Characterization of Avian Leukosis Virus Evolution in Culture

A Dissertation

submitted by

Robbie Narang

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Molecular Microbiology

TUFTS UNIVERSITY
Sackler School of Graduate Biomedical Sciences

Date

February, 2011

ADVISER: Dr. John Coffin

Abstract

HIV infection *in vivo* is characterized by rapid population turnover that, when coupled with a high mutation rate, allows the virus to quickly adapt to the immune system and antiretroviral therapy. To date, all known antiretroviral therapy has been met with the emergence of drug resitance mutations causing therapy to fail. The large number of infected cells imply that drug resistant mutations might preexist in the population and predictions of the emergence or clearance of mutations should be modeled deterministically. The effective population size, a population genetics concept that accounts for factors that make a population more or less sensitive to genetic drift, may be much smaller than the census population size for HIV infections. Directly measuring the effective population size of HIV infection is difficult given the complexities of natural infections.

We developed a simple tissue culture system to examine the relationship between the census population size and effective population size of an evolving retrovirus population. We initiated competition experiments between wild-type ALV subgroup B virus (WT) and two previously described mutants, LS and LSTI, which were found to confer a host range extension phenotype [Rainey et al., 2003, Taplitz and Coffin, 1997]. We developed a measure for genetic drift in these cultures by initiating replicate competition experiments and measuring the variance in the time of emergence of WT virus. We found that under the relatively simple conditions of tissue culture, genetic drift plays a significant role in the evolution of the population, a role that was a

function of the initial starting frequency of WT virus and the strength of selection between WT and mutant virus.

To define the idealized expectations of the competition experiments, we experimentally measured the parameters relevant to the evolution of the viral populations. We found that WT virus is 4.6 fold more fit that LS and 10.9 fold more fit than LSTI viruses under the conditions of the competition experiments. The generation time of the virus, defined as the time that elapses from one point in the replication cycle to the same point in daughter virus, was found to be greater than 96 hours. Target cell doubling times were unaffected by infection with all viruses used in this study. We supplied these parameters to simulations based on our simple model and found that the theoretical expectations underestimate the role of drift. These results highlight the complexities that exist even under the relatively homogenous conditions of tissue culture.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Properties of Retroviruses

Retroviruses are enveloped, plus-strand RNA viruses. Each viral particle carries two copies of a linear, non-segmented genome and is considered to be diploid. The "retro" in retrovirus is due the the reverse transcription step required to generate a double stranded DNA copy of the RNA genome. This DNA copy is integrated into the host cell genome through a process called integration, generating a provirus. Once integrated, the provirus behaves as a host cell gene. It is replicated by host cell replication machinery and is inherited in a mendelian manner.

Retrovirus genomes range in size from about 7 kb to 12 kb and can be classified into two categories, simple and complex. Simple viruses contain the genes *gag, pro, pol,* and *env*. Complex retroviruses contain these genes and additional accessory genes that modulate interactions with the host cell or play regulatory roles in viral replication. Viral particles are about 80-120 nm in diameter.

1.2 Historical Highlights of Retrovirology

Retroviruses were discovered by two independent investigations involving neoplasms in chickens. In 1908, Ellermann and Bang showed transfer of leukemia to healthy chickens inoculated with cell-free filtrates of homogenized spleen, liver, and bone marrow prepared from chickens with leukemia [Ellermann, 1909] and in 1911, Peyton Rous showed that cell-free filtrates of homogenized solid tumors were able to induce tumor formation when administered to healthy chickens [Rous, 1911]. The agent discovered by Ellermann and Bang was eventually named Avian Leukosis Virus (ALV) and is the subject of study in this thesis. The agent discovered by Rous was named Rous sarcoma virus (RSV) and differs from ALV due to the acquisition of a host cell gene *src* which increases the frequency with the which the virus transforms cells.

Subsequent work increased the list of organisms in which oncogenesis due to retroviral infection was observed. It was discovered that the mouse mammary carcinomas which were transmissible through breast milk from mother to offspring are due to mouse mammary tumour virus (MMTV). The identification of murine leukemia viruses (MLVs) provided another valuable model for cancer development. The transforming ability of retroviruses, including their ability to capture oncogenes, pointed towards proteins and pathways important in cancer biology [Coffin, 1997]. Subsequently, major advances in tissue culture assays, such as the focus forming assay of RSV in chick embryo fibroblasts [Temin and RUBIN, 1958] and the fusogenecity assay of MLV infected cells and the XC indicator cell line [Rowe et al., 1970], allowed quantitative studies of infectivity and isolation of clonal populations of virus.

Howard Temin first proposed the theory that retroviruses reverse transcribe their genomes and integrate a DNA copy into the DNA of the host in 1964 [Temin, 1964]. This proposal was based on various lines of evidence such as the sensitivity of viral replication to DNA inhibitors [Temin, 1963]. The independent discovery of reverse

transcriptase by David Baltimore [Baltimore, 1970] and Howard Temin [Temin and Mizutani, 1970] in 1970 confirmed Temin's hypothesis. This discovery changed the landscape of molecular biology not only due to changing the understanding of how information flows between DNA and RNA but by also providing the basic tools for the generation of DNA from RNA *in vitro*.

Human T-cell Leukemia virus (HTLV-1) was discovered as the first human oncogenic retrovirus in the early 1980s [Wong-Staal et al., 1983, Poiesz et al., 1980], relying on the discovery of new techniques to culture human T-cells [Morgan et al., 1976, Ruscetti et al., 1977]. HTLV-1 became the prototype complex retrovirus when it was discovered that it encoded a more complicated gene set than previously discovered retroviruses [Seiki et al., 1983].

The rise of the AIDS epidemic eventually led to the discovery of HIV [Levy and Shimabukuro, 1985, Gallo et al., 1984, Barré-Sinoussi et al., 1983]. Today, the AIDS epidemic remains a significant burden in both developed and developing nations. No cure or effective vaccine has yet been developed.

1.3 Basic Retrovirus Biology

1.3.1 Classification

All retroviruses belong to the family *Retroviridae* [Murphy et al., 1995]. The family is divided into seven genera: Alpharetroviruses, Betaretroviruses, Gammaretroviruses, Deltaretroviruses, Epsilonretroviruses, Lentiviruses, and Spumaretroviruses. Retroviruses were originally sorted into these genera based on characteristic particle morphologies, but this classification system has been replaced by sequence relatedness of reverse transcriptase genes. The classes of particle morphology are still useful as a descriptive characteristic of viral particles. A-type particles are intracellular structures

with a thick shell and hollow, electron-lucent center. These are now known to be an intermediate structure that matures into the other categories of viral particles [Kuff et al., 1972]. B-type particles show a round and eccentric core. C-type particles assemble at the plasma membrane and have a central spherical core. D-type particles assemble in the cytoplasm and, after budding, exhibit a cylindrical core [Knipe and Howley, 2006]. The Alpha-, Beta-, and Gammaretrovirus genera contain simple retroviruses. The Deltaretrovirus, Epsilonretrovirus, Lentivirus, and Spumaretrovirus genera contain complex retroviruses.

Alpharetroviruses. The prototype alpharetrovirus is Rous sarcoma virus (RSV). Alpharetroviruses have simple genomes and have particles with C-type morphology. This genera encompasses the avian retroviruses including Avian leukosis virus, which is the subject of this study.

Betaretroviruses. The prototype betaretrovirus are mouse mammary tumor virus (MMTV) and Mason-Pfizer monkey virus (MPMV). Betaretroviruses are simple retroviruses that have B-type morphology. Betaretroviruses are known to infect mice, primates, and sheep.

Gammaretroviruses. The prototype gammaretroviruses are murine leukemia virus (MLV), feline leukemia viruses (FeLVs), and the gibbon ape leukemia virus (GALV). These are simple retroviruses with C-type morphology. These viruses infect a variety of mammals and have also been found in reptiles and birds.

Delaretroviruses. The prototype deltaretroviruses are human T-lymphotropic virus (HTLV) and bovine leukemia virus (BLV). These viruses are complex retroviruses and have C-type morphology. These viruses express the accessory proteins *rex* and *tax*, which regulate transcription and transport of unspliced forms of viral RNA for use as viral genomes.

Epsilonretroviruses The prototype epsilonretroviruses are walleye dermal sarcoma virus (WDSV) and walleye epidermal hyperplasia virus (HEHV). These are

complex retroviruses with C-type morphology. The only known examples are found in fish and reptiles.

Lentiviruses. The prototype lentivirus is Human immunodeficiency virus (HIV). The genus is named for the slow disease progression caused by infections with lentiviruses. These viruses have complex genomes and exhibit D-type morphology.

Spumaretroviruses. The prototype spumaretrovirus is human foamy virus (HFV). These viruses have a unique morphology with spikes extending from their surface and an uncondensed core.

1.3.2 Genome Organization

All retroviruses share a genomic RNA organization that begins (from 5' to 3') with a ^{7}mG cap, a repeat (R), unique 5' region (U5), primer binding site (PBS), the genes *gag, pro, pol,* and *env*, polypurine tract (PPT), unique 3' region (U3), another repeat (R), and a poly-A tail. Retroviruses may differ in the reading frames of the genes or in the location of the splice sites. The Lentiviruses and Spumaretroviruses contain an additional PPT, which is required for reverse transcription [Telisnitsky and Goff, 1997].

Retroviruses are classified into two general categories, independent of phylogeny, based on genome structure. Simple retroviruses contain the canonical gene set *gag, pro, pol,* and *env* (listed in the order they appear in the genome from 5' to 3') as in the example of ALV (Figure 1.1 A). Complex retroviruses contain these genes as well as a number of accessory genes such as *vif, vpr, vpu, tat, rev*, and *nef* in the example of HIV-1 (Figure 1.1 C). Certain retroviruses have acquired host genes or gene fragments, as is the case with RSV and its acquisition of the gene *src* [Vogt, 1997](Figure 1.1 B).

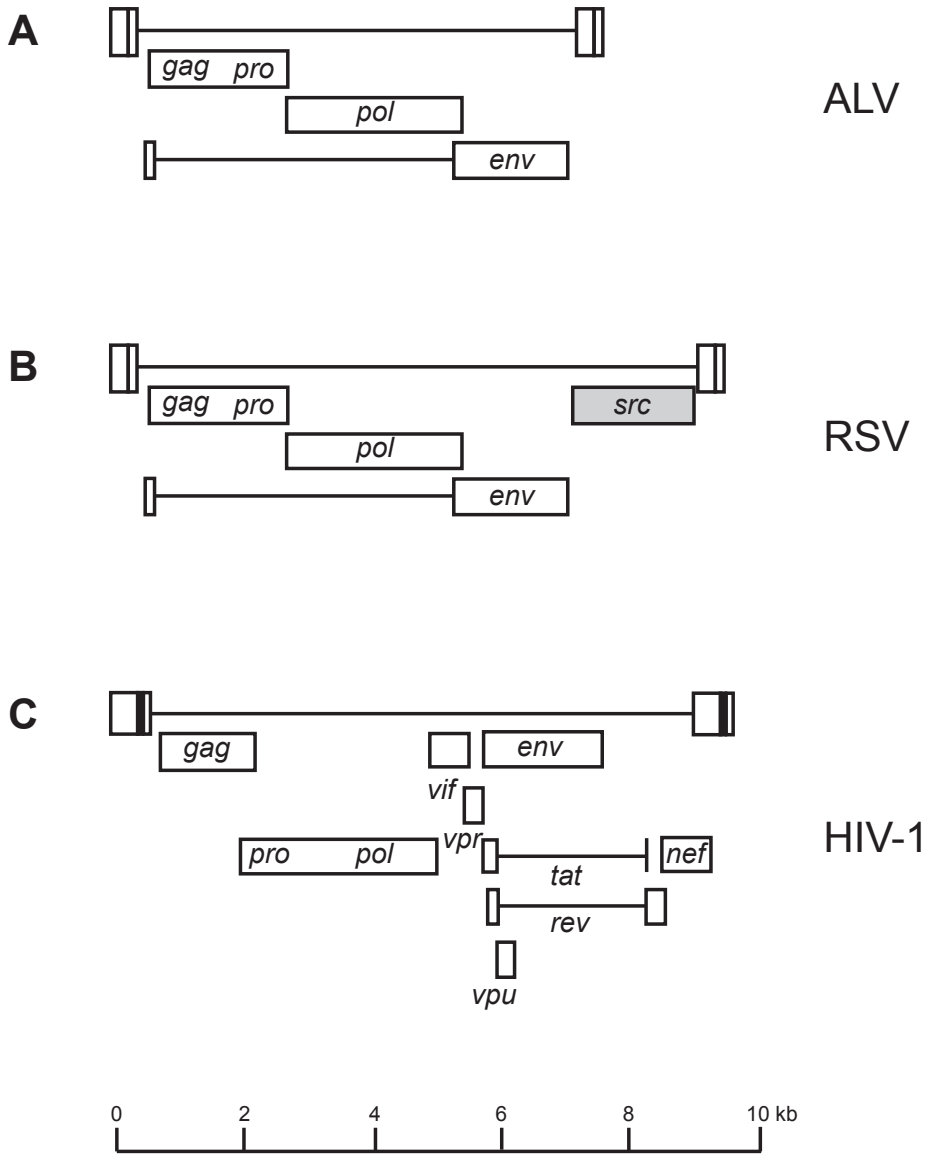

Figure 1.1: Genomes of three typical retroviruses. A) Avian Leukosis Virus (ALV), a simple retrovirus. B) Rous Sarcoma Virus (RSV), a simple retrovirus that picked up an oncogene. C) HIV-1, a complex retrovirus. Adapted from [Coffin, 1997]

1.3.3 Virion Structure

Retroviruses are enveloped plus-sense RNA viruses (Figure 1.2). Two copies of of the RNA genome are complexed with nucleocapsid (NC) and contained with a viral core composed of capsid (CA) protein. This complex is surrounded by a membrane associated spherical assembly of matrix (MA). The envelope is a host-derived lipid bilayer containing the membrane anchored transmembrane subunit of Env (TM) and its surface exposed binding partner (SU).

The viral core can be spherical or conical in shape, depending on the type of retrovirus, and requires proteolytic cleavage shortly after budding to assume its mature conformation. Cleavage is performed by a virally encoded protease, the product of the *pro* gene that is packaged within the viral particle.

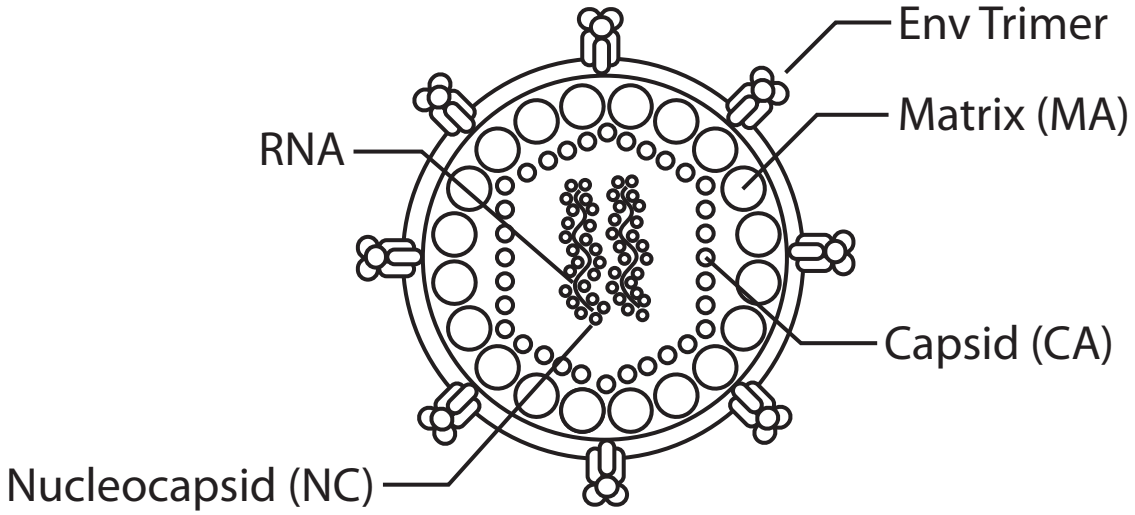

Figure 1.2: **Structure of a Typical Retrovirus.** Two copies of a positive-sense RNA genome are decorated with nucleocapsid protein (NC) and surrounded by a core formed by capsid protein (CA). The core is enveloped by a lipid bilayer, which is host derived, and surrounds a spherical structure composed of matrix protein (MA). The lipid bilayer contains the viral envelope protein (Env) which is a trimer of two subunits, the membrane anchored TM portion, and the surface exposed SU portion.

1.3.4 The Viral Lifecycle

Viral replication occurs in a cycle. Therefore, there is no defined beginning or end to the process. As a matter convention, binding and entry are considered to be the starting point. After entry, the cycle proceeds through the steps of uncoating, reverse transcription, integration, expression, packaging, budding, and maturation. The sum total of the process results in a cell that harbors a provirus in its genome. The provirus provides the information necessary for the expression of the proteins required to assemble viral particles capable of initiating new rounds of infection. Each of these steps are discussed in further detail in the following sections. The entire viral lifecycle is summarized in Figure 1.3.

Binding and Entry

Binding of a retrovirus to a cell is mediated through a specific interaction between the envelope glycoprotein and a host cell receptor. This interaction plays a large role in determining the types of cells a retrovirus can infect. Env consists of two subunits, SU and TM, and it is the SU subunit that makes contact with the cognate host cell receptor. Some viruses require additional interactions with cell surface proteins and these co-receptor interactions may confer additional specificity to the host-range of the virus. HIV requires the co-receptors CC-chemokine receptor-5 (CCR5) or CXC-chemokine receptor-4 (CXCR4) for fusion. Binding of viral envelope SU and host receptor initiates conformational changes within the envelope that ultimately result in extension of TM into the host membrane, mediating fusion between the viral and host membranes. The entry mechanism of entry for ALV, the virus used in this study, is discussed in further detail in Chapter 1.5.2.

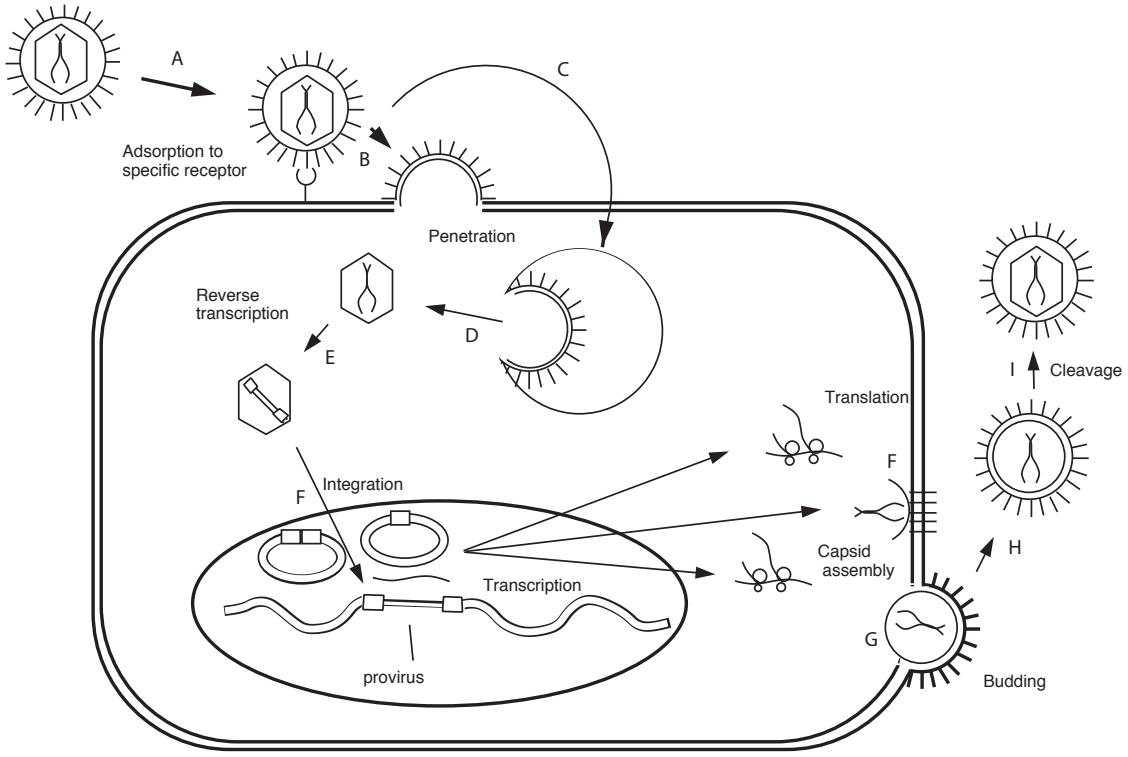

Figure 1.3: The retroviral lifecycle. (A) Viral particles bind specific receptors of target cells and, depending on the type of virus, (B) enter the cell directly at the membrane, or (C) enter an endocytic compartment and (D) enter the cytoplasm upon fusion initiated upon acidification. (E) The single stranded RNA genome is reverse transcribed to DNA and associated with integrase enzyme to form a pre-integration complex. (F) The complex is transported to the nucleus and integration into the host genome generates a provirus. The provirus is transcribed and assembly (F) occurs at the host cell plasma membrane. (G) The virus buds from the membrane and upon release (H) matures into an infectious form through the proteolytic cleavage (I) of precursor polypeptides. See main text for details of each step. Copied with permission from [Coffin, 1996].

Uncoating

The protein core is delivered into the cytoplasm of the cell where uncoats. Studies based on panels of mutagenized target cells [Gao and Goff, 1999, Bruce et al., 2005] and one study investigating cell-free uncoating [Narayan and Young, 2004] indicate that some host factors must be involved in the early steps of replication between fusion and the formation of a reverse transcription complex, but a comprehensive set of factors are yet to be identified.

Reverse Transcription

Reverse transcription is performed by a reverse-transcriptase molecule (a *pol* gene product) that is carried into the host cytoplasm inside the viral core. The substrate for the reverse transcription reaction is the plus-sense single-stranded RNA genome of the virus (packaged into the virion as a dimer) and the product of the reaction is a double-stranded DNA copy of the genome. This process is error-prone (discussed further in Chapter 1.7.1) due to poor fidelity of the polymerase, and often leads to mutations in the viral genome.

Reverse trancriptase initiates polymerization at a partially unwound tRNA, serving to prime the reaction. It is annealed at the primer binding site (PBS), which is situated near the 5' end of the genome. Polymerization proceeds towards the 5' end of the template RNA until the end is reached, forming a small DNA/RNA hybrid termed the minus strand strong stop DNA. In addition to serving as a RNA-dependent DNA polymerase, reverse transcriptase encodes a RNaseH activity which degrades the RNA in the DNA/RNA hybrid. This results in a free single stranded DNA that anneals at the 3' end of the viral genome in a step called the first strand transfer. Now situated at the opposite end of the viral genome, the nascent DNA strand serves as a primer for another DNA polymerization event towards the 5' end of the viral genome. This is

accompanied by RNase-H degradation of a second DNA/RNA hybrid, this time leaving degradation-resistant polypurine tract (PPT) untouched, which serves to prime minus strand DNA synthesis through a short stretch of the tRNA primer, leading to the plus strand strong stop DNA. RNase-H degradation removes the tRNA, leaving another exposed stretch of single stranded DNA that anneals with the primer binding site in the minus-strand DNA, called the second strand transfer. This primes the remaining positive strand DNA synthesis required to generate the double stranded viral genome. Replication results in the generation of characteristic repeats at the 3' and 5' ends known as the long terminal repeats (LTRs) [Telisnitsky and Goff, 1997].

Integration

Following reverse transcription, integration of the viral genome into the host DNA is performed by the viral protein integrase (like reverse transcriptase, a product of the viral gene *pol*). A pre-integration complex is formed by the association of the newly reverse transcribed viral DNA and integrase. Integrase cleaves two bases from the 3' ends of each LTR, leaving exposed 3'-OH groups which are available to attack the phosphodiester bonds of the host DNA in an integrase-catalyzed concerted reaction. This reaction results in two single stranded breaks in the target DNA followed by joining of the viral DNA ends. The net result is a viral genome integrated into the host DNA, flanked by two single stranded gaps that are repaired by host enzymes [Brown, 1997b].

Transcription and Translation

At this point in the replication cycle, the integrated genome behaves as if it were a host gene; it replicates along with the other genes on the chromosome during mitosis (or meiosis in the case of gametes) and is expressed using the same machinery. The viral LTRs contain the regulatory requirements for recruitment and initiation of transcrip-

tion by RNA polymerase II. Transcription initiates in the R sequence of the 5' LTR and continues towards the 3' LTR. Transcripts are 5'-capped and 3'-polyadenylated, behaving just like cellular mRNAs. Full length transcripts of the viral genome are packaged into virions. Subgenomic transcripts, which are splice products of the full length transcripts, serve as the templates for translation of the viral proteins. Retroviruses have evolved different strategies to ensure that both spliced and unspliced (to make viral genomes and structural proteins) transcripts exist in the appropriate proportions. Complex retroviruses encode gene products with the function of regulating splicing and simple retroviruses utilize *cis*-acting sequences to alter the efficiency of splicing [Rabson and Graves, 1997].

Post Translational Modifications

Env proteins contain a short leader peptide which traffics the nascent peptide to the endoplasmic reticulum where it cotranslationally enters into the lumen. Env is heavily N- and O-linked glycosylated, then transported to the Golgi where host proteases cleave Env into surface (SU) and transmembrane (TM) subunits. The proteins oligimerize before final transport to the surface of cell plasma membrane. Gag and Gag-Pro-Pol proteins are cotranslationally myristoylated, which is required for binding of Gag to the plasma membrane [Rein et al., 1986]. The Gag protein of ALV, which is not myristoylated, is modified by an acetyl group [Palmiter et al., 1978]. However, it is unclear what role this modification plays in Gag assembly [Coffin, 1997].

Assembly

The mature viral proteins are almost all derived from translated polyprotein precursors. *env* is the simplest, expressed as a singly spliced product, transported to the ER for post-translational modifications, and cleaved into SU and TM by host cell proteases. *gag* is both translated alone, due to its position at the start of the 5'

end of the genome, and as a *gag-pro-pol* precursor due to occasional slippage [Jacks and Varmus, 1985] or stop codon suppression [Yoshinaka et al., 1985]. This has two important consequences. First, *gag* is expressed much more than the *gag-pro-pol* precursor, providing the correct stoichiometry to assemble the structural features of the nascent viral particles. Second, the *gag-pro-pol* precursor targets *pro* and *pol* to the site of viral assembly due to their association with *gag*.

Gag to Gag associations are considered to be the driving force for particle assembly. Gag contains three domains required for assembly and budding. The membrane binding domain (M) is required for Gag association with the inner leaflet of the plasma membrane. The interacting domain (I) is required for Gag to Gag association. Finally, the late assembly domain (L) is involved in particle release. Two major strategies are used for assembly, depending on the genus of virus under consideration. Most retroviruses assemble particles at the inner leaflet of the plasma membrane. The betaretroviruses and spumaretroviruses assemble particles in the cytoplasm and the particles are transported to the plasma membrane for release. Most retroviruses can assemble viral particles by expression of Gag particles alone. However, these particles are not infectious [Goff, 2001, Swanstrom and Wills, 1997].

Packaging of viral RNA requires Gag [Shields et al., 1978, Sakalian et al., 1994] as well as sequences found in the 5' end of the genome. This sequence, referred to as the ψ packaging signal, interacts with the nucleocapsid (NC) precursor region of Gag. The ψ sequence most likely enhances packaging of viral sequences rather than serving as a requirement, due to the fact packaged cellular RNAs are often observed in retroviral infections. Alpharetroviruses contain direct repeat (DR) sequences found at the 3' end of the genome which reduce the the efficiency of RNA packaging 10-fold when mutated [Aschoff et al., 1999].

Budding and Maturation

Assembled viral particles are thought to force curvature of the host membrane leading to a structure primed for budding. In HIV, Gag contains two L domains that bind cellular proteins Tsg101 and Alix, which recruit the host's fission machinery (ESCRT), facilitating release of virions [Hurley et al., 2010, Dussupt et al., 2010]. Particles are not infectious unless they are processed by the viral protease (Pro) in a step called maturation. Pro, present in the virion as a low activity Gag-Pro or Gag-Pro-Pol fusion, auto-cleaves which initiates a cascade of cleavage events that produce matrix (MA), capsid (CA), and nucleocapsid (NC) from Gag. Matrix and nucleocapsid remain associated with the membrane and RNA respectively. Capsid undergoes a dramatic polymerization event that results in the regular geometric shape that encapsidates the genome, an event that is observable by electron microscopy by a shift from electron lucent to electron dense viral cores. Pol is cleaved into the active forms of reverse transcriptase (RT) and integrase (IN). At this point in the replication cycle, the virus is now infectious and able to initiate a new round of the lifecyle.

1.4 Endogenous Retroviruses

The occasional infection of a germ line cell of an organism can result in progeny organisms where every cell contains a provirus. These proviral elements are called endogenous retroviruses and are inherited in a mendelian manner. Endogenous retroviruses have been found in every vertebrate species examined and it is estimated the human genome consists of 8% endogenous retroelements, highlighting the long relationship between retroviral infections and vertebrate species [Belshaw et al., 2004].

Endogenous retroviruses can be classified into two general categories, ancient and modern. Ancient endogenous retroviruses integrated into the genome of a species before its speciation event and the proviruses will be located in the exact same loca-

tion in the genome for all members of the species. Modern endogenous retroviruses are the results of endogenization after a speciation event and can be heterogenously distributed in a population [Boeke and Stoye, 1997].

Chickens contain two types of endogenous elements. The first element, *evs* were first discovered through the production of replication competent virus from uninfected cells that were chemically stimulated [Weiss et al., 1971]. The sequence of these viruses were closely related to that of exogenous ALVs, however they contained variations in the U3 region of the LTR that reduced transcriptional activity [Tsichlis and Coffin, 1980] as well is in the SU subunit of envelope [Coffin et al., 1983, Coffin et al., 1978]. *ev* loci are only found in domestic chickens and their ancestor, the red jungle fowl, but are not found in other members of the genus *Gallus*, indicating that *ev* integrations are due to relatively recent infections [Frisby et al., 1979]. The ubiquity of *ev* insertions in domestic chicken suggested that they may serve a beneficial role. However, breeding of chickens free of *evs* demonstrated that their role was inessential [Astrin et al., 1979]. Importantly, these chickens eventually led to the Lansing line 0 (EV-O, ev^-) chickens from which the cell line DF1 is derived [Himly et al., 1998]. The DF1 cell line is used extensively in this study. The ev^- genotype of DF1 cells is important to reduce recombination events between the viruses used in this study and the chromosomes of infected cells.

The second type of endogenous elements are endogenous avian viruses (EAVs), which are more widely distributed among avian species. This suggests a more ancient integration than *evs* [Boyce-Jacino et al., 1992, Dunwiddie et al., 1986]. EAV elements generally contain more mutations and deletions than *evs* and are not replication competent.

1.5 Avian Retroviruses

1.5.1 Receptor Usage in Avian Retroviruses

Alpharetroviruses, along with murine leukemia viruses, are unique in that closely related members within these genera can utilize a wide variety of receptors for entry. This variation in receptor usage is observable through several types of experiments. The simplest type of experiment is to determine the range of cells a virus can infect. In another type of experiment, receptor interference patterns provide additional information; cells that are infected with virus that use a particular receptor typically do not support superinfection by another virus that utilizes the same receptor. Interference is most likely due to the pre-engagement of the receptor by Env expressed by the pre-infecting virus [Gilbert et al., 1990]. Finally, neutralizing antibody cross-reactivity can give additional information about shared structural features of viral envelope proteins and provide further information for categorization. Taken together, the data from these types of studies have led to the classification of ALV into 10 subgroups, A-J. Table 1.1 summarizes the classification of typical ALV strains as well as indicates their species of origin and receptor usage [Levy, 1993].

The most well studied subgroups of ALV are subgroups A-E. The SU portion of *env* (gp85) in this set of viruses is highly conserved except for five regions, termed *hr1*, *hr2*, *vr1*, *vr2*, and *vr3*. *hr1* and *hr2* are known to be the determinants of host range specificity (Figure 1.4)([Dorner et al., 1985]). The widely divergent variable regions of alpharetroviruses translate to a wide variety of receptor usage, suggesting that frequent selective pressures by the host have forced the alpharetrovirus genus to maintain regions of plasticity to allow for relatively few mutations to utilize new receptors.

