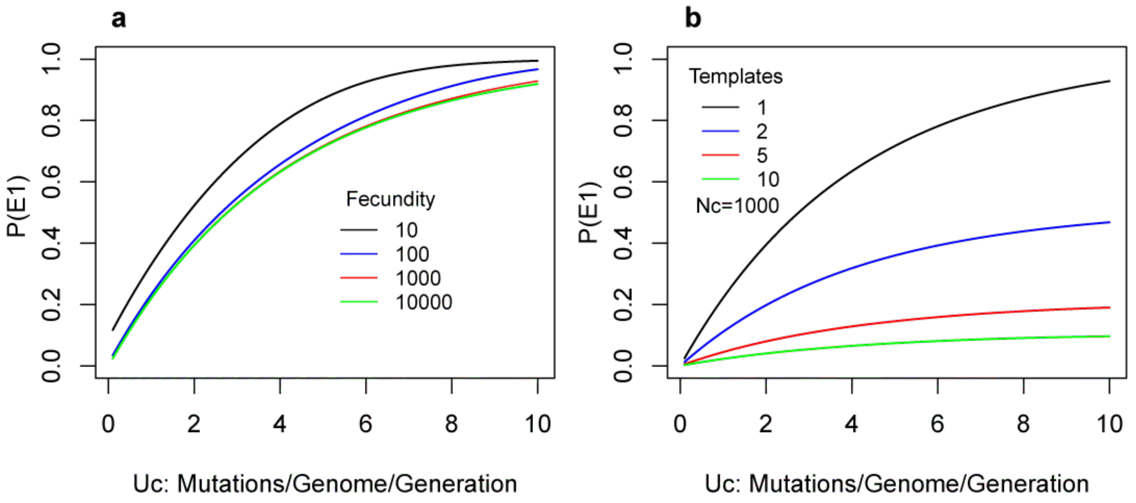

Figure 2.2: Variation in the probabily of extinction under linear replication. a) Effect of fecundity: The value of mean fecundity has only a marginal effect on the probability of extinction, given some mutation rate; b) Effect of template: The number of initial opposite templates utilized in the replication process has a large and significant impact on the probability of extinction.

Consistent with the results depicted in Figure 2.1b, the number of templates involved in replication greatly affects the probability of extinction in a single generation.

Under conditions of non-linear replication, if an increase in the number of rounds of intercellular replication is associated with an overall increase in the mutation rate per cell (U mutations/replication is kept constant), then the probability of extinction also increases (Figure 2.3a). This is expected as an increase in the number of rounds of replication leads to an increase in the mutation rate per generation. On the other hand, increasing the number of rounds of replication while keeping the mutation rate per generation constant also leads to an increase in the probability of extinction, albeit to a lesser extent than the former case. Given a fixed mutation rate per cell, there

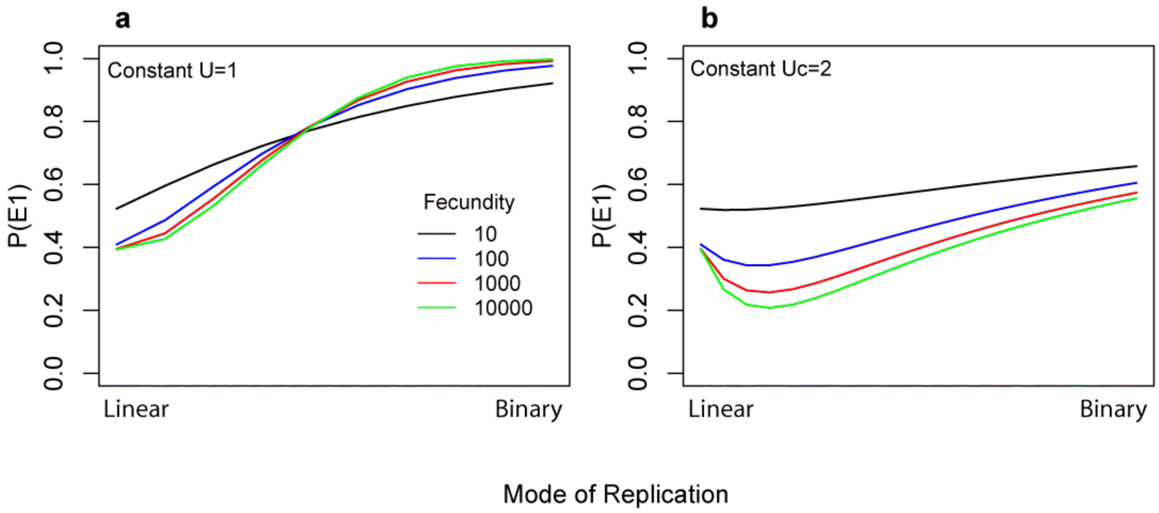

Figure 2.3: Variation in the probabily of extinction under non-linear replication. a) Variation in the probability of extinction with increasing mutation rate per generation; b) variation in the probability of extinction with constant mutation rate per generation. Some non-linear mode of replication minimizes the risk of extinction.

is some non-linear mode of replication that minimizes the probability of extinction (Figure 2.3b).

2.3.3 Extinction in a complex fitness landscape

The results of my simulations, assuming additive fitness effects of individual deleterious mutations, suggest that the mean fitness of viable viral populations is low at mutation rates well below any critical mutational threshold. Drawing from the original source of my empirically derived distribution of fitness effects [Sanjuan et al., 2004], fitness here is defined as the ability of an individual virus to successfully infect another cell. The true rate of extinction of viral populations is therefore likely higher

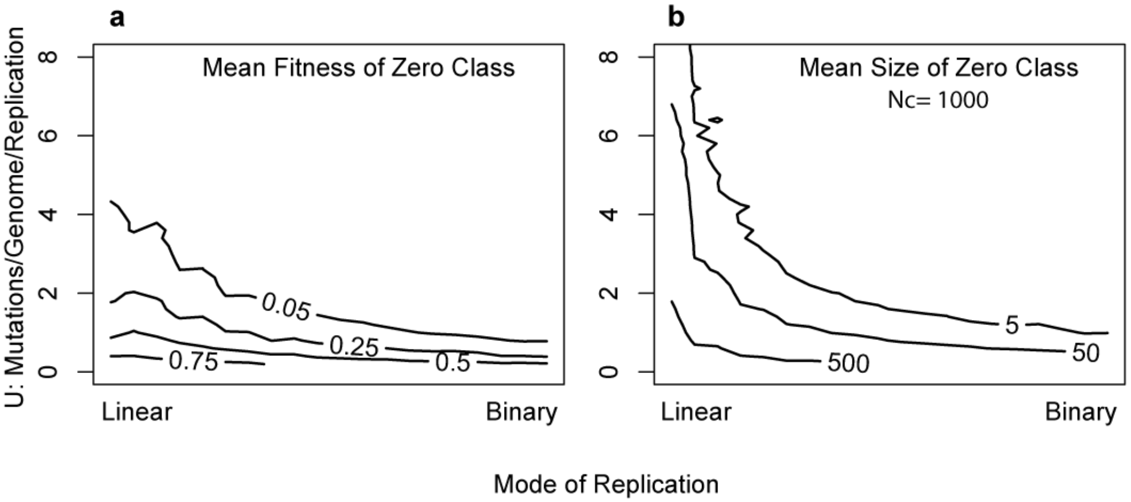

Figure 2.4: Effects of a complex fitness landscape. The presence of non-lethal deleterious mutation affects bother the a) mean fitness and b) size of the zero class.

than what my projections, which only account for lethal mutations, would suggest for a given mutation rate.

2.4 Discussion

2.4.1 A case for $P(E_1)$

If the point of any mathematical exploration of the phenomenon of lethal mutagenesis is to make it more practical for use as a means of within-host control, then an understanding of the rate at which populations will go extinct is ultimately necessary for any practical use. A simple result of branching process theory mirrors the extinction condition described in Equation 2.1 - Populations for which we can assert that the mean number of offspring per individual per replication event is less than 1, will go extinct with probability=1. However, this assertion provides us with no information on how quickly populations go extinct. Among other things, we see from the results depicted in Figure 2.1b, the probability of extinction of a viral population in a single generation at critical mutation rates that satisfy the extinction condition described in Equation 2.1 varies radically with the number of initial templates used in the replication process. This suggests that failing to account for variation in the intercellular replication process of viruses can explain some of the deviations from expectations of extinction in in-vitro tests of lethal mutagenesis [Bull et al., 2013].

The significance of mutations accumulated on the opposite sense templates cannot be overstated. In a sense, templates can act as mutation accumulation bottlenecks, with all subsequent progeny inheriting any lethal mutations, barring the rare reversion or back mutation. In the results depicted in Figure 2.3b, we can see, at least initially, that an increase in the number of rounds of replication (while keeping the mutation rate per generation constant by proportionally reducing the mutation rate per replication) causes a decrease in the probability of extinction. This is because, given a non-lethal initial template, the probability of exinction of a lineage decreases with increasing rounds of replication. However, at some non-linear mode of replication, the effect of accumulated probability of a lethally mutated template superscedes the effect of increase rounds replication in the branching process, beyond which the probability of extinction rises again. This adds to the narrative that what really drives extinction in viral populations is mutation on a template, rather than a mutation rate high enough to prevent lineage replacement.

The linear special case of Equation 2.3 depicted below makes it especially apparent that the probability of extinction is driven by the probability of accumulating some non-zero number of lethal mutations on a template:

$$P(E_1) = (e^{-U.X_L}) \cdot q) + (1 - e^{-U.X_L}) \cdot 1)/x \tag{2.4}$$

As we can see, the probability of extinction given a non-lethal template is expressed in the linear case as q, which is the probability that a template produces no direct decendants, which is very small when the mean number of offspring per individual g is high.

2.4.2 The importance of the intracellular replication process

Given that a formulation for demographic extinction based solely on the expected size of the zero class cannot, by virtue of its structure, account either for variation in the process of intracellular replication or for the effects of mutated templates, we must look to the effects of these processes in order to try and understand deviations from extisting theory in the context of practical applications of lethal mutagenesis.

Based on our understanding of the replication process, it is likely that given some measured mutation rate per generation, viruses with highly non-linear replication are subject to higher rates of extinction than viruses with linear or near-linear modes of replication. We also know that there is some non-linear mode of replication that minimizes the risk of extinction given some measured mutations rate per generation. The exact number of rounds of replication that constitute this *replicative optimum*, of sorts, is inversely proportional the template mutation rate.

Finally, we see that the estimated mean fecundity of the virus in the absence of mutation has only a limited effect on extinction probabilities, to the extent that highly fecund viruses likely replicate with some non-linear mode of replication or utilize a large number of initial opposite sense templates.

2.4.3 Implications for lethal mutagenesis in practice

Estimating or predicting the effects of a complex fitness landscape with multiple, interacting non-lethal deleterious mutations requires information about the nature of interacting mutations. The total fitness effect of multiple mutations depends on whether the fitness effects of individual mutations are additive, or subject to some epistatic interaction. In my simulations, I assumed that fitness effects are purely additive. Apart from the rare cases, having multiple deleterious mutations is worse that having just one. Therefore, it it is likely that demographic extinction of populations could occur at a higher rate than what my projections, which only take lethal mutations into account, might suggest.

Indeed, the results of my simulations of viral intracellular replication with a complex fitness landcape that include non-lethal deleterious mutations also suggest that in addition to affecting the effective fecundity of viral population, high mutation rates could also reduce the mean relative fitness of subsequent viable, non-lethal progeny, in the sense that multiple accumulated deleterious mutations could affect the ability of viral progeny with any lethal mutations to successfully reinfect other susceptible cells in the host, thereby increasing the proability of extinction of viral infection chains.

2.5 Conclusions

We can now begin to put together a better picture of viral evolution, and ways in which the intracellular replication process of viruses can serve to uncouple the rate at which variation is generated from rate at which new variation is observed.

An corollary to the assertion that most mutations are deleterious, is that not all are.

Mutation is a double edged sword, making survival a delicate balancing act for SS-RNA viruses given the inherently higher error rate and lack of proofreading function of most RNA-dependant RNA polymerase enzymes. Survival requires balancing the need for variation to respond to ever-changing environment with the risk of extinction. If my mathematical projection are at all accurate, RNA viruses have several methods of mitigating the risk of extinction given their high mutation rates. In particular, the apparent survival of viruses at mutation rates that, according to formulations of extinction criteria based on the expected size of the zero class, should cause rapid extinction, suggests that such formulations are overly simplistic, and a deeper understanding of the viral replication process and empirical estimates of replicative parameters like those used in my projections are necessary in order to refine methods of viral control by lethal mutagenesis.

2.6 Acknowledgments

I acknowledge the help and assistance of Gavino Puggioni in writing my simulation of intracellular viral replication. This research was supported by NIH grant RO1 AI047498 to LAR and the RAPIDD Program of the Science and Technology Directorate, the Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Chapter 3

Genome Architechture & Adaptive Evolution.

Anand Bhardwaj, David J. Culter & Leslie A. Real

Here, I examine the impact of genome architecture on ability of RNA viruses to increase the amount of available genetic variation on certain parts of the genome on which selection can act, while maintaining essential, conserved regions in other parts of the genome. I use a large dataset of RNA sequences to assess the impact of segmented genomes and reassortment on the ability of viruses to support highly variable genomic regions. I find that this ability is likely limited to segmented viruses or viruses that have high rates of recombination.

3.1 Introduction

RNA viruses evolve rapidly because of their high mutation rates and short generation times [Holland et al., 1982, Elena and Sanjuan, 2005]. However, various ecological and biological factors can serve to uncouple the rates of mutation and evolution, and this study is an attempt at infering whether genome architecture can affect the relationship between mutation rate and the rate of adaptive evolution.

In general, the rate of evolution, measured as substitutions per unit of time, is proportional to the rate of generation of new variation, or mutation rate, and some measure of the probability that these new mutations spread and fix in the population. In the case of selectively neutral mutations, the probability of fixation is the reciprocal of the population size [Kimura, 1962], while in the case of selectively advantageous mutations, the probability of fixation depends on the selective advantage confered by that mutation [Kimura, 1962, Kimura, 1964, Fisher, 1930].

Assuming that mutations are generated at roughly the same rate across the RNA virus genome, any large differences in the rates of adaptive and neutral substitution between genomic regions could therefore be due to differences in the biological factors that affect their spread and fixation, as well as differences in availability of adaptive and netural mutations.

3.1.1 Adaptive evolution in non-recombining asexuals

The presence of linked deleterious mutations slows down the rate of fixation of adaptive mutations and therefore the adaptive substitution rate in non-recombining asexuals [Orr, 2000]. This is due to the fact that for information-dense genomes like in the case of RNA viruses, deleterious mutations are both more common and on average, of larger effect than beneficial mutations [Sanjuan, 2010]. Adaptive evolution, therefore, is largely limited to beneficial mutations that occur on genomes that have not already accumulated any deleterious mutations [Fisher, 1930, Barton, 1995].

In general, the rate of adaptive evolution K_A (substitutions/genome/generation) is proportional to the number of new adaptive mutations generated by the population per generation, the probability that the new adaptive mutation occurs on a deleterious mutation-free genome, and the probability of fixation of the new mutation.

$$K_A \propto U_c \cdot X_A \cdot e^{-U_c \cdot X_D / s_H} \cdot \pi \tag{3.1}$$

Here, U_c is the mean number of mutations per genome per generation, X_A is the fraction of these mutations that are adaptive, X_D is the fraction of mutations that are deleterious, s_H is the harmonic mean of the of fitness effects of new mutations, and

π is the probabilty of fixation of new adaptive mutations. $e^{-U_c \cdot X_D/s_H}$ is the expected fraction of genomes free of deleterious mutations at equilibrium, assuming a Poisson distribution.

The above equation is equivalent to Equation 5 of Orr's publication on the rate of adaptation in asexuals [Orr, 2000]. From the general form of the expression, we can see that there is a non-linear relationship between the rate of mutation and the rate of adaptive evolution. The rate of adaptive evolution increases with increasing mutation rate, until an "optimal" mutation rate $U_{c,opt}$ is reached. The rate of adaptive evolution drops rapidly with any further increase in mutation rate, since any increase in adaptive mutations is negated by the lack of availability of genomes free of deleterious mutations on which adaptive evolution can occur.

Here, $U_{c,opt} = s_H/X_D$. We can see that the optimal mutation rate for a population depends on the fraction of mutations that are deleterious.

Orr's expression for adaptive evolution makes some simplifying assumptions, in the interests of generalizability, about the evolving asexual population in question, sidestepping the issue of evolution measured at a particular locus versus evolution measured over the entire genome. In the case of real evolving asexuals like RNA viruses, the rate of substitution is usually measured by gene, and substitution rate can vary among genes on the same genome [Jenkins et al., 2001].

Prior work on the effects of reassortment on the rate of evolution in RNA viruses suggests that segmented genomes can be advantageous as they could increase the amount of variation in viral genetic information on which natural selection can act [Pressing and Reanney, 1984]. Theoretically, in unsegmented RNA viruses, we can imagine that the presence of a highly conserved gene slows the fixation of adaptive mutations on genes elsewhere on the genome. In contrast, in RNA viruses with segmented genomes, the deleterious mutation rate of a particular segment of the genome could slow down the adaptive substitution rate of that segment, with deleterious mutations occuring on other segments having a limited effect.