The receptors for subgroups A-E virus are encoded by three distinct loci, *tv-a*, *tv-b*, and *tv-c* [Payne and Biggs, 1966, Vogt and Ishizaki, 1965, RUBIN, 1965]. *tv-*

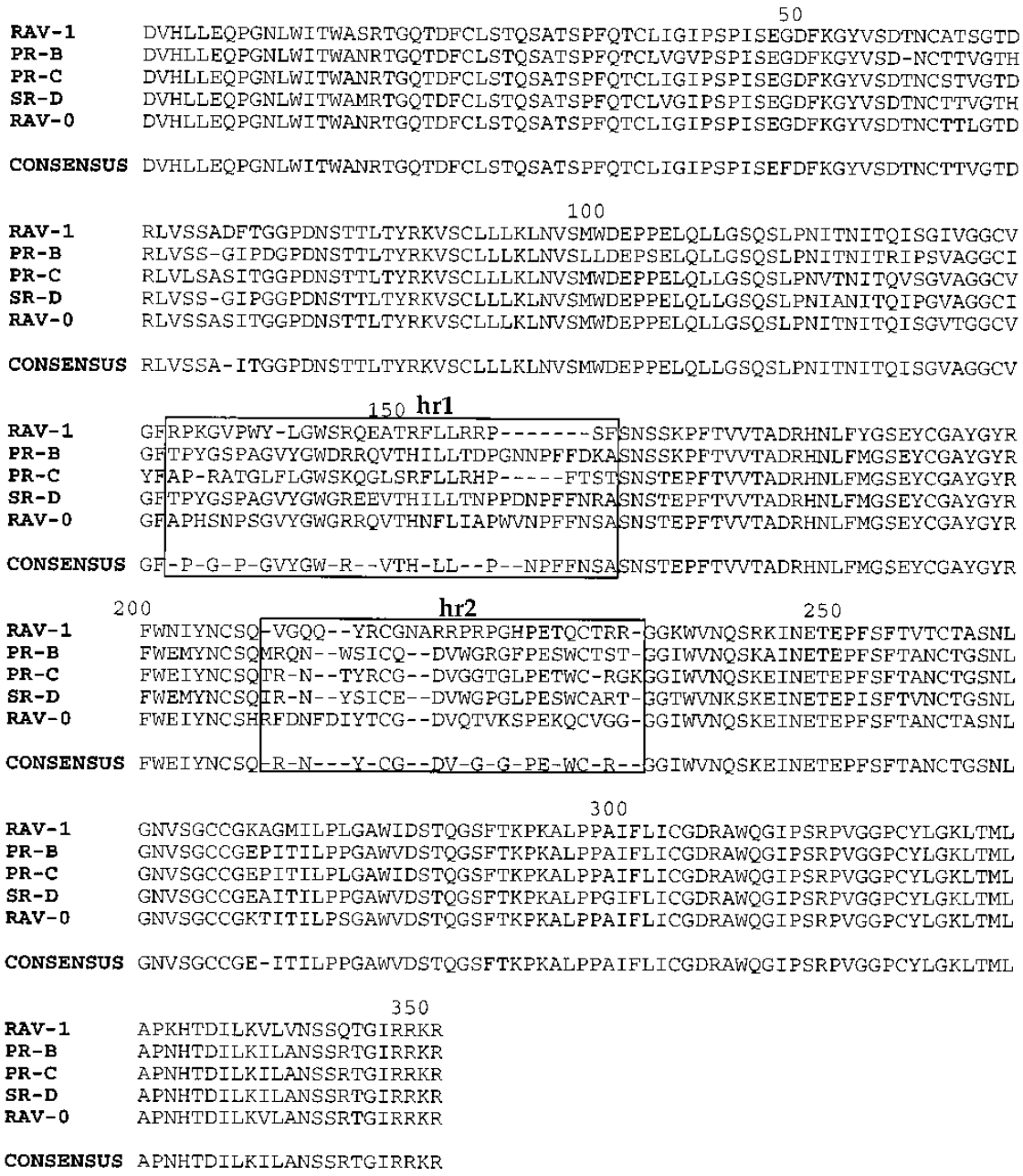

Figure 1.4: The amino acid sequences of subgroup A (RAV-1), B (PR-B), C (PR-C), D (SR-D), and E (RAV-0) ALV viruses show near perfect identity except for two regions, *hr1* and *hr2*. These variable regions are known to be the determinants for host range among the different subgroups. Copied with permission from [Taplitz and Coffin, 1997].

a encodes the receptor for subgroup A viruses and *tv-c* encodes the receptor for subgroup C viruses. Allelic variants of *tv-b* serve as the receptors for subgroup B, D, and E viruses.

Tv-a, encoded by the *tv-a* gene, is a member of the low-density lipoprotein receptor (LDLR) family [Young et al., 1993, Bates et al., 1993]. This membrane-associated protein contains an extracellular domain with LDLR repeats that are the sites of interaction for Env [Rong et al., 1998, Zingler et al., 1995, Bélanger et al., 1995].

The allelic receptors encoded by *tv-b* are members of the tumor necrosis factor receptor (TNFR) family. Chickens have the $tv\text{-}b^{s1}$ and $tv\text{-}b^{s3}$ alleles. Tv-b^{s1} permits infection by subgroup B, D, and E viruses [Adkins et al., 2000]. Preinfection of cells expressing Tv-b^{s1} with subgroup B or D virus prevents superinfection by subgroups B, D, or E. Preinfection with subgroup E virus only prevents superinfection by subgroup E virus, suggesting that Tv-b^{s1} exists in two different conformations on the cell surface [Coffin, 1997]. Tv-b^{s3} permits infection by subgroup B and D viruses. Quail and turkeys express $tv\text{-}b^{q}$ or $tv\text{-}b^{t}$, and permit infection by subgroup E viruses only [Crittenden and Motta, 1975, Crittenden et al., 1973, Adkins et al., 1997]

The most recently identified receptor is *tv-c* [Elleder et al., 2004, Elleder et al., 2005], which is closely related to mammalian butyrophillins. These proteins are similar to the immunoglobulin superfamily. Tv-c contains two immunoglobulin-like domains, IgV and IgC. The IgV domain is known to interact with the envelope protein of subgroup C viruses [Munguia and Federspiel, 2008].

1.5.2 Mechanisms of Entry

Enveloped viruses are strikingly similar in their mechanisms of entry into cells. Influenza virus, Ebola virus, paramyxoviruses, and retroviruses all encode type-1 envelope glycoproteins that are expressed on the surface of the viral particle as a trimer,

Subgroup	**Typical strains**	**Host species of origin**	**Receptor locus**
A	RAV-1, PR-RSV-A	chicken	*tv-a*
B	RAV-2, PR-RSV-B	chicken	*tv-b*
C	RAV-49, B77	chicken	*tv-c*
D	CZAV, SR-RSV-D	chicken	*tv-b*
E	RAV-0, RAV-60	chicken	*tv-b*
F	RPV	Ring-necked pheasant	-
G	GPV	Golden pheasant	-
H	-	Hungarian partridge	-
I	GQV	Gable's (Lophortyx) quail	-
J	HPRS-103	chicken	-

Table 1.1: **Subgroup/Receptor Classification of ALV** Avian retroviruses are classified into subgroups A-J based on host range, interference patterns, and neutralizing antibody cross-reactivity. Copied from [Coffin, 1997].

with each member of the trimer composed of two subunits. One subunit is membrane anchored at its C-terminus and bound to a surface exposed subunit at its N-terminus [Knipe and Howley, 2006]. This structure is considered to be metastable, meaning that it is in a high energy pre-fusion state. The triggers that causes destabilization of the high energy state to a lower energy state, along with massive conformational rearrangements, fall into two categories: those that require low pH for fusion and those that do not.

The fusion process is best described for influenza virus, which is often used as model to understand retroviral fusion. HA2 is the membrane anchored subunit of the envelope and binds HA1, the surface-exposed receptor binding subunit. Both pre-fusion [Wilson et al., 1981] and post-fusion [Bullough et al., 1994] conformations have been visualized by crystal structures and provide evidence of a mechanism of entry. First, HA1 binds sialic acid, which is the cognate receptor for entry, followed by endocytosis of the virus. Next, acidification of the endocytic compartment causes a conformational change in the envelope that forces helices packed in the core of HA2 to extend outward and embed in the cell membrane [Harrison, 2008]. Extension of this model to retroviruses relies largely on similar crystal structures for the core regions of HA2 and the core regions of TM for HIV and MoMLV [Chan et al., 1997]. No complete structure of a retroviral TM, in either pre-fusion or post-fusion states, has been determined to date.

Retrovirus entry is mediated by the two subunits of Env, TM and SU (Figure 1.5). TM is the membrane anchored subunit and binds to the surface exposed subunit SU, which is responsible for interactions with receptor on the host cell surface. TM of ALV contains an approximately 20 amino acid fusion peptide that is inserted into the host cell membrane upon fusion [Earp et al., 2003]. C-terminal to the fusion peptide are two heptad repeats, HR1 and HR2. These repeats collapse during fusion to form a six helix bundle and bring the viral and host membrane into close proximity [Netter

et al., 2004].

Most retroviruses enter by a pH-independent mechanism. ALV was originally thought to fall into the pH-independent category based on evidence that showed pre-treatment of virus with low pH was insufficient to block the replication ability of RSV [Gilbert et al., 1990]. However, it has recently been appreciated that ALV is a special case of pH-dependent entry; only following receptor binding by SU is the virus sensitive to low pH. This was discovered by demonstrating that lysosomotropic agents that prevented acidification of endosomal compartments were able to inhibit ALV subgroup A infectivity [Mothes et al., 2000]. Additionally, treatments of ALV subgroup A with soluble forms of Tv-a receptor were able to make the virus susceptible to inactivation by low pH [Smith et al., 2004].

The complete model for ALV entry is as follows: first, SU binds cognate receptor on the cell surface, which induces conformational changes in TM. These changes result in the extension of the fusion peptide from TM into the cell membrane. After trafficking into an endocytic compartment, acidification causes the heptad repeats to associate, forcing a collapse of TM which brings the viral and cell membranes into close proximity, creating the fusion pore [Barnard et al., 2006, Barnard et al., 2004, Earp et al., 2003, Netter et al., 2004] (Figure 1.6).

1.5.3 Host Range Extension in ALV

A few exceptions have been observed to the patterns of receptor recognition by ALV subgroups A-E. Recombinant forms of RSV have been described that have the host range of both parental strains. The most well studied of these is NTRE-4, which arose from a recombination event between the Prague B strain of RSV, a subgroup B virus, and the endogenous Rous-associated virus 0, a subgroup E virus. NTRE-4 virus exhibited the host range of both viruses, as it was able to infect cells expressing

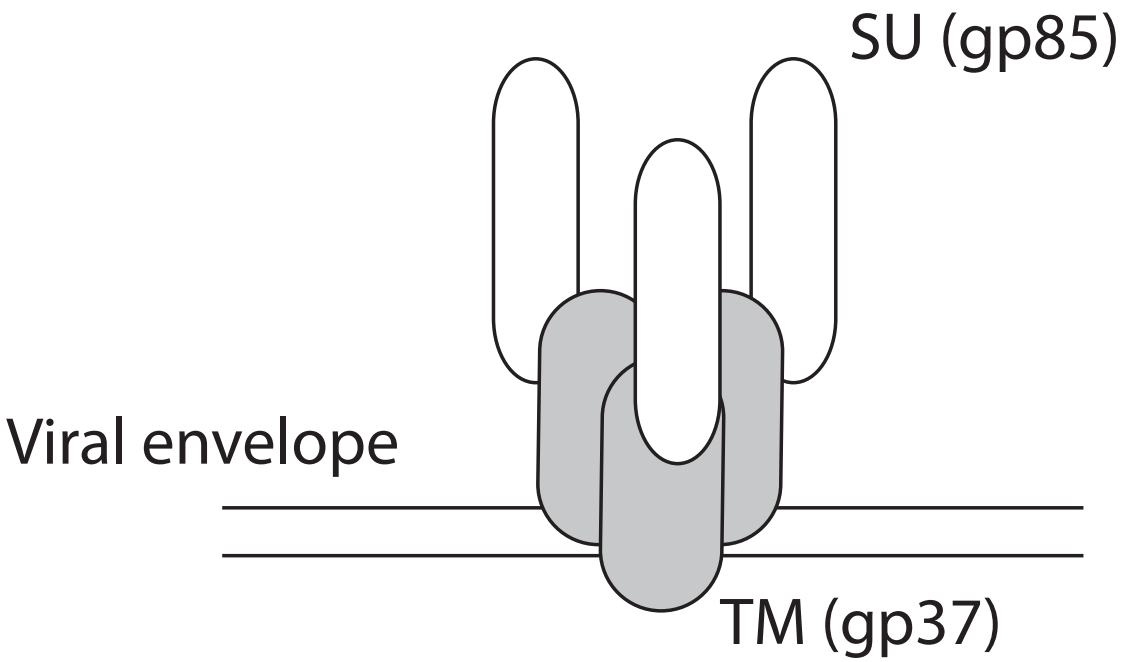

Figure 1.5: **The arrangement of SU and TM in the envelopes of alpharetroviruses.** The envelope subunits SU, for surface subunit, and TM, for transmembrane unit, associate as a trimer of heterodimers at the surface of the viral particle. SU is responsible for binding cognate receptor. TM contains the machinery necessary for fusion.

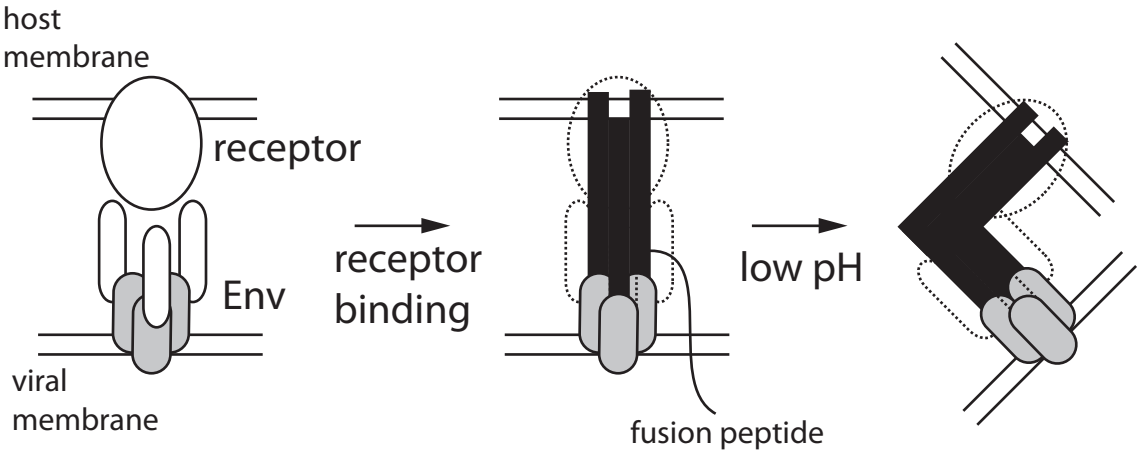

Figure 1.6: **Proposed model for entry of ALV.** Binding of the SU subunit of Env results in a conformation change that exposes a fusion peptide within TM and the fusion peptide buries its N-terminus into the host membrane. Upon trafficking to an endocytic compartment, low pH triggers a conformational change that forces the association of N- and C- terminal heptad repeats in the fusion peptide which brings virus and host membrane together to form a pore.

the subgroup B or E receptor [Coffin et al., 1980]. NTRE-4 was later confirmed to use the same receptors as subgroup B and E viruses based on the presence of receptor interference when cells were preinfected with parental subgroup B or E viruses [Tsichlis et al., 1980]. Construction of a panel of chimeric molecular clones of viruses demonstrated that the combination of the *hr1* region of subgroup B virus and the *hr2* region of subgroup E virus was sufficient to allow host range extension phenotype seen with NTRE-4 [Dorner and Coffin, 1986] (Figure 1.8).

Further work identified specific point mutations in *hr1* that could confer host range extension to subgroup B virus [Taplitz and Coffin, 1997]. These mutants were isolated from an experimental protocol in which C300 cells, permissive for subgroup B virus, and QT6 cells, non-permissive for subgroup B virus, were co-cultured. Subgroup B virus was serially passaged on the mixture of cell types until variants arose that could infect the non-permissive cell line QT6 (Figure 1.7). A mutant that could detectably infect QT6 cells was discovered after 20 passages and was identified to have two point mutations in adjacent codons in the *hr1* region of SU. (Figure 1.8). Two amino acids, L154 and T155 were mutated to generate L154S and T155I as a consequence of the mild selective pressure applied in the co-culture passaging experiment. Both mutations were the product of transitions (T to C, in the case of L154S and C to T in the case of T155I) and both mutations occurred in the first position of each codon. L154S was shown to be sufficient for host-range extension. When T155I was introduced alone it did not confer a host range extension phenotype [Rainey et al., 2003]. The host range extended beyond avian cell types. The mutants L154S, and L154S/T155I could infect human, dog, cat, mouse, rat, and hamster cells. The L154S and L154S/T155I mutants are used in experiments described in this thesis and are referred to as LS or LSTI.

The extended host range phenotype observed with LS and LSTI mutants was particularly interesting when coupled with some additional information about receptor

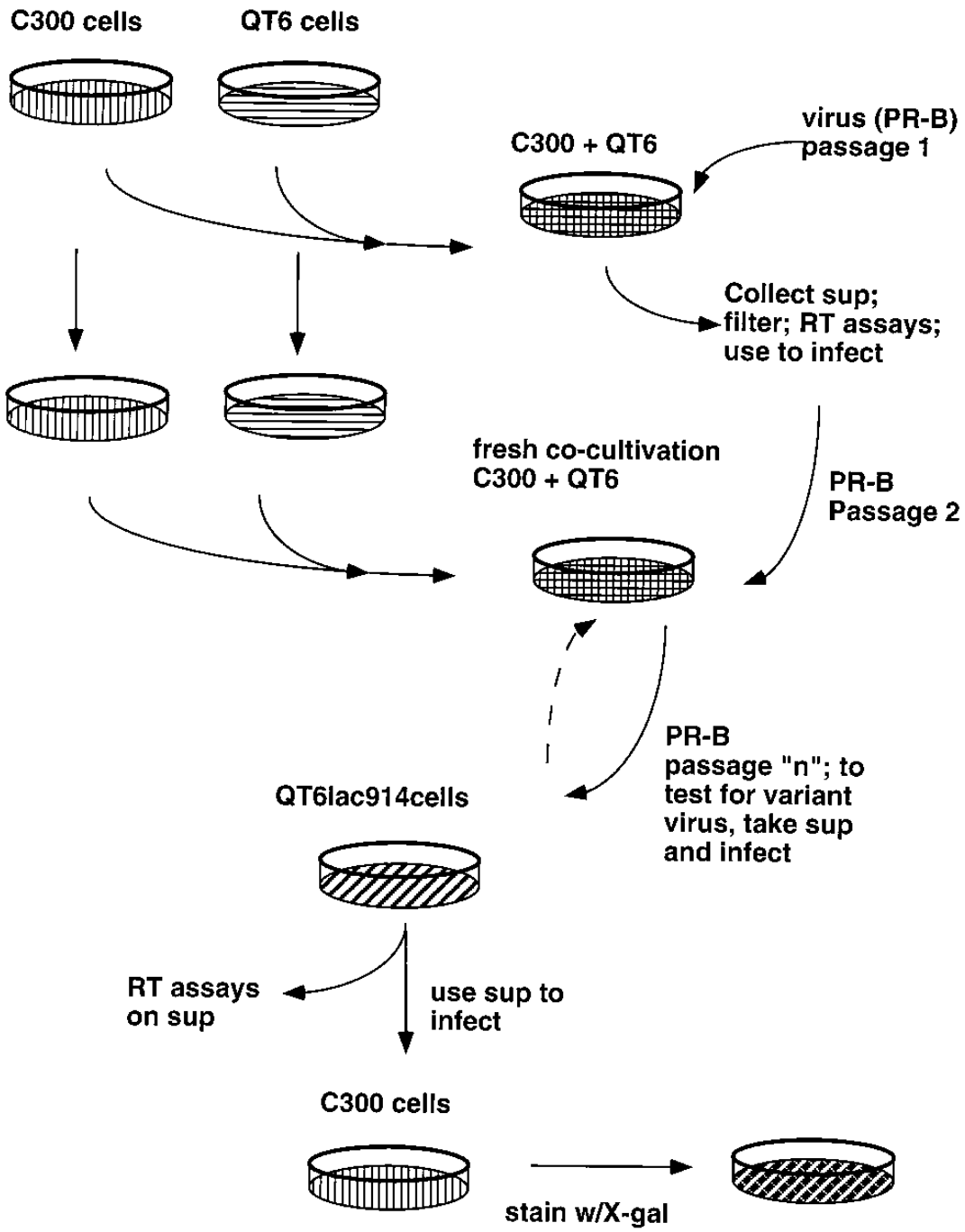

Figure 1.7: **Schematic of passaging experiment to identify additional host range extension mutants.** Wild-type ALV subgroup B virus was passaged in a co-culture of C300 (permissive) and QT6 (non-permissive) cells. At each passage, supernatant was tested for the presence of virus able to infect QT6 cells using the QT6lac914 reporter cell line. Copied with permission from [Taplitz and Coffin, 1997].

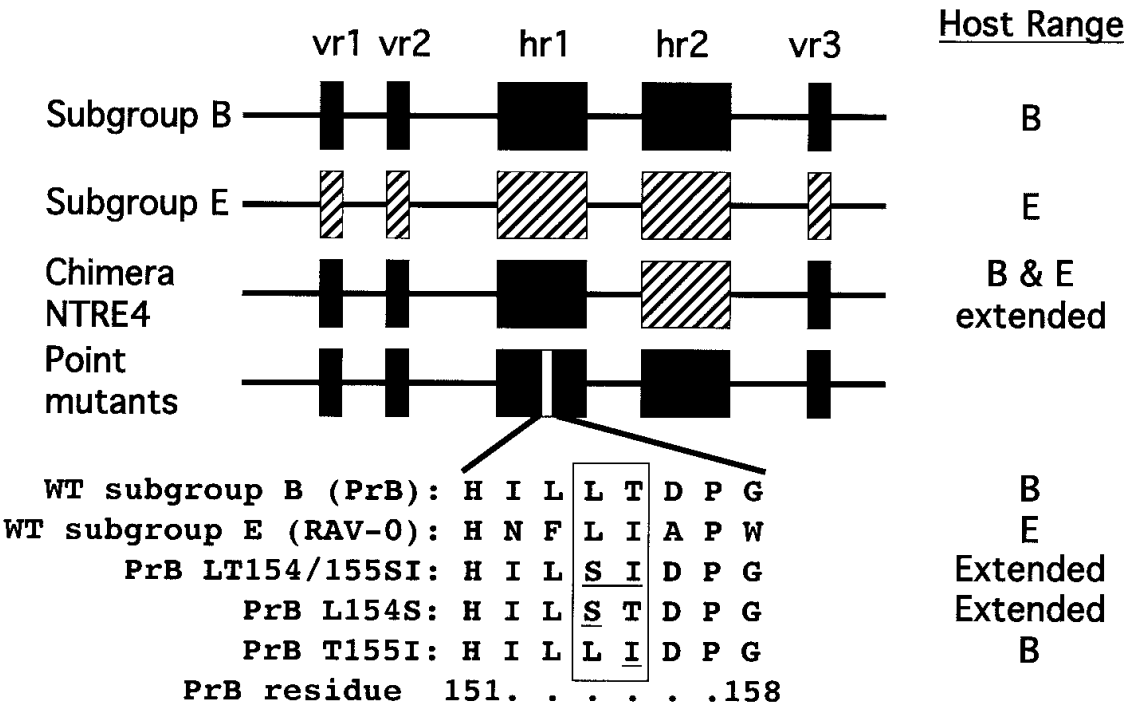

Figure 1.8: **Host range extension mutations in SU.** Host range extension mutants were originally isolated as recombinants between two different viruses, such as NTRE4 which was shown to contain the *hr1* subgroup B viruses and the *hr2* of subgroup E viruses. Specific point mutations were shown to confer host range extension, as in the example of LT154/155SI. Further work with this mutant demonstrated that L154S was sufficient for the host range extension phenotype. T155I was not able to extend host range when introduced alone. Copied with permission from [Rainey et al., 2003].

interference. Pre-infection of QT6 cells with subgroup E virus did not cause receptor interference when challenged by the single or double mutants, suggesting that these mutants were able to enter cells by a receptor independent mechanism. On the other hand, pre-infection of DF1 cells with subgroup B virus caused significant interference when challenged with the mutants, suggesting that these viruses retained the ability to use the parental subgroup B receptor. Further work with these mutants showed that they were inactivated when treated with low pH prior to incubation with cells that did or did not express the appropriate receptor (Daniel Negusse's Thesis, unpublished work). This is in contrast to WT virus, which requires priming by binding a soluble receptor before it demonstrates sensitivity to low pH treatment.

The current model for the receptor-independent entry mechanism is that the host range extension mutants may non-specifically enter non-permissive cells through endocytic pathways, where acidification could lead to a fusion event in the absence of receptor-primer Env. This model is consistent the observation that mutants remain sensitive to the effects lysomotropic agents that prevent acidification of endocytic compartments. Additionally, R99, a peptide inhibitor previously shown to prevent HR1 and HR2 collapse in WT ALV subgroup A [Earp et al., 2003], was able to inhibit the LS and LSTI mutants as efficiently as WT, suggesting that the post entry fusion steps of mutants are similar to WT.

This brought up the question of why the LS and LSTI mutants, which retained the ability to use the cognate subgroup B receptor and had gained the ability to infect many other cell types, were not the wild-type form of the virus. The selective pressure applied by the passaging protocol which gave rise to these mutants suggests that the host range extension phenotype may come at the cost of fitness. The fitness differences between the WT, LS, and LSTI variants under the conditions of tissue culture are a topic of this dissertation.

1.6 Cytopathic Effect of Host Range Extension Mutants

Further work provided the interesting observation that the host range extension mutants, LS and LSTI, were more cytopathic than wild-type subgroup B virus in infections of the chicken cell line DF1. Infection of DF1 cells by LS or LSTI virus resulted in a greater that 50 fold reduction of viable cells 16 to 27 days post infection compared to only a 2 fold reduction with WT subgroup B virus [Rainey and Coffin, 2006]. The most likely mechanism of cytopathic effect of the mutant viruses is the accumulation of viral DNA in the host cell due to superinfection. This phenomenon has been observed for other retroviruses as well and is thought to lead the production of pro-apoptotic signals based on characterizations of nuclear morphology [Temin et al., 1980, Overbaugh et al., 1992, Mullins et al., 1986, Weller et al., 1980].

1.7 Retroviruses and Evolution

Evolution acts through the mechanisms of mutation, recombination, selection, and genetic drift. Mutation and recombination act to diversify populations through the introduction of new alleles or the reassortment of existing alleles. Selection acts against diversification through the unequal contribution of individuals to the next generation by better-fit or less-fit individuals. Genetic drift is the random sampling of alleles causing chance fluctuations in a population.

The complex interplay of mutation, selection, and drift has had a significant impact on the treatment of HIV infected individuals. The relationship between retrovirus populations and each of these evolutionary forces is discussed in the following subsections.

1.7.1 Mutation

Retroviruses are obligated to reverse transcribe their genomes as part of their replication cycle. This step requires the enzyme reverse transcriptase which is shown to be error-prone both as purified enzyme *in vitro* or in single round infections in tissue culture. The high mutation rate of reverse transcription is likely due to the lack of proofreading ability of the enzyme [Battula and Loeb, 1976]. The range of demonstrated error rates is 4.8×10^{-6} mutations per base-pair per cycle for bovine leukemia virus (BLV) [Mansky and Temin, 1994] to 1.4×10^{-4} mutations per base-pair per cycle for Rous sarcoma virus (RSV) [Leider et al., 1988]. HIV lies in the middle of this range at 3.4×10^{-5} per base-pair per cycle [Mansky and Temin, 1995]. However, these error rates were determined for particular sites in the genome. The error-rate of reverse transcriptase likely depends on the particular base of interest [Ricchetti and Buc, 1990, Klarmann et al., 1993] and is sensitive to sequence context and nucleotide pools [Skalka and Goff, 1993].

1.7.2 Recombination

The potential for recombination during the replication of retroviral genomes is well-studied. Recombination occurs as a consequence of copackaging of two different viral genomes in the same virion, which can occur when a single cell is coinfected with two different viruses [Stuhlmann and Berg, 1992, Hu and Temin, 1990]. Estimates for recombination rates are as high as 4% for the DNA reverse transcribed over a 1 kb region [Hu and Temin, 1990].

The copy choice model has been proposed to explain the recombination during reverse transcription. In this model, reverse transcriptase switches from one RNA strand to another RNA strand during minus strand DNA synthesis. The switching of strands, which can happen multiple times per replication cycle, results in a provirus

that is a mosaic of the original two RNA copies of the genome that were copackaged in the virion.

There are two proposed models for strand switching: in the forced copy choice model, reverse transcriptase encounters an RNA break that forces a switch to a homologous region on another RNA molecule [Coffin, 1979]. The second model does not require RNA damage [Xu and Boeke, 1987] and may be a function of sites on the genome that cause reverse transcriptase pausing [Pathak and Temin, 1990, Jones et al., 1994].

1.7.3 Selection and Fitness

Selection is the force of evolution that acts to increase the mean fitness of a population [CHARLESWORTH, 2000]. Selection is calculated from fitness and as such fitness is the more fundamental measurement. Fitness is a measure of an organism's ability to reproduce and contribute offspring to the next generation. When considering a retrovirus, differences in individual fitness can be the consequence of any number of factors. To name a few examples, a virus may contain mutations that change how efficiently its envelope binds to a receptor, how well it is expressed once integrated, or how stable a particle is once packaged. Combinations of these factors lead to viruses that are less-fit and will produce fewer infectious progeny than a virus that is better-fit.

Because it is comparisons between different viruses that are most useful, it is the relative fitness that is considered more often in the literature than individual fitness. The relative fitness of viruses is calculated by normalizing the growth rate of the variants under consideration to that of the fittest virus. In a population consisting of only two variants (such as the populations that are the subject of this study) the better-fit virus will always have a relative fitness of one and the less-fit virus will

have a relative fitness that is equal to the less-fit growth rate divided by the better-fit growth rate.

Another useful metric is that of the selection coefficient (s), which is the relative fitness of a virus - 1. Because selection coefficients are calculated using relative fitness values, it is important to note that a selection coefficient always indicates the replicative capacity of a virus in reference to another a virus. Additionally, selection coefficients only apply to a set of viruses when the viruses are considered in the same environment in which their relative growth rates were originally measured. Changing the context in which the viruses replicate, for instance changing the cell type the cells infect, will often result in different fitness values.

Selection coefficient can be estimated from time-series plots of viral competition experiments [Maree et al., 2000]. When the ratio of two viruses are plotted logarithmically over time, the slope is a measure of the absolute fitness difference between viruses. If we consider a virus that replicates with rate r and a mutant virus that replicates with rate r', their relationship to each other is $r' = (1 + s)r$, where s is the selection coefficient. The absolute fitness difference is $r' - r$ which is equal to rs. Thus, estimating the selection coefficient between viruses from competition experiment data requires knowledge of r. However, mathematical methods exit to estimate s without knowledge of r under the conditions of culture [Bonhoeffer et al., 2002].

1.7.4 Genetic Drift

Genetic drift is due to the random sampling of alleles in a population causing changes in allele frequencies due to chance alone. In the absence of genetic drift, the change in frequency of an allele in a population owes solely to mutation and selection. Populations that are small are more susceptible to genetic drift, and populations that are large are less susceptible to genetic drift. This makes population size an important

parameter when determining how to model a viral population - in the case of very large populations, drift can be ignored and deterministic models are more appropriate. In the case of small populations, stochastic models that include the effects of drift are more appropriate.

The choice of deterministic or stochastic models in HIV infection is controversial [Kouyos et al., 2006]. The number of infected cells in HIV infected individuals are estimated to be on the order of 10^7 to 10^8 cells [Haase et al., 1996]. This population size is considered large when compared to the inverse of the estimated mutation rate of HIV (around 10^{-5}, see Chapter 1.7.1), which means that 100 to 1000 point mutations are predicted to arise at every generation. However, the census population size is not the relevant parameter when supplying a population size to a particular model. A population with a large census size can exhibit the same stochastic behavior as smaller populations depending on the magnitude of variance in progeny number of infected cells, the frequency and intensity of selective sweeps and bottlenecks, and the degree to which the population is organized into metapopulations. Instead, the effective population size (N_e), which takes into account these phenomena, is the relevant parameter to supply to the model (Figure 1.9).

The effective population size (N_e) is a population genetics concept first introduced by Sewall Wright in the early 1930s as a way to rescale mathematical representations of populations so they recreate the behavior observed in natural populations [Wright, 1931]. Much of population genetics was derived from the study of a model population known as the Wright-Fisher population [Wright, 1931,Fisher, 1958]. This population consists of an finite number of randomly mating diploid individuals. The individuals create gametes which randomly encounter the gametes of the other individuals, with each encounter resulting in an offspring, such that the population is hermaphroditic and mating is random. Because each individual is equally likely to contribute offspring to the next generation, this means there is no fitness difference between individuals

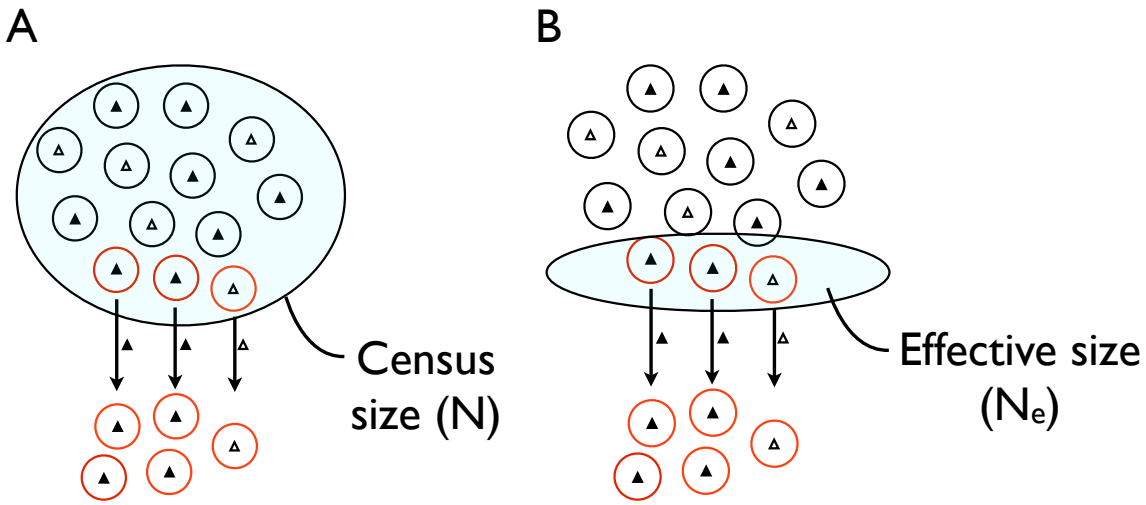

Figure 1.9: **Schematic representation of the difference between census and effective population size.** A) The circled population represents the census population size. In this case, all cells are considered in the size of the population even though only a small subset of cells (those cells that are red) contribute to the next generation. Models using this census size as a parameter will underestimate the role of genetic drift. B) The circled population represents the effective population size. Only the cells that contribute progeny to the next generation are considered.

and selection does not play a role. Generations are discreet, meaning there is no overlap between parent individuals and their offspring.

Natural populations often grossly violate the assumptions of the Wright-Fisher population. Determining N_e for natural populations can be difficult given that one would need to have complete knowledge of population structure, timing and magnitude of selective pressures, and the presence of bottlenecks, among other factors. Depending on the system of study, it may be impossible to observe the role these events play in the evolution of a population. However, estimates of N_e can be made by observing a characteristic of the population that is sensitive to drift and back-calculating the population size through a formula that incorporates the relevant biological details [Charlesworth, 2009]. An example of this method will be presented in the context of HIV populations in Chapter 1.9.1.

In conclusion, the role that genetic drift plays in viral evolution has been studied in many systems, both in the context of long transmission chains of many individuals and within-individual evolution. Genetic drift may play an important role in the dynamics of HIV infection *in vivo*.

1.8 Dynamics of HIV Infection *in vivo*

Evolution of HIV populations in infected individuals clearly affects clinical outcomes. How the forces of selection, mutation, and drift ultimately influence HIV evolution requires an understanding of the dynamics of HIV infection *in vivo*.

A typical HIV infection consists of three phases. First, an early phase that starts a few weeks after infection and is characterized by acute symptoms, high viremia, and large numbers of infected $CD4^+$ T cells. Next, this is followed by a clinically latent phase where the individual mounts an immune response that leads to very low levels of viremia and a gradual decline in $CD4^+$ T cells, a phase that can last for many

years. Finally, the immune system reaches a point of collapse and the individual rapidly returns to high viremia and exhibits symptoms of AIDS [Coffin, 1994].

A clear picture of how the low levels of viremia during the clinically latent phase could lead to the collapse of the immune system was not possible until the steady state infection of clinical latency was disrupted by the first generation of HIV inhibitors [Wei et al., 1995, Ho et al., 1995]. HIV-infected patients treated with these inhibitors were monitored by RT-PCR based assays that allowed sensitive quantification of viral load [Mulder et al., 1994, Piatak et al., 1993]. Treatment with nucleoside reverse transcriptase inhibitors, non-nucleoside reverse transcriptase inhibitors, and protease inhibitors all showed similar results (Figure 1.10). After administration of these inhibitors, the plasma viral load drops rapidly, coincident with a rise in $CD4^+$ T cells. However, this drop is only transient: in all cases the viral load rebounds due predominantly to drug resistant mutants [Wei et al., 1995, Richman et al., 1994].

The kinetics with which infected cells and free virus decline after treatment with inhibitor allow important inferences about the dynamics of infection to be made (Figure 1.11). The steady state observed at clinical latency is actually a very dynamic process where there is rapid turnover of virus and infected cells. The average lifetime of infected cells is estimated to be around 1-2 days, based on the 100-fold decay in detectable virus over a 2 week period following the start of treatment. Based on these estimates, the viral population undergoes approximately 300 replication cycles per year. Circulating virus is produced from recently infected cells and infection drives a rapid turnover of $CD4^+$ T cells, where approximately 10^9 cells die and are replaced each day [Coffin, 1994, Perelson et al., 1996].

The remarkably short time it takes for drug resistant mutants to arise after initiation of therapy implies that drug resistant mutants pre-exist in the population. Taken together with the high mutation rate for HIV (estimated to be on the order of 10^{-5} mutations per base pair per generation) and the high turnover of HIV infected

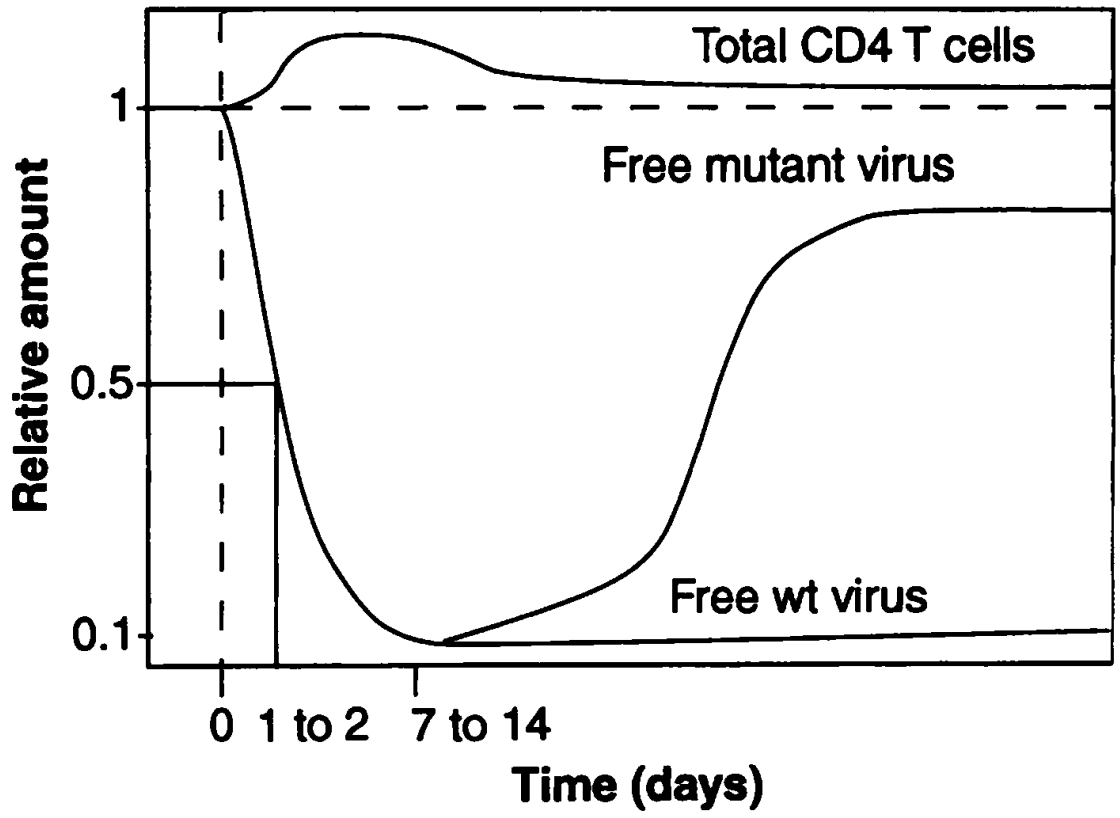

Figure 1.10: **Kinetics of viral and CD4$^+$ T cells in HIV infected individuals.** Individuals treated with HIV inhibitors exhibit changes in viral load and CD4$^+$ T cells with similar kinetics. After the initiation of continuous treatment with drug, indicated with vertical dashed line, plasma viral load decreases and CD4$^+$ T cells number increases. Plasma viral load always rebounds due to the rise of drug resistant mutants. Copied with permission from [Coffin, 1994].

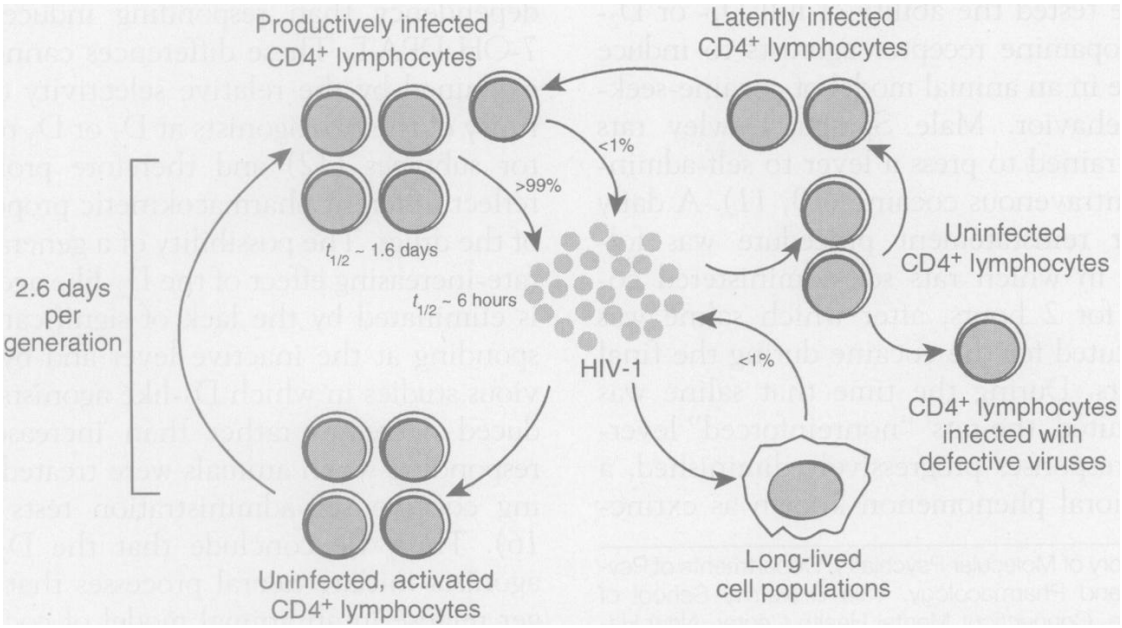

Figure 1.11: **Summary of HIV Infection Dynamics *in vivo*.** The clinically latent phase of HIV infection is characterized by high turnover of infected cells (2.6 days / generation) and virus. The majority of virus (99%) in infected individuals is released from recently infected $CD4^+$ T cells. Latently infected and long lived cell populations contribute only a small percentage (¡1%) of the total virus population. Copied from with permission from [Perelson et al., 1996].

cells suggests that every single point mutation occurs 10^4 - 10^5 times per day. These characteristics have made HIV populations the subject of modeling.

1.9 Effective Population Size of HIV infection

Modeling efforts in the HIV field have been extensive. To date, there have been multiple efforts to make predictions of the expected frequency of drug-resistance mutations before patients are treated with drugs [Ribeiro and Bonhoeffer, 2000, Ribeiro et al., 1998, Bonhoeffer and Nowak, 1997, Bonhoeffer et al., 1997, Rouzine et al., 2001], how long it takes for drug resistance mutations to rise after administration of drugs [Bonhoeffer et al., 1997, Bonhoeffer and Nowak, 1997, Rouzine and Coffin, 2005], and how long it takes for immune escape variants to arise [Liu et al., 2006]. It is important to note that all of these predictions require models in which the population size of HIV is parameter.

1.9.1 How Estimates for N_e Are Made for HIV Infections

Estimates of the effective population size (N_e) are typically made by measuring another value, a calibration quantity, that is under the same forces of evolution as the value one wants to predict (reviewed in [Kouyos et al., 2006]). After a measurement for the calibration quantity is made, an estimate for N_e can be made by determining the population size that one must supply to the model to generate the same value as the observed calibration quantity (Figure 1.12). Most studies that estimate N_e for HIV infection use models based on neutral evolution and measure sequence diversity in *env* [Shriner et al., 2004, Achaz et al., 2004, Brown, 1997a, Nijhuis et al., 1998, Rodrigo et al., 1999, Seo et al., 2002]. These estimates place the N_e on the order of 10^3. One criticism of this approach is that *env* is not evolving neutrally, based on high nonsynonymous to synonymous substitution ratios observed in *env* [Bonhoeffer et al.,

1995, Seibert et al., 1995, Yamaguchi and Gojobori, 1997]. This would have the effect of underestimating the N_e.

A single study developed a test incorporating selection, based on observing patterns linkage disequilibrium, and estimated N_e on an order of magnitude of at least 10^5 [Rouzine and Coffin, 1999]. These estimates vary widely and have sparked debate as to what the appropriate size of N_e is in HIV infection [Kouyos et al., 2006].

The study performed by Rouzine, et al., developed a model that incorporates the effects of selection through a one locus, two-allele model [Rouzine et al., 2001](Figure 1.13). This model dealt with selection and mutation at one site in the genome of the retrovirus. This site has a choice between two alleles: one allele is better-fit and the other allele is less-fit. If a mutation event occurs within a genome, this genome switches from a better-fit variant to a less-fit variant, or vice-versa. As noted before, recombination need not be explicitly dealt with in this model due to the fact that recombination between a better-fit genome and a less-fit genome will simply cause them to swap identities. This model is used to generate the simulations used in this thesis.

1.9.2 Assays for HIV Diversity

Many HIV modeling efforts rely on sequence data from HIV infected individuals. A variety of techniques are used to gain large amounts of sequence data, or monitor specific sequences over longitudinal samples. These assays are described in the following sections.

Bulk Sequencing

Bulk sequencing involves generating an RT-PCR product from a pool of viral RNA, usually extracted from patient plasma. The bulk DNA product is sequenced

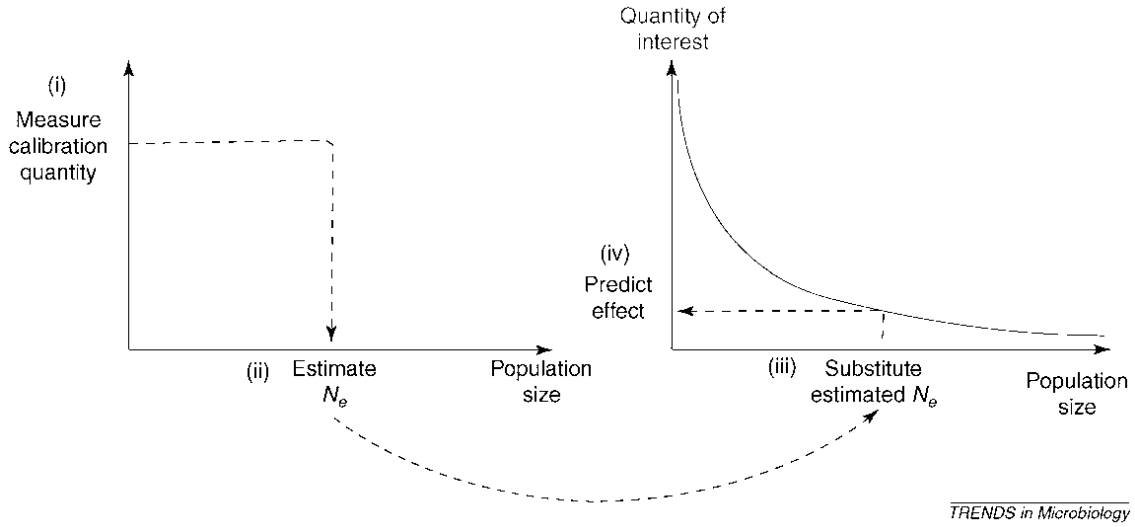

Figure 1.12: **Estimating N_e through a calibration quantity.** An illustration of the concept of and use of N_e. The estimation of N_e is based on an idealized mathematical model in which population size is a parameter. To estimate a quantity of interest that is difficult to measure in a natural population, the approach is taken as follows: (i) One decides on an easily measurable quantity (calibration quantity) such as genetic diversity. (ii) N_e is defined as the population size for which the calibration quantity in the idealized model has the same value as that measured in the natural population. (iii) N_e is then substituted in the model to predict the value of the quantity of interest in the natural population (iv). Copied with permission from [Kouyos et al., 2006].

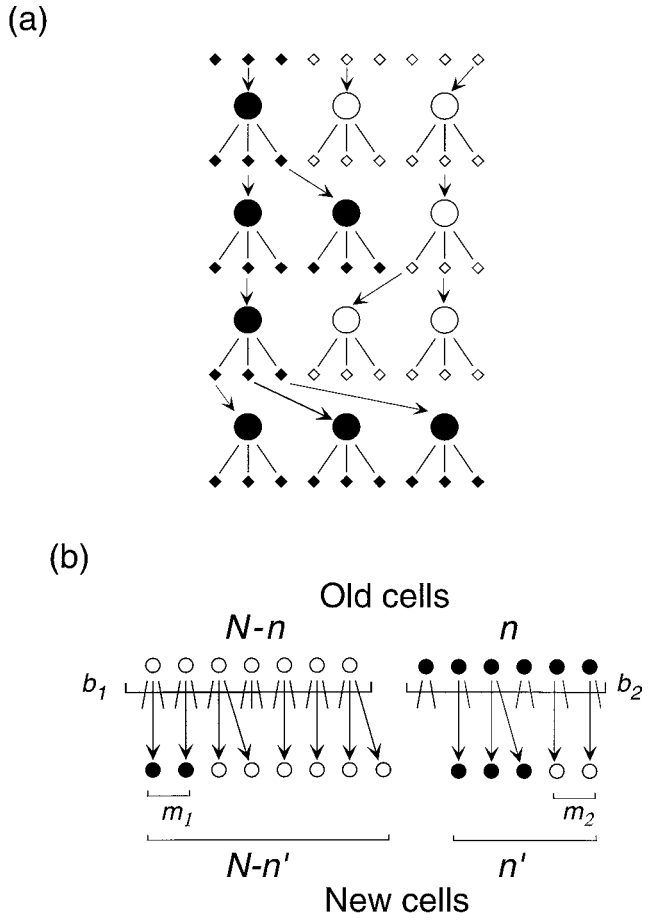

Figure 1.13: (a) Drift of genetic composition due to random sampling of infecting virions. Circles denote infected cells, and small diamonds show free virus particles. Black and white denote virus genetic variants. (b) Full virus population model including random drift, selection, and mutation. Two consecutive generations of infected cells are shown. Lines radiating from a cell denote virions, some of which, as shown by arrows, infect new cells. Mutant cells yield fewer progeny per cell. A small fraction of infecting virions, m_1 and m_2, mutate to the other variant. (Figure and caption copied with permission from [Rouzine et al., 2001].)

[Günthard et al., 1998] and the presence of mutations are determined by presence of mixed bases in the chromatogram. This technique requires that the mutation of interest be a sufficient fraction of the population for detection (around 10% [Hance et al., 2001]) and does not give any information about linkage between mutations.

Allele-Specific PCR

Allele specific PCR is identical to quantitative PCR (qPCR) except primers are designed to discriminate between sequences that differ by as little as one base pair [Palmer et al., 2006]. This is done by introducing mismatches in a primer that anneals at the variable site, increasing the specificity of the reaction at the cost of efficiency. This technique cannot be used to discover the presence of new mutations as it requires prior knowledge of the sequence to generate the appropriate primers. Allele specific PCR was used extensively for this thesis and is therefore discussed further in Chapter 3.1.

Single Genome Sequencing

Single genome sequencing was developed to allow sensitive detection of arbitrary viral variants [Palmer et al., 2006]. The technique relies on limiting dilution of viral templates followed by PCR amplification. Many replicates of each member of a dilution series are PCR amplified, and the dilutions that generate approximately 30% positive PCR reactions are likely to contains product derived from a single genome as template with an 80% probability [Palmer et al., 2005]. The advantages over bulk sequencing are increased sensitivity and information about linked mutations.

Next Generation Sequencing

Finally, next generation sequencing encompasses the current technologies that allow massively parallel sequencing. The most commonly used sequencing technology in the

HIV field is pyrosequencing, which sequences single immobilized fragments of DNA by detecting the release of pyrophosphate during synthesis [Wang et al., 2007]. The advantage of this technique are the large volume of sequences that can be obtained in a single sequencing run allowing detection of rare variants. Sequencing reads are currently limited to 300-500 base pairs and subject to a high background mutation rate that may complicate detection of very rare frequency mutations.

1.10 Overview and Aims of this Thesis

Retroviruses exhibit an extraordinary ability to adapt to selective pressures. This characteristic has made all efforts to cure HIV infection unsuccessful, because the virus evolves to resist the host immune system and antiretroviral therapy. Moreover, efforts to describe the evolution of HIV *in vivo* are further complicated by a lack of information about the initial population at the time of transmission and the apparent complexity of the infection. The population most likely faces variable selective pressures, segregates into separate biological and spatial compartments, and faces bottlenecks of unknown intensity. These properties of the viral population have led to the question, what is the effective population size of HIV infections?

Estimates of the effective of population of HIV rely on models of evolution that may oversimplify the system. In order to address the issue, we designed a tissue culture based system that came as close as possible to meeting the assumptions of a simple model of evolution. This system used the chicken cell line DF1 and ALV subgroup B virus. Previous work in our lab led to the identification [Taplitz and Coffin, 1997] and characterization [Rainey et al., 2003, Rainey and Coffin, 2006] of mutants of ALV with extended host range. Here, we utilize these mutants to perform pair-wise competition experiments in a simple tissue-culture system that provides constant selective pressure in well-mixed conditions.

Measurements were then made to define the initial conditions of the infection and characterize the parameters important to the evolution of the population. Finally, we compare the results of the tissue culture experiments with the results of simulation based on a model where most the assumptions are tightly held by the experimental protocol. We determine the relationship between the effective population size and the census population size in this system.

Chapter 2

Materials and Methods

2.1 Routine Tissue Culture

DF-1 cells are a spontaneously immortalized chicken cell line [Schaefer-Klein et al., 1998,Himly et al., 1998]. DF1 cells were supplied Dubelcco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum and penicillin/ streptomycin. All cultures used in this study were supplemented to contain serum from a single lot. Cells were incubated at 37 °C and 5% CO_2. Viable cells were counted by trypan blue exclusion and a haemocytometer. Cells were maintained in either 60 mm or 100 mm dishes and subcultured when the cultures became 100% confluent (approximately every 3-4 days).

2.2 Transfections and Virus Harvest

Transfections were performed by lipofectamine reagent (Invitrogen) according to the manufacturer's directions. DF1 cells were plated in 60 mm dishes at a concentration that would result in 70-80% confluent cultures after 24 hours of incubation. DNA/Plus complexes were created by mixing 0.25mL Dulbecco's Modified Eagle

Medium, 10 μg plasmid DNA, and 10 μL Plus reagent. This mixture was incubated for 15 minutes. 15 μL Lipofectamine was diluted in Dulbecco's Modified Eagle Medium, mixed with the DNA/Plus mixture, and incubated for 15 minutes. Medium was removed from DF1 cultures, 0.5 mL DNA/Plus/Lipofectamine mixture was overlaid on the cells, and the culture was incubated for 2-5 hours at 37°C and 5% CO_2 after which 4 mL Dulbecco's Modified Eagle Medium was supplied to the culture. Transfected cultures were left to incubate for 48 hours and then split in a 1:6 ratio and left to incubate for another 72 hours.

Virus was harvested by pipetting supernatant from cultures and passing the supernatant through a 0.22 μm filter (Millipore) to remove DF1 cells. Harvested supernatants were either used immediately for infections or stored at 4°C.

2.3 Construction of GFP-Labeled LS and LSTI Molecular Clones

A 1075 base pair fragment of the SU region of *env* was generated by digesting plasmid WT-PrB-GFP with SalI and KpnI. This fragment was gel purified and ligated into pGEM-T Easy Vector (Promega, Madison, Wisconsin, USA) to generate plasmid WT-PrB-GFP-Tvec. Site-directed mutagenesis was performed using forward primer mutPrB-LS-sense (3'-AGACAGGTTACTCACATCCTTTCGACCGACCCAGG-5') and reverse primer mutPrB-LS-antisense (3'-TCTGTCCAATGAGTGTAGGAAAGCTGGCTGGGTCC-5') to introduce the LS mutation to SU. Forward primer mutPrB-LSTI-sense (3'-GAGACAGGTTACTCACATCCTTTCGATCGACCCAGGGAA-5') and reverse primer mutPrB-LSTI-antisense (3'-CTCTGTCCAATGAGTGTAGGAAAGCTAGCTGGGTCCCTT-5') were used

to simultaneously introduce the LS and TI mutations to SU. Reaction conditions for one reaction were 2.5 μL 10x reaction buffer, 4 μL WT-PrB-GFP-Tvec, 1.25 μL forward primer, 1.25 μL reverse primer, 0.5 μL dNTP mix, 0.5 μL Pfu Ultra enzyme, and 15 μL H_2O. Cycling conditions were 1 cycle at 95°C for 30 s followed by 17 cycles at 95°C for 30 s, 52°C for 1 min, and 68°C for 5 min. PCR products were digested with DpnI for 2 hours, to digest non-mutagenized template DNA, and 2.5 μLdigested product was used to transform E. coli DH5α. Mutagenesis reactions were verified by sequencing the PCR products using primer RCG1 (3'-GATGTCACCCAAAAGGATGAGG-5').

The mutagenized 1075 base pair fragment of SU was digested from pGEM T-Easy using SalI and KpnI, the desired fragment was gel purified. The desired SalI/KpnI digested fragment from WT-PrB-GFP was treated with calf intestinal phosphatase (CIP) and gel purified. The mutagenized SU fragments were ligated to the RCAS backbone under the following reaction conditions: 1 μL RCAS backbone, 4 μL mutagenized insert (either LS or LSTI), 0.5 μLT4 ligase, and 4.5 μLH_2O. Ligation reactions were incubated at 16°C overnight. 2.5 μLof ligatation product was used to transform E. coli DH5α by electroporation.

2.4 Flow Cytometry

24 hours before infection, 3×10^5 cells were seeded as replicate cultures in 6-well plates. On the day of infection, 0.5 mL supernatant harvested from transfected cells was applied to cells aspirated of medium and washed with PBS. After 1 hour, infectious supernatants were aspirated, the cultures were washed with PBS, and 2 mL medium was supplied to the cells. One replicate culture was not infected and instead used to count viable cells on the day of infection by trypan blue exclusion and a haemocytometer.

48 hours post infection, medium was aspirated from tissue culture vessels and cells were washed with PBS before trypsinization for 5 minutes at 37°C. Cells were resuspended in PBS for analysis by flow cytometry. Live cells were then gated by forward scatter and side scatter. An uninfected culture was used to gate for GFP-negative cells. All flow cytometry measurements were made by counting 10,000 events. Viral titers were calculated using the following formula:

$$\frac{\text{IU}}{\text{mL}} = \frac{1}{\text{d}} \times \frac{1}{\text{f}} \times \text{cells at time of infection} \times \frac{\text{GFP-positive cells by FACS}}{\text{live cells by FACS}} \qquad (2.1)$$

Where d is the dilution factor and f is the fraction of 1 mL used to infect the culture.

2.5 Passaging Virus in Culture

Infections were initiated by measuring the RNA copy number of supernatants by real-time PCR. First, serial dilutions were made of mutant viruses (LS and LSTI) and defined mixtures of WT and LS or LSTI were made. Next, all infections were initiated with 0.5 mL mixture containing 5 $\times 10^9$ RNA copies of less-fit virus (LS or LSTI) and either 5 $\times 10^6$, 5 $\times 10^5$ or 5 $\times 10^4$ copies of better-fit virus (WT).

Experiments are referred to by their RNA copy ratio of less fit to better fit. For example an experiment initiated with and LS:WT RNA copy ratio of 10^3:1 was named LS103WT. Infections were performed at 37°C and 5% CO_2 for 1 hour at which time the cultures were washed with PBS and supplied with 2 mL of culture medium.

Virus containing supernatants were passaged twice per week. One passage was 3 days long and the other 4 days long. At the end of each passage, culture supernatant was harvested by pipetting and filtering through a 0.22 μm filter. 0.5 mL of filtered

culture supernatant was used to initiate the infection for the next passage, the remaining filtrate was reserved for population frequency quantification. All infections were initiated with a target cell density of 3×10^5 cells / culture.

2.6 RT-PCR and Allele-Specific PCR

Viral RNA was extracted from harvested supernatants at each passage using the Qiagen Viral Mini Kit (Qiagen, Valencia, California, USA) according the the manufacturer's protocol. RNA was stored at -20°C until used for quantification by allele-specific PCR.

2.6.1 Generation of RNA Standards

A 913 base pair fragment of SU was amplified from plasmid WT-PrB-GFP using primers T1F (5'-AAGACCCGGAGAAGACACCC-3') and T1R (5'-AGAATCGTGATCGGTTCTCC-3'). PCR amplification was performed for 1 cycle at 95°C for 1 min followed by 41 cycles of 94°C for 30 s, 50°C for 30 s, 72°C for 1 min, and a final single cycle of 72°C for 10 min. The DNA product was ligated to pGEM-T Easy Vector (Promega, Madison, Wisconsin, USA) according to the manufacturer's protocol. DH5α were transformed with the ligated product by electroporation described before.

RNA standards were created by *in vitro* transcription from plasmid WT-PrB-TVec using the Ambion MEGAscript reverse transcription kit (Applied Biosystems/Ambion, Austin, Texas, USA). WT-PrB-TVec was linearized by SpeI digestion. RNA was quantified spectrophotometrically by absorbance at 260 nmand using an average ribonucleotide monophosphate molecular weight of 339.5 g/mol to convert to RNA copies. Stocks of 10^{12} copies/μL were aliquoted and stored at -80°C until needed to generate RNA standards the same day as the real-time PCR.

2.6.2 Generation of DNA Standards

DNA standards were generated by quantifying plasmid preparations of WT-PrB-GFP spectrophotometrically and generating 10-fold serial dilutions. Standards ranged from 10^2 to 10^8 copies/μL.

2.6.3 Allele-Specific Primer Validation

Primers for allele-specific PCR amplification were validated for specificity by generating 10-fold serial dilutions from 10^2 to 10^7 copies of the target template of interest, either WT, LS, or LSTI, and spiking each of the reactions with 10^7 of non-specific template. For instance, the WT-specific primer was validated by generating PCR reactions that contained 10^2 - 10^7 copies of WT plasmid and each reaction was spiked with either 10^7 copies LS plasmid, 10^7 copies LSTI plasmid, or H_2O (not spiked with a competitor).

2.6.4 First Round cDNA Synthesis and RNA Quantification

RNA was extracted from supernatants harvested from passaging experiments using the Viral Mini Kit (Qiagen, Valenica, California, USA) according to the manufacturer's protocol. RNA was eluted in 60 μL buffer AVE. 5 μL of eluate was used as template for RT-PCR using the Applied Biosystems RNA-to-C_t kit (Applied Biosystems, Carlsbad, California, USA). The reaction condition for each reaction was 5 μL eluant from RNA extraction, 1.25 μL primer RT1F, 1.25 μL RT1R, 12.5 μL PowerSYBR enzyme, 0.2 μLreverse-transriptase enzyme, and 2.3 μL H_2O. Cycling conditions were 1 cycle of 50°Cfor 30 min then 95°C for 15 min followed by 51 cycles of 94°Cfor 30 s, 50°C for 30 s, 72°C for 1 min, 76°C for 1 s, and reading of the plate.

All reactions were performed in triplicate along with controls that lacked reverse transcriptase or template. The RNA copies in the fraction of passage transferred

from an infected culture to an uninfected culture was calculated with the following formula:

$$\frac{\text{RNA copies}}{\text{passage}} = \frac{\text{RNA copies}}{\text{eluant for qPCR}} \times \frac{\text{total eluant volume}}{\text{vol of sup used to prep RNA}} \times \frac{\mu\text{L}}{\text{passage}} \quad (2.2)$$

2.6.5 Second Round Allele-Specific qPCR

cDNA product from first round qRT-PCR was diluted 10,000-fold (to approximately 10^7 copies/mL) and supplied as a template to a second-round allele-specific PCR (AS-PCR). The forward primer for all allele-specific PCR reactions was F5723 (5'-CTCACCTATCGGAAGGTTTCATG-3'). The reverse primers varied by the desired specificity at each mutant site. For wild-type specific amplification at the LS site, the reverse primer was WT-AS1 (5'-TTGTTCCCTGGGTCTGTGA-3'). For mutant-specific amplification at the LS site, the reverse primer was LS-AS2 (5'-TTGTTCCCTGGGTCGGTGG-3'). For wild-type specific amplification at the TI site, the reverse primer WT-AS2 (5'-GGATTGTTCCCTGGGTCCG-3'). For mutant-specific amplification at the TI site, the reverse primer TI-AS1 (5'-GGATTGTTCCCTGGGTCCA-3').

The reaction condition for each reaction was 2.5 μL forward primer F5723, 2.5 μL reverse primer (WT-AS1, WT-AS2, LS-AS2, or TI-AS1 depending on the desired specficity), 12.5 μL SYBR enzyme, and 6.5 μL H_2O. Cycling conditions were 1 cycle at 95°C for 1 min followed by 51 cycles of 94°C for 30 s, 51°C for 45 s, 72°C for 1 min, 76°C for 30 s, followed by reading of the plate. As before, all reactions were performed in triplicate along with controls without template to exclude contamination.

Frequencies of variants were calculated with copy numbers determined from reactions run on the same plate using the same master mixes. For example, to determine

the frequency of WT in a competition against LS, both WT-specific reactions and LS-specific reactions were performed on the same plate using the same master mixes.

2.7 ddI Inhibition and Cytotoxicity

2.7.1 ddI Inhibition of Viral Replication

Supernatants containing WT, LS, or LSTI virus were harvested from transfections of DF1 cells and passed through a 0.22μm filter. 3×10^5 DF1 cells were plated in 6-well plates 24 hours before infection. Infections were performed by aspirating medium from cells, washing the cells with PBS, and incubating the cells with filtered infectious supernatants for 1 hour at 37°C and 5% CO_2. After adsorption, unbound virus was aspirated and washed from the cultures and fresh medium containing 0.01μM to 100μM 2'-3'-dideoxyinosine (ddI) was added in 10-fold increments to replicate cultures. The frequency of infected cells was scored 48 hour post infection by flow cytometry.

2.7.2 ddI Cytotoxicity on DF1 Cells

3×10^5 DF1 cells were seeded in each well of a 6-well plate. 1.5mL of medium containing ddI (0.01μM to 100μM, in 10-fold increments) was added to individual cultures. 72 hour post-plating, viable cells were scored by trypan blue exclusion and a haemocytometer.

2.8 Measurement of Initial MOI and Census Population Size

The initial infectious mixtures of better-fit and less-fit virus were created based on measurements of viral RNA copy number (see section 2.5). Total viral RNA was quantified at subsequent passages. The relationship between viral RNA copy number and viral titer was determined by harvesting WT, LS, or LSTI virus containing supernatants from transfected cells, generating 10-fold serial dilutions of each (from undiluted to 10^{-3}), measuring the titer of each serial dilution by FACS (see section 2.4), and RNA quantification by qPCR (see section 2.6.4).