In this context, since K_A is measured by gene, and X_A refers to the available beneficial mutations for that particular gene, is the rate of adaptive evolution of any given gene limited by the rate of deleterious mutation $U_c \cdot X_D$ on that gene, or by the rate of deleterious mutation over the entire genome? Does the presence of genes on a genome that code for essential functions, like the highly conserved RNA polymerase gene [Poch et al., 1990] - where we can assume that the value of X_D is high - slow down the rate of adaptation in other, distant parts of the genome that are not themselves under comparable evolutionary constraint?

In this study, I use a large dataset of genetic sequences from four genes each from Dengue and H3N2 Influenza viruse Type A to try and answer the questions alluded to in the preceding sections: Does genome architecture play a role in whether the rate of adaptive evolution of a gene is limited by the deleterious mutations on that gene, or on the entire genome?

3.2 Methods

100 population-level estimates of gene-specific substitution were made from populations of Dengue Virus 1, 2 and 3 and Seasonal H3N2 Influenza type A, from a total of 10,662 genetic sequences obtained from Genbank (see SI for accession numbers) [Benson et al., 2013]. Whole genome sequences from each population of virsuses were aligned using Geneious and the sequences of the genes of interest were extracted for substitution rate estimation [Drummond et al., 2011] so that the individual estimates of gene-specific substitution rate all came from the same set of individuals for each population of each viral species. Estimates of substitutions/nucleotide/year were made using BEAST, a Bayesian MCMC program that uses phylogenies with dated tips to estimate parameters [Drummond et al., 2012].

Codon-Position Substitution Rates: Substitution rate estimates were further divided into substitution rate on the 1^{st}, 2^{nd} and 3^{rd} codon positions for each gene. The substitution rate on the 3^{rd} codon position was used as a proxy for neutral substitution rate, since the degeneracy of the genetic code means that only a small fraction of these substitutions are expressed as amino acid changes [Lagerkvist, 1978]. While evidence suggests that synonymous substitutions are not always selectively neutral, at the very least, substitutions on the 3^{rd} codon position are *more neutral* than substitutions on the 1^{st} and 2^{nd} codon position [Schoniger et al., 1994]. Substitutions that occur on the 1^{st} and 2^{nd} codon positions, by virtue of very often being non-synonymous substitutions, were used as proxies of adaptive substitution rate [Xia, 1998], under the assumption that mutations that coded for amino acid changes would be unlikely to show up in representative samples of the viral genome if the associated amino acid change was selectively deleterious.

All subsequent analyses of the general and codon position-specific estimates of substitution rate were done using the STATS package in the R programming language [R Development Core Team, 2013].

3.2.1 Data

RNA sequences from four genes each from Dengue and Influenza were selected for use in this study. For each viral system, I selected two genes that were described in the literature as coding for essential functional enzymes under evolutionary constraint and two genes that code for structural proteins or genes under relatively less constraint.

Dengue and Influenza were selected as exemplars of unsegmented and segmented ssRNA viruses, in particular because of the relative abundance of freely available sequence data annotated with temporal and geographic information from these two viral systems. Sequences associated with lab strains, or any other non-representative sequences were not used in this study.

In Dengue type 1, 2 and 3, an unsegmented positive sense ssRNA virus and member of the viral family *Flaviviridae*, the Env gene, which codes for an structural surface protein, and NS1, a membrane associated glycoprotein with a role in combating host immune response were selected as less constrained genes [Schlesinger et al., 1990]. The genes NS3 - a protease-helicase, and NS5 - a methyl transferase-polymerase, were selected as genes that were potentially under relatively high evolutionary constraint because of their vital function [Perera and Kuhn, 2008].

In H3N2 seasonal Influenza type A, a segmented negative sense ssRNA viruses amd member of the viral family *Orthomyxoviridae*, the HA haemagglutinin and NA neuraminidase genes, both surface structural proteins, were selected as less constrained genes. The NP gene, which codes for a multifunctional nucleoprotein with a role in genome packaging and transport, and the PA gene, which codes for a polymerase, were selected as genes that are under high constraint [McCauley and Mahy, 1983, Portela and Digard, 2002]. Each of the four genes selected for Influenza occur on separate genome segments.

3.3 Results

3.3.1 Variation in k between species

I found significant differences in estimates of substitutions/nucleotide/year k between influenza and dengue (ANOVA, F(1,98)=191.46, $p << 0.05$), independent of gene.

I also found significant differences in the substitution rate on the 3^{rd} codon position - k_{CP3} between influenza and dengue. (ANOVA, F(1,98)=167.86, $p << 0.05$).

Despite influenza being more variable in overall substitution rate, no significant effects of population on substitution rate were found for influenza (ANOVA, F(12,19)=2.14, $p = 0.06$). In contrast, population-level estimates of dengue substitution rate varied significantly in substitution rate (ANOVA, F(13,46)=8,298, $p << 0.05$). In addition no significant differences in substitution rate were found between estimates of substitution rate from dengue types 1, 2 and 3 (ANOVA, F(2,57)=0.896, $p = 0.414$).

3.3.2 Variation in k by gene

I found siginficant differences between estimates of substitution rate k by gene in the case of segmented influenza (ANOVA, F(3,28)=2.897, $p = 0.03257$), but no significant

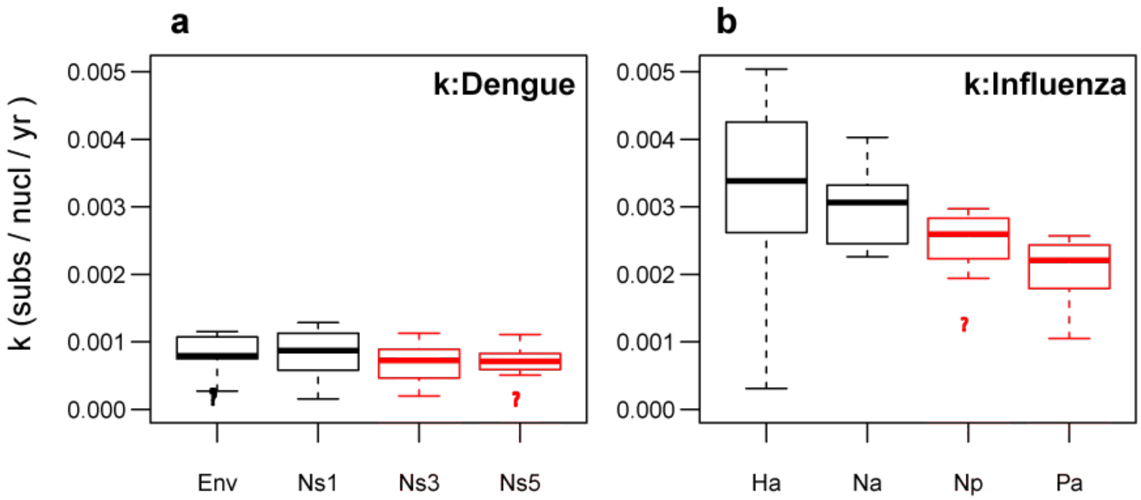

Figure 3.1: Variation in substitution rate by gene in Dengue and H3N2 Influenza. a) No significant pairwise difference between any two pairs of gene in Dengue; b) Significant pairwise differences for across-group pairs of genes in Influenza.

effect of gene in the case of unsegmented dengue (ANOVA, F(3,56)=0.8664, $p = 0.464$). A subsequent post-hoc comparison of influenza gene substitution rates using the Tukey HSD test showed that only substitution rate k of the influenza genes PA and HA differed significantly at the $p < 0.05$ level.

Substitution at CP1: When estimates of substitution rate were partitioned by codon position, I found significant effects of gene on substitution rates of the 1^{st} codon position - k_{CP1} in Dengue (ANOVA, F(3,56)=3.1334, p=0.03946), and influenza (ANOVA, F(3,28)=8.1476, $p << 0.05$).

A post-hoc Tukey HSD test to compare influenza gene-specific substituion rates suggests significant differences between the value of k_{CP1} for genes NP and HA, NP and NA, PA and HA, and PA and NA at the $p < 0.05$ level. No significant differences in k_{CP1} were found between the NA and HA genes and the PA and NP genes. In contrast, the same test suggests no significant pairwise differences between Dengue gene-specific estimates k_{CP1} at the $p < 0.05$ level.

Substitution at CP2: No significant effect of gene was found on k_{CP2} for dengue (ANOVA, F(3,56)=2.2494, $p = 0.092$). A strong effect of gene on k_{CP2} was found for influenza (ANOVA, F(3,28)=17.736, $p << 0.05$).

Post-hoc tests revealed, as in the case of k_{CP1}, highly significant differences in estimates of k_{CP2} for all pairwise combinations of influenza genes except for between the NA and HA, and PA and NP genes at the $p < 0.05$ level. As before, no siginifcant pairwise difference in k_{CP2} were found between any pair of genes in dengue.

Substitution at CP3: Analyses of estimates of substitution at the 3^{rd} codon position, k_{CP3} suggests no significant differences based on gene either in influenza (ANOVA, F(3,28)=0.6363, $p = 0.5979$) or dengue (ANOVA, F(3,56)=0.6636, $p = 0.5779$). A subsequent Tukey HSD post-hoc analysis also suggested no significant differences in estimates of k_{CP3} for any pair of genes either in dengue or influenza at the $p < 0.05$ level.

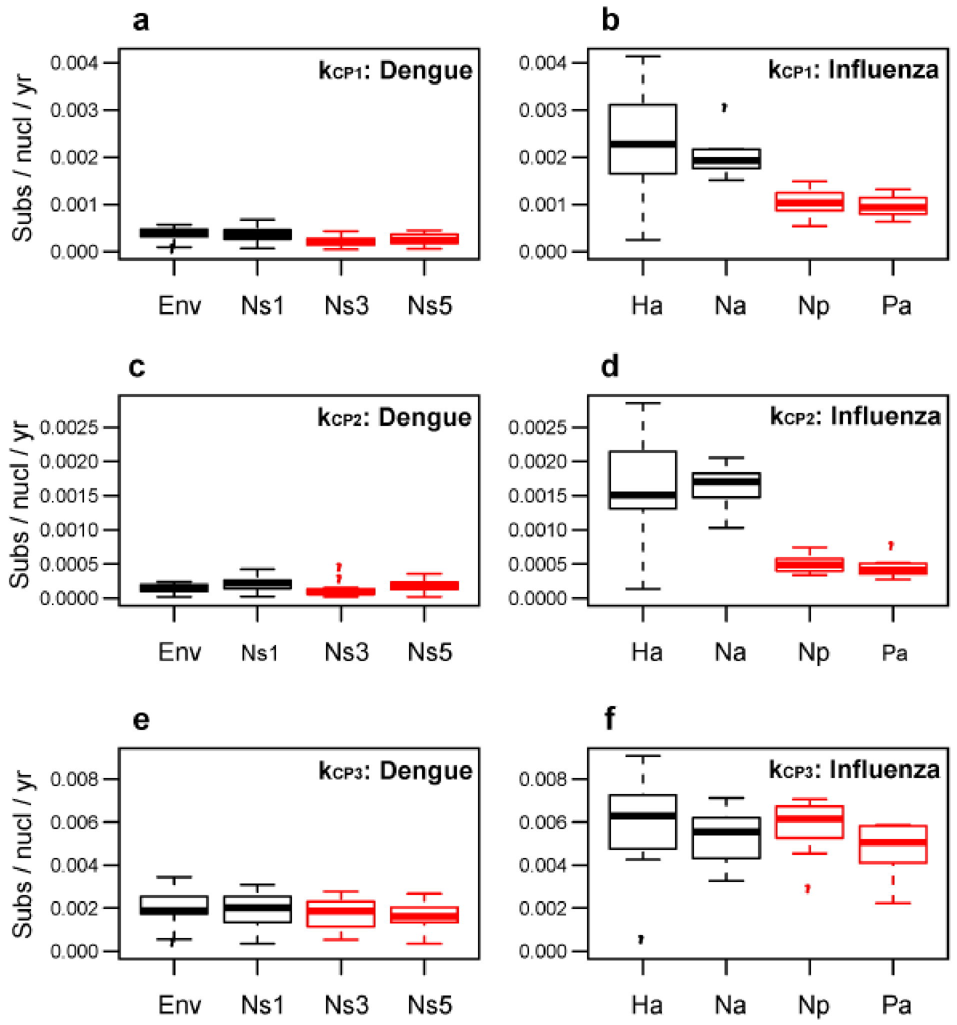

Figure 3.2: Substitution rate on the 1st codon position k_{CP1} for genes from a) Dengue and b) Influenza; Substitution rate on the 2nd codon position k_{CP2} for c) Dengue and d) Influenza; and Substitution rate on the 3rd codon position k_{CP3} for e) Dengue and f) Influenza. Boxplots in red indicate genes that were described in the literature as conserved and under evolutionary constraint.

3.4 Discussion

3.4.1 The effect of deleterious mutations on adaptive substitution

Substitution rate k varies weakly by gene both in the case of influenza and dengue. However, my results show that estimates of substitution rate k are only significantly different across a single pair of genes in the case of influenza, and are not significantly different across any pair of genes in dengue. This stands in contrast to the highly significant pairwise differences in estimates of substitution rate on the 1^{st} and 2^{nd} codon position between several genes, and the complete lack of any significant differences in k_{CP3} across genes in influenza.

Assuming that k_{CP1} and k_{CP2} are proxies for adaptive substitution rate, and k_{CP3} stands in for neutral substitution rate, my results suggest that in the case of segmented viruses, the rate of adaptive substitution varies by gene (or genome segment), but the rate of neutral substitution is unaffected by gene. In the case of unsegmented viruses, neither adaptive nor neutral substitution rate varies by gene.

The relatively large amount of variation in substitution rate k between genes of influenza, when compared to dengue, is easily visualized (Figure 3.1). In addition, the *a priori* classification, based on the literature, of genes into two categories - the highly conserved functional genes (depicted in Figure 3.1 in red) which are likely under purifying selection, and the relatively less conserved genes (either structural or with functions that require genetic variation) is partially validated. In the case of the segmented influenza virus, all significant pairwise differences in estimates of k, k_{CP2} and k_{CP2} were found across these two categories, with no significant pairwise differences found among pairs within each category.

The lack of effect of gene on overall substitution rate, as well as on substitution rates on the 1^{st} and 2^{nd} codon positions in the unsegmented dengue virus suggests that the rate of evolution of genes that we do not typically think of as being under heavy selective constraint, like the Env gene, could in fact be limited by the presense of highly conserved genes like the NS5 polymerase elsewhere on the genome. This is particularly apparent when compared to the high rate of substitution of the influenza structural genes HA and NA, which evolve almost an order of magnitude faster than functional genes NP and PA.

Overall, my results suggest that the presence of highly conserved regions in a genome can slow down the fixation of adaptive mutations on other parts of the genome. However, this effect of linked deleterious mutations likely depends on the strength of linkage between the region of concern and other highly conserved regions of the genome, and is much weaker when the deleterious mutation occurs on a separate gene segment and can be lost by reassortment.

Assuming that most expressed amino acid substitutions are adaptive, we can surmise that the adaptive substitution rate varies by gene in the case of segmented viruses, and either does not vary or varies to a much smaller extent by gene in unsegmented viruses. It is admittedly harder to distinguish, using only phylogenetic data, between high adaptive subsitution rates due to high avalability of adaptive mutations in the fitness landscape around the genome and high adaptive substitution rates due to a relative lack of linked deleterious mutations [Burch and Chao, 2000].

3.4.2 Adaptive optima and mutation rates in nature

The results of this study have implications for the concept and estimation of adaptive optima and our understanding of viral mutation rates in nature. As expressed in the introduction, the expected rate of substitution for a particular gene is an increasing function of the beneficial mutation rate in the context of that gene, and a decreasing function of the deleterious mutation rate over the linked part of the genome. The mutation rate that maximizes the rate of adaptive substitution, or the adaptive optimum U_{opt} is a function of the fraction of all mutations that are deleterious, X_D.

However, as the result of this study suggest, the value of X_D used to calculate adaptive optima depends on the organization of the genome in question. The mutation rate that optimizes the rate of adaptive evolution for a particular gene is likely limited by deleterious mutation rate of genes on the same segment in the case of segmented genomes, and by the deleterious mutation rate of the entire genome for non-recombining unsegmented genomes.

Higher mutation rates mean more variation on which natural selection can act. For segmented viruses, in an ecological context where a high level of variability can be evolutionarily advantageous for a gene on a genome segment, viral populations could exist at a theoretical mutation rate above the optimal mutation rate calculated with respect to a highly conserved functional gene occuring on another gene segment. This would not be possible in non-recombing unsegmented viruses. Given recent data that suggest that viral populations in nature could have mutation rates at or around this theoretical optimum [Sanjuan, 2012], this may explain the relatively larger between-gene variation in substitution rates of the unsegmented influenza virus.