Using these results, a standard curve was generated that compares viral RNA copy number to viral titer expressed as infectious units per mL (IU/mL). Log transformations of RNA copy number and viral titer were plotted on the X and Y axes, respectively. The relationship between the two values was estimated using an ordinary least squares linear regression (Figure 3.5).

The total number of functional viral particles at each passage (the census population size) was estimated using the following formulas:

$$\text{WT Particles} = 10^{(\log \text{WT RNA Copies} \times .7716) - 1.099} \tag{2.3}$$

$$\text{LS Particles} = 10^{(\log \text{LS RNA Copies} \times .6251) - 1.527} \tag{2.4}$$

$$\text{LSTI Particles} = 10^{(\log \text{LSTI RNA Copies} \times .6251) + 0.4936} \tag{2.5}$$

Individual measurements of the WT, LS, or LSTI RNA copy numbers at each passage were estimated by multiplying the total RNA copies measured multiplied by

the frequency of each species in the population.

2.9 Measurement of Target Cell Generation Time

3×10^5 DF1 cells were seeded in replicate cultures and grown under the conditions described in section 2.1. At 24, 48, 72, and 96 hours post-seeding, cultures were trypsinized and assayed for viable cells by trypan blue exclusion. The results were fit to the exponential growth model:

$$Y = Y_0 \times e^{(kX)} \tag{2.6}$$

where Y_0 is the initial number of cells, Y is the number of cells at time X, and k is the rate constant. The doubling time (T_d) was calculated as:

$$T_d = \frac{ln(2)}{k} \tag{2.7}$$

2.10 Measurement of Viral Generation Time

The method of determining the viral generation time is described in in Figure 3.10. $3*10^5$ cells were seeded in 6-well plates and incubated overnight. Virus was titered and used for a 1 hour infection after which unbound virus was washed with PBS. Medium was replaced on cells and cultures were incubated. Supernatants were replaced with medium containing 100μM dideoxyinosine (ddI) every 4 hour for 96 hours to halt viral replication. After 96 hours had elapsed, each culture was trypsinized, resuspended in PBS, and used for FACS analysis.

Medium containing ddI was serially diluted to generate a series of ddI concentrations from 0.01 μM to 100 μM in 10-fold increments. Virus from fresh transfections was incubated for 1 hour, washed with PBS, and then replaced with ddI contain-

ing medium. The cultures were incubated for 48 hour at 37°C and 5% CO_2. Cells were trypsinized and resuspended in PBS. The frequency of infected cells was scored by flow cytometry. Changes in the frequency of GFP-positive cells represent new rounds of infection. Even though the absolute number of GFP-positive cells increases with division of infected cells, the frequency must increase only when new rounds of infection are complete.

2.11 Measurement of Selection Coefficient

Due to the fact that generation times for WT, LS, and LSTI variants are essentially equal (see Chapter 2.10), we were able to use a measure of the relative growth rates of each virus to calculate the selection coefficients for each pairwise competition. The selection coefficient for each virus in a pair-wise competition was calculated with the formula:

$$s = \frac{n'}{n} - 1 \tag{2.8}$$

Where s is the selection coefficient and n' and n are the relative growth rates of the less-fit and better-fit variants, respectively. The relative growth rates for each virus ($\frac{n'}{n}$) was determined by mixing uninfected cells with transfected cells and measuring the frequency of GFP-positive cells at regular intervals post mixture.

2.12 Simulation

To define the theoretical expectations, simulations were constructed using the Ruby programming language. The complete source code for the simulation is provided in Appendix D. The models for the simulations are qualitatively described in Figure 3.19.

Chapter 3

Results

The large number and rapid turnover of infected cells in HIV infected individuals implies that drug resistant mutants may preexist in the population at a low frequency [Coffin, 1995]. The estimated production rate of virus is 10^{11} new viral particles each day [Perelson et al., 1996], a population size that is large enough to imply deterministic evolution of the virus upon administration of an inhibitor. For some types of drugs, reproducible timings in the appearance of drug resistance mutations between patients are observed [Schuurman et al., 1995]. However, for other types of inhibitors, there is significant variability in the time it takes for resistance mutations to appear, indicating that some stochastic factors may be in play [Richman et al., 1990, Richman et al., 1994].

It has been suggested that the effective population size (N_e) of HIV infection may be much lower than the census population size (N). The value N_e takes into account factors that can make a large population more susceptible to the random sampling of alleles, and thus more sensitive to drift than predictions based on census population size (N). An example factor that can affect the effective size of the population is variance in the number of progeny from infected cells; a small proportion of infected cells may produce the bulk of the infectious population. If this is the case, chance

events that occur on this minority population could have drastic effects on the genotypic makeup of the circulating viral population. Another factor may be the spatial separation of populations of infected cells, which can alter access of progeny virus to susceptible target cells. It is unclear how the complexities of HIV infection influence the effective size of the viral population.

In order to address the relationship between N_e and N, we generated a model tissue culture system using avian leukosis virus (ALV). This system simplifies the infection by removing the complexities of heterogenous cell types and spatial structuring of HIV infections.

Our strategy to quantify the role of stochastic effects in this system was to initiate infections with different known frequencies of better and less fit virus in the population for pairwise competitions. We measured the frequency of the better fit virus over many replication cycles by serially passaging the virus population onto fresh uninfected cells and terminated cultures when the better-fit virus became nearly fixed in the population. We quantified the role of stochastic effects by calculating the variance in the time it took for the better-fit virus to dominate the population between replicate cultures.

For these experiments the better-fit virus is always wild-type subgroup B ALV and referred to as WT. The less fit viruses are a single point mutant and a double point mutant of ALV subgroup B known to confer a host range extension phenotype. These mutants are referred to as LS and LSTI for the amino acid changes the mutations confer. LS differs from WT by a single nucleotide change and LSTI differs from WT by two nucleotide changes.

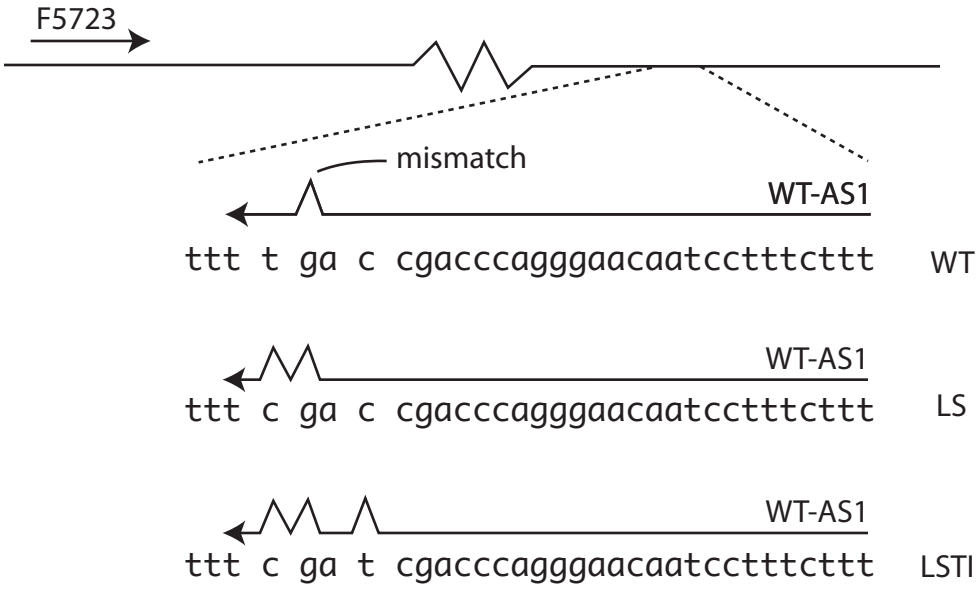

Figure 3.1: **Allele specific PCR primer design strategy.** Primers were designed to discriminate between templates that differed by a single base pair in a quantitative PCR reaction. A common forward primer, F5723, was used in all reactions. Reverse primers contained one mismatch at the penultimate site of the 3' end of the desired template. Reverse primers contained two mismatches against the 3' end of competitor template.

3.1 Allele-Specific PCR Strategy

A quantitative PCR based strategy was designed to determine the frequency of WT, LS, and LSTI variants in culture supernatants. The basic strategy was previously described in [Palmer et al., 2006] and schematically described in Figure 3.1. Since each variant in competition experiments differed by as little as a single base pair, primers designed with 100% complementarity to the target regions of the viral genomes did not provide enough specificity to discriminate between the two variants (data not shown). Instead, a mismatch was introduced at the penultimate residue of each reverse primer of each primer pair. All reactions used a common forward primer (F5723) that annealed at a site with identical sequence in all viruses used for this study.

The primers were validated for specificity by comparing the sensitivity of quantitative PCR reactions against the desired template in the background of excess undesired template. The cycle threshold, defined at a fluorescence intensity at which all PCR reactions have exponentially increasing product, was determined for this panel of reactions. The WT-specific primer sensitively detected WT template at a frequency of 10^{-3} in a background of 10^7 copies of either LS or LSTI template (Figure 3.2 A). WT frequencies of 10^{-4} and 10^{-5} crossed the cycle threshold at the same point as a WT frequency of 10^{-3}, indicating our limit of detection. The LS-specific and LSTI-specific primers both detected the correct templates at a frequency of 10^{-5} in the background of 10^7 copies of WT template (Figure 3.2 B and C).

The frequency of the variants at each were quantified using the following formula:

$$\text{frequency better-fit variant} = \frac{\text{better-fit RNA copies}}{\text{better-fit RNA copies} + \text{less-fit RNA copies}} \quad (3.1)$$

3.2 Viral Competition Results

To observe the role that genetic drift contributes in the evolution of a retrovirus population, we developed a tissue culture based system to continuously maintain a replicating virus population. We performed competition experiments between two viruses, initiating infections with defined mixtures of WT and LS virus or WT and LSTI virus. For each type of competition experiment we initiated infections with three different starting ratios WT to mutant virus, normalized by RNA copies of each virus. The RNA copies in the final solution were either 10^3:1, 10^4:1, or 10^5:1 for LS or LSTI to WT. The total viral RNA copies in the infectious mixture was approximately 10^9. Each experiment was named according to the ratio of mutant to WT RNA copy number, e.g. the experiment with LS to WT starting ratio of 10^3:1

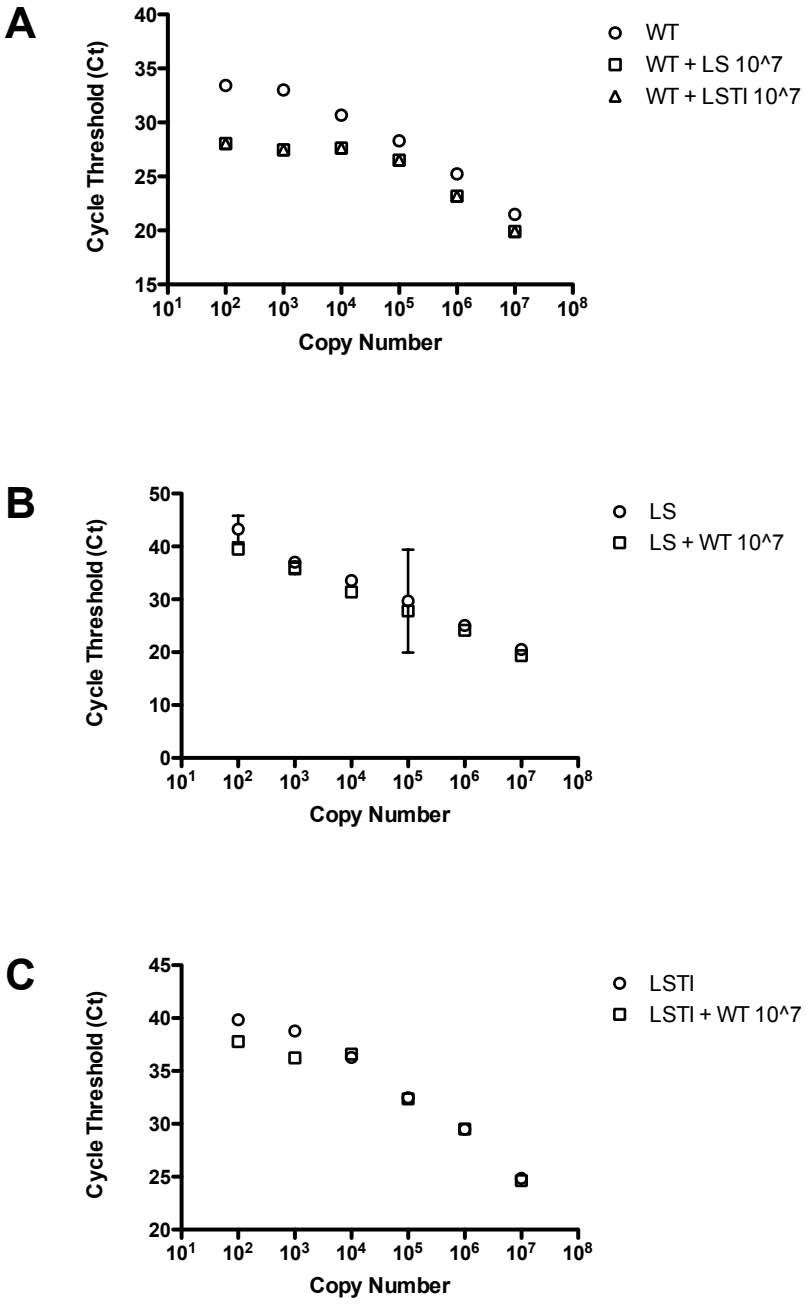

Figure 3.2: **Validation of allele specific primers.** Allele specific primers were validated by measuring the number of PCR cycles for reactions to reach a fluorescent threshold (cycle threshold). Forward primers for all reactions was F5723. The reverse primers were: A) Primer WT-AS1, specific for WT sequence at the LS site. B) Primer WT-AS2, specific for the WT sequence at the TI site. C) LS-AS2, specific for mutant sequence at the LS site. Error bars represent the standard deviation of triplicate samples.

was named LS103WT.

After the initiation of infections on DF1 cells, we washed away unbound virus with PBS before adding fresh medium and incubating the cultures. 25% of the virus in the supernatant was filtered from the cultures and used to initiate a fresh rounds of infections. This passaging protocol was repeated, alternating every 3 or 4 days, such that the virus population was passaged twice per week (Figure 3.3). Each experiment was performed in 6 replicates and the frequency of WT and the frequency of the mutants at each passage were quantified using the following formula:

$$\text{frequency better-fit variant} = \frac{\text{better-fit RNA copies}}{\text{better-fit RNA copies} + \text{less-fit RNA copies}} \quad (3.2)$$

The results of the competition experiments are presented in (Figure 3.4). LS103WT, LS105WT, and LSTI103WT experiments each had 6/6 replicate cultures successfully initiate competitions with WT wins as the outcome. LS104WT, LSTI104WT, and LSTI105WT experiments each had 5/6 replicate cultures successfully initiate competitions with WT wins as the outcome. The sixth replicate in these cases did not have detectable virus after 1 passage and were excluded from this study. The intensity of stochastic effects for each experiment was determined by measuring the variance in the time it took for WT to become 50% of the population. This was chosen as our reference point to measure variance because it represents the point at which the wild-type and mutant specific primers have the highest combined sensitivity (Figure 3.2). Additionally, once the WT frequency reached 50%, it did not drop below this value in any of the replicates, making it a good measure for the shift to the nearly monomorphic WT state. In experiments where WT was in competition against LS, we observed a dependance of both the mean and variance in time to 50% WT on the initial frequency of WT. This same relationship was observed in competitions of

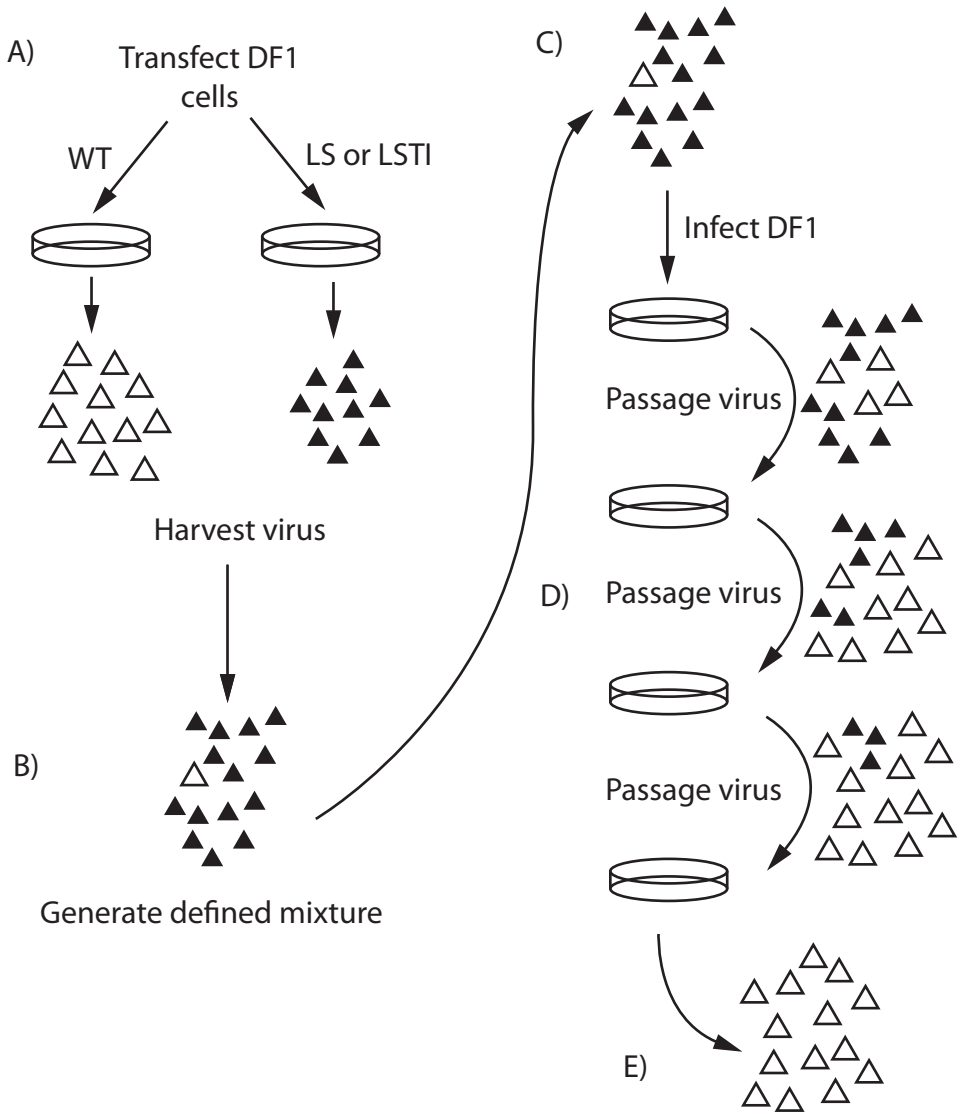

Figure 3.3: **Schematic of viral competition experiments.** Competition experiments between better-fit (WT) and less-fit (LS or LSTI) virus were initiated by individually transfecting DF1 cells with virus (A) and then generating defined mixtures of virus (B). The mixture was used to infect a culture of DF1 cells (C). Alternating every 3 or 4 days, the cell-free virus was harvested from 25% supernatant and used to infect a fresh culture of DF1 cells (D). The experiments were ended when the less-fit virus became undetectable in the population, indicating that the better-fit virus had won the competition (E).

WT against LSTI, however the effect was less pronounced (Table 3.1).

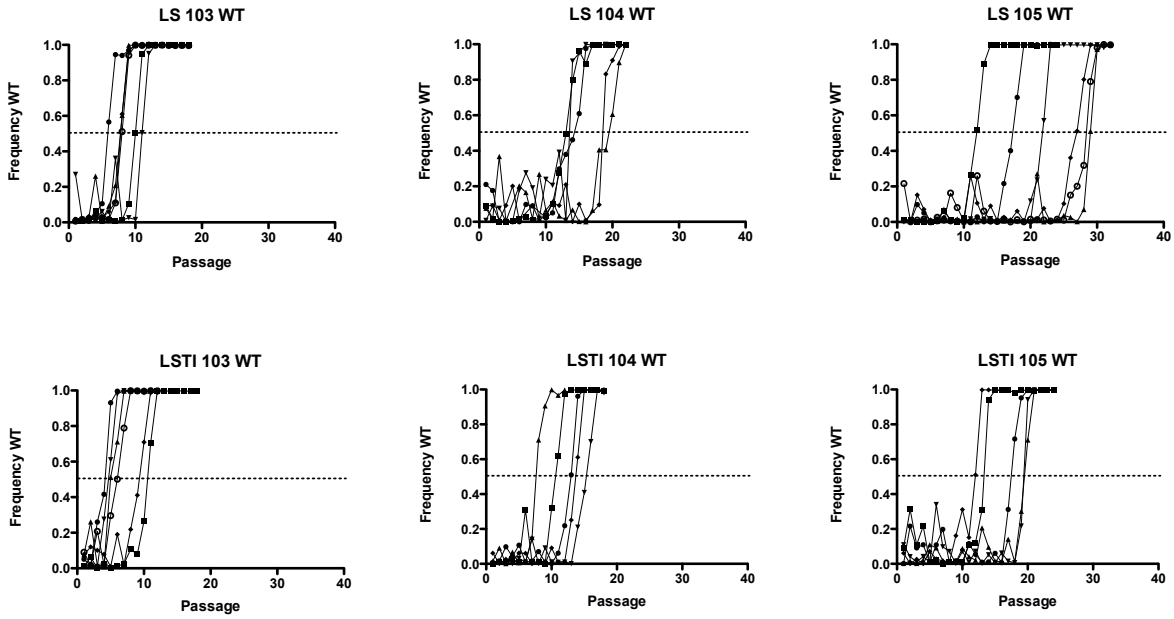

Figure 3.4: **Frequency of wild-type variant for WT vs LS or WT vs LSTI experiments in tissue culture.** (A) Competition experiments of WT vs. LS virus in tissue culture at the indicated starting frequencies of WT (B) Competition experiments of WT vs. LSTI virus in tissue culture at the indicated starting frequencies of WT. All experiments were performed by serial passaging the viral population (see Figure 3.3) and frequency of WT was measured by allele-specific PCR (see Figure 3.1). Dotted lines cross the competition curves at the passage where WT is 50% of the population.

The RNA copy number of each virus at each passage provides information about the relative population sizes of better-fit and less-fit viral particles without discriminating between functional and defective particles. To determine the relationship between RNA copy number and functional particles, we measured the number of infectious units of virus and the number of RNA copies of virus in serial dilutions of supernatants harvested from individual WT, LS, or LSTI transfections. The number of infectious units was measured by flow cytometry, scoring for GFP-positive cells, and the number of RNA copies was measured by quantitative PCR (Figure 3.5). The results were fit to an ordinary least squares regression, which we used to estimate the initial frequencies of better-fit functional particles at the start of each experiment (Table 3.2).

Interestingly, the LS and LSTI viruses produced approximately 100-fold less infectious particles for the same amount of viral RNA as WT. This suggests that the mechanism of selection between WT and mutant viruses may be, at least in part, due to lower stability of mutant viral particles.

We estimated the census population size at each passage by first measuring the total viral RNA in each passaged supernatant (Figure 3.6), estimating the contribution of WT or mutant virus to this pool of RNA using the frequencies of each determined by allele-specific PCR, then transforming the RNA copy numbers to infectious particles using the linear regression described above. The estimated size of the viral populations fluctuated over a 2-3 log range for all the replicates (Figure 3.7), demonstrating that population sizes varied greatly for the duration of the experiments. Some replicates contained RNA levels that exceeded 10^8 copies. These samples contained high levels of viral DNA contamination, evident from control quantitative PCR reactions that lacked reverse transcriptase. It is unlikely that this contamination is due to viral DNA carried over from transfections, due to the appearance of the contaminant after passages where quantitative PCR reactions did not indicate signs DNA contamina-

Experiment	**Mean Time to 50% WT** (passages)	**Variance in Time to 50% WT**
LS103WT	8.83	3.36
LS104WT	16.4	16.4
LS105WT	22.8	22.8
LSTI103WT	7.00	9.2
LSTI104WT	12.4	9.3
LSTI105WT	17.0	11.0

Table 3.1: **Summary statistics for competition experiments.** Experiments are named according the ratio of less-fit (LS or LSTI) to better-fit (WT) RNA copies in the initial infection, e.g. LS103WT is the name of the experiment that was initiated with a ratio of 10^3 LS RNA copies to 1 WT RNA copy.

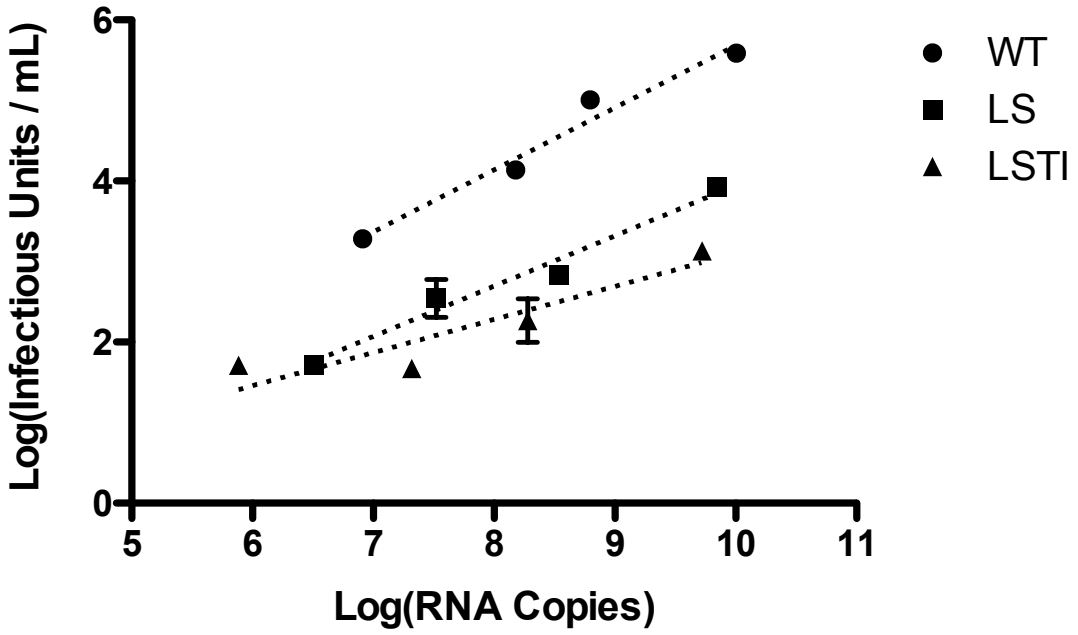

Figure 3.5: **Relationship of RNA Copy Number to Viral Titer.** Viral RNA and viral titer were quantified for 10-fold serial dilutions of supernatants harvested from DF1 cells transfected with WT, LS, or LSTI virus. Dotted lines represent the linear regression of Log-transformed data. R^2 values are 0.97, 0.94, and 0.81 for WT, LS, and LSTI data, respectively. Vertical error bars represent the range of duplicate titering experiments.

Experiment	Mutant (IU)	WT (IU)	Total IU	Initial Freq. WT	Initial MOI
LS103WT	1.25×10^4	3.39×10^3	1.60×10^4	0.213	0.053
LS104WT	1.25×10^4	574	1.31×10^4	0.043	0.043
LS105WT	1.25×10^4	97.1	1.27×10^4	0.008	0.042
LSTI103WT	1.73×10^4	3.39×10^3	2.07×10^4	0.164	0.069
LSTI104WT	1.73×10^4	574	1.79×10^4	0.032	0.060
LSTI105WT	1.73×10^4	97.1	1.74×10^4	0.006	0.058

Table 3.2: **Initial conditions for competition experiments.** Defined mixtures of better-fit (WT) and less-fit (LS or LSTI) virus were created by normalizing against RNA copies of each virus. Experiments were named according to the initial ratio of RNA of each competitor, i.e. LS 10^3:WT 1 was named LS103WT. The number of infectious units of better-fit or less-fit virus in each mixture was determined by generating a standard curve of viral RNA copies to infectious units.

tion. These specific samples are not likely representative of the viral population size and were excluded from further analysis.

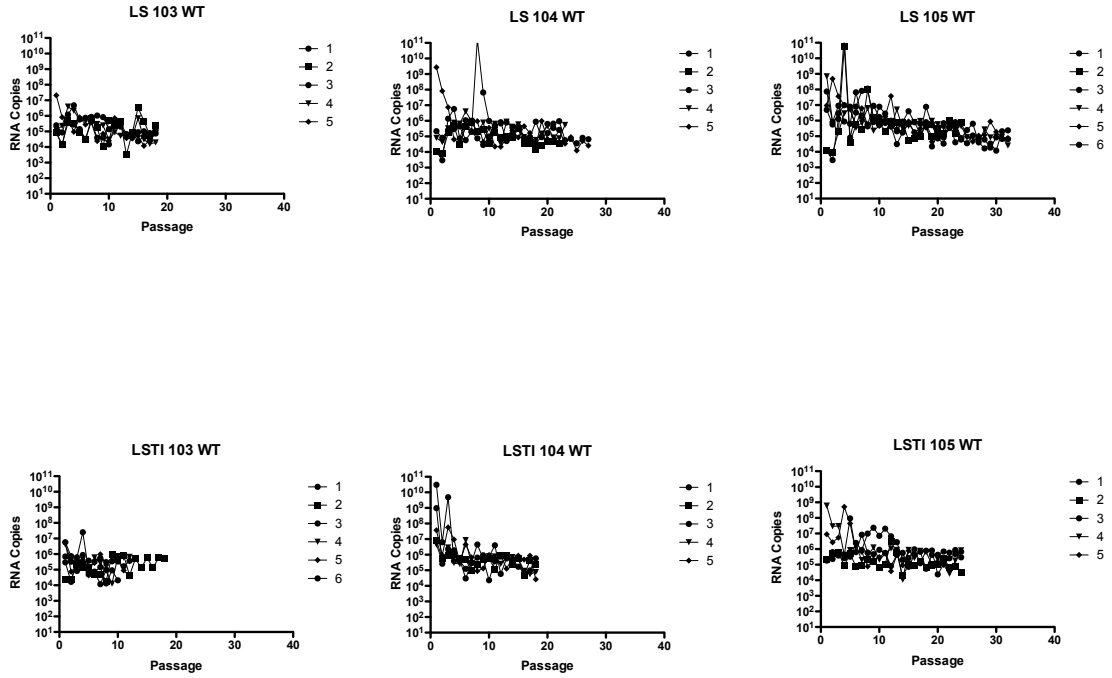

Figure 3.6: **Copies of Viral RNA at Each Passage for WT vs LS or WT vs LSTI experiments.** The total copies of viral RNA was quantified by qPCR for A) WT vs LS competition experiments and B) WT vs LSTI competition experiments. Each curve represents a single replicate.

3.3 Measurement of viral evolution parameters

In order to determine if level of genetic drift observed in the passaging experiments is more or less than what is expected given the size of the population, we measured the relevant parameters of evolution to supply to a model that describes the system. This model is discussed further in Chapter 3.4.

Effect of Adsorption Time on Infectivity

Due to the highly parallel nature of the passaging experiments, many replicates were infected and/or passaged at the same time. All cultures were subject to a minimum of 1 hour of adsorption to virus on DF1 cells, however some cultures were left to incubate for at most 30 minutes longer. To test the role of adsorption time on DF1 cells we initiated infections with WT, LS, or LSTI on DF1 cells and measured the change in GFP-positive cells 48-hours post infection after varying lengths of adsorption time. The results are summarized in Table 3.4. No significant increase in infectivity was observed for adsorption times greater than 1 hour, indicating that most infectious particles of all three variants bind cells within the first hour of adsorption.

Target Cell Growth Rate

Previous work with ALV subgroup B virus showed that the LS and LSTI mutants exhibited a strong cytopathic effect on DF1 cells [Rainey and Coffin, 2006]. We tested if infection by the mutants caused different growth rates of DF1 cells when compared to WT, an effect that might contribute to fitness differences between the three viruses. We infected cells with WT, LS, and LSTI and monitored viable cells at regular intervals post-infection. The doubling time of DF1 cells was approximately the same for uninfected cells and cells with infected with LS, WT, or LSTI variants (Figure 3.8). The doubling time is 15.9 hour for uninfected cells, 17.1 hour for WT-infected cells,

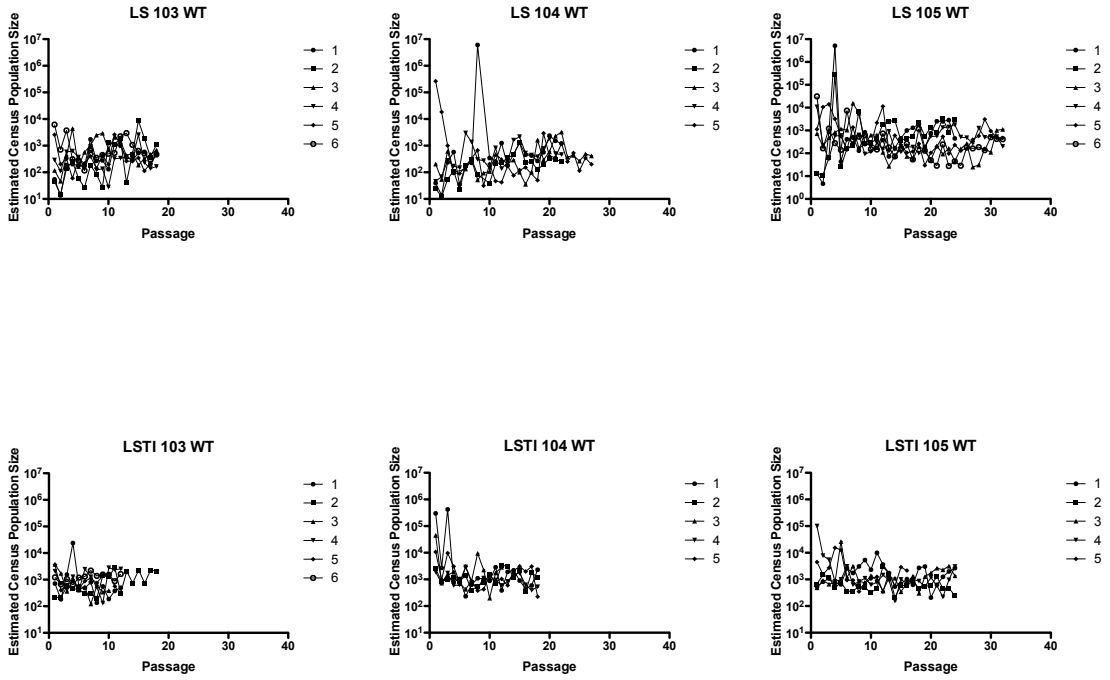

Figure 3.7: **Estimated census population size at each passage.** The census size of the virus population at each passage was estimated by first measuring the total RNA at each passage, then calculating the number of RNA copies of better-fit and less-fit virus using the frequency of each variant at each passage, and finally, converting the number of RNA copies to infectious units by the relationship of RNA copy number to infectious particles described in Figure 3.5. The census number of viral particles is the sum of each variant at each passage.

Experiment	Replicate	Mean	S.D.	Min	Max
LS103WT	1	430.5	438.5	13.6	1731.3
	2	978.6	2012.7	14.3	8634.0
	3	1058.9	1180.5	46.2	4384.2
	4	435.5	581.7	28.1	2590.1
	5	575.7	774.9	59.7	2618.3
	6	1226.4	1596.4	116.2	6191.2
LS104WT	1	599.2	621.2	13.8	2394.4
	2	242.3	274.0	11.9	1309.8
	3	509.7	753.2	35.7	3269.5
	4	717.8	793.7	45.8	3002.6
	5	1161.2	3761.6	31.4	18581.3
LS105WT	1	771.7	865.6	4.6	2856.6
	2	1246.7	1592.0	10.7	6724.4
	3	992.1	2858.5	25.0	15615.8
	4	801.9	1864.6	55.8	10743.9
	5	1702.3	3523.3	29.8	14227.0
	6	1484.1	5571.4	27.7	31172.6
LSTI103WT	1	2498.0	6755.0	136.2	23914.7
	2	1040.1	889.3	202.7	1040.1
	3	926.4	1101.8	122.6	3782.6
	4	1431.1	970.1	131.6	2812.7
	5	1064.9	914.4	304.9	3669.2
	6	1271.9	437.7	637.1	2192.0
LSTI104WT	1	1250.4	853.6	233.7	2855.0
	2	1385.5	899.1	344.6	3419.6
	3	5636.5	12630.0	198.5	44971.6
	4	1054.3	660.8	390.5	2919.6
	5	2468.3	3006.0	226.2	9749.6
LSTI105WT	1	2060.9	2088.2	208.7	9920.4
	2	754.9	615.6	211.3	3064.6
	3	2308.7	5213.3	300.8	26486.7
	4	1508.3	1794.1	156.4	7801.5
	5	3123.5	4441.4	347.1	15345.1

Table 3.3: **Summary statistics for estimated census population size.** The mean estimated population size across all passages for each replicate experiment. S.D. = standard deviation of the mean

	Adsorption Time on DF1 Cells (Hours)				
	1	2	3	4	5
WT	63.65% ± 7.21	64.9% ± 9.74	65.1% ± 6.30	65.9% ± 10.1	67.0% ± 8.82
LS	7.78% ± 1.26	8.98% ± 1.29	6.52% ± 2.00	8.33% ± 1.53	8.68% ± 1.77
LSTI	2.84% ± 0.67	3.01% ± 0.62	3.08% ± 0.65	3.19% ± 0.79	2.93% ± 0.51

Table 3.4: **Effect of adsorption time on infectivity.** WT, LS, or LSTI virus was adsorbed to DF1 cells for the indicated number of hours before being washed with PBS. 48 hour post infection, the frequency of GFP-positive cells was scored by flow cytometry.

17.42 hour for LS-infected cells, and 18.28 hour for LSTI-infected cells. Cells grew exponentially over a 96 hour observation period and showed no evidence of negative-feedback (s-shaped logistic growth) due to contact inhibition, waste accumulation, or resource limitation.

Viral Generation Time

The viral generation time is defined as the time that elapses from a particular step in the viral replication cycle to the same step reached by the daughter virus. We tested if WT, LS and LSTI viruses exhibited different generation times, another factor that would contribute to fitness differences between viruses due to differential production of progeny virus as a result of more or less replication cycles.

We used expression of virus as our reference point and scored the elapsed time between expression of consecutive generations by a flow cytometry based assay. Since all viruses used in this study were GFP labeled, after the start of infection we expected to see an initial rise in GFP-positive cells corresponding to the expression of the supplied virus followed by a plateau representing the steps of the viral lifecycle post expression, including packaging, particle release, encounter of an uninfected cell, and entry. Finally, we expected to observe a second increase in the frequency of GFP-positive cells, representing expression of the daughter virus from the initial infectious dose (Figure 3.9).

The absolute number of GFP-positive cells was too low to directly measure by flow cytometry at the early time points of infection (data not shown). Rather than directly measure the frequency of GFP positive cells at each time point, we initiated replicate infections and halted replication by removing the culture supernatant and adding a viral replication inhibitor at different time points post-infection in each replicate. The frequency of GFP-positive cells for all replicates was measured 96-hours post infection, allowing for infected cells to replicate and amplify the GFP signal. By

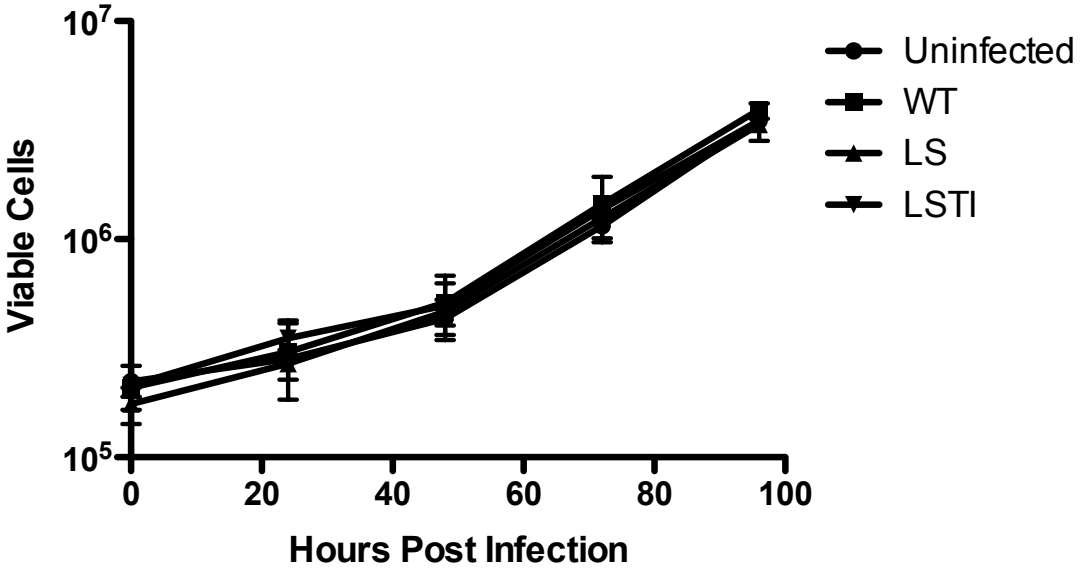

Figure 3.8: **Growth rate of DF1 cells.** DF1 cells were plated and 24 hours later infected with WT, LS, LSTI, or left uninfected. At regular intervals post infection, viable cells were scored by trypan blue exclusion. Error bars represent the standard deviation of triplicate experiments.

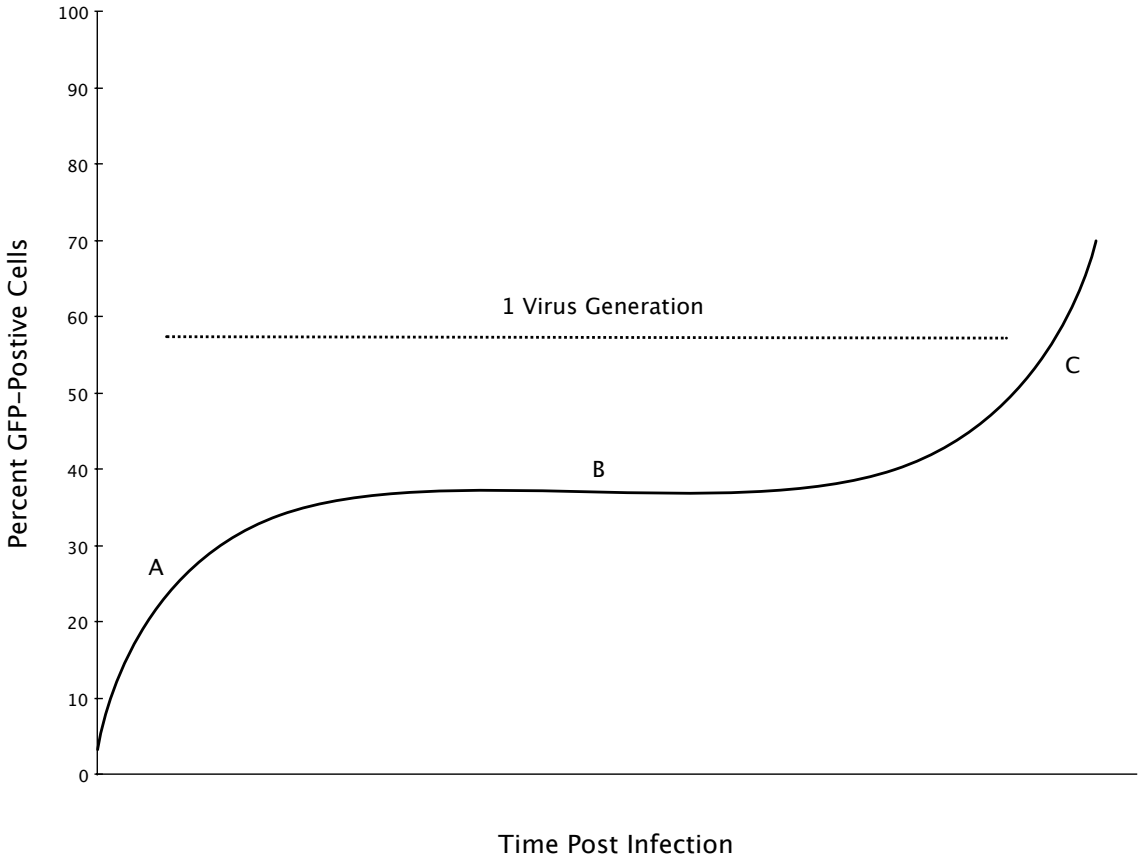

Figure 3.9: **Measurement of Viral Generation Time by Flow Cytometry** Simulated results of viral generation time experiment described in Figure 3.10. One viral generation is defined as the time that elapses from the initial rise of GFP-positive cells (A), which correspond to the expression of the initial infectious dose of virus to the second rise of GFP-positive cells (C), which corresponds to the expression of daughter virus. The plateau (B) represents the steps of viral lifecycle after expression including packaging, particle release, encounter of an uninfected cell, and entry.

this method, each replicate culture would serve as a "snapshot" of the frequency of GFP-positive cells at each time post-infecton that inhibitor was added (Figure 3.10).

2'-3'-dideoxyinosine (ddI) was previously shown to be an effective inhibitor of murine retrovirus infectivity [Stair et al., 1991] and is an inhibitor of HIV infection *in vivo* [Yarchoan et al., 1989]. ddI is a nucleoside analog of adenosine and acts as a chain terminator of reverse transcription. We tested the efficacy of ddI as an inhibitor of ALV subgroup B replication by preincubating DF1 cells for 24 hours with different concentrations of ddI and then infecting with WT, LS, or LSTI virus. The frequency of GFP-positive cells was scored by flow cytometry 48 hours post infection.

We found an identical dose response for all three viruses with a ddI concentration of 100 μM halting viral replication completely (50% inhibitory concentration for WT, LS, and LSTI of approximately 3.62 μM, 1.42 μM, and 1.42 μM, respectively)(Figure 3.11). We then tested if ddI effected infectivity through cytotoxic killing of DF1 cells by scoring viable DF1 cells after a 96 hour incubation with different concentrations of inhibitor. We found that at all concentrations of ddI tested, approximately equal numbers of viable cells remained in the culture indicating that 100 μM ddI did not alter the replication rate of DF1 cells, (Figure 3.12). Finally, we tested if 100 μM ddI was able to suppress replication when added post adsorption of virus (as would be necessary for the generation time experiment described in Figure 3.10) for the entire duration of a 96 hour incubation. This was done by adsorbing WT, LS, or LSTI virus for 1 hour, adding 100 μM ddI, then scoring GFP-positive cells at regular intervals post infection. We found this concentration of ddI to suppress viral replication until the end of a 96 hour window (Figure 3.13).

Using 100 μM ddi as an inhibitor, we observed the viral generation time using the protocol described in Figure 3.10. A second round of infection was not observed over a 96 hour time period (Figure 3.15). To confirm that progeny virus was indeed shed within a 96 hour time period and available for a second round of infection, super-

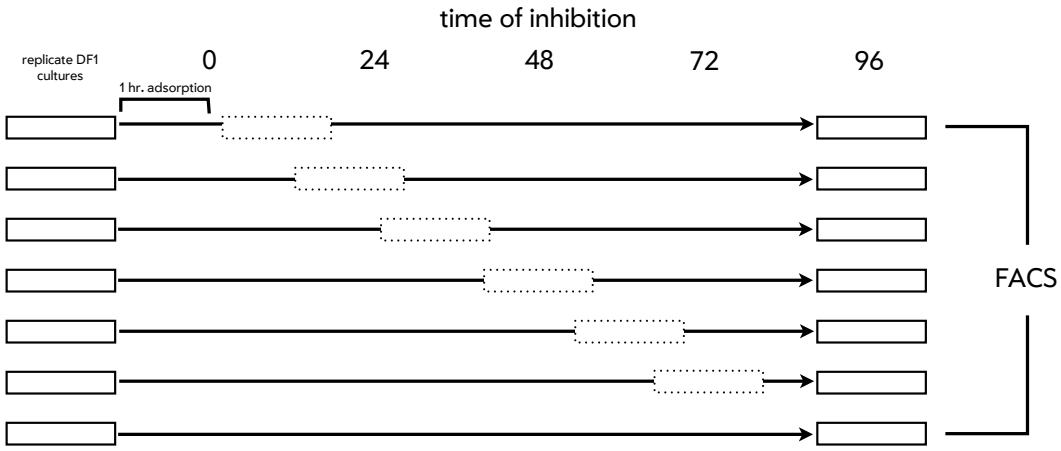

Figure 3.10: **Schematic representation of generation time experiment.** Replicate DF1 cultures were each subjected to 1 hour of adsorption to virus. Unbound virus was washed from cells and replaced with medium. At regular intervals post infection, supernatants were harvested from cultures and replaced with medium containing 100 μM ddI to halt viral replication, taking a "snapshot" of the replication that had occurred to that point. 96 hours post infection, all replicates were tested from the frequency of GFP-positive cells.

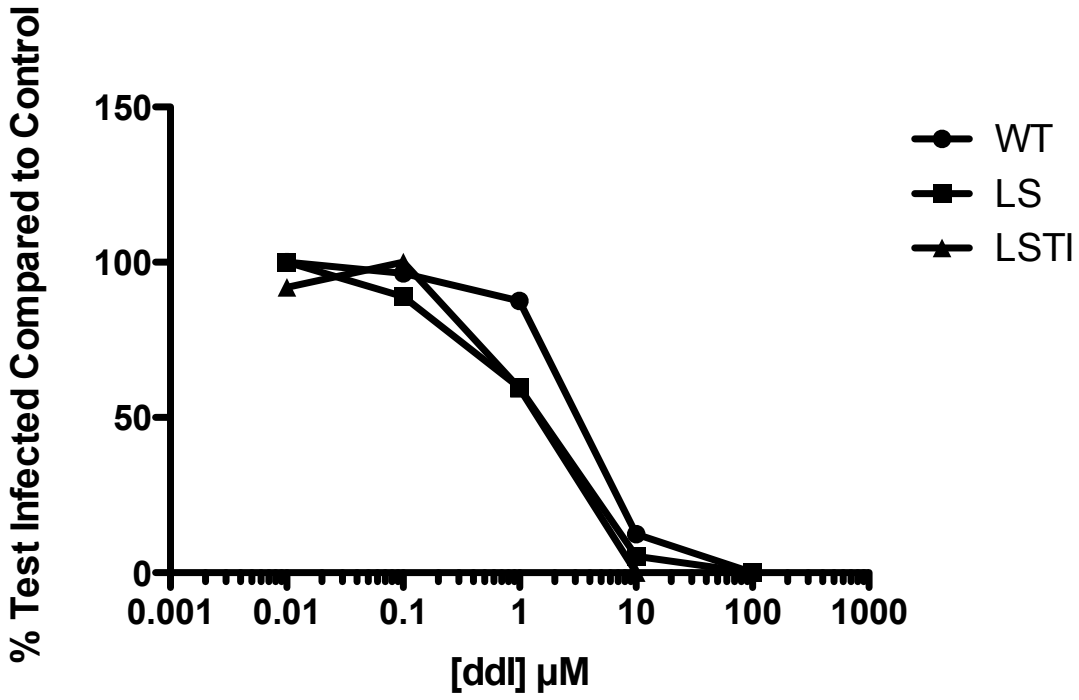

Figure 3.11: **Effective dideoxyinosine on viral infectivity** DF1 cells were pre-incubated for 24 hours in medium containing the indicated concentrations of dideoxyinosine. WT, LS, or LSTI virus were adsorbed to the pretreated cells for one hour. Unbound virus was washed away with PBS and cells were incubated in medium containing the indicated concentrations of dideoxyinosine. 48 hours post-infection, the frequency of GFP-positive cells was determined by flow cytometry. Results are normalized to infection frequency on cells not treated with inhibitor.

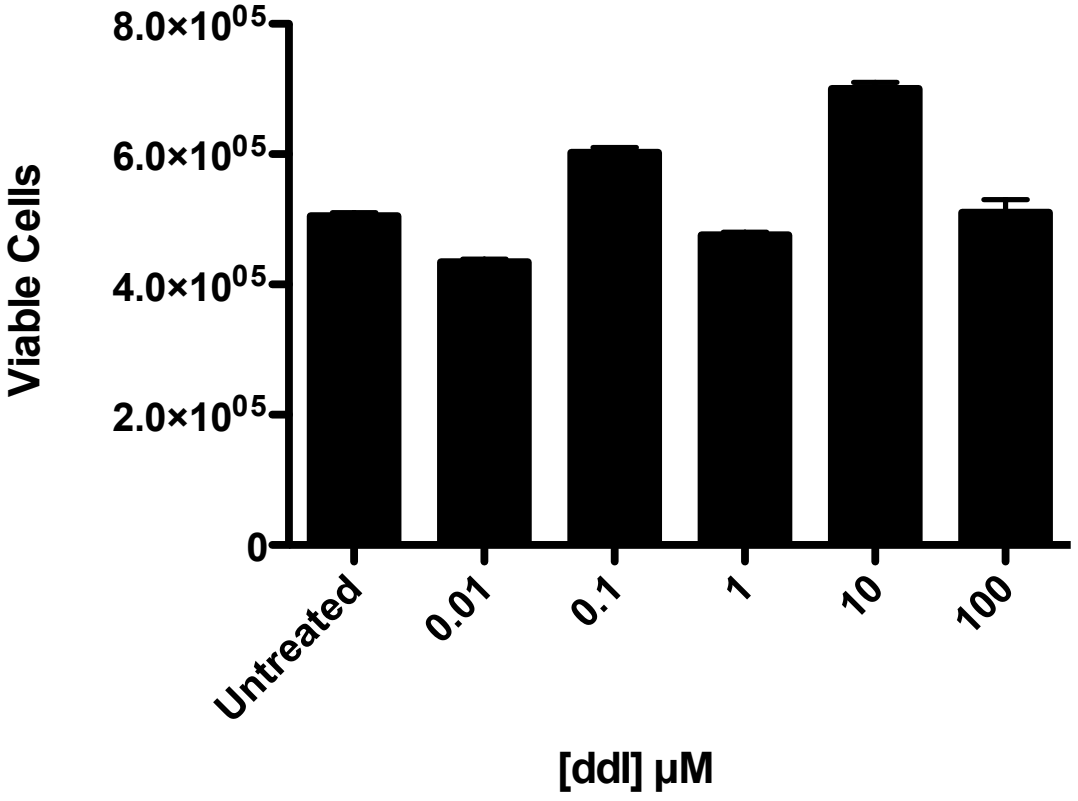

Figure 3.12: **Cytotoxicity of dideoxyinosine of DF1 cells.** 3×10^5 DF1 cells were seeded in the presence of the indicated concentrations of dideoxyinosine. 96 hours post-addition of drug, viable cells were scored by trypan blue exclusion. Error bars represent the range of duplicates.

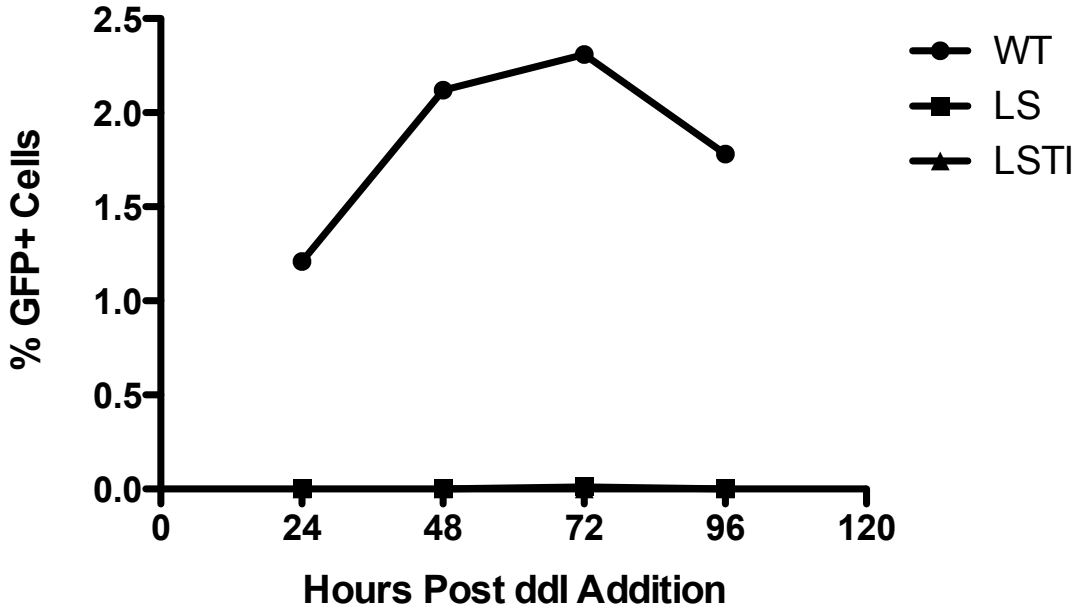

Figure 3.13: **Inhibition of viral replication by ddI added post infection.** WT, LS, or LSTI virus were adsorbed to DF1 cells for 1hour. Immediately post adsorption, 100 μM ddI was added to each culture and the frequency of GFP positive cells was scored at regular intervals post infection.

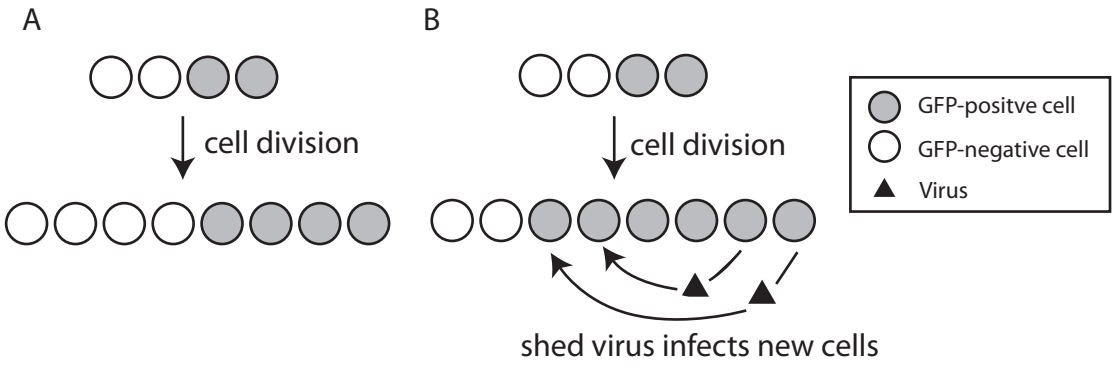

Figure 3.14: **Schematic representation of assay for new rounds of infection.** In panel A, 50% of cells are infected at the start of the experiment. After one cell division and no new rounds of infection, 50% of cells are infected, resulting in no change in GFP-positive cell frequency. In panel B, 50% of cells are infected at the start of the experiment. After one round of cell division and a round of infection, 75% of cells are infected, resulting in a 25% change in the frequency of GFP-positive cells.

natants were reserved at each point during the timecourse when medium containing ddI was added. The supernatants were tested for the presence of infectious virus by initiating infections on another set of uninfected DF1 cultures. At 44 hours post infection virus was confirmed to be present in the supernatant (Figure 3.16). The titer of this virus increased for the remainder of the experiment and the virus shed at 96 hour post infection was able to infect approximately 8% of the cells in a secondary infection.

It is possible that we did not observe a second rise in the frequency of GFP-positive cells because not all DF1 cells are capable of supporting an infection. To rule out the possibility that our initial inoculation saturated all infectable cells, we repeated the experiment at a 10-fold lower multiplicity of infection and found the same behavior to hold true (Figure 3.15 B).

To further test if all cells in the culture are infectable, we measured the frequency of GFP-positive cells in a co-culture of WT-transfected cells with uninfected cells at regular intervals post-mixture. Transfected cells should have no lag in expression of virus and immediately release virus into the supernatant upon placement into culture. Ratios of 1:1, 1:2, and 1:4 of transfected cells to uninfected cells all resulted in 100% GFP-positive cells at the end of a 96 hour incubation (Figure 3.17), indicating that no subpopulations of DF1 cells exist that are incapable of supporting viral entry or expression.

Taken together, these data show that the viral generation time is greater than 96 hour under the tissue culture conditions tested. In the competition experiments, virus was passaged in alternating intervals of 72 hours and 96 hours per passage, i.e. there were two passages per week. Given that the generation time is longer than the length of the passages, the progeny virus at each passage is the daughter virus of the initial infectious dose. Put another way, each passage represents one virus generation.

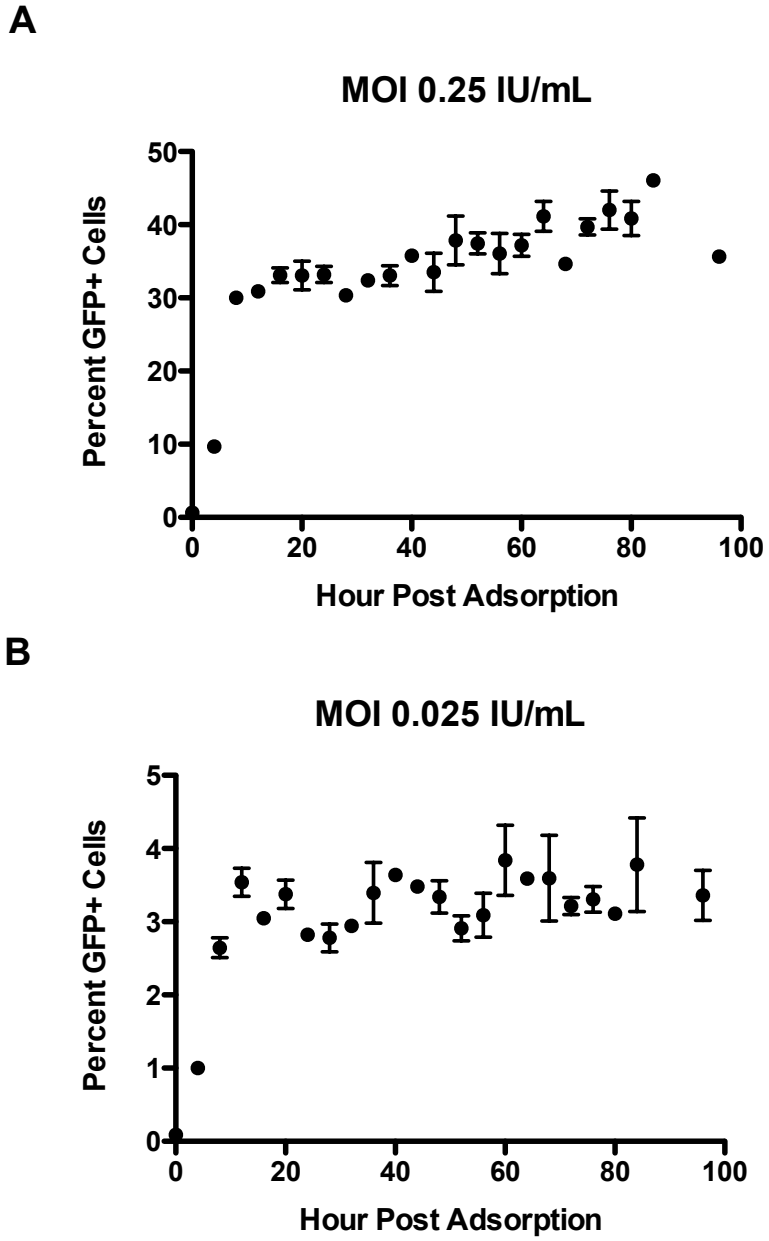

Figure 3.15: **Measurement of viral generation time.** Replicate cultures of DF1 cells were infected with WT virus. At regular intervals post-infection, medium was replaced on a replicate with medium containing 100 μM dideoxyinosine. 96 hour post-infection, replicates were analyzed for the frequency of GFP-positive cells by flow cytometry. Each point represents the mean of duplicate experiments. Results are normalized to cultures that were not treated with the dideoxyinosine. Error bars represent the range of duplicates.

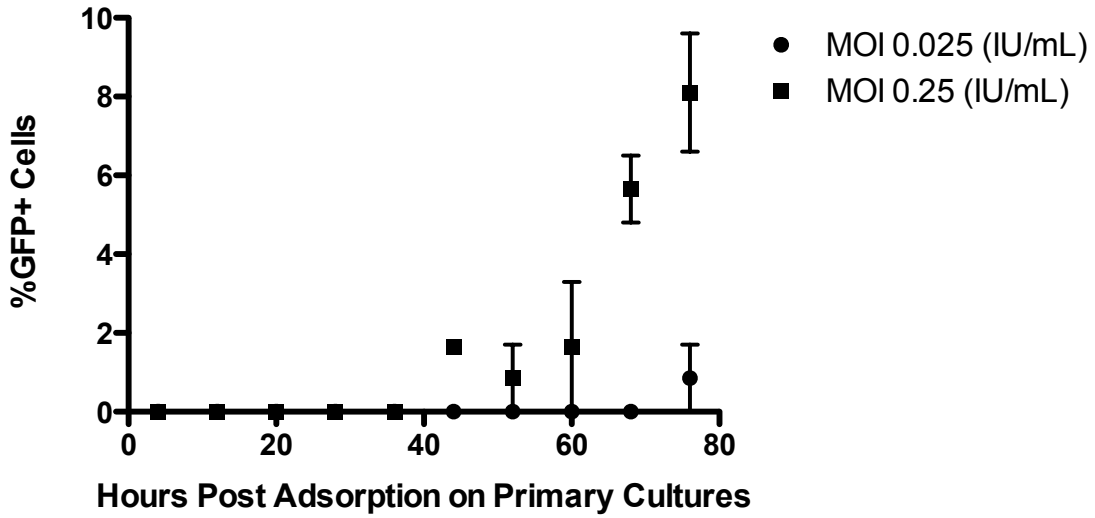

Figure 3.16: **Secondary infections from supernatants harvested from viral generation time experiment.** Supernatants were reserved at regular intervals post infection of cultures in Figure 3.15 and used to infect fresh uninfected DF1 cultures. 48 hour post infection, the frequency of GFP-positive cells were analyzed by flow cytometry. Error bars represent the range of duplicates.

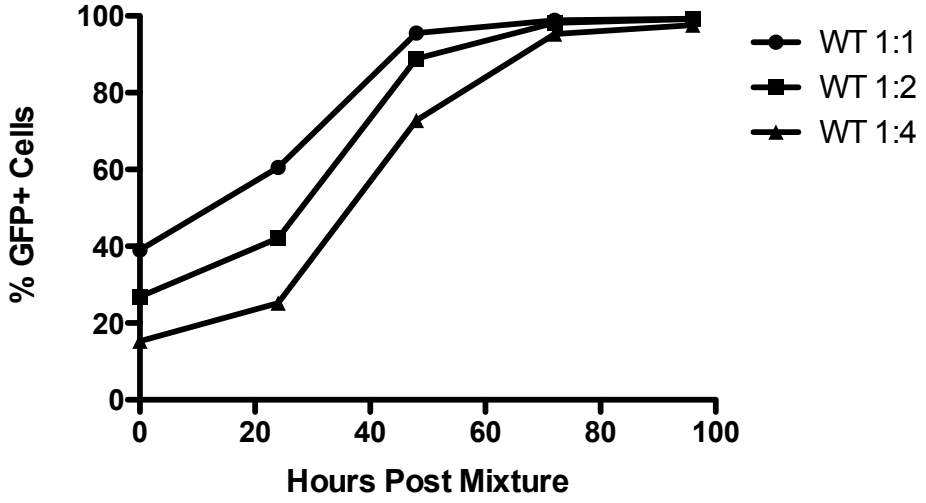

Figure 3.17: **Assay for a subpopulation of DF1 cells that can't support infection.** Uninfected DF1 cells were co-cultured with the indicated dilutions of WT-transfected DF1 cells. At regular intervals post-infection, the frequency of GFP-positive cells was scored by flow-cytometry. For all starting conditions, 100% of the cells were infected by 96 hours, indicating that no uninfectable subpopulations of DF1 cells exist. Each point represents the mean of duplicate experiments. Error bars represent the range of duplicate experiments. The range of error across all experiments was 0.0 - 2.4%, making error bars undetectable on the scale shown.