3.4.3 The effect of deleterious mutations on neutral substitution

As mentioned in the results section, there is no comparable effect of linked deleterious mutation rates on the rate of neutral substitution. This is supported by long-standing theoretical conclusions that the rate of neutral evolution is unaffected by linked deleterious mutations [Birky and Walsh, 1988]. This also serves as an effective rebuttal to recent claims that neutral mutation rates are affected by deleterious mutations, leading to a misperception that there is such a thing as a neutral optimum, or a neutral mutation error threshold in nature. Such claims are unsound, and I will elaborate on this in the next chapter of this dissertation.

3.5 Conclusions

My results suggest that adaptive substitution is largely limited to new adaptive mutations that occur on deleterious mutation-free genomes. However, in the case of segmented genomes, adaptive substition on a gene on a genome segment is relatively unaffected by deleterious mutations occuring on other genome segments. The effect of linked deleterious mutations on adaptive substitution is therefore largely dependant on the strength of the linkage.

This can allow a virus with a segmented genome to have some rapidly evolving genes, while retaining highly conserved functional genes in other distant parts of the genome, something that would not be possible in the case of unsegmented genomes with low or no recombination. This is particularly relevant when it comes to segmented viruses like influenza which routinely evolve to escape the immune respose and control methods of their hosts.

These findings should inform our models of adaptive evolution for viruses. In particular the manner in which we paramerterize our models should depend on the genome architecture of the virus in question, since segmented and unsegmented virsues display different patterns of between-gene variation in substitution rate, which affect where we think the viral populations exist in parameter space. Our data also suggest that, in the case of segmented viruses, estimates of subsitution rate made based on single genes may not be represtative of the rate of evolution of the entire genome. Indeed, the very idea of a single value for the rate of evolution of a segmented RNA virus genome is inconsistent with our findings.

3.6 Acknowledgments

This research was supported by NIH grant RO1 AI047498 to LAR and the RAPIDD Program of the Science and Technology Directorate, the Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Chapter 4

The Fallacy of Neutral Optima

Anand Bhardwaj & David J. Cutler

In this comment, I criticize a recent peer-reviewed publication in which the author uses a flawed mathematical formulation for neutral evolution in order to make a case for so-called "neutral optima" and the idea of an error threshold in nature based on the alleged relationship between mutation rate and the rate of neutral evolution.

4.1 Neutral Optima: A scientifically inconsistent theory

Recent theoretical work on variation in the rate of neutral evolution suggests that the presence of deleterious mutations can also slow down the rate of fixation of neutral muations [Sanjuan, 2012]. This theoretical result is extrapolated from Orr's conclusions on adaptive evolution described in the previous section and extended to neutral evolution. This study also presents data that suggests that some viral populations in nature exist at or near the mutation rate that maximizes the neutral mutation rate, which happens to be mathematically identical, according to this theory, to Orr's formulation of the adaptive optimum U_{opt}.

$$K_N \propto U_c \cdot X_N \cdot e^{-U_c \cdot X_D / s_H} \tag{4.1}$$

.

However, the idea of neutral mutation rates being affected by linked deleterious mutations runs in contrast to a long-standing theoretical result that neutral substitution rates are directly proportional to the rate of generation of neutral variation and are unaffected by the rate and strength of linked deleterous mutations [Birky and Walsh, 1988]. Since Sanjuan's neutral optimum is mathematically identical to Orr's adaptive optimum, it begs the question of whether Sanjuan's data support his own theory, or Orr's.

4.2 Rebuttal and Discussion

If, as Sanjuan's work suggests, the relationship between neutral substitution rate and mutation rate is similar to the relationship betwen adaptive substitution rate and mutation rate, we would expect to see similar results for variation in adaptive substitution and neutral substitution. However, this is not the case, as the data pertaining to k_{CP3} is quite different from k_{CP1} and k_{CP2}, as shown in the previous chapter.

Sanjuan's formulation of a neutral mutational optimum, beyond which the rate of neutral substitution rapidly drops off with any increase in mutation rate, hinges on an inconsistent treatment of individuals that have accumulated deleterious mutations. Assuming mutation-selection equilibrium, $P_0 = e^{-U_c.X_D/s_H}$ is the expected fraction of individuals in the population that have not accumulated any deleterious mutations. We expect these individuals to be rapidly lost. Sanjuan includes this fraction when estimating the number of new neutral mutations arising in the population per unit of time, but fails to correct for this loss of individuals in the term for the probability of fixation of new mutations, in effect, erroneously assuming that the probability of fixation of neutral mutations with linked deleterious mutation is $1/N$ and not nearly zero, as we assume in this famework, like the deleterious mutations to which they are linked.

The rate of neutral substitution is proportional to the rate of generation of new neutral variation, the probability that these new neutral mutations occur on deleterious mutation-free genomes, and the probability of fixation of these new mutations, as follows:

$$K_N = (N.U_C.X_N).(P_0).(1/N.P_0) \tag{4.2}$$

Or

$$K_N = U_c.X_N \tag{4.3}$$

Here, K_N is the neutral substitution rate and X_N is the fraction of all mutations that are neutral.

The above, in combination with the data presented in the previous chapter of this dissertation, suggest that neutral substitution rates increase monotonically with mutation rate, are not affected by the presence of linked deleterious mutations, and do not have optima with respect to mutation rates. This is consistent with the long-standing theoretical interpretation [Birky and Walsh, 1988].

Chapter 5

Intracellular Dynamics & Adaptive Evolution

Anand Bhardwaj, David J. Cutler & Leslie A. Real

In this chapter I examine the effects of variation in the intracellular replication process and viral demographics on the rate of adaptive evolution ssRNA *viruses. I highlight demographic conditions under which rapid fixation of small-effect adaptive mutations is possible.*

5.1 Introduction

Viral diseases are ubiquitous and are the cause of significant health and economic burdens. Among viruses, those with RNA genomes are of particular interest because the lack of a proofreading mechanism in RNA polymerase leads to error rates several orders of magnitude higher than in DNA viruses and other DNA based organisms [Holland et al., 1982]. This, in combination with their rapid replication rate allows for a relatively fast rate of evolution [Holmes, 2009], making it more likely that the control of RNA viral diseases will be balanced by the emergence of new viral diseases in comparable time scales [Lederberg, 1998]. It is in this context that I must investigate some of the factors that affect this high rate of evolution in RNA viruses.

The rate of evolution, measured as substitutions per unit of time, is proportional to the rate of generation of new variation, or mutation rate, and some measure of the the probability that these new mutations spread and fix in the population.

We know that the rate of fixation of neutral mutation is inversely proportional to population size, while the number of new neutral mutations entering the population per unit of time is directly proportional to population size [Kimura, 1962, Kimura, 1964]. These terms cancel each other out to make neutral evolution a function of the mutation rate, and only weakly dependant on population size.

$$K_N \propto U_c \cdot X_N$$

Here, K_N is the number neutral substitutions per generation, U_c is the number of mutations per generation and X_N is the fraction of these mutations that are neutral. The question of adaptive substitution and evolution is far more complex, and deserves further exploration.

Our historic understanding of adaptive evolution in asexuals with low rates of recombination suggests that genomes free of deleterious mutations are required in order for adaptive evolution to proceed [Fisher, 1930]. This rests on the assumption that deleterious mutations are, on average, more numerous and of typically larger effect than adaptive mutations.

For genomes of unsegmented ssRNA viruses and genome segments of segmented ssRNA viruses, these are reasonable assumptions. We know from recent studies on the distribution of fitness effects of mutations in RNA viruses that some significant fraction of mutations are selectively neutral, a larger fraction are either deleterious or lethal, while a very small fraction are adaptive or beneficial and are typically of small effect [Sanjuan et al., 2010, Sanjuan, 2010].

[Orr, 2000] proposes a mathematical expression for the expected adaptive substitution rate that reveals the main factors that drive the rate of adaptive evolution: In simple terms, the rate of adaptive evolution is proportional to the number of new adaptive mutations arising in the population per generation, the probability that this new adaptive mutation does not co-occur with a deleterious mutation, and the probability of fixation of this new mutation.

$$K_A \propto N \cdot U_c \cdot X_A \cdot e^{-U_c \cdot X_D / s_H} \cdot \pi \tag{5.1}$$

Here, K_A is the rate of adaptive substitution per generation, N is a measure of population size, X_A is the fraction of mutations that are adaptive or beneficial, X_D is the fraction of mutations that are deleterious, s_H is the harmonic mean of selection coefficients of all mutations, and $\pi = 2 \cdot s_A$ is some measure of the probability of fixation of new adaptive mutations where s_A is the mean selective advantage of adaptive mutations.

Both neutral and adaptive evolution depend heavily on the fitness landscape of the virus in question, on the availablity of adaptive and neutral mutations, and (in the case of adaptive evolution) on the probability of these mutations not being linked to deleterious mutations.

We can see now from the general structure of the mathematical framework used to describe the rate of adaptive evolution that there are two major classes of factors that affect the rate of adaptive evolution:

Characteristics of Deleterious Mutations:

Since the primary requirement for adaptive evolution is the presence of genomes free of deleterious mutations on which adaptive evolution can occur, the rate and nature of linkage of deleterious mutations is important to examine. As we explored in the two previous chapters, while the rate of adaptive evolution is measured empirically by gene, the availability of genomes without deleterious mutations depends on the rate of deleterious mutation over the entire genome in the case of unsegmented viruses, or over the entire genome segment in the case of segmented mutation. Therefore, the presence of highly conserved regions on a genome can potentially limit the rate of adaptive evolution elsewhere on the genome, depending on the genome architecture of the virus under examination. In the context of mathematical frameworks like the one described above, terms like X_A and X_D take on a special significance. In a sense, the rate of deleterious mutations per generation can give us an idea of the effective *adaptable* population size of a virus, by allowing us to estimate the expected fraction of genomes in the population that are free of deleterious mutations.

Characteristics of Adaptive Mutations:

While the empirically derived distributions of fitness effects of point mutations mentioned earlier do suggest that some fraction of mutations are selectively advantageous, this empirical distribution was derived from stable laboratory populations of viruses, and does not account for the effects of environmental disruption, host immunity, and the co-evolutionary history between virus and host. In addition, the mathematical framework described above makes certain simplifying assumptions about the probability of fixation of new mutations, and the initial frequency of new mutations that, while effective as a general way of thinking about adaptive evolution, might be inaccurate in the context of true viral evolution.

The aim of this study is to build a mathematical framework for exploring the effects of variation in the mode of viral intracellular replication and demographic parameters like viral fecundity on the probabiliy of fixation and therefore the rate of adaptive substitution.

5.2 Limitations of Orr's framework

:

The rate of adaptive evolution is proportional to two component characteristics of adaptive mutations - the number of new adaptive mutations arising in the population on genomes free of deleterious mutations, and the probability of fixation of these new mutations.

The mathematical framework descibed in the previous section cannot accurately account for either of these two components in the case of viral populations for the following reasons:

The Tiered Nature of Viral Populations:

Viral populations can be conceptually subdivided into intracellular and extracellular populations. The general scientific consensus is that viral genomes are not under selection during the intracellular replication process [Drake and Holland, 1999]. They are only subject to selection when packaged in proteins and released into the extracellular environment. A typical viral cell infection cycle involves a single or very small number of viral particles entering a cell and replicating a large number of daugher progeny that can either be released all at once by lysing the cell or gradually, by budding off the cell membrane. The tiered nature of viral populations is critical to adaptive evolution because of the potential effect of simultaneously releasing multiple copies of new adaptive mutants into the environment. This violates some common simplifying assumptions about the initial frequency and the probability of fixation of new mutations made in the general mathematical framework for adaptive evolution mentioned in the previous section.

Variation in Viral Intracellular Replication: There is considerable variation in the process of SSRNA viral intracellular replication. In general, after the initial infection of a cell, all subsequent daughter progeny are direct decendants of one or more initial opposite sense templates. Some of these daughter progeny can be the basis for further opposite sense templates and secondary and tertiary daughter progeny [Drake, 1993]. The mode of viral replication can therefore vary all the way from completely linear - where all daughter progeny are produced in a single round of replication, to binary - where the size of the intracellular viral population doubles in each round of replication until some target fecundity is reached [Duffy et al., 2002]. This has implications for the the initial frequency and probability of fixation of new mutations as well. In some ways, the initial frequency ρ_0 of a new mutation is a function of how early along the viral replication process the mutation occurs, i.e., a mutation occuring during the production of an initial opposite sense template gets copied on to all progeny of that template, and potentially, the entire product of a cell infection cycle [Loverdo et al., 2012]. This assumes that any replication event within the cell is equally likely to incur a mutation. This is a reasonable assumption for single stranded RNA viruses where the template and progeny replications are performed by RNA-dependant RNA polymerase, and less reasonable in the context of DNA viruses or retroviruses, where there are DNA replication steps that involve other, less error-prone polymerase enzymes [Holmes, 2009].

Viral Fecundity:

If the per-cell fecundity of the virus is large, this could have a significant effect on initial frequency. If we're interested in an isolated population of an organism (where N can reasonably be assumed to be much smaller than the true "global" N), in the context of evolution of a viral population within a single host or an emerging pathogen where global N itself is small, changes in initial frequency of new mutations could have a significant effect on the probability of fixation and on the rate of adaptive evolution.

What remains to be sees is whether any of the above will matter. Developing this mathematical framework will allow us to explore the areas of parameter space where mode of replication and demographics play a large role in accelerating or decelarating the rate of adaptive evolution. Ultimately, the mutational fitness landscape, i.e. - $U_c \cdot X_A$ and $U_c \cdot X_D$, could play the dominant role in the biologically realistic parts of parameter space, in which case none of the above would matter. The aim of this study is to develop the tools to find out.

5.3 The Number of New Adaptive Mutations per Generation

As mentioned in the previous section, the tiered nature of viral populations means that viruses are not subject to selection when they are within a cell. The allows new genomes to be introduced into the populations in batches, with multiple copies of new mutations. Assuming equilibrium dynamics and constant population size, the population term in Equation 5.1 can then be unpacked into the mean size of an intracellular viral population, or viral fecundity N_c, and some measure of the number of infected cells N to give us an estimate of the total viral populations size $N \cdot N_c$.

The number of viral particles released per cell infection cycle or generation without any deleterious mutations is a product of viral fecundity and the expected fraction of deleterious mutation-free genomes (i.e., the probability of zero mutations), assuming that mutations/genome/generation are Poisson distributed with mean $U_c \cdot X_D$ is $N_c \cdot e^{-U_c \cdot X_D}$.

$N \cdot N_c \cdot e^{-U_c \cdot X_D}$ then gives us the expected number of genomes free of deleterious mutations, or the effective *adaptable* population size. $U_c \cdot X_A$ is the number of new adaptive mutations per genome per generation. Accounting for multiple adaptive mutations on the same genome, and retaining the assumption of Poisson distribution, $(1 - e^{-U_c \cdot X_A})$ gives us the expected fraction of each "generation" or cell infection cycle that have some non-zero number of adaptive mutations. If the value of $U_c \cdot X_A$ is very low, it is a good approximation of $(1 - e^{-U_c \cdot X_A})$. Therefore, the expected number of new adaptive mutant genomes introduced into the population on genomes free of deleterious mutations per generation is:

$$E[A] = N \cdot N_c \cdot (1 - e^{-U_c \cdot X_A}) \cdot e^{-U_c \cdot X_D} \tag{5.2}$$

5.4 The Probability of Fixation

One of the major assumptions of the mathematical expression for adaptive evolution described in Equation 5.2 is that the probability of fixation π for new adaptive mutations is approximately $2 \cdot s_A$. This approximation is derived from Kimura's famous result for the probability of fixation of new small-effect adaptive mutations in large, constant-size populations [Kimura, 1962]:

$$\pi = \frac{1 - e^{-2 \cdot N \cdot \rho_0 \cdot s_A}}{1 - e^{-2 \cdot N \cdot s_A}} \tag{5.3}$$

Under the implicit assumption that population size N is large, mean selective advantage s_A is small and initial frequency of new mutations $\rho_0 = 1/N$, the approximation $\pi \approx 2 \cdot s_A$ is a reasonable one. However, as suggested in the previous section, the tiered nature of viral populations means that new adaptive mutations are not always singly introduced into the generation population.

To look at the effect of treating initial frequency as a random variable, equation 5.3 must be modified to reflect the tiered nature of viral populations. We now have:

$$\pi = \frac{1 - e^{-2 \cdot N \cdot N_c \cdot \rho_0 \cdot s_A}}{1 - e^{-2 \cdot N \cdot N_c \cdot s_A}} \tag{5.4}$$

The above expression accounts for the lack of selection during the viral intracellular replication process. Within this framework, the initial frequency of new mutations ρ_0 can range from $1/N \cdot N_c$, when a single copy of a new mutation is introduced into the population, to $1/N$, where all viral progeny of a single cell infection cycle have a copy of the new adaptive mutation.

However, as mentioned earlier in this chapter, and repeatedly throughout this dissertation, adaptive evolution can only occur on genomes free of deleterious mutations. To account for this, the expected fraction of genomes in the population with some non-zero number of deleterious mutations can be incorporated into the expression for the probability of fixation of new adaptive mutations described in equation 5.4; yielding:

$$\pi = \frac{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \cdot \rho_0 \cdot s_A}}{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \cdot s_A}} \tag{5.5}$$

Modeling ρ_0 as a random variable:

For $\rho_0 = 1/N \cdot N_c$, often used in the literature to describe the initial frequency of new mutations, the mutation has to occur at the very last replication event. Mutations occuring at any earlier replication event would yield a higher initial frequency. This concept is reflected in prior studies suggesting that non-linear replication can enhance the rate of spread and fixation of mutations [Sardanyes et al., 2009, French and Stenger, 2003, Thebaud et al., 2010] and can even lead to a dampening of standing variation.

For any viral population with some fecundity N_c, the total number of replication events R that occur within the cell is related to the mode of replication, and the number of rounds of replication within the cell t_c as follows:

For $t_c = 1$: Under the simplest conditions, with purely linear replication, and assuming a single initial opposite sense template, 1 replication event produces the initial opposite sense template and N_c replication events to produce the progeny.

Here, $R = 1 + N_c$ replication events occur within the cell.

For all $t_c > 1$: Under conditions of nonlinear replication, the number of replication events can be generalized as follows:

Here, $R = 1 + 2[N_c^{1/t_c} + N^{2/t_c}... + N_c^{(t_c-1)/t_c}] + N_c$ replication events occur within the cell.

If the initial frequency of new mutations ρ_0 can be rewritten as $x/(N \cdot N_c)$ where x refers to the expected number of copies of a new mutation released per generation or cell infection cycle, we can think of x as some function of how early in the replication process the mutation occurs. For example, a mutation occuring on the very last round of replication, or the replication event that produces a progeny genome that is released into the extracellular environment, $x = 1$ and the associated initial frequency of that new mutations $\rho_0 = 1/N \cdot N_c$. Let a be a measure of when in the replication process the mutation occurs. The value of a therefore ranges from 0, representing a mutation occuring during the production of the initial opposite-sense template, to t_c, representing a mutation occuring duing the very last replication event that produces a daughter progeny that is released into the extracellular environment.

We know that $x \in [1 : N_c]$. This can be re-written as $x = (N_c)^{t-a/t}$ with $a \in [0 : t]$ such that $x = 1$ when $a = t$ and $x = N_c$ when $a = 0$. Smaller values of a are associated with larger values of x.

In general:

$$\text{When } a = 0 \text{ and } Pr[x = N_c^{(t_c-a)/t_c}] = Pr[x = N_c] = \tfrac{1}{R}$$

$$\text{When } a = t_c \text{ and } Pr[x = N_c^{(t_c-a)/t_c}] = Pr[x = 1] = \tfrac{N_c}{R}$$

$$\text{when } 0 > a > t_c;\ Pr[x = N_c^{(t_c-a)/t_c}] = \tfrac{2.N_c^{a/t_c}}{R}$$

Under conditions of linear replication linear ($t_c = 1$):

$$Pr[x = N_c] = Pr[x = N_c^{(t_c-0)/t_c}] = \tfrac{1}{R} = \tfrac{1}{1+N_c}$$

$$Pr[x = 1] = Pr[x = N_c^{(t_-t)/t}] = \tfrac{N_c}{R} = \tfrac{N_c}{1+N_c}$$

Under conditions of nonlinear replication ($t_c > 1$):

$$Pr[x = N_c] = Pr[x = N_c^{(t-0)/t}] = Pr[x = N_c] = \tfrac{1}{R} = \tfrac{1}{1+2[N_c^{1/t}+N^{2/t}...N_c^{(t-1)/t}]+N_c}$$

$$Pr[x = N_c^{(t-a)/t}] = \tfrac{2.N_c^{a/t}}{R} = \tfrac{2.N_c^{a/t}}{1+2[N_c^{1/t}+N^{2/t}...N_c^{(t-1)/t}]+N_c}$$

$$Pr[x = 1] = Pr[x = N_c^{(t-t)/t}] = Pr[x = 1] = \tfrac{N_c}{R} = \tfrac{N_c}{1+2[N_c^{1/t}+N^{2/t}...N_c^{(t-1)/t}]+N_c}$$

The above calculations do not account for the effect of linked deleterious mutations.

5.4.1 The Linear case

Here I explore how π (where the initial frequency is a random variable) compares with π' from the literature (where the initial frequency is constant), under conditions of linear replication.

Under the assumption of purely linear replication with one initial opposite sense template, where a total of $R = 1 + N_c$ replication events occurs, a mutation can occur on the template with probability $\frac{1}{1+N_c}$ leading to a high initial frequency $\frac{N_c \cdot e^{U_c \cdot X_D}}{N \cdot N_c \cdot e^{-U_c \cdot X_D}} = \frac{1}{N}$ or that mutation can occur on a progeny with probability $\frac{N_c}{1+N_c}$ leading to the initial frequency of $\frac{1}{N \cdot N_c \cdot e^{-U_c \cdot X_D}}$.

$$\pi = \frac{1}{1+N_c} \cdot \frac{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \frac{1}{N} \cdot s_a}}{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \cdot s_a}} + \frac{N_c}{1+N_c} \cdot \frac{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \frac{1}{N \cdot N_c \cdot e^{U_c \cdot X_D}} \cdot s_a}}{1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} s_a}} \tag{5.6}$$

This can be rewritten as follows:

$$\pi = \frac{1 - e^{2 \cdot N_c \cdot e^{-U_c \cdot X_D} \cdot s_A} + N_c \cdot (1 - e^{-2 \cdot s_A})}{(1 + N_c)(1 - e^{-2 \cdot N \cdot N_c \cdot e^{-U_c \cdot X_D} \cdot s_A})}$$

The structure of the above expressions suggest that the probability of fixation π where initial frequency ρ_0 is a random variable is not quite equivalent to the probability of fixation π‘ where the initial frequency is constant. This is explored in Figure 5.1 where we can see that we underestimate the probabiity of fixation if we treat ρ_0 as a constanct. The extent to which we underestimate π depends on the ratio of fecundity

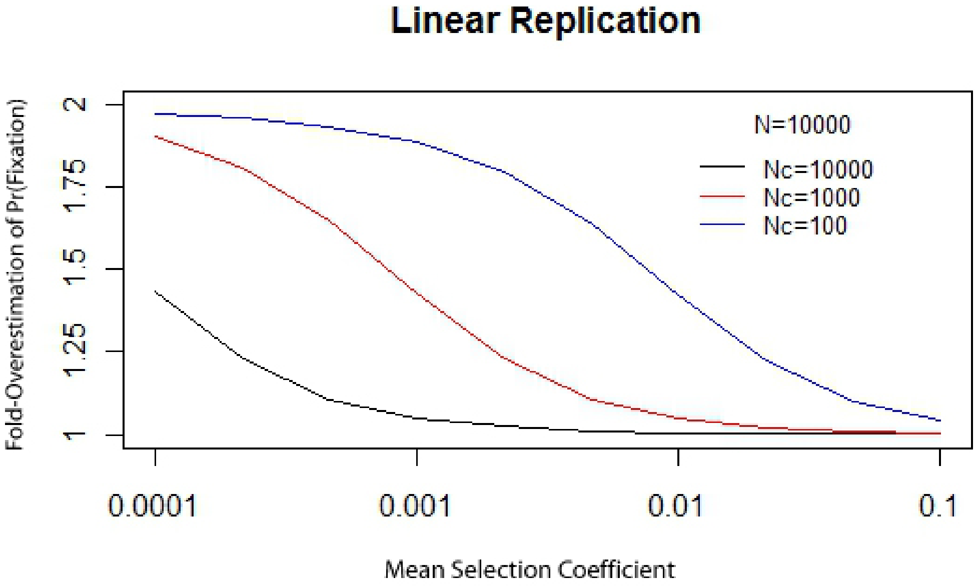

Figure 5.1: The effect of mean selective advantage s_A on the extent to which we underestimate the probability of fixation when we do not account for variation in the initial frequency of new adaptive mutations.

to global population size N_c/N as well as the magnitude of s_A. We can see that for adaptive mutations of large effect, the number of copies of new mutations released into the population has a relatively small effect on the probability of fixation.

Figure 5.1 might be a bit misleading, however, since higher values of s_A always lead to higher π (Figure 5.2). In general, $2 \cdot s_A$ is always a good approximation of π', while it is a good approximation of π only for some values of s_A, depending on N_c.

Deviations due to multiple templates: Assume that m initial opposite sense templates are used to produce the entire progeny of a cell:

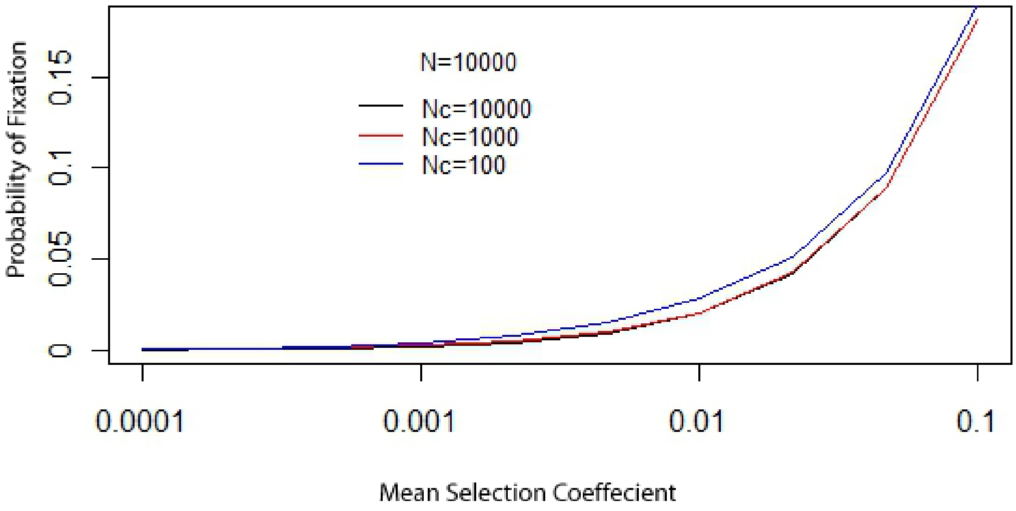

Figure 5.2: Change in the probability of fixation with selective advantage s_A.

Prior to this, I assume that viral replication utilizes only one initial opposite sense templated. Here, $t_c = 1$, a can either be 0, where the mutation occurs on the template step or 1, where the mutation occurs on the progeny step.

$$\pi = Pr[a = 0] \cdot (1 - e^{-2 \cdot N_c^{(1-0)/1} \cdot s_A}) + Pr[a = 1] \cdot (1 - e^{-2 \cdot N_c^{(1-1)/1} \cdot s_A})$$

$$\pi = \frac{1}{1+N_c} \cdot (1 - e^{-2 \cdot N_c \cdot s_A}) + \frac{N_c}{1+N_c} \cdot (1 - e^{-2 \cdot s_A})$$

Given reasonably high values of N_c, the above can be approximated as follows:

$$\pi = \frac{1+N_c \cdot 2 \cdot s_A}{1+N_c} \approx 2 \cdot s_A$$

The probability of fixation is generally approximated as $2 \cdot s_A$ (it is typically slightly

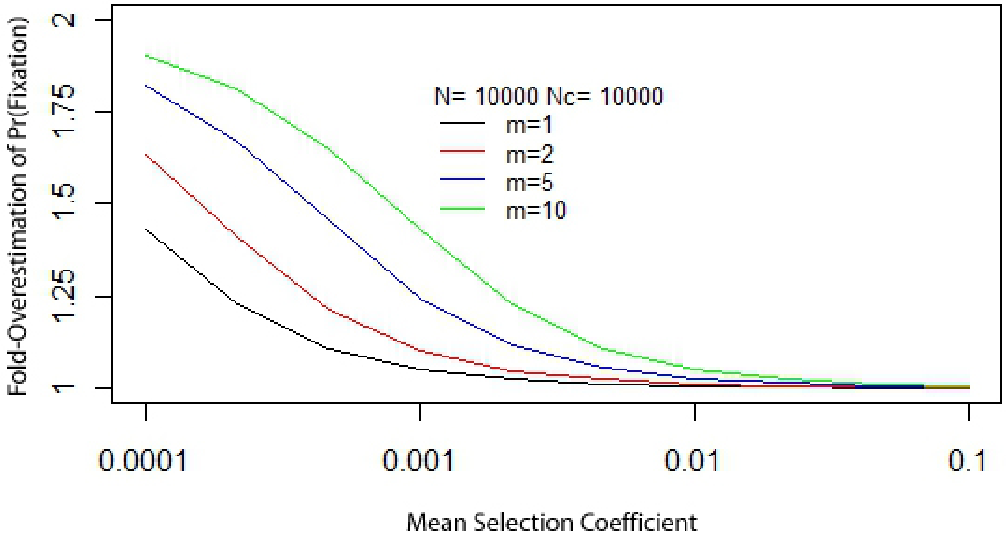

Figure 5.3: The effect of number of templates m on the relative increase in the probability of fixation with random ρ_0 when compared to the probability of fixation wih constant ρ_0. Relatively small effect of template for large values of selective advantage s_A.

lower than $2 \cdot s_A$) but we see that this is a good approximation only under very specific circumstances; i.e.,

- When the mode of replication is purely linear, and

- When only a single initial opposite sense template is used to produce the entire progeny of a single viral cell infection cycle.

Assuming multiple templates: N_c is replaced with N_c/m with m being the total number of template strands used to produce the viral progeny.

$$\pi = \tfrac{m}{m+N_c} \cdot \frac{1 - e^{-2 \cdot N \cdot N_c \cdot \frac{(N_c/m)}{N \cdot N_c} \cdot s_A}}{1 - e^{-2 \cdot N \cdot N_c \cdot s_A}} + \tfrac{N_c}{m+N_c} \cdot \frac{1 - e^{-2 \cdot N \cdot N_c \cdot \frac{1}{N \cdot N_c} \cdot s_A}}{1 - e^{-2 \cdot N \cdot N_c \cdot s_a}}$$

Again, given reasonably high values of N_c, the above can be approximated as follows:

$$\pi = \frac{m}{m + N_c} \cdot (1 - e^{-2 \cdot (N_c/m) \cdot s_A}) + \frac{N_c}{m + N_c} \cdot (1 - e^{-2 \cdot s_A}) \tag{5.7}$$

In this case, if m is high enough for the assumption that $1 - e^{-2 \cdot (N_c/m) \cdot s_A} \approx 1$ to no longer be true, we see deviations from the expectation that $\pi \approx 2 \cdot s_A$. In particular, $\pi > 2 \cdot s_A$ when $1 < N_c/m < 10$.

5.4.2 The Non-linear Case

Under conditions of non-linear replication, the number of replication events and initial frequency of new mutations is dependant on the mode of replication and the number of rounds of replication t_c. With non-linear replication, the initial opposite sense strand copied from the infecting strand is not the direct template for the viral progeny that are released from the infected cell. Instead, the viral progeny released from the cell are the product of opposite sense templates that are themselves copied from progeny of earlier opposite sense templates, and have as such undergone several rounds of potentially error-accumulating replication within the cell.

The frequency of a new mutation can range from $\frac{1}{N \cdot N_c}$ to $\frac{N_c}{N \cdot N_c}$ depending on how early in the replication process the mutation occurs. If a is a measure of how early along in the process the mutation has occured, $a = 0$ signifies a mutation that occured during the production of the initial opposite template (leading to an initial frequency of $\frac{N_c}{N \cdot N_c}$) and $a = t_c$ signifying a mutation that occured during the production of the final progeny (leading to an initial frequency of $\frac{1}{N.N_c}$).

In general, the initial frequency of the new mutation is a function of a such that:

$$\rho_0 = \frac{N_c^{(a-t_c)/t_c}}{N \cdot N_c}$$

Assuming a branching process model with non-overlapping generations for viral intracellular replication:

Let $R = 1+N_c+2(N_c^{1/t_c}+N_c^{2/t_c}...N_c^{(t_c-1)/t_c})$ be the total number of replication events in the cell. The closer to purely linear replication, the more the values of R and N_c converge.

At each round of replication, a single template produces N_c^{1/t_c} progeny. A mutation that occurs at time a gets passed on to $N_c^{(t_c-a)/t_c}$ progeny.

- $a = 0$ for $\frac{1}{R}$ of all events. $\rho_0 = \frac{N_c^{(t_c-a)/t_c}}{N{\cdot}N_c} = \frac{N_c}{N{\cdot}N_c}$.
- $0 > a > t_c$ for $\frac{2{\cdot}[N_c^{1/t_C}+N_c^{2/t_c}...N_c^{(t_c-1)/t_c}]}{R}$ of all events. $\rho_0 = \frac{N_c^{(t_c-a)/t_c}}{N{\cdot}N_c}$

(There are $2(N_c^{a/t_c})$ events for each value of $0 > a > t_c$)

- $a = t_c$ for $\frac{N_c}{R}$ of all events. $\rho_0 = \frac{N_c^{(t_c-a)/t_c}}{N{\cdot}N_c} = \frac{1}{N{\cdot}N_c}$

Clearly, with increasing t_c, the probability of acquiring a mutation with $\rho_0 > \frac{1}{N{\cdot}N_c}$ increases. As suggested in the linear case above, this will cause deviations from the π' with constant ρ_0 from the literature.