Selection Coefficients

Because the WT, LS, and LSTI variants did not demonstrate a second round of infection in each passage, calculating the selection coefficients was simply done by comparing the growth rates of each virus. By co-culturing transfected cells (which shed virus immediately upon placement in culture) with uninfected cells, we were able to measure the rate of spread of each mutant (Figure 3.18) in culture. The selection coefficients were then calculated using the formula described in the materials and methods section (Chapter 2.11) as follows: 0.785 for WT vs. LS and 0.908 for WT vs. LSTI.

3.4 Theoretical Expectations

The variance in time to fixation for each set of initial conditions was compared to those predicted by Monte Carlo simulations using the model described in Figure 3.19. This model is an adaptation of a previously described model for viral evolution that incorporates the effects of selection [Rouzine et al., 2001]. The introduction of selection is made by considering the simplest case of selection acting at a single locus that has a choice of two alleles. One allele is better-fit and produces more progeny per generation than the less fit allele. Because this model considers only a single site in the genome, the effects of recombination are ignored. Selective pressure is constant and serves to increase the frequency of the better-fit allele at each generation.

We consider a finite population of cells of which an initial fraction are infected according to the number of infectious units of virus supplied to the culture (I_0). These cells divide with a doubling time of t_c, and at the end of each passage (alternating 3 and 4 days, as in the experimental protocol) the cells shed virus in proportion to the selection coefficients of the two competitor viruses. A fraction of the virus released (F_p) is randomly sampled to account for the random sampling of virus due to the

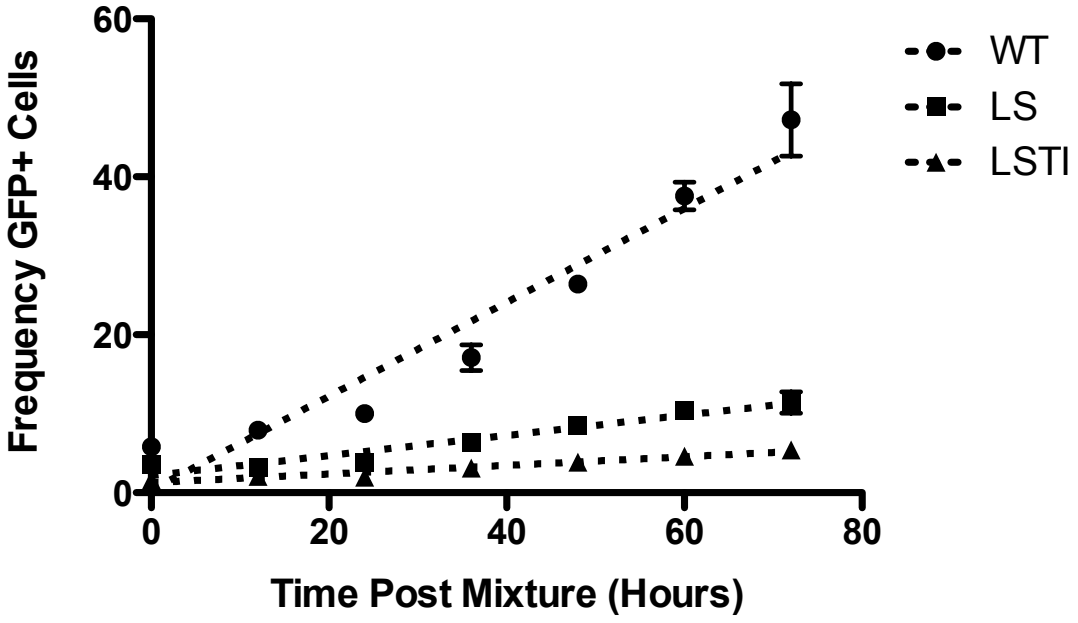

Figure 3.18: **Rate of increase of GFP-positive cells in co-cultures of transfected and uninfected cells.** DF1 cells that were transfected with WT, LS, or LSTI virus were mixed in a 1:1 mixture with uninfected DF1 cells. At regular intervals post mixture, the frequency of GFP-positive cells were measured by flow cytometry. Error bars represent the range of duplicate experiments.

experimental protocol. The sampled viral population is set to a constant size at the end of each passage (N_e) and used to infect a set of uninfected cells. At each infection, the virus has a chance μ to mutate from wild-type to mutant or mutant to wild-type. The forward and reverse mutation rate are equal, meaning that mutations that swap the identities of mutant and wild-type are equally likely.

3.4.1 Simulation Results

The parameters supplied to the model are summarized in Table 3.5. All parameters were measured in this study except for the mutation rate, which was set to 10^{-5} mutations per base pair per generation, which represents an estimate of the ALV mutation rate. Simulations were run supplying population sizes over a range of 50 - 1000, in 50 replicates per simulation. The outcomes are summarized in Table 3.6. Representative results are shown for WT vs LS in Figure 3.20 and WT vs LSTI in Figure 3.21.

3.4.2 Estimation of N_e

Over the range of population sizes supplied to the model, 50-1000, we did not observe simulation outcomes that describe the experimental results. The lowest population size supplied to the model (50) generated simulated results that exhibited little variance and underestimated the mean time to 50% WT in the tissue culture experiments (Table 3.6). The poor fit of modeling results to experiments results suggests a value of N_e under 50. The applicability of the model to the tissue culture system is discussed further in Chapter 4.1.

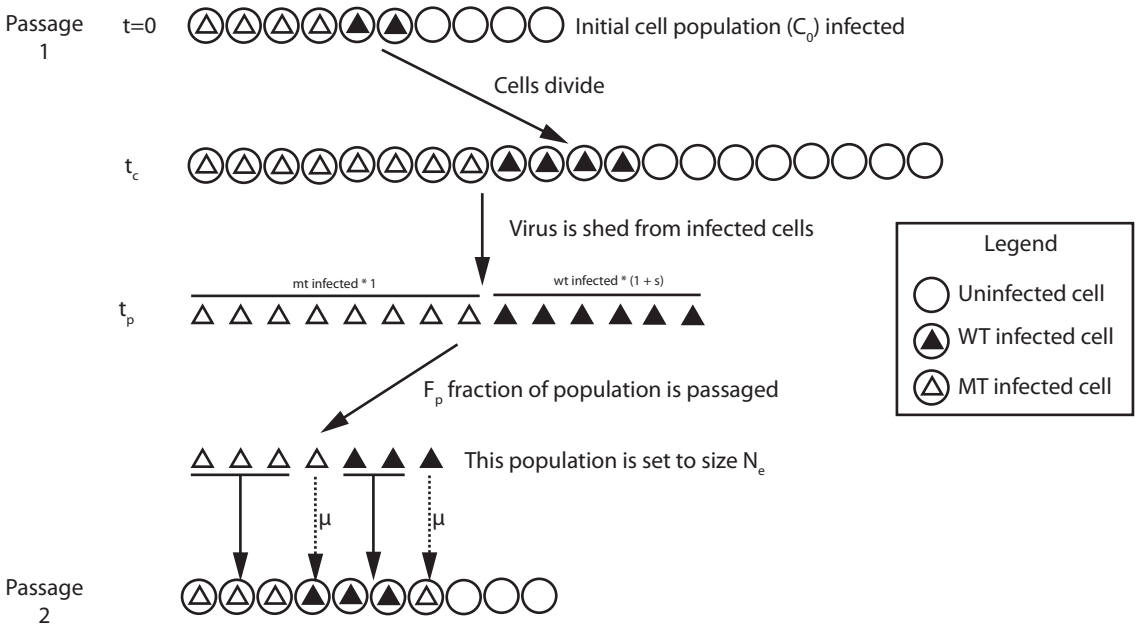

Figure 3.19: **Model of viral evolution.** At t=0 of passage 1, C_0 cells are infected with a mixture of better-fit and less-fit virus. Cells divide ever t_c hours, uninfected cells divide into daughter cells that are uninfected and infected cells divide into daughter cells that are infected. At the end of each passage of length t_p, virus is shed from infected cells in proportion to the relative growth rates of the variants. A fraction (F_p) is randomly sampled, a population of size N_e, with the same frequency of better-fit and less-fit variants as the sampled population, is used to infect the next cells of the next passage. Each virus has a probability μ of mutating to the other allele at each infection.

Variables	Symbol	Value	Description
Effective Population Size	N_e	$50 - 1000$	The population size of virus used to infect at each passage in the simulation.
Parameters	Symbol	Value	Description
Initial MOI	I_0	0.05	The fraction of cells infected at the start of the experiment.
Selection coefficient	s	-0.785 for WT vs. LS, -0.908 for WT vs. LSTI	The selection coefficient calculated from the relative fitness of WT and LS or LSTI
Cell doubling time	t_c	18 hours	The amount of time that elapses from one cell's division to its daughter's division.
Viral generation time	t_v	$>$ 96 hour, Set to t_p	The amount of time that elapses from one point in the replication cycle of a virus to the same point in its daughter virus.
Fraction of virus passaged	f_p	0.25	The fraction of supernatant from the end of one passage used to infect at the start of the next passage.
Initial frequency of better-fit allele	m	0.213, 0.043, and 0.008 for LS vs. WT. 0.164, 0.032, and 0.006 for LSTI vs. WT.	The frequency of the better-fit allele at the start of a competition experiment.
Number of cells at start of passage	C_0	3×10^5	The initial number of uninfected cells at the start of a passage.
Mutation rate	μ	10^{-5}	The frequency of viral mutations per site per generation.
Passage length	t_p	Alternating 72 hours and 96 hours	The time post-infection of one passage to infection of the next passage.

Table 3.5: **Summary of parameters used for viral evolution model**

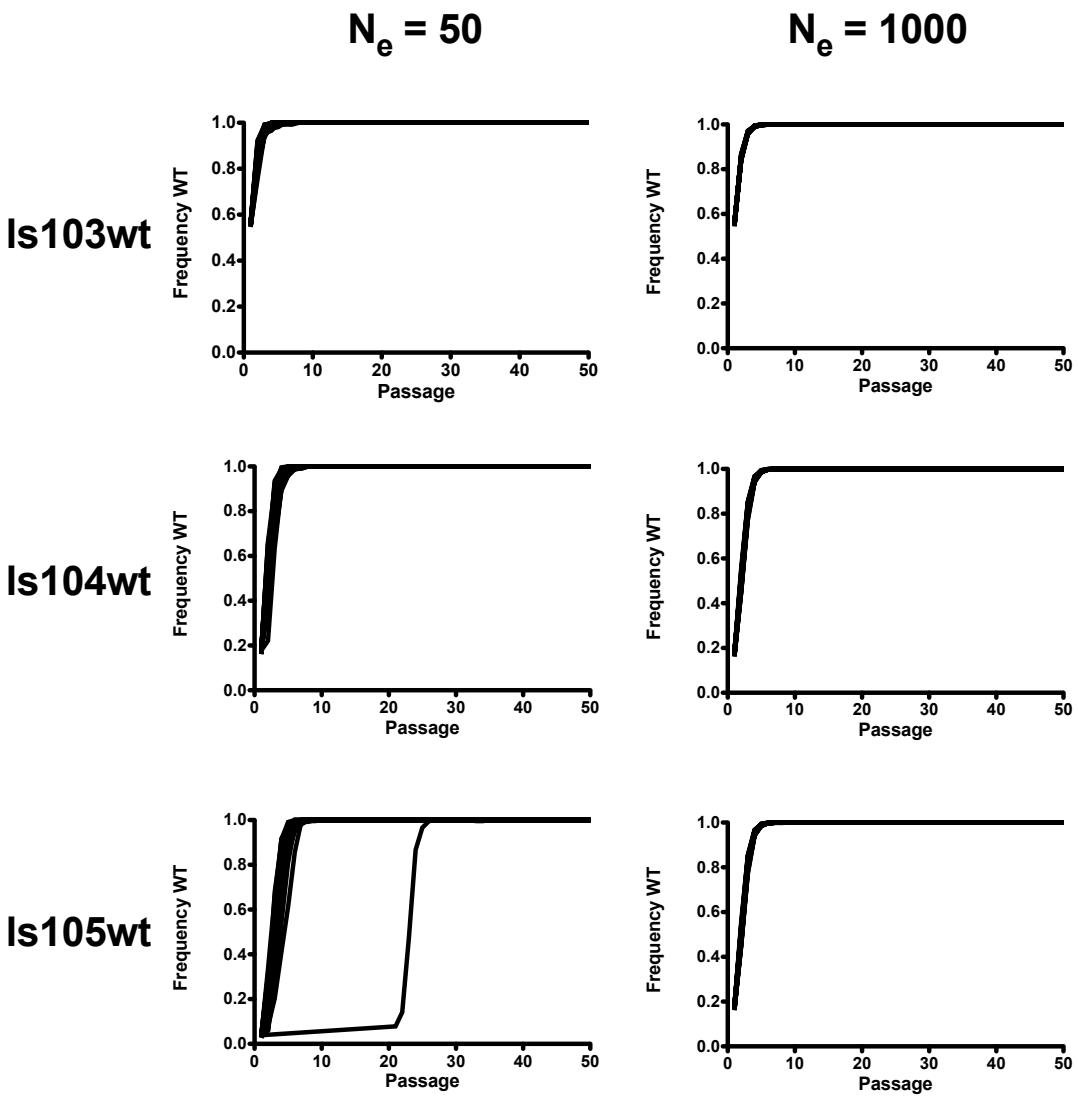

Figure 3.20: **Simulated results of competition experiments between WT and LS.** Each panel represents 50 replicate simulations of competition experiments based on the model described in Figure 3.19. The starting parameters for the simulation are: initial MOI (I_0) = 0.05; selection coefficient (s) = -0.785; cell doubling time (t_c) = 18 hours; viral generation time (t_v) = 1 per passage; fraction of virus passaged (f_p) = 0.25; initial frequency better fit allele (m) = 0.213, 0.043, or 0.008 for LS103WT, LS104WT, or LS105WT, respectively; number of cells at start of passage (C_0) = 3×10^5; mutation rate (μ) = 10^{-5}. Panels on the right indicate the results of a population size of 1000. Panels on the left indicate the results for a population size of 50.

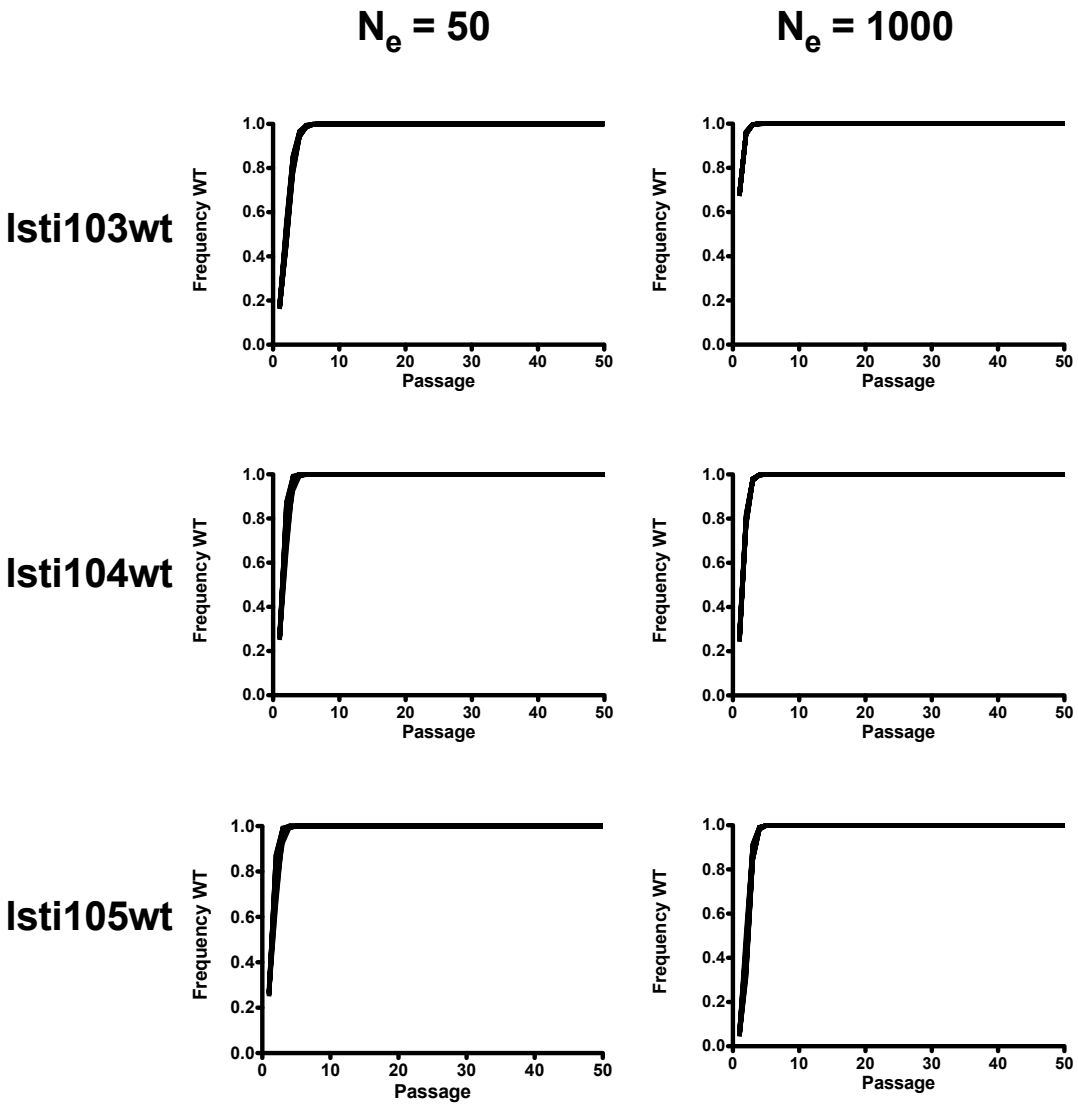

Figure 3.21: **Simulation of competition experiments between WT and LSTI.** Each panel represents 50 replicate simulations of competition experiments based on the model described in Figure 3.19. The starting parameters for the simulation are: initial MOI (I_0) = 0.05; selection coefficient (s) = -0.908; cell doubling time (t_c) = 18 hours; viral generation time (t_v) = 1 per passage; fraction of virus passaged (f_p) = 0.25; initial frequency better fit allele (m) = 0.164, 0.032, or 0.006 for LSTI103WT, LSTI104WT, or LSTI105WT, respectively; number of cells at start of passage (C_0) = 3×10^5; mutation rate (μ) = 10^{-5}. Panels on the right indicate the results of a population size of 1000.Panels on the left indicate the results for a population size of 50. Panels on the right indicate the results of a population size of 1000.

Simulation	**Mean Time to 50% WT** (passages)	**Variance in Time to 50% WT**
LS103WT	1.0	0
LS104WT	2.52	0.25
LS105WT	4.35	12.23
LSTI103WT	2.0	0
LSTI104WT	2.0	0
LSTI105WT	2.58	0.25

Table 3.6: **Summary statistics for simulation results at population size 50.** Each simulated experiment represents the results of 50 runs of the simulation.

Chapter 4

Discussion

Our rationale in designing this tissue culture system was to generate the simplest case possible that allows observation of the evolution of a population under selective pressure. Since evolution can continue on indefinitely, competition experiments were chosen because they provide a clearly defined end to the experiment. We chose competitors that differed by only one or two nucleotide changes to simplify the role that mutation plays in the experiments. We performed the experiments in the relatively homogenous conditions of tissue culture, infecting only a single cell line.

As competitor viruses, we chose WT Avian leukosis virus subgroup B and two mutants previously described and identified in our lab, LS and LSTI. These mutants differed from WT by only one and two mutations, respectively. We determined the relative growth rates of the viruses in culture and used these values to calculate selection coefficients for the competition experiments.

We observed the contribution of genetic drift to the evolution of a viral population in cell culture. We quantified genetic drift by measuring the variance in the time it takes for a better-fit virus, starting at low frequency, to reach 50% of the culture in a competition experiment with less-fit virus. Making the better-fit virus increasingly rare at the start of the experiment caused an increase in the variance, demonstrating

that the amount of drift observed is a function of the initial frequency of better-fit virus. Additionally, the amount of drift was inversely dependent on the magnitude of the fitness difference between better-fit and less-fit virus; a larger fitness difference resulted in a lower amount of drift.

The cell culture system relied on passaging the virus from infected cell cultures to uninfected cells every 3-4 days. At each passage, 25% of the culture supernatant was used to initiate infections on the next culture, randomly sampling a portion of viruses from the population. When the experimental results are compared to a simple model, designed to take into the account the relevant biological details of the system including selection and the random sampling due to passage, we observe a large difference in the mean and variance in time for WT to reach 50% of the population.

Despite all these efforts to simplify the system and approach the assumptions of our model, our model did not describe the system. There are, however, some gross observations that can guide the refinement of future models and tissue culture experiments. These are discussed in the following sections.

4.1 Which factors can account for excess drift observed in tissue culture?

Asynchronous Infections

It is possible that the asynchronous adsorption upon initiation of infection could increase the observed level of drift. Although we observed that adsorption times from 1 hour to 6 hours had little effect on the total infectivity with all three virus, it is possible that WT, LS and LSTI viruses differentially adsorb within the first hour. If a fraction of virus binds cells during the early minutes of a 1 hour adsorption, this virus may have a significant advantage in total replication by the end of the passage

over the late binding population. If the size of this fraction is small when compared to the total virus population, it would introduce another random sampling event that would increase the amount of drift observed in the experiment.

This effect has been observed in tissue culture with HIV. Voronin et al. initiated replicate infections with 1:1 mixtures of HIV-1 variants harboring different neutrally selected alleles. After all the cells in the culture were infected, they measured the frequencies of each allele for each culture, and then calculated the variance between the cultures as a measurement of drift. They found a 10-fold excess of variance than what would be expected for an ideal, neutrally evolving population. They were able to reduce this effect by pre-binding the virus to an uninfectable cell type and then co-culturing these cells with infectable cells. This essentially synchronized the infection by uniformly presenting the viral particles to target cells. The improved synchroniztion reduced, but did not eliminate, the excess drift observed compared to an ideal population by 3-fold [Voronin et al., 2009].

We can observe some asynchronicity of infections in the generation time experiment presented in Figure 3.15. At both MOIs tested, there is a 12 hour period post adsorption of virus before the maximum number of GFP-expressing cells is reached. This indicates that there is some asynchrony in the infections but it is impossible to know from this experiment how much is due to binding of virus or variations in the expression times of GFP post entry.

It may be possible to gain additional information about the early binding kinetics of the variants used in this study by using a similar approach. 1:1 mixtures of WT and mutant viruses can be used to infect DF1 cells. The adsorption time will be varied over a range of just a few minutes to 1 hour before unbound virus is washed from the cultures. The cultures will incubate for 3-4 days and then frequencies of WT and mutant RNA in the supernatant will be measured by allele-specific PCR. The variance in frequencies between many identical replicates may give a clue to the

level of drift that is associated with stochasticity in early binding.

Fluctuations in Population Size

The population of virus fluctuated over a large range at the time of each passage (Figure 3.7). This estimate was made by measuring the amount of viral RNA at each passage and transforming this value to a number of functional viral particles using a standard curve that relates RNA copy number to infectious units (Figure 3.5). However, it is possible that we are overestimating the census size by this methodology. The first detectable virus in the cultures appears at 44 hours post infection (Figure 3.16) but may not remain infectious at the time of passage. The virus in the infectious dose for the next passage most likely represents recently produced virus. However, the RNA from defective virus could still be available to serve as a template in the quantitative PCR reaction, increasing the total RNA observed by quantitative PCR. This effect may be partially accounted for due to the fact that virus harvested from the transfection for standard curve generation would be subject to the same effect of accumulating RNA from defective particles, but the differences in the culture conditions of transfection and the passaging experiments might be important.

Testing the stability of viral particles under the relevant tissue culture conditions will give additional information about the ratio of RNA that is from infectious particles to the RNA that is from inactivated particles. A half life for viral particles can be measured by harvesting supernatants from transfected or infected cells, filtering them of cell debris, and incubating the supernatants at 37 °C. At regular intervals, aliquots of supernatant can be used to initiate titering experiments, allowing us to observe particle stability as a function of time in medium conditioned by cell growth.

Even if the census population size here is an overestimate, the relative amount of viral RNA in the supernatant should still serve as a good measure of the relative populations sizes. The relative levels indicate large fluctuations regardless of the

absolute number of viral particles, summarized in Table 3.3. This could contribute to the excess drift we observed in the passaging experiments. The effects of bottlenecking have been experimentally observed for a variety of RNA viruses including bacteriophage phi 6 [Chao, 1990], Tobacco etch virus [de la Iglesia and Elena, 2007], Vesicular stomatitis virus [Novella et al., 1999], and HIV-1 [Yuste et al., 1999]. In all studies, bottlenecking reduced the mean fitness of the population, increasing the role of genetic drift.

Our model does not consider the effects of fluctuating population size, however, it will be valuable to compare the experimental results to the results from modified simulations that fluctuate in population size between a defined range and at a defined frequency. It is unclear to what level HIV serially bottlenecks within an infected individual, but bottlenecking along a person to person transmission chain is observed and thought to be the result of infection by a single virus [Fischer et al., 2010]. Modeling refinements that accurately account for the role of bottlenecking may be valuable for making predictions under these circumstances.

It may also be possible to use the LS and LSTI mutants for competition experiments under the conditions of constant population size. In early work during this thesis, we developed a QT6 cell line that was able to divide under the conditions of suspension culture. The protocol to select for non-adherent cell types involved a combination of step-wise serum reduction, as well as sub-culturing of unattached cells. It may be possible for this cell line to support an infection of LS and LSTI variants in a steady state system that supplies fresh uninfected cells to the culture at rate equal to the removal of cells in the culture. These would have the effect of maintaining a constant population size. This system will not be suitable for competition experiments that include the WT virus used in this study because QT6 cells do not express the subgroup B receptor.

The Role of Multiple Mutations

Our experimental system was designed to meet a simple model that considers selection due to mutation at a single site in the genome. The experiments initiated with WT competing against LS met the assumptions of the model due to the fact that WT and LS differed by only a single base pair at a site known to be responsible for a fitness difference (Figure 1.8 and Figure 3.18). Mutations that accumulate at sites other than the LS position are not considered by our model, however it is possible that random mutations could appear anywhere in the genome due to reverse transcriptase errors. These mutations could potentially alter the fitness difference between WT and mutant virus, increasing or decreasing the amount of drift we observe.

While we cannot rule out the possibility that these mutations occurred in our passaging experiments, there are some arguments against their presence. We determined that the generation time of the viruses used in this study is greater than the length of each passage. The longest competition experiment took 30 passages for WT to reach 50% (Figure 3.4, LS105WT), which means that the virus at passage 30 is only 30 replication cycles away from the virus that initiated the experiment. There is only one reverse transcription step for each replication cycle at which to introduce mutations. Additionally, mutations that occur on the WT virus are equally likely to occur on the mutant competitor virus.

Given the relatively low number of passages it took for WT to emerge across experiments, it may be useful to sequence many full length *env* sequences from the passaging experiments to directly observe if the WT virus detected at the end of the passaging experiments accumulated mutations at different sites.

4.2 The relative fitness of WT, LS, and LSTI variants

In the original selection experiment [Taplitz and Coffin, 1997], the isolated host-range extension mutants always had the LS mutation in the background of the TI mutation. This mutation extended the host range of the virus to cells that didn't express the cognate subgroup B receptor. It was subsequently shown that the LS mutation alone was sufficient to confer host-range extension on QT6 cells and the TI mutation alone did not have an effect [Rainey et al., 2003]. These observations suggest that the LS mutation could only be selected in the background of the TI mutation and the TI mutation serves to attenuate the fitness cost of the LS mutation. Another explanation is that the TI mutation confers some additional efficiency to receptor-independent entry of non-permissive cells. This conclusion was drawn from the fact that the LSTI mutant replicates about twice as fast on QT6 cells (Jonah Rainey's thesis, unpublished).

On DF1 cells, which express the subgroup B receptor, the LS mutant is more fit than the LSTI mutant (Figure 3.18). This result suggests that while the TI mutation in the context of the LS mutation provides additional fitness in receptor-independent entry, and this benefit may come at the cost of subgroup B receptor dependent entry. LSTI is capable of weak interference when preinfecting QT6 cells and challenged with subgroup E virus. The LS mutant was not, indicating that the TI mutation might provide additional binding to subgroup E receptor, which would be of no use on DF1 cells.

The allele-specific strategy we developed in this study could be used to make sensitive measurements of the frequencies of the LS and TI mutations in the original co-culture experiment in which they were selected. Of interest would the order in which the mutations arose in the passaging experiments, which would give some

additional information about the role that TI plays in the fitness of this mutant.

Our results show that for the equivalent amount of viral RNA shed into the supernatant, the mutant viruses produce approximately 100 fold fewer infectious particles (Figure 3.5) than WT. One possibility for the difference is that the WT particles are more stable and produce a higher productive to defective particle ratio. There is at least one line of evidence to support this. Incubation of WT, LS, or LSTI virus at 41 °C prior to infection showed that the mutant virus was more sensitive to the temperature shift than WT, indicating particle stability may play some role in the fitness difference between variants (unpublished results, Jonah Rainey's thesis). Another possibility is that the WT virus is more efficient at infecting DF1 cells than mutant.

4.3 The Viral Generation Time

Our measurements for the viral generation time indicate that one generation is greater than 96 hours (Figure 3.15). This measurement was for WT virus. When we tried to use this protocol for LS and LSTI virus, infectivity was too poor to see the initial rise in GFP-positive cells (data not shown). It is unlikely, however, that the generation times of LS and LSTI are less than 96 hours given that LS and LSTI exhibit a replication defect compared to WT (Figure 3.18).

The observation that only a single generation elapses at each passage is important for two reasons. First, the population in culture behaves as if generations are discrete, which upholds an assumption made to simplify modeling. Second, each infectious particle that survives the bottleneck imposed at the time of passaging is only one reverse transcription step removed from the virus that originally infected the culture. Again, this simplifies modeling by justifying the method by which mutation is introduced into the simulation. Our measurements do not indicate at which point in the viral replication cycle the passage is terminated. Progeny virus may enter

uninfected cells and reach any point in the replication cycle just before expression. Additionally, because no new rounds of infection occur during the duration of a passage, the frequency of better-fit and less-fit infected cells are expected to be the same on day one of the passage as on the last day of passage.

It is interesting that a second round of infection does not occur under the tissue culture conditions used in this study. DF1 cells infected at an MOI of 0.25 show that supernatants harvested 44 hours post infection contain enough WT virus to infect approximately 2% of a fresh culture (Figure 3.16). By 72 hours, the amount of virus available for infection grows to 8%. The cultures from which they are harvested contain only 30% infected cells, which means there is no shortage of infectable cells (Figure 3.15).

The fact that virus is present in the supernatant of infected cultures, but not able to infect the uninfected cells in that culture is puzzling. It is possible that at later time points after the initial infection, the medium and/or cells are conditioned in a way that make the cells less susceptible to infection. This is consistent with the fact that virus harvested from the supernatants of infected cultures is able to initiate a new round of infection on freshly plated cells. One experiment to test this hypothesis would be to seed DF1 cells and incubate the cells for 72 hours before applying virus for infection.

Another possibility for the lack of a second round of infection in culture is that released virus may remain close to the cell from which it was produced. This could be due to either a direct association, or simply the lack of significant migration after particle release. This could explain why the physical act of passaging virus is able to initiate a new round of infection, but virus shed in the same culture as uninfected cells does not produce a detectable level of new infection. This hypothesis can be tested by repeating the generation time experiments under conditions where the cultures are agitated for the duration of the incubation.

The LS and LSTI mutants were previously shown to have a strong cytopathic effect on DF1 cells [Taplitz and Coffin, 1997]. We did not observe this cytopathic effect under the conditions of our experiments (Figure 3.8). This is consistent with results about the windows of cytopathic effect observed previously, which was shown to be 16-27 days post infection. The proposed mechanism for the cytopathic effect is the accumulation of viral DNA forms in cells due to superinfection. This accumulation leads to apoptosis of cells. Results here show that it takes at least 60 hours from the time of infection to observe retrovirus shed into the supernatant, and greater than 96 hours for a second round of infection to be observed.

4.4 Selection in the Context of Passaging

The fitness differences between WT and mutant viruses may be different when considered in the context of the passaging protocol. While experiments to determine the number of viral generations that elapse during one passage indicate that only a single generation of virus is shed, it is unclear which point in the replication cycle the progeny viruses reaches. If some of the progeny virus reaches at least the point of receptor binding on an uninfected cell, then the higher affinity of WT for receptor could reduce the fitness difference observed between WT and mutant viruses seen in a single culture. Since the passaging virus requires virus to be unbound and available for harvest in the supernatant, efficient receptor binding may sequester virus at the surface of cells.

4.5 Applications

Selection of drug resistant mutations in tissue culture has been performed by in at least one study [Doyon et al., 2005]. In this experiment, HIV was serially passaged

on C8166 cells in the continuous presence of Tipranavir, a protease inhibitor. After 9 months of continuous passage, virus was selected containing 10 mutations in the protease, conferring 87-fold reduced susceptibility to Tipranavir. The emergence of these mutations may have happened at a much more rapid rate if performed in a tissue culture system in which population size was constantly large and rapidly turning over. Increasing the rate with which the types of mutants that arise in response to drug inhibition may be valuable when considering design decisions for novel inhibitors.

4.6 Concluding Remarks

The drastic deviation of our tissue culture experiments with those of a simple model designed to describe them highlight the complexities that must exist in natural infections. Future work describing the relatively basic tissue culture system described here may be valuable in guiding modeling efforts in the future. Gaining an understanding of the stochastic events that occur at the sub-cellular level, as opposed to treating infection as simply the encounter of virus with cell, may provide further refinements to current models, improving their predictive capacity.

Appendix A

Glossary of Terms

Allele - A reference to one of two or more forms of a gene. In this thesis, three alleles are considered for the env gene of avian leukosis virus, wild-type, L154S (LS), and L154S T155I (LSTI).

Allele-specific PCR (AS-PCR) - A polymerase chain reaction (PCR) where the primers are designed to specifically generate a product by amplifying one allele among a template population that contains more than one allele of a particular gene.

Avian Leukosis Virus (ALV) - An alpharetrovirus that infects chickens and other fowl. Along with Rous sarcoma virus (RSV), ALV is one of the first retroviruses discovered and is the subject of study in this dissertation.

Bottleneck - A population genetics term that refers to a period or episode in which a large percentage of a population is killed, removed from the population, or otherwise prevented from reproducing.

Calibration quantity - A calibration quantity is an experimentally measurable phenomenon that is also predicted by a model. It is used to determine the value of something that is difficult to measure, and is also a parameter in the same model, by determining the value of the difficult-to-measure parameter that one must supply to the model to generate the experimentally observed calibration quantity.

Census population size (N) - The census size of a population is the total number of individuals in a population, regardless of whether they equally contribute progeny to the next generation of the population or not.

Cycle threshold (C_t) - The cycle threshold is a value for fluorescence intensity that is chosen to be above background and in the exponential phase of amplification in a real-time PCR reaction. The cycle threshold can be compared between real-time PCR reactions to determine relative amounts of starting template DNA or it can be compared to a standardized set of reactions to determine absolute quantities of starting template DNA in a reaction.

Deterministic model - A deterministic model does not take into account random events and allows precise determinations of the future state of a system.

DF1 cells - DF1 cells are spontaneously immortalized chicken cell line derived from chicken fibroblasts. They are from chickens to that are bred to be free of endogenous retrovirus elements.

Dideoxyinosine (ddI) - Also referred to as 2'-3'-didexoyinosine. ddI is a nucleoside analog that serves as a chain terminator during DNA polymerization by either reverse transcriptase or DNA polymerase.

Effective population size (N_e) - The effective population size is a population genetics concept that establishes the relationship between an idealized mathematical representation of a population and a natural biological population. Specifically, the ideal population may not account for all the factors that make the biological population susceptible to genetic drift. The effective population size accounts for these factors by scaling the size of the ideal population so that it experiences the same level of drift as the biological population.

Fixed allele - An allele becomes fixed when only one version of the allele is present in the population.

Genetic drift - Genetic drift, sometimes referred to as drift, is the fluctuation of

alleles in a population due to random sampling.

Generation time - The amount of time that elapses from one point in a the replication cycle of a virus to the same point in the progeny of this virus. For instance, if viral expression is the reference point under consideration, the generation time is time that elapses from expression of virus to the expression of daughter virus in another infected cell. There is no universal generation time for a particular type of virus. The generation time depends on the conditions of the infection, such as the availability of uninfected target cells, the density of target cells, the degree to which virus and cells are mixed, etc.

Linkage disequilibrium - The non-random association of two or more alleles, not necessarily on the same chromosome.

Locus - The location of a gene or DNA sequence on a chromosome.

LS virus - The avian leukosis virus subgroup B variant harboring the L154S mutation in the surface subunit (SU) of the envelope gene Env. This mutation confers a host-range extension phenotype.

LSTI virus - The avian leukosis virus subgroup B variant harboring the L154S and T155I mutations in the surface subunit (SU) of the envelope gene Env. These mutations confer a host-range extension phenotype.

Metapopulation - A group of populations of the same species that interact, but are spatially separated.

Monte Carlo simulation - A type of simulation that relies on the repeated random sampling of numbers to simulate a stochastic process.

Multiplicity of infection (MOI) - The ratio of infectious viral particles to cells that are able to support an infection.

Mutation - A change in the DNA or RNA sequence of the genome of an organism.

Neutral evolution - Neutral evolution is the theory that the vast majority of mutations to the genome of an organism do not result in a fitness difference.

Passage virus - The act of collecting an aliquot of virus from the supernatant of a culture and using this aliquot to initiate a new round of infection on new culture of uninfected cells.

Recombination - The breaking of DNA or RNA from one molecule and joining to another molecule.

Rous Sarcoma Virus (RSV) - Rous sarcoma virus is one of the first retroviruses to be discovered. It is oncogenic, causing sarcomas in chickens. RSV is an alpharetrovirus and harbors the oncogene *src*.

Selection - Selection is the evolutionary process by which an allele becomes more or less prevalent in a population. Different versions of an allele confer reproductive benefits or disadvantages that result in the differential contribution of progeny to the next generation.

Selection coefficient - The selection coefficient is a measure of how strong the force of selection is between two organisms.

Steady state - In terms of a viral infection, steady state refers to a virus population that does not change in size over time. HIV infections are thought to exist in a steady state during the chronic phase of infection, also known as the period of clinical latency.

Stochastic model - A stochastic model is a mathematical model that accounts for random events. Stochastic models make predictions in probabilistic terms such that the chance of observing a future state is predicted rather than the absolute future state itself.

Variance - A measure of how far a set of numbers are spread out from each other. Variance is calculated through the formula:

$$s_N^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{A.1}$$

s_N^2 is the variance. N is the number of samples or observations. $\bar{x}$ is the average if the samples or observations.

Wright-Fisher population - A theoretical population model used first introduced by Sewall Wright and R.A. Fisher. This population consists of an finite number of randomly mating diploid individuals. The individuals create gametes which randomly encounter the gametes of the other individuals, with each encounter resulting in an offspring, such that the population is hermaphroditic and mating is random. Because each individual is equally likely to contribute offspring to the next generation, this means there is no fitness difference between individuals and selection does not play a role. Generations are discreet, meaning there is no overlap between parent individuals and their offspring.

WT virus - The avian leukosis virus subgroup B variant that harbors no mutations in the surface subunit (SU) of the envelope gene Env.

Appendix B

Summary of Calculations

Cell Doubling Time

Calculation of the DF1 cell doubling time was done by fitting growth curves of uninfected cells or cells infected with WT, LS, or LSTI virus to the exponential growth model:

$$Y = Y_0 \times e^{(kX)} \tag{B.1}$$

where Y_0 is the initial number of cells, Y is the number of cells at time X, and k is the rate constant. The doubling time (T_d) was calculated as:

$$T_d = \frac{ln(2)}{k} \tag{B.2}$$

For uninfected, WT infected, LS infected, and LSTI infected, k equals 0.04347, 0.04054, 0.03979, and 0.03791, respectively.

Initial Frequency Better-Fit Allele

The initial frequency the better-fit allele (always WT in the experiments presented here) was estimated from the starting ratio of WT to mutant RNA copies. Impor-

tantly, the RNA copy number of WT or mutant virus was first used to estimate the number of functional particles of each virus. This was done by experimentally determining the number of functional particles per unit of RNA and generating a standard curve (Figure 3.5). From this standard curve, the following equations were used to estimate the number of functional particles of each viral variant used in this study:

$$\text{WT Particles} = 10^{(\log \text{WT RNA Copies} \times .7716) - 1.099} \tag{B.3}$$

$$\text{LS Particles} = 10^{(\log \text{LS RNA Copies} \times .6251) - 1.527} \tag{B.4}$$

$$\text{LSTI Particles} = 10^{(\log \text{LSTI RNA Copies} \times .6251) + 0.4936} \tag{B.5}$$

The initial frequency of the WT allele was then calculated using the formula:

$$\text{WT frequency} = \frac{\text{WT particles}}{\text{WT particles} + \text{Mutant particles}} \tag{B.6}$$

The mutant particles are the number of either LS or LSTI particles, depending on the competition experiment.

Selection Coefficients

Selection coefficients were calculated using the formula:

$$s = \frac{n'}{n} - 1 \tag{B.7}$$

Where s is the selection coefficient and n' and n are the relative growth rates of the less-fit and better-fit variants, respectively. The selection coefficient for the LS virus in the LS vs. WT competitions is:

$$s_{ls} = \frac{0.1275}{0.5954} - 1 \tag{B.8}$$

$$s_{ls} = -0.785 \tag{B.9}$$

The selection coefficient for the LSTI virus in the LSTI vs. WT competitions is:

$$s_{lsti} = \frac{0.05464}{0.5943} - 1 \tag{B.10}$$

$$s_{lsti} = -0.908 \tag{B.11}$$

Appendix C

Alternate Simulations

Simulations with parameter sets different from those presented in the main body of the thesis are presented in this section. The purpose of these simulations is to provide a more complete description of the behavior of the model used for simulations. In Figure C.1 and Figure C.2, the WT vs. LS and WT vs. LSTI experiments are simulated using starting frequencies of WT equal to the initial RNA copy frequency of WT. These frequencies are 10^{-3}, 10^{-4}, and 10^{-5} - which are far lower than the estimated initial frequencies of infectious WT particles.

The mean time to 50% WT and the variance in the mean time to 50% WT for these experiments are summarized in Figure C.3 and Figure C.4. Along with the summary statistics for the simulations, the summary statistics of the biological experiments are plotted for comparison (open circles in both figures). In both the LS vs. WT and LSTI vs. WT cases, the model is better predictor of the mean time to 50% WT of the biological experiments when the population size is set to 100. However, the variance in time to 50% is overestimated in both cases.

Finally, a lower effective population size than was tested in the main body of the thesis is shown in Figure C.5. At a population size 25, the model still predicts very low variance between replicates for both LS vs. WT and LSTI vs. WT experiments, with

all other parameters equal to the values used in the main body of the dissertation.

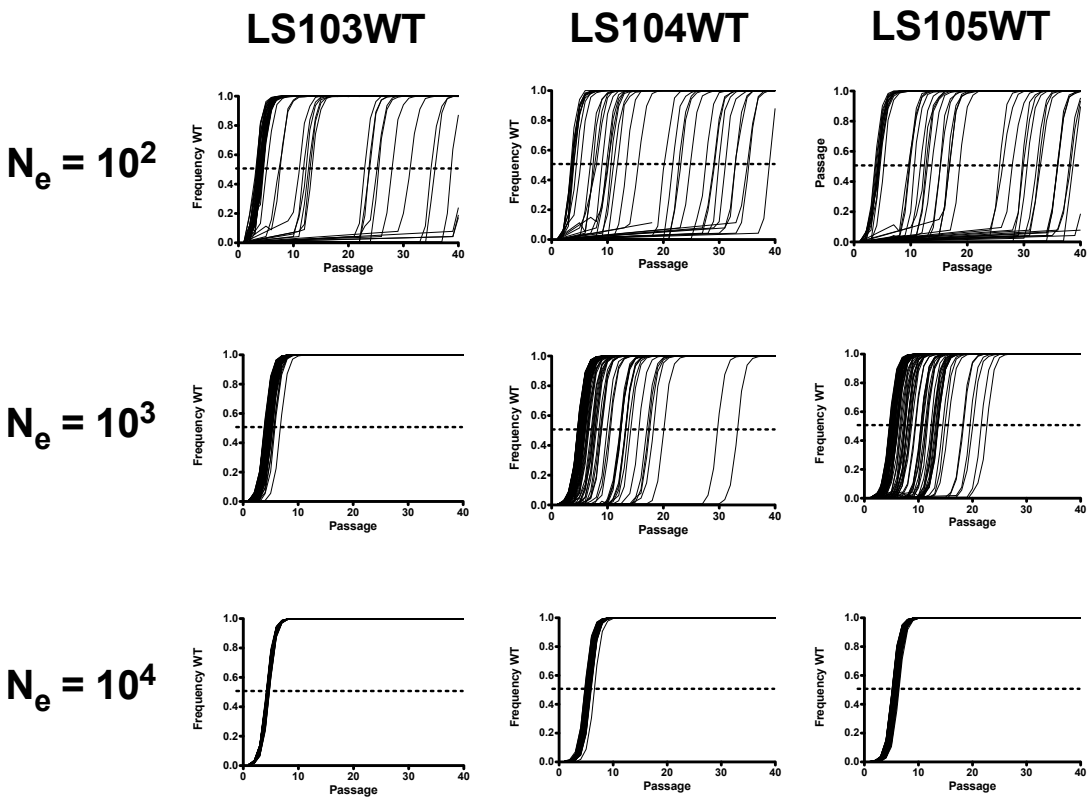

Figure C.1: **Alternate simulation results of competition experiments between WT and LS.** Each panel represents 50 replicate simulations of competition experiments based on the model described in Figure 3.19. The starting parameters for the simulation are: initial MOI (I_0) = 0.05; selection coefficient (s) = -0.785; cell doubling time (t_c) = 18 hours; viral generation time (t_v) = 1 per passage; fraction of virus passaged (f_p) = 0.25; initial frequency better fit allele (m) = 10^{-3}, 10^{-4}, or 10^{-5} for LS103WT, LS104WT, or LS105WT, respectively; number of cells at start of passage (C_0) = 3×10^5; mutation rate (μ) = 10^{-5}.

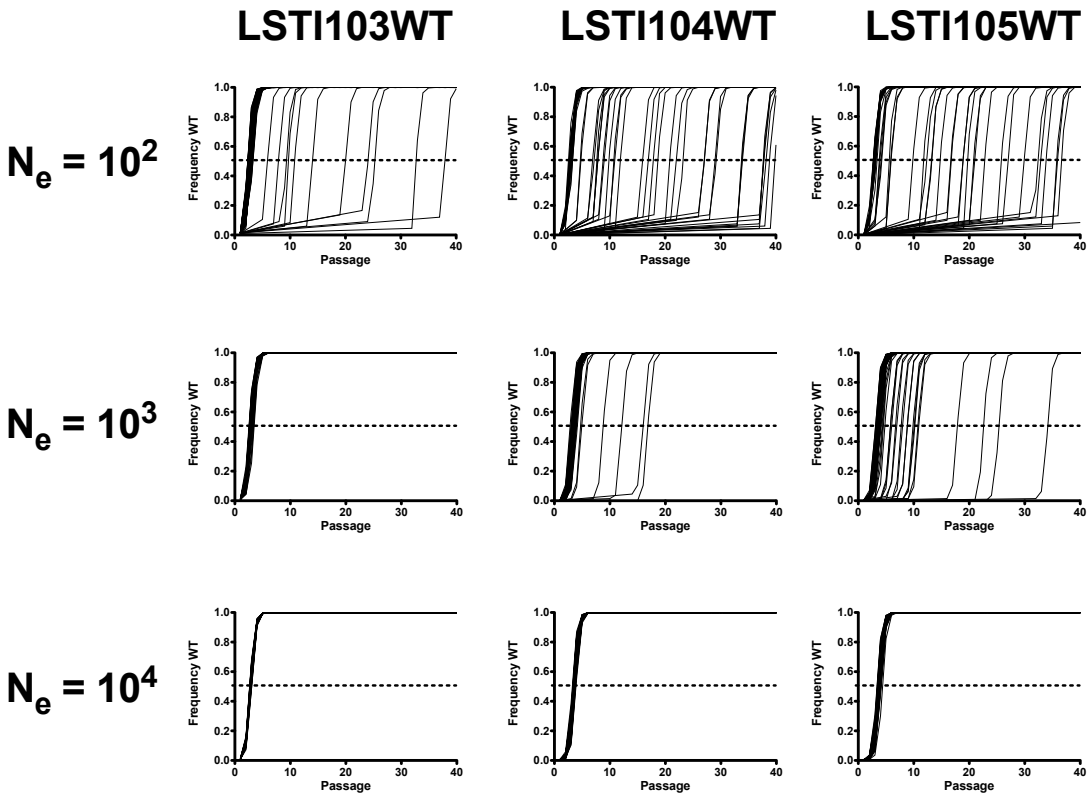

Figure C.2: **Alternate simulation results of competition experiments between WT and LSTI.** Each panel represents 50 replicate simulations of competition experiments based on the model described in Figure 3.19. The starting parameters for the simulation are: initial MOI (I_0) = 0.05; selection coefficient (s) = -0.908; cell doubling time (t_c) = 18 hours; viral generation time (t_v) = 1 per passage; fraction of virus passaged (f_p) = 0.25; initial frequency better fit allele (m) = 10^{-3}, 10^{-4}, or 10^{-5} for LSTI103WT, LSTI104WT, or LSTI105WT, respectively; number of cells at start of passage (C_0) = 3×10^5; mutation rate (μ) = 10^{-5}.

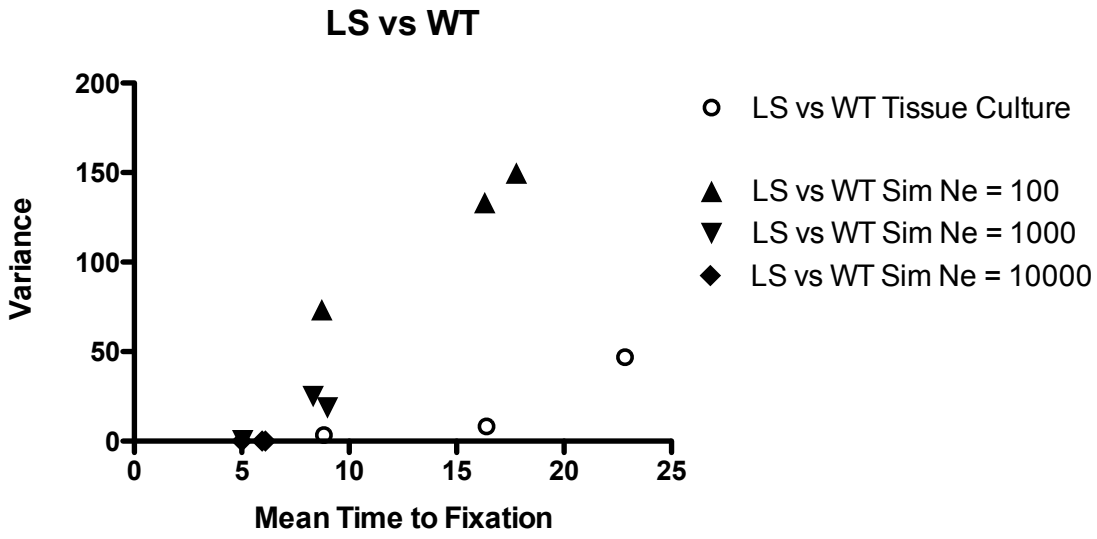

Figure C.3: **Summary of alternate simulation of competition experiments between WT and LS.** Summary of the variance and mean time to fixation of the WT allele for the alternate simulations presented in Figure C.1. Each symbol appears three times in the figure. From right to left, these symbols represent the decreasing initial frequency of WT described in the legend of Figure C.1.

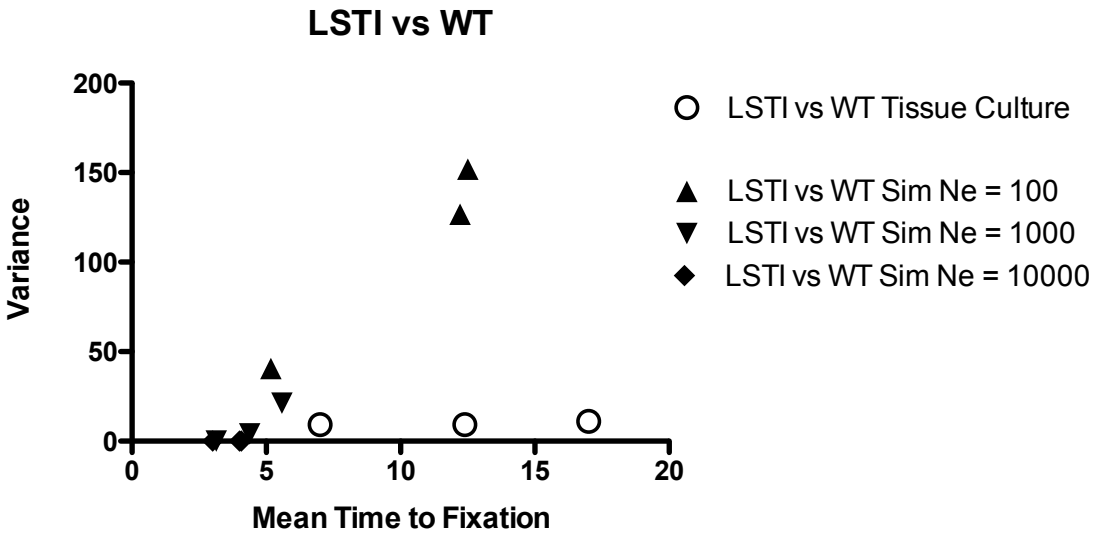

Figure C.4: **Summary of alternate simulation of competition experiments between WT and LSTI.** Summary of the variance and mean time to fixation of the WT allele for the alternate simulations presented in Figure C.2. Each symbol appears three times in the figure. From right to left, these symbols represent the decreasing initial frequency of WT described in the legend of Figure C.2

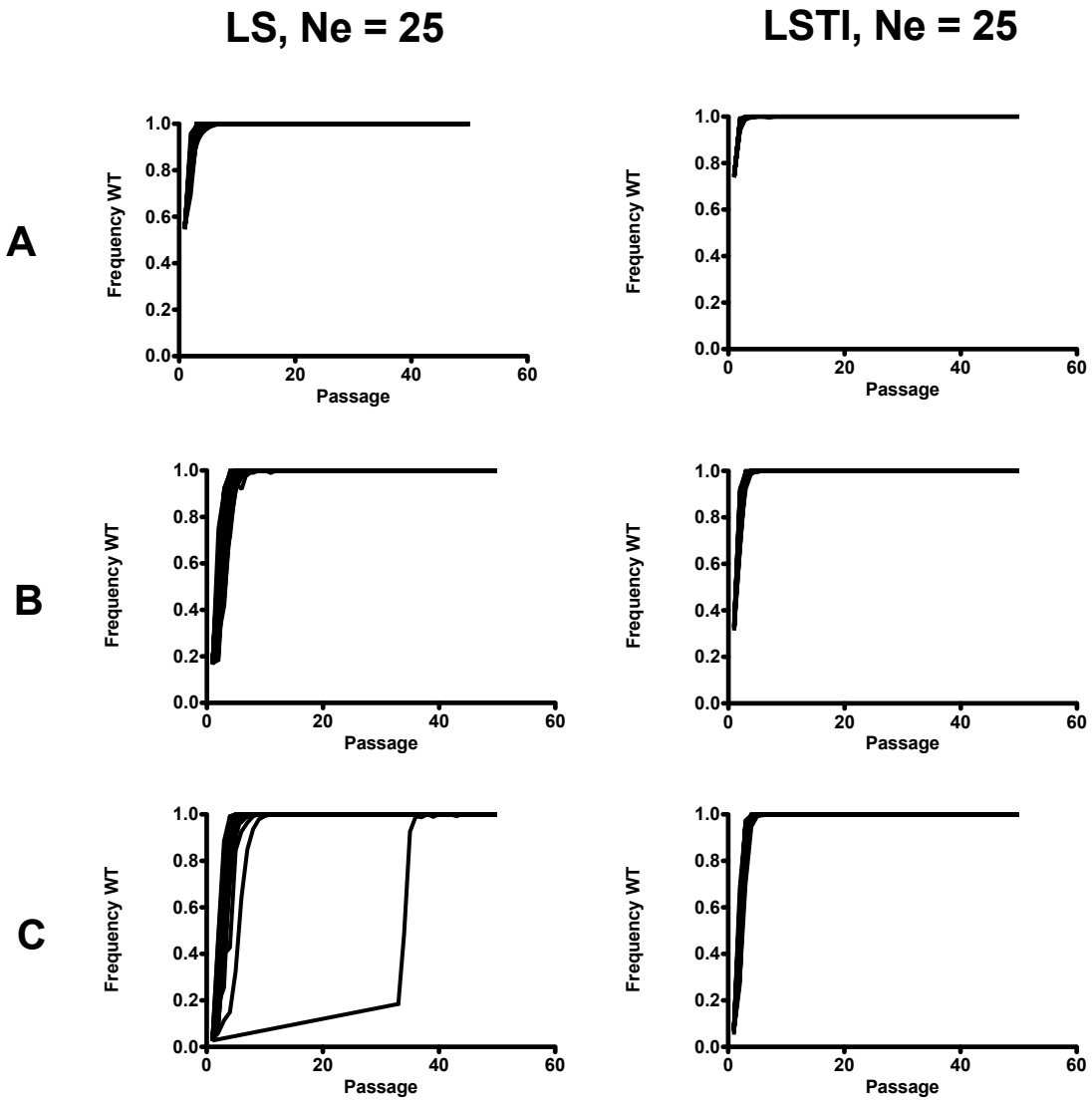

Figure C.5: **Results of alternate simulations with $\mathbf{N}_e$ = 25.** Each panel represents 50 replicate simulations of competition experiments based on the model described in Figure 3.19. The starting parameters for the simulation are: initial MOI (I_0) = 0.05; For left panels, selection coefficient (s) = -0.785, for right panels, s = -0.908; cell doubling time (t_c) = 18 hours; viral generation time (t_v) = 1 per passage; fraction of virus passaged (f_p) = 0.25; initial frequency better fit allele (m) = 0.213, 0.043, or 0.008 for left panels A, B. and C, respectively; initial frequency better fit allele (m) = 0.164, 0.032, or 0.006 for right panels A, B. and C; number of cells at start of passage (C_0) = 3×10^5; mutation rate (μ) = 10^{-5}. The population size in all panels is 25.

Appendix D

Source Code for Simulations

All simulations used for this work were written in the Ruby programming language. Listed below is the source code for the one-locus two-allele model.

```
===begin client.rb===
require 'Culture'
require 'yaml'

# Read experiment parameters from the file "config.yaml"
config = YAML::load(File.read('config.yaml'))
options = { :replicates => config['replicates'].to_i,
            :cells => config['cells'].to_i,
            :cell_doubling_time => config['cell_doubling_time'].to_i,
            :majority_allele => config['majority_allele'],
            :minority_allele => config['minority_allele'],
            :freq_minority_allele => config['freq_minority_allele'].to_f,
            :initial_moi => config['initial_moi'].to_f,
            :time_step => config['time_step'].to_i,
            :viral_generation_time => config['viral_generation_time'].to_i,
            :passages_per_week => config['passages_per_week'].to_i,
            :number_of_passages => config['number_of_passages'].to_i,
            :effective_moi => config['effective_moi'].to_f,
            :fraction_to_passage => config['fraction_to_passage'].to_f,
            :effective_population_size => config['effective_population_size'].to_i,
            :selection_coefficient => config['selection_coefficient'].to_f,
            :mutation_rate => config['mutation_rate'].to_f }

data = Array.new(options[:replicates]){[]}

# Create a new data file.  Overwrite existing datafile if it already exists.
File.open("data.out", "w") do | f |
  f.puts "Start of experiment: #{Time.now}"
end

replicate = 0
```

```
while replicate < options[:replicates]
  # Initialize the experiment with the parameters.  experiment is a Passage
  # object.
  experiment = Culture.new(options)

  #puts "client.rb: Working on replicate #{replicate + 1} of #{options[:replicates]}"

  experiment.add_cells(options[:cells]) #Adds cells to each culture, uses :cells number of cells in options
  experiment.add_virus(options[:freq_minority_allele], options[:initial_moi]) #Performs initial infection using :initial_moi and :freq_minority_allele

  total_time_steps = (options[:number_of_passages] / options[:passages_per_week]) * 168 / options[:time_step]
  short_passage_time_steps = 72 / options[:time_step]
  long_passage_time_steps = 96 / options[:time_step]

  # This while loop controls the alternation between a short 3 day passage and
  # a long 4 day passage.  This is how the experiment was done, two passages a
  # week, one short and one long.
  time_step = 0
  passage = 0
  short_passage = true

  while time_step < total_time_steps
    if short_passage
      puts "client.rb: replicate: #{replicate + 1} passage: #{passage + 1}"
      #puts "in short passage..."
      short_time_step = 0
      short_passage_time_steps.times do
        time_step += 1
        #puts "client.rb: timestep: #{time_step * options[:time_step]}"
        experiment.increment_time(options[:time_step])
        short_time_step = short_time_step + options[:time_step]

        #This if block tests if it is the last passage of the short passage, and
        #if so, calls the shed_virus method which causes all infected cells to
        #shed virus.  This code had to be added because the viral generation
        #time is longer than the passage length, meaning there are no new rounds
        #of infection, so the test "is it time to shed virus" that each cell has
        #never gets executed.  We need to force the issue.
        if short_time_step == 72
          experiment.shed_virus
          #puts "client.rb: Just shed virus"
        end
      end
      experiment.add_cells(options[:cells])
      experiment.passage!(options[:effective_population_size], options[:effective_moi], options[:fraction_to_passage])
      data[replicate][passage] = experiment.get_data
      #puts "client.rb: Just put data into array in short passage\n"
      experiment.infected_cells #This puts the number of infected cells""
      short_passage = false
      passage += 1
    else
      puts "client.rb: replicate: #{replicate + 1} passage: #{passage + 1}"
      #puts "client.rb: in long passage..."
      long_time_step = 0
```

```
      long_passage_time_steps.times do
        time_step += 1
        #puts "client.rb: timestep: #{time_step * options[:time_step]}"
        experiment.increment_time(options[:time_step])
        long_time_step = long_time_step + options[:time_step]

        #This if block tests if it is the last passage of the short passage, and
        #if so, calls the shed_virus method which causes all infected cells to
        #shed virus.  This code had to be added because the viral generation
        #time is longer than the passage length, meaning there are no new rounds
        #of infection.
        if long_time_step == 96
          experiment.shed_virus
          #puts "client.rb: Just shed virus"
        end
      end
      experiment.add_cells(options[:cells])
      experiment.passage!(options[:effective_population_size], options[:effective_moi], options[:fraction_to_passage])
      data[replicate][passage] = experiment.get_data
      #puts "client.rb: Just put data into array in long passage\n"
      experiment.infected_cells
      short_passage = true
      passage += 1
    end
  end
  replicate += 1
end

File.open("data.out", "a") do | f |
  passage = 0
  while (passage < options[:number_of_passages])
    replicate = 0
    while (replicate < options[:replicates])
      f.print "#{data[replicate][passage]}\t"
      replicate += 1
    end
    f.puts "\n"
    passage += 1
  end
end


File.open("data.out", "a") do | f |
  f.puts "\nEnd of experiment: #{Time.now}"
end
===end client.rb===

===begin culture.rb===
require 'Cell'

class Culture

  # Initialize
  def initialize(options)
```

```
 @options = options
 @cell_array = []
 @virus_array = []
 @@cell_doubling_time = options[:cell_doubling_time]
 @time_left_to_divide = options[:cell_doubling_time]
 @mutation_rate = options[:mutation_rate]
 @majority_allele = options[:majority_allele]
 @minority_allele = options[:minority_allele]
 @freq_minority_allele = options[:freq_minority_allele]
 @@viral_generation_time = options[:viral_generation_time]
 @@effective_moi = options[:effective_moi]
 @@fraction_to_passage = options[:fraction_to_passage]
 @@number_of_cells = options[:cells]
 @@effective_population_size = options[:effective_population_size]
 @majority_allele_count = 0.0
 @minority_allele_count = 0.0
 @sampled_majority_allele_count = 0.0
 @sampled_minority_allele_count = 0.0
 @sampled_freq_minority_allele = 0.0
 @virus_just_shed = FALSE
 @@selection_coefficient = options[:selection_coefficient]
end

# Initializes cell_array and then loads cell_array with number_of_cells
# number of cells
def add_cells(number_of_cells)
  #puts "culture.rb: seeding sells..."
  @cell_array = []
  number_of_cells.times do
    @cell_array << 0
  end
end

# Randomly infect a user definded number or percentage of cells at the
# start of the experiment.  This loop can get stuck if all of the cells in
# the culture are infected.  TODO: add a check to see if all of the cells
# are infected and break from the program with an error.
def add_virus(freq_minority_allele, initial_moi)
  number_of_cells = @cell_array.length
  puts "culture.rb: add_virus: freq_minority_allele: #{@freq_minority_allele}"
  (initial_moi * number_of_cells).to_i.times do

    index = rand(number_of_cells)

    until @cell_array[index] == 0
      index = rand(number_of_cells)
    end

    if rand >= @freq_minority_allele
      if rand <= @mutation_rate #Mutate virus at frequency of mutation_rate
        @cell_array[index] = @minority_allele
      else
        @cell_array[index] = @majority_allele #If no mutation, infect with original identity virus
      end
    else
```

```
      if rand <= @mutation_rate #Mutate virus at frequency of mutation_rate
        @cell_array[index] = @majority_allele
      else
        @cell_array[index] = @minority_allele #If no mutation, infect with orginal identity virus
      end
    end

  end
end

# Iterates through the cell array and calls the increment time method on
# each cell object.  Next, it tests two things.  First, if it is time to
# divide, the cell object is deep copied my marshalling and added to the end
# of the cell array.  Second, it tests if it is time for a cell to shed
# virus.  The way that uninfected cells are found is very inefficient.
# TODO: store infected and uninfected cells in different arrays.  That way,
# we don't have to search for uninfected cells until we find one.  We can
# just infect, then move to the infected array.
def increment_time(time_step)
  #puts "culture.rb: time_step: #{time_step}"
  @time_left_to_divide = @time_left_to_divide - time_step
  #puts "culture:rb: time_left_to_divide: #{@time_left_to_divide}"
  number_of_cells = @cell_array.length
  if @time_left_to_divide <= 0
    #puts "culture.rb: cells dividing..."
    i = 0
    number_of_cells.times do
      @cell_array << @cell_array[i]
      i += 1
    end
    @time_left_to_divide = @@cell_doubling_time
  end
end

#This
def passage!(effective_population_size, effective_moi, fraction_to_passage)
  #Set the number of cells = to the total number of cells in the cell_array
  #This is necessary because the cells are dividing and we need to know how
  #many we have at the end of the passage
  number_of_cells = @cell_array.length

  #Calculate the frequency of the minority allele.  This is done by looking
  #at the allele counts set in shed_virus
  @freq_minority_allele = @minority_allele_count.to_f / (@minority_allele_count.to_f + @majority_allele_count.to_f)
  puts "culture.rb: minority allele count: #{@minority_allele_count}"
  puts "culture.rb: majority allele count: #{@majority_allele_count}"
  #puts "culture.rb: freq minority allele: #{@freq_minority_allele}"
  #puts "culture.rb: number of cells: #{number_of_cells}"

  #Randomly sample fraction_to_passage * effective_population_size of the
  #population.  Sampling is done without replacement.
  (@majority_allele_count + @minority_allele_count * @@fraction_to_passage).to_i.times do
    if rand >= @freq_minority_allele
      @majority_allele_count -= 1
      @sampled_majority_allele_count += 1
```

```
    else
      @minority_allele_count -= 1
      @sampled_minority_allele_count += 1
    end
    @freq_minority_allele = @minority_allele_count.to_f / (@minority_allele_count.to_f + @majority_allele_count.to_f)
  end

  #Calculate the allele frequency of randomly sampled 25% of original viral
  #population.
  @sampled_freq_minority_allele = @sampled_minority_allele_count.to_f / (@sampled_minority_allele_count.to_f + @sampled_majority_allele_count.to_f)
  puts "culture.rb: sampled freq minority allele: #{@sampled_freq_minority_allele}"

  #The following line was changed to infect a set # of cells at each
  #passage.  That number is the effective_moi * number_of_cells.  We use
  #the sampled_freq_minority_allele because only 25% of the population was
  #used to infect the next culture
  (@@effective_moi * @@effective_population_size).to_i.times do

     index = rand(number_of_cells)
     until @cell_array[index] == 0
       index = rand(number_of_cells)
     end

     if rand >= @sampled_freq_minority_allele
       if rand <= @mutation_rate
         @cell_array[index] = @minority_allele
       else
         @cell_array[index] = @majority_allele
       end
     else
       if rand <= @mutation_rate
         @cell_array[index] = @majority_allele
       else
         @cell_array[index] = @minority_allele
       end
     end

  end
  @majority_allele_count = 0
  @minority_allele_count = 0
  @sampled_majority_allele_count = 0
  @sampled_minority_allele_count = 0
end

def shed_virus
  number_of_cells = @cell_array.length
  i = 0
  while (i < number_of_cells )
    #puts "testing cell #{i}"
    if @cell_array[i] == @majority_allele
      #puts "in majority infected shed"
      @majority_allele_count = @majority_allele_count + 1.0
    elsif @cell_array[i] == @minority_allele
     #puts "in minority infected shed"
      @minority_allele_count = @minority_allele_count + (1.0 / (1.0 - @@selection_coefficient))
```

```
    end
    i += 1
  end
  @freq_minority_allele = @minority_allele_count / ( @minority_allele_count + @majority_allele_count)
  puts "culture.rb: shed_virus: minority allele_count: #{@minority_allele_count}"
end


def cells
  @cell_array.length
end

def infected_cells
  count = 0
  @cell_array.each do |x|
    if x != 0
      count += 1
    end
  end
  count
end

# majority_freq returns the frequency of the majority allele.  In this
# function, the cell_array is iterated for all the cells infected with the
# @majority_allele and then the cell_array is iterated again for all the
# cells that are infected.  This can probably be done more efficiently, with
# one iteration.
def majority_freq
  major_allele_count = 0
  @cell_array.each do |x|
    if x == @majority_allele
      major_allele_count += 1
    end
  end

  infected_count = 0
  @cell_array.each do |x|
    if x == TRUE
      infected_count += 1
    end
  end
  major_allele_count.to_f / infected_count.to_f
end

def get_data
  #This block truncates the float to 5 decimal places.
  #floated_freq_minority_allele = (@freq_minority_allele * 100000).round.to_f / 100000 #Reduce number of decimal points for clean output
  @freq_minority_allele
end

private

  # Test if all cells are infected before trying to infect
  def all_cells_infected?
    total_cells = @cell_array.length
```

```
      total_infected_cells = @cell_array.inject(0) do | sum, cell |
        if cell.infected?
          sum += 1
        else
          sum
        end
      end
      total_cells == total_infected_cells
    end
end
===end culture.rb==

===begin cell.rb===

class Cell

  attr_accessor :divide, :shed_virus

  def initialize(options)
    @infected = FALSE
    @allele = 'none'
    @@cell_doubling_time = options[:cell_doubling_time]
    @time_to_division = options[:cell_doubling_time]
    @divide = FALSE
    @shed_virus = FALSE
    @@viral_generation_time = options[:viral_generation_time]
  end

  #TODO
  def infect(freq_minority_allele, majority_allele, minority_allele)
    @infected = TRUE
    @time_to_shed_virus = @@viral_generation_time #This line is not necessary now that virus is shed once / passage
    if rand >= freq_minority_allele
      @allele = majority_allele
    else
      @allele = minority_allele
    end
  end

  def time_to_shed_virus?
    @shed_virus
  end

  def increment_time(time_step)
    @time_to_division -= time_step
    if @time_to_division <= 0
      @time_to_division = @@cell_doubling_time
      @divide = TRUE
    end

    #if @time_to_shed_virus
    #  @time_to_shed_virus -= time_step
    #  if @time_to_shed_virus <= 0
    #    @time_to_shed_virus = @@viral_generation_time
    #    @shed_virus = TRUE
```

```
    #  end
    #end
  end

  def time_to_divide?
    @divide
  end

  def infected?
    @infected
  end

  def allele?
    @allele
  end

end

==end cell.rb===
```