The distribution of a gives us the distribution of initial frequencies of new mutations. Note that linear replication is a special case of this where $t_c = 1$ and a is either 0 or 1.

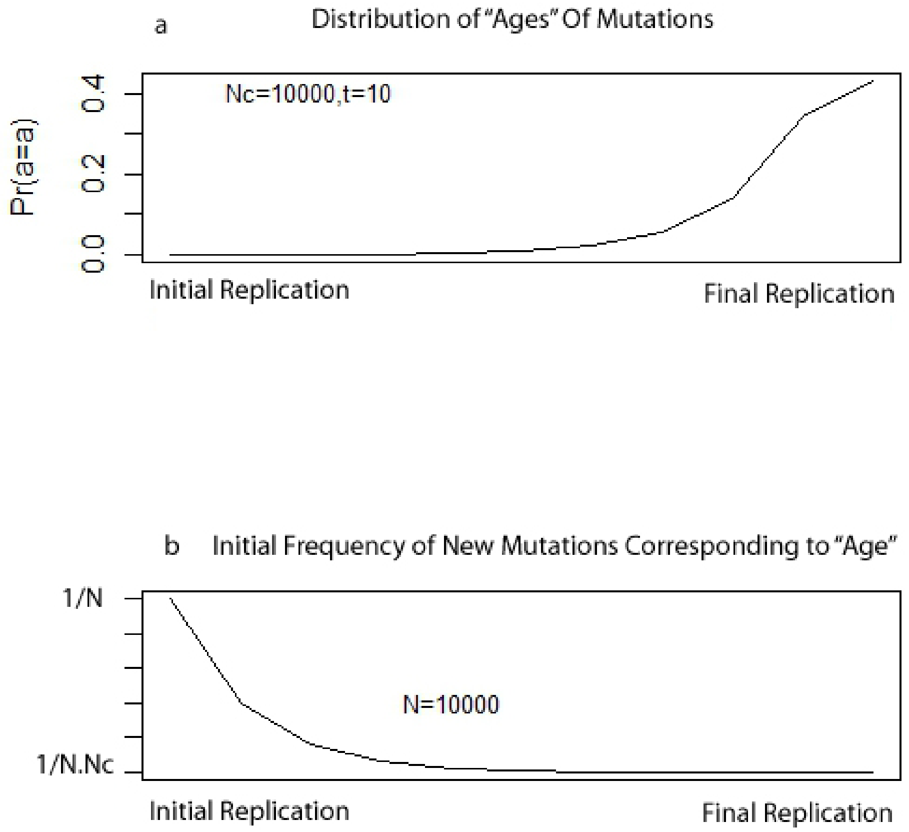

Figure 5.4: Explicitly modeling ρ_0. a) Distribution of possible "ages" of mutations, with lower values of a being associated with larger initial frequencies. b)Distribution of initial frequencies of new mutations.

In figure 5.4b, $N_c = 10,000$ and $t_c = 10$. Clearly, the initial frequency of new mutations varies depending on when along the replication process the mutation occured.

As in the linear case, we expect to see an effect of mode of replication on the initial frequency, and by consequence, the probability of fixation of new beneficial mutations. For any given mutation rate and fitness landscape, the probability of fixation of new adaptive mutations, and therefore the rate of adaptive evolution increases as replication gets more binary. The effect of mode of replication on the rate of adaptive evolution is mitigated at high mutation rates (whether these are biologically relevant mutations rates requires investigation). The same measured mutation rate per generation can lead to different rates of adaptive substitution. These differences can be explained by mode of replication.

All the above depend on the associated rate of linked deleterious mutations. This can change substantially with recombination, and between segmented and unsegmented viruses, as has been explored in previous chapters of this dissertation.

5.5 A New Formulation for Adaptive Evolution in ssRNA viruses

We can now incorporate information about variation in ssRNA intracellular replication dynamics for a new mathematical framework of adaptive evolution. We can see that the summed probability of fixation of new adaptive mutations is a function of the distribution of initial frequencies of new mutations.

$$K_A = N\cdot N_c\cdot(1-e^{-U_c\cdot X_A})\cdot e^{-U_c\cdot X_D}\cdot \sum_{p=1/N\cdot N_c\cdot e^{-U_c\cdot X_D}}^{1/N} Pr(\rho_0 = p)\cdot \frac{1-e^{-2\cdot N\cdot N_c\cdot e^{-U_c\cdot X_D}\cdot\rho_0\cdot s_A}}{1-e^{-2\cdot N\cdot N_c\cdot e^{-U_c\cdot X_D}\cdot s_A}} \tag{5.8}$$

5.6 Results and Discussion

5.6.1 Probability of Fixation

Under our model, the expected probability of fixation of new adaptive mutations increases with increasing rounds of replication. For any mutation rate, the value of

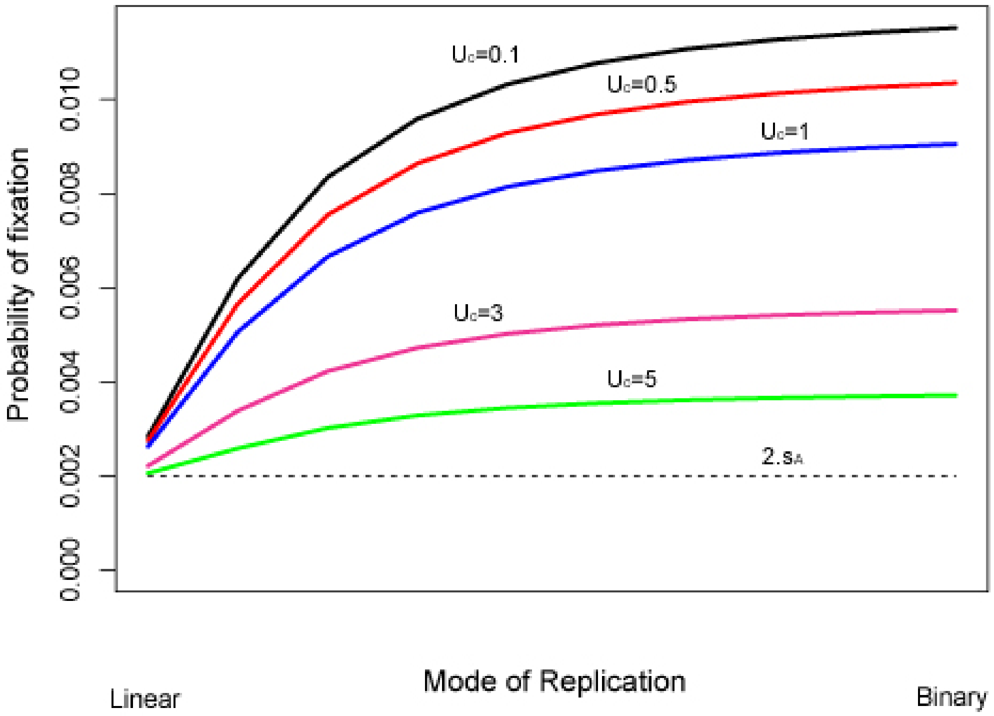

Figure 5.5: Variation in the expected probabilty of fixation due to the mode of replication. Non-linear and binary replication maximizes the probability of fixation given mutation rate and fitness landscape. On the X-axis, the number of rounds of replication goes from 1 to $Log_2(N_c)$.

π increases as the mode of replication becomes more linear, with the probability of fixation maximized when the mode of replication is purely binary.

If we assume that the initial frequency of new mutations is constant, we almost always underestimate the probability of fixation of new mutations. An interesting corollary to this result is that the difference bewteen the probability of fixation of new adaptive mutations under linear replication and under binary replication is inversely proportional to the mutation rate. This is consistent with the established scientific consensus (French and Stenger 2003; Sardanyes et al. 2009; Thebaud et al. 2010).

This result is likely due to the fact that at extremely high mutation rates, the expected fraction of genomes free of any deleterious mutations produced per generation is quite low. Under these conditions, the difference between having a single copy of a new adaptive mutation and having all viable viral progeny contain a copy of the new adaptive mutation is not very large.

The above result suggests that the probability of fixation of new adaptive mutations is some function of the mean selective advantage of new mutations, the mode of replication, and **the linked deleterious mutation rate** of the viral genome. Again, as suggested in previous chapters of this dissertation, the genome architecture of the virus and its effects on the strength of linkage between regions of the genome must be considered in order to accurately assess the effect of linked deleterious mutation on both the probability of fixation of new mutations, as well as the rate of adaptive substitution.

5.6.2 Rate of Adaptive Evolution

Our results suggest that viral intracellular replication dynamics can have a large effect on the rate of adaptive evolution in ways that cannot be captured in simple models of evolution. Complex replication strategies allow viruses to uncouple the rate at which variation is generated from the rate at which variation is observed.

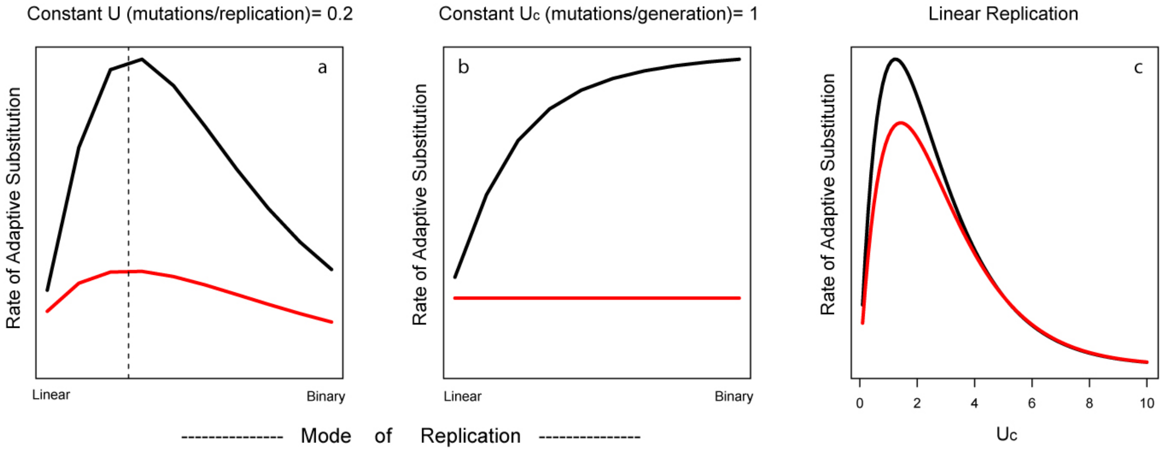

Figure 5.6: Variation in the expected rate of adaptive evolution in ssRNA viruses. Black lines are our projections and red lines are projections based on simpler models of evolution, like Orr's. a) The effect of mode of replication while keeping the biological mutation rate per replication constant, b) The effect of the mode of replication while keeping the mutation rate per generation constant, c) The effect of mutation rate per generation on the rate of adaptive evolution.

In Figure 5.6.2b, we can clearly see the effects of variation in the mode of replication on the rate of adaptive evolution. The same measured mutation rate per generation can be associated with radically different rates of adaptive evolution, the difference explained entirely by variation in the mode of replication. In general, for any mutation rate per generation, non-linear replication can amplify the rate of adaptive evolution.

In Figure 5.6.2c, we can see that the concept of the adaptive optimum [Orr, 2000] is recreated. We almost always underestimate the rate of adaptive evolution given a mutation rate if we treat the initial frequency of new adaptive mutations as a constant. Importantly, the extent to which we underestimate this rate is the greatest near this adaptive optimum.

5.7 Conclusions

My results suggests that simplifying assumptions ignoring the complexity of viral intracellular replication can lead to misleading conclusions about the rate of evolution and evolvability of viral populations. My models suggest that under the right replicative and demographic conditions, viruses could be capable of extremely rapid evolution, suggesting future empirical studies to verify this phenomenon.

We can see that many standard models of evolution apply to viruses only under particular conditions that do not reflect current understanding of viral replication, i.e., under conditions of purely linear replication with a single initial opposite sense template. For viruses with large values of fecundity, these assumptions are unrealistic.

We expect to see an effect of mode of replication on the initial frequency, and by consequence, the probability of fixation of new beneficial mutations. For any given mutation rate and fitness landscape, the probability of fixation of new adaptive mutations, and therefore the rate of adaptive evolution increases as replication becomes more binary. The effect of mode of replication on the rate of adaptive evolution is mitigated at high mutation rates (whether these are biologically relevant mutations rates is up for question). The same measured mutation rate per generation can lead to different rates of adaptive subsitution. All the above depend on the associated rate of linked deleterious mutations. This can change substantially with recombination, and between segmented and unsegmented viruses.

The next step would be to design empirical studies to verify the major claims of this theoretical work, i.e., whether the non-linear replication amplifies the rate of adaptive evolution, and whether certain demographic conditions can lead to rapid evolution, particularly conditions that reflect the early stages of viral infection of a new and naive host. The models presented in this study should also be modified to reflect the replication dynamics of double-stranded viruses and retroviruses, and refined to reflect new and current understanding of viral replication processes.

5.8 Acknowledgements

This research was supported by NIH grant RO1 AI047498 to LAR and the RAPIDD Program of the Science and Technology Directorate, the Department of Homeland Security and the Fogarty International Center, National Institutes of Health.

Chapter 6

Conclusions: Portrait of a Virus

We can now begin to put together a picture of the conditions under which viruses can survive in nature despite their high mutation rates, and under which conditions they can be pushed to extinction.

Summary

In **Chapter 2**, I examined the impact of typical simplyfying assumptions in existing models of lethal mutagenesis. My findings suggest that under a very limited subsets of conditions (linear replication, single initial opposite sense templates, and very high values of mean fecundity), simple models can provide accurate predictions of mutation rates sufficient to drive a viral population to demographic extinction. For viruses with non-linear replication and multiple templates, however, improvements are needed. I draw a distinction between mathematical conditions that suggest extinction conditions, and mathematical models that describe the rate at which populations will go extinct, and why the latter provide more focused reference to in-vitro tests of lethal mutagenesis, and hence, the drug development pipelines. I develop an explicit model of viral intracellular replication based on a Walton-Gaston branching process to derive an expression for the rate of extinction of viral populations under artificially elevated mutation rates. My models suggest that the presence of multiple initial opposite sense templates greatly mitigates the risk of extinction. I also show that there is some non-linear mode of replication that minimizes the risk of extinction, given a particular mutation rate per generation. An important caveat is that my models only account for lethal mutations, and so for a viral population that exists in a real, complex fitness landscape, the critical mutation rates required to cause demographic extinction are likely lower than those predicted by my models.

Chapters 3 and 4 explore the effects of genome architecture and the effective *adaptable* populations of viruses. Under the assumption that adaptive mutations are on average rarer and of small effect than deleterious mutations, adaptive evolution likely only occurs in regions of the genome that are free of linked deleterious mutations. In most single stranded RNA viruses, recombination rates are very low, and therefore this effect manifests itself differently in segmented and unsegmented viruses. Most viral genomes include highly conserved regions, i.e., genes that code for critical functional enzymes like RNA polymerase. Within these conserved regions, most expressed substitutions are likely deleterious. In unsegmented viruses, the presence of these conserved regions limits the rate of adaptive evolution everywhere else on the genome. In contrast, in segmented viruses, the rate of adaptive evolution of any segment is uncoupled from the deleterious effects of mutations on other segments. In effect, viruses with segmented genomes can support highly variable genes while maintaining critical functional genes on other genome segments. Segmented viruse, therefore, are better suited to surviving in environments that impose directional and disruptive selection on parts of the genome, in addition to the inevitable purifying selection on other parts of the genome. While this effect is clearly manifested in gene-specific estimates of adaptive substitution rates from segmented and unsegmented viruses, no comparable effect is seen on neutral subsitution rates. My results call into question recent results suggesting that neutral evolution is also subject to the limiting effects of linked deleterious mutations. My models clarify assumptions and provide a link to a long-standing theoretical result [Birky and Walsh, 1988] suggesting that the rate of neutral evolution is unaffected by linked deleterious mutations.

I develop a novel mathematical framework for describing the expected rate of adaptive evolution in **Chapter 5**. Existing models of adaptive evolution in non-recombining asexuals make certain simplifying assumptions that are unrealistic in the case of single stranded RNA viruses. My models explicitly account for the tiered nature of viral populations - the current scientific consensus is that viral genomes are not under selection while they are still within a host cell, with selection only being imposed when viral particles are released into the extracellular environment. Mutation accumulation during the intracellular replication means that many adaptive mutations are released into the environment in multiple copies, which can affect the rate at which these adaptive mutations spread through the population. My model illustrates the effect of mode of replication and mutation-accumulation bottlenecks during template replication both on the number of new adaptive mutant genomes produced per generation, and the mean probability of fixation of these new adaptive mutations. I show that the same mutation rate per generation can lead to radically different rates of adaptive evolution, with the differences explained by the mode of replication of the virus. Non-linear modes of replication serve to amplify the rate of adaptive evolution. Under certain demographic conditions, when the mean fecundity of an infected cell is on the same order as the number of infected cells, extremely rapid evolution is possible. These conditions are best approximated when looking at evolution at the scale of a single host.

Over the course of this body of work, we can get an idea of a “best practice” of sorts for a single stranded RNA virus. Some non-linear mode of replication can serve both to minimize the risk of extinction due to deleterious and lethal mutations, and amplify the rate of adaptive evolution. The all-or-nothing nature of demographic extinction suggests that viruses that utilize multiple initial opposite sense templates during the intracellular replication can tolerate much higher rates of deleterious mutation, and could serve to explain the failure of existing predictive models of lethal mutagenesis, even with viruses with little or no replication. In addition, viruses that have segmented genomes can support highly variable regions of the genome in a way that viruses with unsegmented viruses cannot.

In all, we get the picture of a multidimensional fitness peak that extends far beyond the traditional "mutational" fitness landscape. Whether viruses in the wild are near this peak remains to be seen, as we lack estimates of replicative parameters like template number, mean fecundity, and replication mode (as well as better spatially and temporally annotated sequence data) needed to do so. I hope that this body of work serves as a call for more data, and serves to show the enormous impact of these parameters on virus evolution and extinction.

After all, knowledge about where these viral populations are in relation to this hypothetical peak will tell us just how much we need to push in order to toss them over the cliff.

Chapter 7

Bibliography, Indices, and Supplements

Substitution Rate Estmates for Chapter 3

#	Pop	Gene	Seq	k	k_{CP1}	k_{CP2}	k_{CP3}
1	Nicaragua	Env	45	0.00079	$4.0e-04$	$1.4e-04$	0.00181
2	Nicaragua	Ns1	45	0.00116	$3.2e-04$	$2.6e-04$	0.00290
3	Nicaragua	Ns3	45	0.00113	$3.9e-04$	$2.8e-04$	0.00272
4	Nicaragua*	Ns5	45	0.00070	$3.4e-04$	$3.6e-04$	0.00140
5	Mexico*	Env	70	0.00087	$4.9e-04$	$2.2e-04$	0.00190
6	Mexico*	Ns1	70	0.00059	$1.3e-04$	$2.2e-04$	0.00143
7	Mexico*	Ns3	70	0.00035	$9.9e-05$	$1.1e-04$	0.00086
8	Mexico	Ns5	70	0.00069	$2.7e-04$	$3.5e-04$	0.00143
9	Cambodia	Env	63	0.00076	$4.0e-04$	$9.5e-05$	0.00178
10	Cambodia	Ns1	63	0.00087	$3.8e-04$	$2.0e-04$	0.00203
11	Cambodia	Ns3	63	0.00059	$1.7e-04$	$5.4e-05$	0.00154
12	Cambodia	Ns5	63	0.00068	$2.1e-04$	$1.9e-04$	0.00162
13	Venezuela	Env	60	0.00016	$5.1e-05$	$3.7e-05$	0.00040
14	Venezuela	Ns1	60	0.00016	$6.2e-05$	$5.2e-05$	0.00035
15	Venezuela	Ns3	60	0.00020	$4.5e-05$	$2.2e-05$	0.00053
16	Venezuela	Ns5	60	0.00015	$5.7e-05$	$3.3e-05$	0.00035
17	Vietnam	Env	80	0.00103	$5.2e-04$	$1.7e-04$	0.00240

18	Vietnam	Ns1	80	0.00108	$4.4e-04$	$3.2e-04$	0.00250
19	Vietnam*	Ns3	80	0.00080	$2.4e-04$	$6.4e-05$	0.00209
20	Vietnam*	Ns5	80	0.00076	$2.3e-04$	$1.5e-04$	0.00189
21	Thailand	Env	60	0.00021	$7.2e-05$	$1.9e-05$	0.00055
22	Thailand	Ns1	60	0.00028	$8.3e-05$	$2.4e-05$	0.00072
23	Thailand*	Ns3	60	0.00025	$1.1e-04$	$1.8e-05$	0.00062
24	Thailand	Ns5	60	0.00015	$6.2e-05$	$1.9e-05$	0.00037
25	Nicaragua	Env	190	0.00079	$3.6e-04$	$2.1e-04$	0.00181
26	Nicaragua	Ns1	190	0.00110	$6.3e-04$	$4.2e-04$	0.00225
27	Nicaragua	Ns3	190	0.00088	$2.8e-04$	$1.3e-04$	0.00224
28	Nicaragua	Ns5	190	0.00078	$4.4e-04$	$2.6e-04$	0.00162
29	US	Env	148	0.00082	$3.8e-04$	$1.9e-04$	0.00188
30	US	Ns1	148	0.00078	$2.9e-04$	$1.5e-04$	0.00190
31	US	Ns3	148	0.00070	$1.7e-04$	$8.9e-05$	0.00185
32	US	Ns5	148	0.00087	$3.6e-04$	$1.6e-04$	0.00208
33	Vietnam*	Env	140	0.00112	$4.5e-04$	$1.2e-04$	0.00278
34	Vietnam	Ns1	140	0.00128	$5.1e-04$	$2.7e-04$	0.00308
35	Vietnam*	Ns3	140	0.00095	$3.4e-04$	$1.5e-04$	0.00235
36	Vietnam	Ns5	140	0.00094	$3.5e-04$	$1.9e-04$	0.00226
37	Cambodia*	Env	44	0.00115	$4.1e-04$	$1.5e-04$	0.00288

38	Cambodia	Ns1	44	0.00124	$4.8e-04$	$1.9e-04$	0.00305
39	Cambodia	Ns3	44	0.00090	$2.1e-04$	$4.5e-04$	0.00241
40	Cambodia	Ns5	44	0.00111	$4.5e-04$	$2.0e-04$	0.00267
41	Brazil	Env	56	0.00074	$3.4e-04$	$1.9e-04$	0.00169
42	Brazil	Ns1	56	0.00057	$3.7e-04$	$1.2e-04$	0.00123
43	Brazil	Ns3	56	0.00057	$1.9e-04$	$8.7e-05$	0.00145
44	Brazil	Ns5	56	0.00051	$1.4e-04$	$1.0e-04$	0.00127
45	Nicaragua	Env	71	0.00115	$5.7e-04$	$2.4e-04$	0.00345
46	Nicaragua	Ns1	71	0.00094	$3.4e-04$	$2.4e-04$	0.00224
47	Nicaragua	Ns3	71	0.00088	$2.8e-04$	$1.3e-04$	0.00224
48	Nicaragua	Ns5	71	0.00071	$2.2e-04$	$1.8e-04$	0.00173
49	Venezuela	Env	89	0.00077	$2.9e-04$	$1.5e-04$	0.00188
50	Venezuela	Ns1	89	0.00084	$3.7e-04$	$2.3e-04$	0.00193
51	Venezuela	Ns3	89	0.00073	$2.5e-04$	$6.8e-05$	0.00186
52	Venezuela	Ns5	89	0.00079	$2.3e-04$	$1.6e-04$	0.00199
53	US*	Env	94	0.00027	$9.4e-05$	$7.4e-05$	0.00065
54	US*	Ns1	94	0.00036	$2.3e-04$	$1.3e-04$	0.00070
55	US*	Ns3	94	0.00023	$7.9e-05$	$3.8e-05$	0.00057
56	US*	Ns5	94	0.00016	$5.8e-05$	$4.3e-05$	0.00038
57	Cambodia	Env	58	0.00115	$5.3e-04$	$2.4e-04$	0.00268

58	Cambodia	Ns1	58	0.00121	$6.8e-04$	$3.4e-04$	0.00262
59	Cambodia	Ns3	58	0.00111	$4.3e-04$	$1.4e-04$	0.00277
60	Cambodia	Ns5	58	0.00102	$4.4e-04$	$2.9e-04$	0.00232
61	Aus-West	Ha	61	0.00460	$4.1e-03$	$2.4e-03$	0.00731
62	Aus-West	Mp	61	0.00384	$2.4e-03$	$1.1e-03$	0.00808
63	Aus-West	Na	61	0.00403	$3.0e-03$	$1.9e-03$	0.00713
64	Aus-West	Np	61	0.00260	$1.1e-03$	$5.1e-04$	0.00614
65	Aus-West	Pa	61	0.00257	$1.3e-03$	$5.2e-04$	0.00587
66	HongKong*	Ha	160	0.00504	$3.2e-03$	$2.9e-03$	0.00908
67	HongKong	Mp	160	0.00178	$1.6e-03$	$6.5e-04$	0.00313
68	HongKong*	Na	160	0.00328	$2.2e-03$	$1.5e-03$	0.00615
69	HongKong*	Np	160	0.00282	$1.5e-03$	$3.9e-04$	0.00657
70	HongKong	Pa	160	0.00234	$9.0e-04$	$3.3e-04$	0.00580
71	Nicaragua*	Ha	102	0.00031	$2.5e-04$	$1.3e-04$	0.00055
72	Nicaragua*	Mp	102	0.00185	$1.3e-03$	$7.3e-04$	0.00349
73	Nicaragua*	Na	102	0.00226	$1.8e-03$	$1.7e-03$	0.00327
74	Nicaragua*	Np	102	0.00126	$5.4e-04$	$3.4e-04$	0.00289
75	Nicaragua*	Pa	102	0.00181	$8.5e-04$	$5.0e-04$	0.00407
76	USA-BOS	Ha	103	0.00291	$1.7e-03$	$1.3e-03$	0.00576
77	USA-BOS	Mp	103	0.00210	$1.6e-03$	$8.6e-04$	0.00385

78	USA-BOS	Na	103	0.00317	$1.5e-03$	$1.7e-03$	0.00626
79	USA-BOS*	Np	103	0.00284	$8.6e-04$	$7.4e-04$	0.00692
80	USA-BOS	Pa	103	0.00253	$9.8e-04$	$7.6e-04$	0.00585
81	USA-CA	Ha	177	0.00386	$2.9e-03$	$1.9e-03$	0.00681
82	USA-CA*	Mp	177	0.00194	$1.7e-03$	$5.0e-04$	0.00361
83	USA-CA	Na	177	0.00296	$1.9e-03$	$1.8e-03$	0.00523
84	USA-CA	Np	177	0.00252	$9.2e-04$	$4.7e-04$	0.00616
85	USA-CA*	Pa	177	0.00227	$1.2e-03$	$3.8e-04$	0.00520
86	USA-NY*	Ha	301	0.00243	$1.7e-03$	$1.3e-03$	0.00425
87	USA-NY	Mp	301	0.00188	$1.4e-03$	$7.2e-04$	0.00350
88	USA-NY	Na	301	0.00251	$1.7e-03$	$1.0e-03$	0.00481
89	USA-NY*	Np	301	0.00194	$8.8e-04$	$4.0e-04$	0.00455
90	USA-NY*	Pa	301	0.00105	$6.3e-04$	$2.7e-04$	0.00223
91	USA-TX	Ha	71	0.00391	$3.0e-03$	$1.5e-03$	0.00722
92	USA-TX	Mp	71	0.00159	$1.1e-03$	$5.4e-04$	0.00311
93	USA-TX	Na	71	0.00336	$2.2e-03$	$2.1e-03$	0.00584
94	USA-TX	Np	71	0.00297	$1.3e-03$	$5.6e-04$	0.00707
95	USA-TX	Pa	71	0.00214	$1.1e-03$	$4.2e-04$	0.00492
96	Viet Nam*	Ha	143	0.00281	$1.6e-03$	$1.5e-03$	0.00527
97	Viet Nam	Mp	143	0.00182	$1.1e-03$	$7.5e-04$	0.00363

98	Viet Nam*	Na	143	0.00240	$2.0e-03$	$1.4e-03$	0.00383
99	Viet Nam	Np	143	0.00259	$1.2e-03$	$5.9e-04$	0.00600
100	Viet Nam	Pa	143	0.00177	$7.4e-04$	$4.0e-04$	0.00416

GENBANK Accession Numbers for Data in Chapter 3

AY679147	CY000001	CY000009	CY000017	CY000025
CY000033	CY000041	CY000049	CY000057	CY000065
CY000073	CY000081	CY000089	CY000097	CY000105
CY000113	CY000129	CY000137	CY000153	CY000161
CY000169	CY000177	CY000185	CY000193	CY000201
CY000209	CY000217	CY000225	CY000233	CY000241
CY000249	CY000257	CY000265	CY000281	CY000289
CY000313	CY000321	CY000329	CY000337	CY000345
CY000353	CY000361	CY000369	CY000377	CY000385
CY000417	CY000425	CY000433	CY000441	CY000473
CY000481	CY000489	CY000497	CY000505	CY000513
CY000521	CY000529	CY000537	CY000545	CY000553
CY000561	CY000569	CY000584	CY000585	CY000625
CY000753	CY000761	CY000777	CY000785	CY000793
CY000865	CY000901	CY000909	CY000933	CY000941
CY000957	CY000965	CY000973	CY001013	CY001021
CY001029	CY001037	CY001045	CY001061	CY001064

CY001080 CY001088 CY001096 CY001104 CY001112

CY001128 CY001144 CY001152 CY001160 CY001168

CY001197 CY001205 CY001213 CY001221 CY001229

CY001261 CY001285 CY001293 CY001301 CY001309

CY001317 CY001333 CY001405 CY001421 CY001512

CY001552 CY001632 CY001648 CY001720 CY001728

CY001736 CY002000 CY002008 CY002016 CY002040

CY002048 CY002056 CY002064 CY002072 CY002128

CY002176 CY002184 CY002192 CY002208 CY002216

CY002224 CY002232 CY002248 CY002256 CY002328

CY002344 CY002424 CY002432 CY002440 CY002448

CY002456 CY002464 CY002488 CY002520 CY002592

CY002712 CY002720 CY002728 CY002736 CY002784

CY002816 CY003032 CY003056 CY003072 CY003080

CY003088 CY003096 CY003104 CY003120 CY003123

CY003136 CY003144 CY003152 CY003160 CY003168

CY003176 CY003184 CY003192 CY003200 CY003208

CY003408 CY003416 CY003424 CY003640 CY003648

CY003656 CY003664 CY003680 CY003777 CY006076

CY006084 CY006092 CY006291 CY006371 CY006379

CY008164 CY008884 CY009260 CY012792 CY013216

CY013224 CY013232 CY013805 CY014159 CY015676

CY015684 CY015692 CY015700 CY015708 CY015716

CY015724 CY015732 CY015740 CY015748 CY015756

CY015764 CY015772 CY015780 CY015788 CY015796

CY015804 CY015812 CY015820 CY015828 CY015836

CY015844 CY015852 CY015860 CY015868 CY015876

CY015884 CY015892 CY015900 CY015908 CY015916

CY015924 CY015932 CY015940 CY015948 CY015956

CY015964 CY015972 CY015980 CY015988 CY015996

CY016004 CY016012 CY016020 CY016028 CY016036

CY016044 CY016220 CY016979 CY016987 CY016995

CY017083 CY017091 CY017099 CY017107 CY017355

CY017797 CY018925 CY019141 CY019149 CY019157

CY019165 CY019173 CY019181 CY019189 CY019245

CY019253 CY019261 CY019269 CY019285 CY019293

CY019301 CY019309 CY019317 CY019325 CY019333

CY019811 CY019819 CY019827 CY019835 CY019843

CY019851 CY019859 CY019931 CY019939 CY020005

CY020053 CY020061 CY020069 CY020077 CY020085

CY020093 CY020101 CY020109 CY020117 CY020125

CY020133 CY020357 CY020365 CY020533 CY021989

CY025421 CY025485 CY025643 CY025715 CY025731

CY025739 CY025747 CY025835 CY025843 CY025851

CY026195 CY026275 CY026555 CY026787 CY027563

CY027579 CY028475 CY031563 CY032429 CY032437

CY032445 CY032453 CY032461 CY032469 CY032477

CY032485 CY032493 CY032501 CY032517 CY032525

CY032533 CY032541 CY032549 CY033638 CY034084

CY034092 CY034414 CY035086 CY035094 CY035102

CY036967 CY037359 CY037543 CY037607 CY037631

CY037711 CY037743 CY038511 CY038519 CY038527

CY038543 CY038559 CY038567 CY038583 CY038591

CY038607 CY038615 CY038623 CY038631 CY038639

CY038647 CY038663 CY038671 CY038679 CY038695

CY038703 CY038711 CY038727 CY038735 CY038743

CY038751 CY038791 CY038815 CY038911 CY038935

CY038943 CY038951 CY038959 CY038975 CY038983

CY038991 CY039007 CY039015 CY039023 CY039031

CY039039 CY039047 CY039055 CY039063 CY039159

CY039167 CY039175 CY039183 CY039207 CY039215

CY039223 CY039231 CY039239 CY039247 CY039439

CY039487 CY039495 CY039503 CY040298 CY040306

CY040314 CY040322 CY040338 CY040346 CY040354

CY043744 CY043752 CY043760 CY043768 CY044333

CY044381 CY044397 CY044429 CY044445 CY044453

CY044461 CY044469 CY044476 CY044492 CY044500

CY044508 CY044540 CY044548 CY044572 CY044580

CY044588 CY044596 CY044604 CY044612 CY044620

CY044628 CY044636 CY044644 CY044652 CY044668

CY044676 CY044692 CY044708 CY044716 CY044724

CY044732 CY044740 CY044748 CY044756 CY044772

CY044780 CY044788 CY044796 CY044804 CY044812

CY044820 CY044828 CY044844 CY044852 CY050452

CY050460 CY050468 CY050492 CY050500 CY050508

CY050532 CY050540 CY050556 CY050564 CY050572

CY050580 CY050588 CY050596 CY050604 CY050620

CY050628 CY050636 CY050652 CY050668 CY050676

CY050684 CY050708 CY050716 CY050724 CY050732

CY050788 CY050796 CY050820 CY050828 CY050836

CY055091 CY055099 CY058756 CY058764 CY058772

CY058780 CY058796 CY058804 CY064815 CY064823

CY064847 CY064855 CY064863 CY064879 CY064887

CY066519 CY067197 CY067205 CY067213 CY067221

CY067229 CY067237 CY067245 CY067253 CY067921

CY067929 CY067937 CY067945 CY067953 CY067961

CY067969 CY067977 CY067985 CY067993 CY068001

CY068009 CY068017 CY068025 CY068033 CY068041

CY068049 CY068057 CY068065 CY068073 CY068081

CY068089 CY068097 CY068105 CY068113 CY068121

CY068129 CY068137 CY068145 CY068153 CY068161

CY068169 CY068177 CY068185 CY068193 CY068201

CY068209 CY068217 CY068225 CY068233 CY068241

CY068249 CY068257 CY068265 CY068273 CY068281

CY068289 CY068297 CY068305 CY068313 CY068321

CY068329 CY068337 CY068353 CY068361 CY068377

CY068385 CY068393 CY068401 CY068409 CY068417

CY068425 CY068433 CY068441 CY068449 CY068457

CY068465 CY068473 CY068481 CY068489 CY068497

CY068505 CY068513 CY068521 CY068529 CY068537

CY068545 CY068553 CY068561 CY068569 CY068577

CY068585 CY068593 CY068601 CY068609 CY068617

CY068625 CY068633 CY068678 CY068686 CY068694

CY068702 CY068710 CY068718 CY068726 CY068734

CY068742 CY068750 CY068758 CY068766 CY068774

CY068782 CY068790 CY068798 CY068806 CY068814

CY068822 CY068830 CY068838 CY068846 CY068854

CY068862 CY068870 CY068878 CY070919 CY070927

CY070935 CY070943 CY070951 CY070959 CY070967

CY072190 CY072198 CY072206 CY072214 CY073757

CY073869 CY074675 CY074683 CY074691 CY074699

CY074707 CY074715 CY074723 CY074731 CY074739

CY074747 CY074755 CY074763 CY074771 CY074779

CY074787 CY074795 CY074803 CY074811 CY074819

CY074827 CY074835 CY074843 CY074851 CY074859

CY074867 CY074875 CY074883 CY074891 CY074899

CY074907 CY074915 CY074923 CY074931 CY077425

CY080459 CY080467 CY080475 CY080483 CY080491

CY081025 CY084334 CY084385 CY084393 CY084401

CY084409 CY084417 CY088774 CY088782 CY088790

CY088843 CY088851 CY088859 CY088867 CY088875

CY088883 CY088891 CY088899 CY088907 CY088915

CY088923 CY088931 CY088939 CY088947 CY088955

CY088963 CY088971 CY088979 CY088987 CY088995

CY089003 CY089011 CY089019 CY089027 CY089540

CY089629 CY089733 CY089741 CY089749 CY089765

CY089773 CY090941 CY090957 CY091013 CY091021

CY091037 CY091053 CY091069 CY091101 CY091133

CY091173 CY091213 CY091221 CY091309 CY091429

CY091461 CY091501 CY091509 CY091517 CY091525

CY091533 CY091541 CY091549 CY091557 CY091581

CY092241 CY092249 CY092257 CY092265 CY092273

CY092281 CY092289 CY092297 CY092305 CY092313

CY092329 CY092353 CY092361 CY092369 CY092377

CY093117 CY093248 CY093343 CY098065 CY098073

CY098081 CY104076 CY104084 CY104092 CY104100

CY104108 CY104116 CY104124 CY104132 CY104140

CY104148 CY104156 CY104164 CY104172 CY104180

CY104188 CY104196 CY104204 CY104212 CY104220

CY104228 CY104236 CY104252 CY104260 CY104268

CY104316 CY104324 CY104332 CY104340 CY104348

CY104356 CY104364 CY104372 CY104380 CY104388

CY104396 CY104404 CY104412 CY104420 CY104428

CY104436 CY104444 CY104452 CY104460 CY104468

CY104476 CY104484 CY104492 CY104500 CY104508

CY104516 CY104524 CY104532 CY104540 CY104548

CY104622 CY104630 CY104638 CY104646 CY104678

CY105190 CY105206 CY105214 CY105238 CY105246

CY105254 CY105262 CY105270 CY105278 CY105286

CY105294 CY105302 CY105310 CY105318 CY105326

CY105334 CY105342 CY105358 CY105366 CY105374

CY105382 CY105390 CY105398 CY105406 CY105414

CY105422 CY105430 CY105438 CY105446 CY105454

CY105462 CY105470 CY105478 CY105486 CY105494

CY105502 CY105510 CY105518 CY105526 CY105534

CY105542 CY105550 CY105558 CY105566 CY105574

CY105582 CY105590 CY105598 CY105606 CY105614

CY105622 CY105630 CY105638 CY105646 CY105654

CY105662 CY105670 CY105678 CY105686 CY105694

CY105702 CY105710 CY105718 CY105726 CY105734

CY105742 CY105750 CY105758 CY105766 CY105774

CY105782 CY105790 CY105798 CY105806 CY105814

CY105822 CY105830 CY105838 CY105846 CY105854

CY105862 CY105870 CY105878 CY105886 CY106576

CY106584 CY106592 CY106600 CY106608 CY106616

CY106624 CY106632 CY106640 CY106648 CY106656

CY106664 CY106672 CY106680 CY106688 CY106696

CY106704 CY106712 CY106720 CY106728 CY106736

CY106744 CY106752 CY106760 CY106768 CY106776

CY106784 CY106792 CY106800 CY106808 CY106816

CY106824 CY106832 CY106840 CY106848 CY106856

CY106864 CY106872 CY106880 CY106888 CY106896

CY106904 CY106912 CY106920 CY106928 CY106936

CY106944 CY106952 CY106960 CY106968 CY106984

CY106992 CY111126 CY111134 CY111142 CY111150

CY111158 CY111166 CY111174 CY111182 CY111190

CY111198 CY111214 CY111222 CY111230 CY111238

CY111246 CY111270 CY111302 CY111310 CY111318

CY111326 CY111334 CY111342 CY111350 CY111358

CY111366 CY111382 CY111390 CY111406 CY111414

CY111422 CY111430 CY111438 CY111446 CY111454

CY111462 CY111470 CY111478 CY111546 CY114373

CY115464 CY115472 CY115480 CY115488 CY115496

CY115504 CY115512 CY115520 CY115528 CY115536

CY115544 CY115552 CY115560 CY115568 CY115576

CY115584 CY115592 CY115600 CY115608 CY115616

CY115624 CY115632 CY115640 CY115648 CY115656

CY115664 CY115672 CY115680 CY115688 CY115696

CY115704 CY115712 CY115720 CY115728 CY115736

CY115744 CY115752 CY115760 CY115768 CY115776

CY115784 CY115792 CY115800 CY115808 CY115816

CY115824 CY116699 CY116707 CY116715 CY117581

CY117589 CY121125 CY121736 CY125717 CY125791

CY134481 CY134505 CY134513 CY134529 CY134545

CY134561 CY134577 CY134593 CY134601 CY134609

CY134732 CY134740 CY134748 CY134756 CY134764

CY134772 CY134780 CY134788 CY134796 CY134804

CY134812 CY134820 CY134828 CY134836 CY134844

CY134852 CY134860 CY134868 CY134876 CY134884

CY134892 CY134900 CY134908 CY134916 CY134924

CY134932 CY134940 CY134948 CY134956 CY134964

CY134972 CY134980 CY134988 CY134996 CY135004

CY135012 CY135020 CY135028 CY135036 CY135044

CY135052 CY135060 CY135068 CY135076 CY135084

CY135092 CY135100 CY135124 CY135132 CY135140

CY135148 CY135156 CY135164 DQ181797 DQ181798

DQ181799 DQ181800 DQ181801 DQ181802 DQ181803

DQ181804 DQ181805 DQ181806 EF629366 EF629367

EF629368 EF629369 EF629370 EF629373 EF643017

EU482444 EU482446 EU482448 EU482450 EU482463

EU482464 EU482465 EU482466 EU482467 EU482468

EU482469 EU482470 EU482471 EU482472 EU482473

EU482474 EU482475 EU482541 EU482542 EU482543

EU482544 EU482545 EU482546 EU482547 EU482548

EU482549 EU482550 EU482551 EU482552 EU482553

EU482554 EU482555 EU482556 EU482557 EU482558

EU482559 EU482560 EU482561 EU482562 EU482563

EU482564 EU482565 EU482566 EU482568 EU482569

EU482570 EU482571 EU482572 EU482573 EU482574

EU482575 EU482576 EU482577 EU482578 EU482579

EU482580 EU482581 EU482582 EU482583 EU482584

EU482585 EU482586 EU482587 EU482588 EU482589

EU482590 EU482593 EU482594 EU482595 EU482596

EU482597 EU482598 EU482599 EU482600 EU482601

EU482602 EU482603 EU482609 EU482610 EU482611

EU482612 EU482613 EU482614 EU482615 EU482616

EU482617 EU482618 EU482619 EU482620 EU482621

EU482622 EU482623 EU482624 EU482625 EU482626

EU482627 EU482628 EU482629 EU482630 EU482631

EU482632 EU482633 EU482634 EU482635 EU482636

EU482637 EU482638 EU482639 EU482640 EU482641

EU482642 EU482643 EU482644 EU482645 EU482646

EU482647 EU482648 EU482649 EU482650 EU482651

EU482652 EU482653 EU482654 EU482655 EU482656

EU482657 EU482658 EU482659 EU482660 EU482661

EU482662 EU482663 EU482664 EU482665 EU482666

EU482667 EU482668 EU482669 EU482670 EU482671

EU482672 EU482673 EU482674 EU482675 EU482676

EU482677 EU482678 EU482679 EU482680 EU482681

EU482682 EU482683 EU482684 EU482685 EU482686

EU482687 EU482688 EU482689 EU48269@ EU482690

EU482691 EU482692 EU482693 EU482694 EU482695

EU482697 EU482698 EU482699 EU482700 EU482701

EU482702 EU482703 EU482704 EU482705 EU482719

EU482720 EU482721 EU482722 EU482723 EU482724

EU482725 EU482726 EU482727 EU482728 EU482729

EU482730 EU482731 EU482732 EU482733 EU482734

EU482735 EU482736 EU482737 EU482738 EU482739

EU482740 EU482741 EU482742 EU482743 EU482744

EU482745 EU482746 EU482747 EU482748 EU482749

EU482750 EU482751 EU482752 EU482753 EU482754

EU482755 EU482756 EU482757 EU482758 EU482759

EU482760 EU482761 EU482762 EU482763 EU482766

EU482769 EU482770 EU482771 EU482772 EU482773

EU482774 EU482775 EU482776 EU482777 EU482778

EU482779 EU482780 EU482781 EU482782 EU482783

EU482784 EU482785 EU482786 EU482787 EU482788

EU529683 EU529684 EU529685 EU529686 EU529687

EU529688 EU529689 EU529690 EU529691 EU529692

EU529693 EU529694 EU529695 EU529696 EU529697

EU529698 EU529699 EU529700 EU529701 EU529702

EU529703 EU529704 EU529705 EU529706 EU569688

EU569689 EU569690 EU569691 EU569692 EU569693

EU569694 EU569695 EU569696 EU569697 EU569698

EU569699 EU569700 EU569701 EU569702 EU569703

EU569704 EU569705 EU569706 EU569707 EU569708

EU569709 EU569710 EU569711 EU569712 EU569713

EU569714 EU569715 EU569716 EU569717 EU569718

EU569719 EU569720 EU569721 EU596483 EU596484

EU596485 EU596486 EU596487 EU596488 EU596489

EU596490 EU596491 EU596492 EU596493 EU596494

EU596495 EU596496 EU596497 EU596498 EU596499

EU596500 EU596501 EU596502 EU596503 EU596504

EU621672 EU660398 EU660399 EU660400 EU660404

EU660405 EU660406 EU660413 EU660414 EU660415

EU660416 EU660417 EU660420 EU677137 EU677138

EU677141 EU677142 EU677143 EU677144 EU677145

EU677146 EU677147 EU677148 EU677149 EU687196

EU687197 EU687198 EU687199 EU687212 EU687213

EU687214 EU687215 EU687216 EU687217 EU687218

EU687219 EU687221 EU687222 EU687223 EU687224

EU687225 EU687226 EU687227 EU687228 EU687229

EU687230 EU687231 EU687232 EU687233 EU687234

EU687235 EU687236 EU687237 EU687238 EU687239

EU687240 EU687241 EU687242 EU687243 EU687244

EU687245 EU687246 EU687248 EU687249 EU687250

EU726767 EU726768 EU726769 EU726770 EU726771

EU726772 EU726773 EU726774 EU726776 EU781135

EU781136 EU781137 EU854291 EU854292 EU854298

EU932687 EU932688 FJ024423 FJ024452 FJ024454

FJ024458 FJ024461 FJ024465 FJ024466 FJ024467

FJ024468 FJ024469 FJ024470 FJ024471 FJ024478

FJ024479 FJ024480 FJ024481 FJ024482 FJ024483

FJ024484 FJ024485 FJ182002 FJ182004 FJ182005

FJ182006 FJ182007 FJ182008 FJ182009 FJ182010

FJ182011 FJ182013 FJ182014 FJ182015 FJ182037

FJ182038 FJ182039 FJ182040 FJ182041 FJ205870

FJ205871 FJ205877 FJ205878 FJ205879 FJ205880

FJ205885 FJ226066 FJ373299 FJ373300 FJ373301

FJ373302 FJ373303 FJ373304 FJ373306 FJ390371

FJ390372 FJ390373 FJ390375 FJ390376 FJ390377

FJ390384 FJ390385 FJ390387 FJ390390 FJ390391

FJ410176 FJ410177 FJ410178 FJ410193 FJ410195

FJ410200 FJ410202 FJ410208 FJ410215 FJ410217

FJ410219 FJ410221 FJ410223 FJ410224 FJ410228

FJ410233 FJ410237 FJ410241 FJ410259 FJ410288

FJ410290 FJ410291 FJ432720 FJ432721 FJ432724

FJ432726 FJ461305 FJ461309 FJ461311 FJ461314

FJ461321 FJ478455 FJ478456 FJ478459 FJ547064

FJ547067 FJ547068 FJ547069 FJ547070 FJ547071

FJ547072 FJ547073 FJ547074 FJ547075 FJ547076

FJ547077 FJ547078 FJ547079 FJ547080 FJ547081

FJ547082 FJ547083 FJ547084 FJ547085 FJ547088

FJ547089 FJ547090 FJ562098 FJ562104 FJ562107

FJ639669 FJ639670 FJ639671 FJ639672 FJ639673

FJ639674 FJ639675 FJ639676 FJ639677 FJ639678

FJ639679 FJ639680 FJ639681 FJ639682 FJ639683

FJ639684 FJ639685 FJ639686 FJ639687 FJ639688

FJ639689 FJ639690 FJ639691 FJ639692 FJ639693

FJ639694 FJ639695 FJ639696 FJ639697 FJ639698

FJ639699 FJ639700 FJ639701 FJ639702 FJ639703

FJ639704 FJ639705 FJ639706 FJ639707 FJ639708

FJ639709 FJ639710 FJ639711 FJ639712 FJ639713

FJ639714 FJ639715 FJ639716 FJ639717 FJ639718

FJ639719 FJ639720 FJ639721 FJ639722 FJ639723

FJ639724 FJ639725 FJ639726 FJ639727 FJ639728

FJ639729 FJ639730 FJ639731 FJ639735 FJ639740

FJ639741 FJ639743 FJ639759 FJ639760 FJ639761

FJ639762 FJ639763 FJ639765 FJ639766 FJ639767

FJ639768 FJ639769 FJ639770 FJ639771 FJ639772

FJ639774 FJ639775 FJ639776 FJ639777 FJ639778

FJ639779 FJ639780 FJ639781 FJ639782 FJ639784

FJ639785 FJ639786 FJ639787 FJ639789 FJ639790

FJ639791 FJ639792 FJ639793 FJ639794 FJ639795

FJ639796 FJ639797 FJ639798 FJ639799 FJ639800

FJ639801 FJ639802 FJ639803 FJ639804 FJ639805

FJ639806 FJ639807 FJ639808 FJ639810 FJ639811

FJ639812 FJ639813 FJ639814 FJ639815 FJ639816

FJ639817 FJ639818 FJ639819 FJ639820 FJ639821

FJ639823 FJ639824 FJ639825 FJ639826 FJ639827

FJ639828 FJ639829 FJ639830 FJ639831 FJ639832

FJ639833 FJ639834 FJ639835 FJ639836 FJ639837

FJ687434 FJ687435 FJ687436 FJ687437 FJ687438

FJ687439 FJ687440 FJ687441 FJ687442 FJ687443

FJ687444 FJ687445 FJ687446 FJ687447 FJ744701

FJ744702 FJ744703 FJ744704 FJ744705 FJ744706

FJ744707 FJ744708 FJ744709 FJ744710 FJ744711

FJ744712 FJ744713 FJ744714 FJ744715 FJ744716

FJ744717 FJ744718 FJ744719 FJ744720 FJ744721

FJ744722 FJ744723 FJ744724 FJ744725 FJ744741

FJ744742 FJ744743 FJ744744 FJ744745 FJ810409

FJ810410 FJ810411 FJ810412 FJ810415 FJ810416

FJ810418 FJ810419 FJ850048 FJ850049 FJ850050

FJ850051 FJ850052 FJ850053 FJ850054 FJ850055

FJ850056 FJ850060 FJ850061 FJ850062 FJ850063

FJ850064 FJ850065 FJ850066 FJ850067 FJ850069

FJ850079 FJ850080 FJ850083 FJ850086 FJ850089

FJ850092 FJ850094 FJ850096 FJ850097 FJ850098

FJ850099 FJ850100 FJ850101 FJ850102 FJ850103

FJ850104 FJ850109 FJ850110 FJ850111 FJ850113

FJ850114 FJ850115 FJ850116 FJ850117 FJ850118

FJ850119 FJ850120 FJ850121 FJ859028 FJ873808

FJ873809 FJ873810 FJ873811 FJ873812 FJ873813

FJ873814 FJ882576 FJ882577 FJ882578 FJ882579

FJ882593 FJ882594 FJ898432 FJ898433 FJ898434

FJ898435 FJ898436 FJ898437 FJ898446 FJ898447

FJ898452 FJ898468 FJ898469 FJ898470 FJ898471

FJ898472 FJ898473 FJ898474 FJ898475 FJ898476

FJ898477 FJ898478 FJ898479 FJ906956 FJ906957

FJ906958 FJ906960 FJ906961 FJ906962 FJ913015

GQ199771 GQ199772 GQ199789 GQ199790 GQ199791

GQ199792 GQ199793 GQ199794 GQ199795 GQ199796

GQ199797 GQ199798 GQ199799 GQ199800 GQ199801

GQ199802 GQ199803 GQ199804 GQ199805 GQ199806

GQ199807 GQ199808 GQ199809 GQ199810 GQ199811

GQ199812 GQ199813 GQ199814 GQ199815 GQ199816

GQ199817 GQ199818 GQ199819 GQ199820 GQ199821

GQ199822 GQ199823 GQ199824 GQ199825 GQ199826

GQ199827 GQ199828 GQ199829 GQ199830 GQ199831

GQ199832 GQ199833 GQ199834 GQ199835 GQ199836

GQ199837 GQ199838 GQ199839 GQ199840 GQ199841

GQ199842 GQ199843 GQ199844 GQ199845 GQ199846

GQ199847 GQ199848 GQ199849 GQ199850 GQ199851

GQ199852 GQ199853 GQ199854 GQ199855 GQ199856

GQ199857 GQ199858 GQ199859 GQ199860 GQ199861

GQ199862 GQ199863 GQ199864 GQ199865 GQ199866

GQ199867 GQ199868 GQ199869 GQ199870 GQ199871

GQ199872 GQ199873 GQ199874 GQ199875 GQ199877

GQ199886 GQ199895 GQ199896 GQ199897 GQ199898

GQ252678 GQ868498 GQ868499 GQ868500 GQ868501

GQ868502 GQ868503 GQ868504 GQ868505 GQ868506

GQ868507 GQ868508 GQ868509 GQ868510 GQ868511

GQ868512 GQ868513 GQ868514 GQ868517 GQ868518

GQ868519 GQ868520 GQ868521 GQ868522 GQ868523

GQ868524 GQ868525 GQ868526 GQ868527 GQ868528

GQ868529 GQ868530 GQ868531 GQ868532 GQ868533

GQ868534 GQ868535 GQ868536 GQ868537 GQ868538

GQ868539 GQ868542 GQ868543 GQ868544 GQ868545

GQ868546 GQ868547 GQ868548 GQ868586 GQ868587

GQ868591 GQ868604 GQ868605 GQ868606 GQ868607

GQ868608 GQ868609 GQ868610 GQ868611 GQ868612

GQ868613 GQ868614 GQ868615 GQ868618 GQ868619

GQ868620 GQ868621 GQ868622 GQ868623 GQ868624

GQ868625 GQ868626 GQ868627 GQ868628 GQ868629

GQ868631 GQ868634 GQ868638 GQ868646 GU056029

GU056030 GU056031 GU056032 GU056033 GU131832

GU131833 GU131834 GU131835 GU131836 GU131837

GU131838 GU131839 GU131840 GU131841 GU131842

GU131844 GU131845 GU131846 GU131847 GU131848

GU131849 GU131850 GU131851 GU131852 GU131853

GU131854 GU131855 GU131856 GU131857 GU131858

GU131859 GU131860 GU131861 GU131862 GU131865

GU131866 GU131867 GU131868 GU131869 GU131870

GU131871 GU131872 GU131873 GU131874 GU131875

GU131876 GU131877 GU131878 GU131886 GU131896

GU131897 GU131898 GU131899 GU131900 GU131901

GU131902 GU131903 GU131904 GU131905 GU131906

GU131907 GU131908 GU131909 GU131910 GU131911

GU131912 GU131913 GU131914 GU131915 GU131916

GU131917 GU131918 GU131924 GU131927 GU131928

GU131929 GU131930 GU131931 GU131932 GU131933

GU131934 GU131935 GU131936 GU131937 GU131938

GU131939 GU131940 GU131941 GU131942 GU131943

GU131944 GU131945 GU131946 GU131956 GU131957

GU131958 GU131960 GU131961 GU131962 GU131963

GU131964 GU131965 GU131966 GU131967 GU131968

GU131969 GU131970 GU131971 GU131972 GU131973

GU131976 GU131977 GU131978 GU131979 GU131980

GU131981 GU131982 GU131983 GU131984 HM181933

HM181934 HM181935 HM181936 HM181937 HM181938

HM181939 HM181940 HM181941 HM181942 HM181943

HM181944 HM181945 HM181946 HM181947 HM181948

HM181949 HM181950 HM181951 HM181952 HM181953

HM181954 HM181955 HM181956 HM181957 HM181958

HM181959 HM181972 HM181973 HM181974 HM181975

HM181976 HM181977 HM181978 HM488255 HM631852

HM631853 HM631854 HM631855 HM631856 HM631857

HM631858 HM631859 HM631860 HM631861 HM631862

HM631863 HM631864 HM631865 HM631866 HM631867

HM631868 HM631869 HM756274 HM756275 HM756276

HM756277	HM756278	HM756280	HM756281	HM756282
HQ166030	HQ166031	HQ166032	HQ166033	HQ166034
HQ166035	HQ166036	HQ166037	HQ235027	HQ541785
HQ541786	HQ541787	HQ541788	HQ541789	HQ541790
HQ541791	HQ541792	HQ541793	HQ541794	HQ541795
HQ541797	HQ541798	HQ541799	HQ541802	HQ541804
HQ541805	HQ634199	HQ671176	HQ671177	HQ705609
HQ705611	HQ705612	HQ705613	HQ705614	HQ705615
HQ705616	HQ705617	HQ705618	HQ705619	HQ705620
HQ705621	HQ705623	HQ705624	HQ705625	HQ733861
HQ891025	JF295012	JF357905	JF357906	JF357907
JF730044	JF730045	JF730046	JF730047	JF730048
JF730049	JF730050	JF730051	JF730052	JF730053
JF730054	JF730055	JF937635	JF937644	JF937645
JF937651	JN697379	JN796245	JN819402	JN819403
JN819405	JN81941	JN819410	JN819411	JN819412
JN819413	JN819414	JN819415	JN819416	JN819420
JN819421	JN819423	JN819424	JN819425	JQ287664
JQ287665	JQ287666	JX079688	JX079690	JX079691
JX079694	KC882479	KC882639	KC882908	KC892437

List of Equations

List of Figures

Bibliography

[Ahlquist, 2002] Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science*, 296(5571):1270–1273.

[Barton, 1995] Barton, N. (1995). Linkage and the limits to natural selection. *Genetics*, 140:821–841.

[Beibricher and Eigen, 2005] Beibricher, C. and Eigen, M. (2005). The error threshold. *Virus Res*, 107(2):117–127.

[Benson et al., 2013] Benson, D., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D., Ostell, J., and Sayers, E. (2013). Genbank. *Nucleic Acids Res.*, 41:D36–D42.

[Birky and Walsh, 1988] Birky, C. and Walsh, J. (1988). Effects of linkage on rates of molecular evolution. *Proc. Natl. Acad. Sci. USA*, 85(17):6414–6418.

[Bull et al., 2013] Bull, J., Joyce, P., Gladstone, E., and Molineaux, A. (2013). Empirical complexities in the genetic foundations of lethal mutagenesis. *Genetics*, 195(2):541–552.

[Bull et al., 2007] Bull, J., Sanjuan, R., and Wilke, C. (2007). Theory of lethal mutagenesis for viruses. *J Virol*, 81(6):2930–2939.

[Burch and Chao, 2000] Burch, C. and Chao, L. (2000). Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature*, 406:625–628.

[Chao et al., 2002] Chao, L., Rang, C., and Wong, L. (2002). Distribution of spontaneous mutants and inferences about the replication mode of RNA bacteriophage phi6. *J Virol*, 76(7):3276–3281.

[Crotty et al., 2001] Crotty, S., Cameron, C., and Andino, R. (2001). RNA virus error catastrophe, direct molecular test by using ribavirin. *Proc Natl Acad Sci USA*, 98:6895–6900.

[Crotty et al., 2000] Crotty, S., Maag, D., Arnold, J., Zhong, W., Lau, J., Hong, Z., Andino, R., and Cameron, C. (2000). The broad-spectrum antiviral ribunucleoside ribavirin is an RNA virus mutagen. *Nat Med*, 6(12):1375–1379.

[Domingo et al., 2001] Domingo, E., Biebricher, C., Eigen, M., and Holland, J. (2001). *Quasispecies and RNA Virus Evolution: Principles and Consequences.* Landes Biosciences.

[Domingo et al., 2005] Domingo, E., Escarmis, C., Lazaro, E., and Manrubia, S. (2005). Quasispecies dynamics and RNA virus extinction. *Virus Res*, 107:129–139.

[Drake, 1993] Drake, J. (1993). Rates of spontaneous mutation among RNA viruses. *Proc Natl Acad Sci USA*, 90:4171–4715.

[Drake and Holland, 1999] Drake, J. and Holland, J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci USA*, 96(24):13910–13913.

[Drummond et al., 2011] Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A. amd Duran, C., Field, M., Heled, J. amd Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., and Wilson, A. (2011). Geneious v5.4.

[Drummond et al., 2012] Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with beauti and the beast 1.7. *Molecular Biology and Evolution*.

[Duffy et al., 2002] Duffy, S., Shackelton, L., and Holmes, E. (2002). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*, 9(4):267–276.

[Eigen, 1971] Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523. 10.1007/BF00623322.

[Eigen, 2002] Eigen, M. (2002). Error catastrophe and antiviral strategy. *Proc Natl Acad Sci USA*, 99(21):13374–13376.

[Elena and Sanjuan, 2005] Elena, S. and Sanjuan, R. (2005). Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol*, 79(18):11555–11558.

[Feller, 1968] Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Third Edition. John Wiley and Sons.

[Fisher, 1930] Fisher, R. (1930). The evolution of dominance in certain polymorphic species. *Am Nat*, 64(694):384–406.

[French and Stenger, 2003] French, R. and Stenger, D. (2003). Evolution of wheat streak mosaic virus: Dynamics of population growth within plants may explain limited variation. *Ann Rev Phytopathol*, 44:199–214.

[Garcia-Villada and Drake, 2012] Garcia-Villada, L. and Drake, J. (2012). The three faces of riboviral spontaneous mutation: spectrum, mode of genome replication, and mutation rate. *PLoS Genet*, 8(7).

[Holland et al., 1990] Holland, J., Domingo, E., de la Torre, J., and Steinhauer, D. (1990). Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J Virol*, 64(8):3960–3962.

[Holland et al., 1982] Holland, J., Spindler, K., Grabau, E., Nichol, S., and VandePol, S. (1982). Rapid evolution of RNA genomes. *Science*, 215(4540):1577–1585.

[Holmes, 2009] Holmes, E. (2009). *The Evolution and Emergence of RNA Viruses*. Oxford Series in Ecology and Evolution. Oxford University Press.

[Jenkins et al., 2001] Jenkins, G., Rambaut, A., Pybus, O., and Holmes, E. (2001). Rate of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J Mol Evol*, 54:156–165.

[Kimura, 1962] Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics*, 47:713–719.

[Kimura, 1964] Kimura, M. (1964). Diffusion models in populations genetics. *J Appl Probab*, 1:177–232.

[Lagerkvist, 1978] Lagerkvist, U. (1978). ”Two out of Three”: An alternative method for codon reading. *Proc Natl Acad Sci USA*, 75(4):1759–1762.

[Lederberg, 1998] Lederberg, J. (1998). Emerging infection: An evolutionary perspective. *Emerg Infect Dis*, 4(3):366–371.

[Loverdo et al., 2012] Loverdo, C., Park, M., Schreiber, S., and Lloyd-Smith, J. (2012). Influence of viral replication mechanisms on within-host evolutionary dynamics. *Evolution*, 66(11):3462–3471.

[McCauley and Mahy, 1983] McCauley, J. and Mahy, B. (1983). Structure and function of the influenza virus genome. *Biochem J*, 211:281–294.

[Orr, 2000] Orr, H. (2000). The rate of adaptation in asexuals. *Genetics*, 155:961–968.

[Orr, 2003] Orr, H. (2003). The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4):1519–1526.

[Perera and Kuhn, 2008] Perera, R. and Kuhn, R. (2008). Structural proteomics of dengue virus. *Curr Opin Microbiol*, 11:369–377.

[Poch et al., 1990] Poch, O., Bloomberg, B., Bougueleret, L., and Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: Theoretical assignment of functional domains. *J Gen Virol*, 71(5):1153–1162.

[Portela and Digard, 2002] Portela, A. and Digard, P. (2002). The influenza virus nucleoprotein: a multifunctinal RNA-binding protein pivotal to virus replication. *J Gen Virol*, 83:723–734.

[Pressing and Reanney, 1984] Pressing, J. and Reanney, D. (1984). Divided genomes and intrinsic noise. *Journal of Molecular Evolution*, 20(2):135–146.

[Pybus and Rambaut, 2009] Pybus, O. and Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet*, 10:540–550.

[R Development Core Team, 2013] R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Sanjuan, 2010] Sanjuan, R. (2010). Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci*, 365(1548):1975–1982.

[Sanjuan, 2012] Sanjuan, R. (2012). From molecular genetics to phylodnmamics: Evolutionary relevance of mutation rates across viruses. *PLoS Pathog*, 8(5).

[Sanjuan et al., 2004] Sanjuan, R., Moya, A., and Elena, S. (2004). The distribution of fitness effects caused by single nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA*, 101(22):8396–8401.

[Sanjuan et al., 2010] Sanjuan, R., Nebot, M., Chirico, N., Mansky, L., and Belshaw, R. (2010). Viral mutation rates. *J Virol*, 84(19):9733–9748.

[Sardanyes et al., 2009] Sardanyes, J., Sola, R., and Elena, S. (2009). Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. *J Virol*, 83(23):12579–12589.

[Schlesinger et al., 1990] Schlesinger, J., Brandriss, M., Putnak, J., and Walsh, E. (1990). Cell surface expression of yellow fever virus non-structural glycoprotein NS1: consequences of interaction with antibody. *J Gen Virol*, 71:553–599.

[Schoniger et al., 1994] Schoniger, M., Janke, A., and von Haeseler, A. (1994). *Studies in Classification, Data Analysis and Knowledge Organization, Chapter: How to deal with Third Codon Positions in Phylogenetic Analysis.* Information Systems and Data Analysis. Springer Berlin Heidelberg.

[Thebaud et al., 2010] Thebaud, G., Chadeouf, J., Morelli, M., McCauley, J., and Haydon, D. (2010). The relationship between mutation frequency and replication strategy in positive-sense single-stranded rna viruses. *Proc R Soc B*, 277(1602):809–817.

[Xia, 1998] Xia, X. (1998). The rate heterogeneity of nonsynonymous substitution in mammalian mitochondrial genes. *Mol Biol Evol*, 15(3):336–344.