Bibliography

[Achaz et al., 2004] Achaz, G., Palmer, S., Kearney, M., Maldarelli, F., Mellors, J. W., Coffin, J. M., and Wakeley, J. (2004). A robust measure of hiv-1 population turnover within chronically infected individuals. *Mol Biol Evol*, 21(10):1902–12.

[Adkins et al., 1997] Adkins, H. B., Brojatsch, J., Naughton, J., Rolls, M. M., Pesola, J. M., and Young, J. A. (1997). Identification of a cellular receptor for subgroup e avian leukosis virus. *Proc Natl Acad Sci USA*, 94(21):11617–22.

[Adkins et al., 2000] Adkins, H. B., Brojatsch, J., and Young, J. A. (2000). Identification and characterization of a shared tnfr-related receptor for subgroup b, d, and e avian leukosis viruses reveal cysteine residues required specifically for subgroup e viral entry. *J Virol*, 74(8):3572–78.

[Aschoff et al., 1999] Aschoff, J. M., Foster, D., and Coffin, J. M. (1999). Point mutations in the avian sarcoma/leukosis virus 3' untranslated region result in a packaging defect. *J Virol*, 73(9):7421–9.

[Astrin et al., 1979] Astrin, S. M., Buss, E. G., and Haywards, W. S. (1979). Endogenous viral genes are non-essential in the chicken. *Nature*, 282(5736):339–41.

[Baltimore, 1970] Baltimore, D. (1970). Rna-dependent dna polymerase in virions of rna tumour viruses. *Nature*, 226(5252):1209–11.

[Barnard et al., 2006] Barnard, R. J. O., Elleder, D., and Young, J. A. T. (2006). Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. *Virology*, 344(1):25–9.

[Barnard et al., 2004] Barnard, R. J. O., Narayan, S., Dornadula, G., Miller, M. D., and Young, J. A. T. (2004). Low ph is required for avian sarcoma and leukosis virus env-dependent viral penetration into the cytosol and not for viral uncoating. *J Virol*, 78(19):10433–41.

[Barré-Sinoussi et al., 1983] Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a t-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (aids). *Science (New York, NY)*, 220(4599):868–71.

[Bates et al., 1993] Bates, P., Young, J. A., and Varmus, H. E. (1993). A receptor for subgroup a rous sarcoma virus is related to the low density lipoprotein receptor. *Cell*, 74(6):1043–51.

[Battula and Loeb, 1976] Battula, N. and Loeb, L. A. (1976). On the fidelity of dna replication. lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus dna polymerase. *J Biol Chem*, 251(4):982–6.

[Bélanger et al., 1995] Bélanger, C., Zingler, K., and Young, J. A. (1995). Importance of cysteines in the ldlr-related domain of the subgroup a avian leukosis and sarcoma virus receptor for viral entry. *J Virol*, 69(2):1019–24.

[Belshaw et al., 2004] Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., and Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA*, 101(14):4894–9.

[Boeke and Stoye, 1997] Boeke, J. D. and Stoye, J. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. (plainview, n.y., cold spring harbor laboratory press). pages 343–435.

[Bonhoeffer et al., 2002] Bonhoeffer, S., Barbour, A., and Boer, R. D. (2002). Procedures for reliable estimation of viral fitness from time-series data. *Proc Biol Sci*, 269(1503):1887–93. (0) Journal Article England.

[Bonhoeffer et al., 1995] Bonhoeffer, S., Holmes, E. C., and Nowak, M. A. (1995). Causes of hiv diversity. *Nature*, 376(6536):125.

[Bonhoeffer et al., 1997] Bonhoeffer, S., May, R. M., Shaw, G. M., and Nowak, M. A. (1997). Virus dynamics and drug therapy. *Proc Natl Acad Sci USA*, 94(13):6971–6.

[Bonhoeffer and Nowak, 1997] Bonhoeffer, S. and Nowak, M. A. (1997). Preexistence and emergence of drug resistance in hiv-1 infection. *Proc Biol Sci*, 264(1382):631–7.

[Boyce-Jacino et al., 1992] Boyce-Jacino, M. T., O'Donoghue, K., and Faras, A. J. (1992). Multiple complex families of endogenous retroviruses are highly conserved in the genus gallus. *J Virol*, 66(8):4919–29.

[Brown, 1997a] Brown, A. J. (1997a). Analysis of hiv-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc Natl Acad Sci USA*, 94(5):1862–5.

[Brown, 1997b] Brown, P. (1997b). Integration. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. (plainview, n.y., cold spring harbor laborotary press). pages 161–204.

[Bruce et al., 2005] Bruce, J. W., Bradley, K. A., Ahlquist, P., and Young, J. A. T. (2005). Isolation of cell lines that show novel, murine leukemia virus-specific blocks to early steps of retroviral replication. *J Virol*, 79(20):12969–78.

[Bullough et al., 1994] Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the ph of membrane fusion. *Nature*, 371(6492):37–43.

[Chan et al., 1997] Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the hiv envelope glycoprotein. *Cell*, 89(2):263–73.

[Chao, 1990] Chao, L. (1990). Fitness of rna virus decreased by muller's ratchet. *Nature*, 348(6300):454–5.

[CHARLESWORTH, 2000] CHARLESWORTH, B. (2000). ... . a complete variorum edition. by ra fisher (edited with foreword and notes by jh bennett). oxford university press. 1999. isbn 0-19-850440-3. xxi+ 318 pages. .... *Genetics Research*.

[Charlesworth, 2009] Charlesworth, B. (2009). Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.*, 10(3):195–205.

[Coffin, 1995] Coffin, J. (1995). Hiv population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, 267(5197):483–9.

[Coffin, 1997] Coffin, J. (1997). Retroviruses.

[Coffin et al., 1980] Coffin, J., Tsichlis, P., Barker, C., Voynow, S., and Robinson, H. (1980). Variation in avian retrovirus genomes. *Ann N Y Acad Sci*, 354:410–25.

[Coffin, 1979] Coffin, J. M. (1979). Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol*, 42(1):1–26.

[Coffin, 1994] Coffin, J. M. (1994). Hiv population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, 267(5197):483–9.

[Coffin, 1996] Coffin, J. M. (1996). Retroviridae: The viruses and their replication. in fields virology, b.n. fields, d.m. knipe, and p.m. howley, eds. (new york philadelphia, raven press lippincott-raven). pages 1767–1847.

[Coffin et al., 1978] Coffin, J. M., Champion, M., and Chabot, F. (1978). Nucleotide sequence relationships between the genomes of an endogenous and an exogenous avian tumor virus. *J Virol*, 28(3):972–91.

[Coffin et al., 1983] Coffin, J. M., Tsichlis, P. N., Conklin, K. F., Senior, A., and Robinson, H. L. (1983). Genomes of endogenous and exogenous avian retroviruses. *Virology*, 126(1):51–72.

[Crittenden and Motta, 1975] Crittenden, L. B. and Motta, J. V. (1975). The role of the tvb locus in genetic resistance to rsv(rav-o). *Virology*, 67(2):327–34.

[Crittenden et al., 1973] Crittenden, L. B., Wendel, E. J., and Motta, J. V. (1973). Interaction of genes controlling resistance to rsv(rav-o). *Virology*, 52(2):373–84.

[de la Iglesia and Elena, 2007] de la Iglesia, F. and Elena, S. F. (2007). Fitness declines in tobacco etch virus upon serial bottleneck transfers. *J Virol*, 81(10):4941–7.

[Dorner and Coffin, 1986] Dorner, A. and Coffin, J. (1986). Determinants for receptor interaction and cell killing on the avian retrovirus glycoprotein gp85. *Cell*, 45(3):365–74.

[Dorner et al., 1985] Dorner, A., Stoye, J., and Coffin, J. (1985). Molecular basis of host range variation in avian retroviruses. *J Virol*, 53(1):32–9.

[Doyon et al., 2005] Doyon, L., Tremblay, S., Bourgon, L., Wardrop, E., and Cordingley, M. G. (2005). Selection and characterization of hiv-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. *Antiviral Res*, 68(1):27–35.

[Dunwiddie et al., 1986] Dunwiddie, C. T., Resnick, R., Boyce-Jacino, M., Alegre, J. N., and Faras, A. J. (1986). Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. *J Virol*, 59(3):669–75.

[Dussupt et al., 2010] Dussupt, V., Sette, P., Bello, N. F., Javid, M. P., Nagashima, K., and Bouamr, F. (2010). Basic residues in the nucleocapsid domain of gag are critical for late events of hiv-1 budding. *J Virol*.

[Earp et al., 2003] Earp, L. J., Delos, S. E., Netter, R. C., Bates, P., and White, J. M. (2003). The avian retrovirus avian sarcoma/leukosis virus subtype a reaches the lipid mixing stage of fusion at neutral ph. *J Virol*, 77(5):3058–66.

[Elleder et al., 2004] Elleder, D., Plachý, J., Hejnar, J., Geryk, J., and Svoboda, J. (2004). Close linkage of genes encoding receptors for subgroups a and c of avian sarcoma/leucosis virus on chicken chromosome 28. *Anim Genet*, 35(3):176–81.

[Elleder et al., 2005] Elleder, D., Stepanets, V., Melder, D. C., Senigl, F., Geryk, J., Pajer, P., Plachý, J., Hejnar, J., Svoboda, J., and Federspiel, M. J. (2005). The receptor for the subgroup c avian sarcoma and leukosis viruses, tvc, is related to mammalian butyrophilins, members of the immunoglobulin superfamily. *J Virol*, 79(16):10408–19.

[Ellermann, 1909] Ellermann, V. (1909). Experimentelle leukämie bei hühnern. ii. *Medical Microbiology and Immunology*.

[Fischer et al., 2010] Fischer, W., Ganusov, V. V., Giorgi, E. E., Hraber, P. T., Keele, B. F., Leitner, T., Han, C. S., Gleasner, C. D., Green, L., Lo, C.-C., Nag, A., Wall-

strom, T. C., Wang, S., McMichael, A. J., Haynes, B. F., Hahn, B. H., Perelson, A. S., Borrow, P., Shaw, G. M., Bhattacharya, T., and Korber, B. T. (2010). Transmission of single hiv-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One*, 5(8):e12303.

[Fisher, 1958] Fisher, R. A. (1958). The genetical theory of natural selection (2. rev. ed.).. page 291.

[Frisby et al., 1979] Frisby, D. P., Weiss, R. A., Roussel, M., and Stehelin, D. (1979). The distribution of endogenous chicken retrovirus sequences in the dna of galliform birds does not coincide with avian phylogenetic relationships. *Cell*, 17(3):623–34.

[Gallo et al., 1984] Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., and Safai, B. (1984). Frequent detection and isolation of cytopathic retroviruses (htlv-iii) from patients with aids and at risk for aids. *Science (New York, NY)*, 224(4648):500–3.

[Gao and Goff, 1999] Gao, G. and Goff, S. P. (1999). Somatic cell mutants resistant to retrovirus replication: intracellular blocks during the early stages of infection. *Mol Biol Cell*, 10(6):1705–17.

[Gilbert et al., 1990] Gilbert, J. M., Mason, D., and White, J. M. (1990). Fusion of rous sarcoma virus with host cells does not require exposure to low ph. *J Virol*, 64(10):5106–13.

[Goff, 2001] Goff, S. P. (2001). Retroviridae: The viruses and their replication. in field's virology, b.n. fields, d.m. knipe, and p.m. howley, eds. (philadelphia, lippincott, williams, & wilkins). pages 1871–1939.

[Günthard et al., 1998] Günthard, H. F., Wong, J. K., Ignacio, C. C., Havlir, D. V., and Richman, D. D. (1998). Comparative performance of high-density oligonu-

cleotide sequencing and dideoxynucleotide sequencing of hiv type 1 pol from clinical samples. *AIDS Res Hum Retroviruses*, 14(10):869–76.

[Haase et al., 1996] Haase, A. T., Henry, K., Zupancic, M., Sedgewick, G., Faust, R. A., Melroe, H., Cavert, W., Gebhard, K., Staskus, K., Zhang, Z. Q., Dailey, P. J., Balfour, H. H., Erice, A., and Perelson, A. S. (1996). Quantitative image analysis of hiv-1 infection in lymphoid tissue. *Science (New York, NY)*, 274(5289):985–9.

[Hance et al., 2001] Hance, A. J., Lemiale, V., Izopet, J., Lecossier, D., Joly, V., Massip, P., Mammano, F., Descamps, D., Brun-Vézinet, F., and Clavel, F. (2001). Changes in human immunodeficiency virus type 1 populations after treatment interruption in patients failing antiretroviral therapy. *J Virol*, 75(14):6410–7.

[Harrison, 2008] Harrison, S. C. (2008). Viral membrane fusion. *Nat Struct Mol Biol*, 15(7):690–8.

[Himly et al., 1998] Himly, M., Foster, D. N., Bottoli, I., Iacovoni, J. S., and Vogt, P. K. (1998). The df-1 chicken fibroblast cell line: transformation induced by diverse oncogenes and cell death resulting from infection by avian leukosis viruses. *Virology*, 248(2):295–304.

[Ho et al., 1995] Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. (1995). Rapid turnover of plasma virions and cd4 lymphocytes in hiv-1 infection. *Nature*, 373(6510):123–6.

[Hu and Temin, 1990] Hu, W. S. and Temin, H. M. (1990). Genetic consequences of packaging two rna genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci USA*, 87(4):1556–60.

[Hurley et al., 2010] Hurley, J. H., Boura, E., Carlson, L.-A., and Różycki, B. (2010). Membrane budding. *Cell*, 143(6):875–87.

[Jacks and Varmus, 1985] Jacks, T. and Varmus, H. E. (1985). Expression of the rous sarcoma virus pol gene by ribosomal frameshifting. *Science (New York, NY)*, 230(4731):1237–42.

[Jones et al., 1994] Jones, J. S., Allan, R. W., and Temin, H. M. (1994). One retroviral rna is sufficient for synthesis of viral dna. *J Virol*, 68(1):207–16.

[Klarmann et al., 1993] Klarmann, G. J., Schauber, C. A., and Preston, B. D. (1993). Template-directed pausing of dna synthesis by hiv-1 reverse transcriptase during polymerization of hiv-1 sequences in vitro. *J Biol Chem*, 268(13):9793–802.

[Knipe and Howley, 2006] Knipe, D. and Howley, P. (2006). Fields virology.

[Kouyos et al., 2006] Kouyos, R., Althaus, C., and Bonhoeffer, S. (2006). Stochastic or deterministic: what is the effective population size of hiv-1? *Trends Microbiol*, 14(12):507–11.

[Kuff et al., 1972] Kuff, E. L., Leuders, K. K., Ozer, H. L., and Wivel, N. A. (1972). Some structural and antigenic properties of intracisternal a particles occurring in mouse tumors (complement fixation-immunodiffusion-neuroblastoma-plasma-cell tumor). *Proc Natl Acad Sci USA*, 69(1):218–22.

[Leider et al., 1988] Leider, J., Palese, P., and Smith, F. (1988). Determination of the mutation rate of a retrovirus. *J Virol*, 62(9):3084–91.

[Levy, 1993] Levy, J. A. (1993). The retroviridae. page 458.

[Levy and Shimabukuro, 1985] Levy, J. A. and Shimabukuro, J. (1985). Recovery of aids-associated retroviruses from patients with aids or aids-related conditions and from clinically healthy individuals. *J Infect Dis*, 152(4):734–8.

[Liu et al., 2006] Liu, Y., Mullins, J. I., and Mittler, J. E. (2006). Waiting times for the appearance of cytotoxic t-lymphocyte escape mutants in chronic hiv-1 infection. *Virology*, 347(1):140–6.

[Mansky and Temin, 1995] Mansky, L. and Temin, H. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol*, 69(8):5087–94.

[Mansky and Temin, 1994] Mansky, L. M. and Temin, H. M. (1994). Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. *J Virol*, 68(1):494–9. BLV mutation rate per bp per cycle: 4.8*10^{-}6.

[Maree et al., 2000] Maree, A., Keulen, W., Boucher, C., and Boer, R. D. (2000). Estimating relative fitness in viral competition experiments. *J Virol*, 74(23):11067–72.

[Morgan et al., 1976] Morgan, D. A., Ruscetti, F. W., and Gallo, R. (1976). Selective in vitro growth of t lymphocytes from normal human bone marrows. *Science (New York, NY)*, 193(4257):1007–8.

[Mothes et al., 2000] Mothes, W., Boerger, A. L., Narayan, S., Cunningham, J. M., and Young, J. A. (2000). Retroviral entry mediated by receptor priming and low ph triggering of an envelope glycoprotein. *Cell*, 103(4):679–89.

[Mulder et al., 1994] Mulder, J., McKinney, N., Christopherson, C., Sninsky, J., Greenfield, L., and Kwok, S. (1994). Rapid and simple pcr assay for quantitation of human immunodeficiency virus type 1 rna in plasma: application to acute retroviral infection. *J Clin Microbiol*, 32(2):292–300.

[Mullins et al., 1986] Mullins, J. I., Chen, C. S., and Hoover, E. A. (1986). Disease-specific and tissue-specific production of unintegrated feline leukaemia virus variant dna in feline aids. *Nature*, 319(6051):333–6.

[Munguia and Federspiel, 2008] Munguia, A. and Federspiel, M. J. (2008). Efficient subgroup c avian sarcoma and leukosis virus receptor activity requires the igv domain of the tvc receptor and proper display on the cell membrane. *J Virol*, 82(22):11419–28.

[Murphy et al., 1995] Murphy, F., Fauquet, C., Bishop, D., Ghabrial, S., Jarvis, A., Martelli, G., Mayo, M., and Summers, M. (1995). Virus taxonomy: Sixth report of the internation committee on the taxonomy of viruses.

[Narayan and Young, 2004] Narayan, S. and Young, J. A. T. (2004). Reconstitution of retroviral fusion and uncoating in a cell-free system. *Proc Natl Acad Sci USA*, 101(20):7721–6.

[Netter et al., 2004] Netter, R. C., Amberg, S. M., Balliet, J. W., Biscone, M. J., Vermeulen, A., Earp, L. J., White, J. M., and Bates, P. (2004). Heptad repeat 2-based peptides inhibit avian sarcoma and leukosis virus subgroup a infection and identify a fusion intermediate. *J Virol*, 78(24):13430–9.

[Nijhuis et al., 1998] Nijhuis, M., Boucher, C. A., Schipper, P., Leitner, T., Schuurman, R., and Albert, J. (1998). Stochastic processes strongly influence hiv-1 evolution during suboptimal protease-inhibitor therapy. *Proc Natl Acad Sci USA*, 95(24):14441–6.

[Novella et al., 1999] Novella, I. S., Quer, J., Domingo, E., and Holland, J. J. (1999). Exponential fitness gains of rna virus populations are limited by bottleneck effects. *J Virol*, 73(2):1668–71.

[Overbaugh et al., 1992] Overbaugh, J., Hoover, E. A., Mullins, J. I., Burns, D. P., Rudensey, L., Quackenbush, S. L., Stallard, V., and Donahue, P. R. (1992). Structure and pathogenicity of individual variants within an immunodeficiency disease-inducing isolate of felv. *Virology*, 188(2):558–69.

[Palmer et al., 2006] Palmer, S., Boltz, V., Maldarelli, F., Kearney, M., Halvas, E., Rock, D., Falloon, J., Davey, R., Dewar, R., Metcalf, J., Mellors, J., and Coffin, J. (2006). Selection and persistence of non-nucleoside reverse transcriptase inhibitor-resistant hiv-1 in patients starting and stopping non-nucleoside therapy. *Aids*, 20(5):701–10.

[Palmer et al., 2005] Palmer, S., Kearney, M., Maldarelli, F., Halvas, E. K., Bixby, C. J., Bazmi, H., Rock, D., Falloon, J., Davey, R. T., Dewar, R. L., Metcalf, J. A., Hammer, S., Mellors, J. W., and Coffin, J. M. (2005). Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J Clin Microbiol*, 43(1):406–13.

[Palmiter et al., 1978] Palmiter, R. D., Gagnon, J., Vogt, V. M., Ripley, S., and Eisenman, R. N. (1978). The nh2-terminal sequence of the avian oncovirus gag precursor polyprotein (pr76gag). *Virology*, 91(2):423–33.

[Pathak and Temin, 1990] Pathak, V. K. and Temin, H. M. (1990). Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. *Proc Natl Acad Sci USA*, 87(16):6024–8.

[Payne and Biggs, 1966] Payne, L. N. and Biggs, P. M. (1966). Genetic basis of cellular susceptibility to the schmidt-ruppin and harris strains of rous sarcoma virus. *Virology*, 29(2):190–8.

[Perelson et al., 1996] Perelson, A., Neumann, A., Markowitz, M., Leonard, J., and Ho, D. (1996). Hiv-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, 271(5255):1582–6.

[Piatak et al., 1993] Piatak, M., Saag, M. S., Yang, L. C., Clark, S. J., Kappes, J. C., Luk, K. C., Hahn, B. H., Shaw, G. M., and Lifson, J. D. (1993). High levels of hiv-1

in plasma during all stages of infection determined by competitive pcr. *Science (New York, NY)*, 259(5102):1749–54.

[Poiesz et al., 1980] Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D., and Gallo, R. C. (1980). Detection and isolation of type c retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous t-cell lymphoma. *Proc Natl Acad Sci USA*, 77(12):7415–9.

[Rabson and Graves, 1997] Rabson, A. and Graves, B. (1997). Synthesis and processing of viral rna. in retroviruses. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. pages 205–262.

[Rainey and Coffin, 2006] Rainey, G. and Coffin, J. (2006). Evolution of broad host range in retroviruses leads to cell death mediated by highly cytopathic variants. *J Virol*, 80(2):562–70. (0) Journal Article United States.

[Rainey et al., 2003] Rainey, G., Natonson, A., Maxfield, L., and Coffin, J. (2003). Mechanisms of avian retroviral host range extension. *J Virol*, 77(12):6709–19.

[Rein et al., 1986] Rein, A., McClure, M. R., Rice, N. R., Luftig, R. B., and Schultz, A. M. (1986). Myristylation site in pr65gag is essential for virus particle formation by moloney murine leukemia virus. *Proc Natl Acad Sci USA*, 83(19):7246–50.

[Ribeiro and Bonhoeffer, 2000] Ribeiro, R. M. and Bonhoeffer, S. (2000). Production of resistant hiv mutants during antiretroviral therapy. *Proc Natl Acad Sci USA*, 97(14):7681–6.

[Ribeiro et al., 1998] Ribeiro, R. M., Bonhoeffer, S., and Nowak, M. A. (1998). The frequency of resistant mutant virus before antiviral therapy. *AIDS*, 12(5):461–5.

[Ricchetti and Buc, 1990] Ricchetti, M. and Buc, H. (1990). Reverse transcriptases and genomic variability: the accuracy of dna replication is enzyme specific and sequence dependent. *EMBO J*, 9(5):1583–93.

[Richman et al., 1990] Richman, D. D., Grimes, J. M., and Lagakos, S. W. (1990). Effect of stage of disease and drug dose on zidovudine susceptibilities of isolates of human immunodeficiency virus. *J Acquir Immune Defic Syndr*, 3(8):743–6.

[Richman et al., 1994] Richman, D. D., Havlir, D., Corbeil, J., Looney, D., Ignacio, C., Spector, S. A., Sullivan, J., Cheeseman, S., Barringer, K., and Pauletti, D. (1994). Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *J Virol*, 68(3):1660–6.

[Rodrigo et al., 1999] Rodrigo, A. G., Shpaer, E. G., Delwart, E. L., Iversen, A. K., Gallo, M. V., Brojatsch, J., Hirsch, M. S., Walker, B. D., and Mullins, J. I. (1999). Coalescent estimates of hiv-1 generation time in vivo. *Proc Natl Acad Sci USA*, 96(5):2187–91.

[Rong et al., 1998] Rong, L., Gendron, K., and Bates, P. (1998). Conversion of a human low-density lipoprotein receptor ligandbinding repeat to a virus receptor: identification of residues important for ligand specificity. *Proc Natl Acad Sci USA*, 95(15):8467–72.

[Rous, 1911] Rous, P. (1911). A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *J Exp Med*, 13(4):397–411.

[Rouzine and Coffin, 1999] Rouzine, I. and Coffin, J. (1999). Linkage disequilibrium test implies a large effective population number for hiv in vivo. *Proc Natl Acad Sci U S A*, 96(19):10758–63.

[Rouzine and Coffin, 2005] Rouzine, I. and Coffin, J. (2005). Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics*, 170(1):7–18.

[Rouzine et al., 2001] Rouzine, I., Rodrigo, A., and Coffin, J. (2001). Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol Mol Biol Rev*, 65(1):151–85.

[Rowe et al., 1970] Rowe, W. P., Pugh, W. E., and Hartley, J. W. (1970). Plaque assay techniques for murine leukemia viruses. *Virology*, 42(4):1136–9.

[RUBIN, 1965] RUBIN, H. (1965). Genetic control of cellular susceptibility to pseudotypes of rous sarcoma virus. *Virology*, 26:270–6.

[Ruscetti et al., 1977] Ruscetti, F. W., Morgan, D. A., and Gallo, R. C. (1977). Functional and morphologic characterization of human t cells continuously grown in vitro. *J Immunol*, 119(1):131–8.

[Sakalian et al., 1994] Sakalian, M., Wills, J. W., and Vogt, V. M. (1994). Efficiency and selectivity of rna packaging by rous sarcoma virus gag deletion mutants. *J Virol*, 68(9):5969–81.

[Schaefer-Klein et al., 1998] Schaefer-Klein, J., Givol, I., Barsov, E. V., Whitcomb, J. M., VanBrocklin, M., Foster, D. N., Federspiel, M. J., and Hughes, S. H. (1998). The ev-o-derived cell line df-1 supports the efficient replication of avian leukosis-sarcoma viruses and vectors. *Virology*, 248(2):305–11.

[Schuurman et al., 1995] Schuurman, R., Nijhuis, M., van Leeuwen, R., Schipper, P., de Jong, D., Collis, P., Danner, S. A., Mulder, J., Loveday, C., and Christopherson, C. (1995). Rapid changes in human immunodeficiency virus type 1 rna load and appearance of drug-resistant virus populations in persons treated with lamivudine (3tc). *J Infect Dis*, 171(6):1411–9.

[Seibert et al., 1995] Seibert, S. A., Howell, C. Y., Hughes, M. K., and Hughes, A. L. (1995). Natural selection on the gag, pol, and env genes of human immunodeficiency virus 1 (hiv-1). *Mol Biol Evol*, 12(5):803–13.

[Seiki et al., 1983] Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult t-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell dna. *Proc Natl Acad Sci USA*, 80(12):3618–22.

[Seo et al., 2002] Seo, T.-K., Thorne, J. L., Hasegawa, M., and Kishino, H. (2002). Estimation of effective population size of hiv-1 within a host: a pseudomaximum-likelihood approach. *Genetics*, 160(4):1283–93.

[Shields et al., 1978] Shields, A., Witte, W. N., Rothenberg, E., and Baltimore, D. (1978). High frequency of aberrant expression of moloney murine leukemia virus in clonal infections. *Cell*, 14(3):601–9.

[Shriner et al., 2004] Shriner, D., Shankarappa, R., Jensen, M., Nickle, D., Mittler, J., Margolick, J., and Mullins, J. (2004). Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. *Genetics*, 166(3):1155–64.

[Skalka and Goff, 1993] Skalka, A. and Goff, S. (1993). *Reverse transcriptase*.

[Smith et al., 2004] Smith, J. G., Mothes, W., Blacklow, S. C., and Cunningham, J. M. (2004). The mature avian leukosis virus subgroup a envelope glycoprotein is metastable, and refolding induced by the synergistic effects of receptor binding and low ph is coupled to infection. *J Virol*, 78(3):1403–10.

[Stair et al., 1991] Stair, R. K., Nelson, C. J., and Mellors, J. W. (1991). Use of recombinant retroviruses to characterize the activity of antiretroviral compounds. *J Virol*, 65(11):6339–42.

[Stuhlmann and Berg, 1992] Stuhlmann, H. and Berg, P. (1992). Homologous recombination of copackaged retrovirus rnas during reverse transcription. *J Virol*, 66(4):2378–88.

[Swanstrom and Wills, 1997] Swanstrom, R. and Wills, J. W. (1997). Synthesis, assembly, and processing of viral proteins. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. (plainview, n.y., cold spring harbor laboratory press). pages 263–334.

[Taplitz and Coffin, 1997] Taplitz, R. and Coffin, J. (1997). Selection of an avian retrovirus mutant with extended receptor usage. *J Virol*, 71(10):7814–9.

[Telisnitsky and Goff, 1997] Telisnitsky, A. and Goff, S. (1997). Reverse transcriptase and the generation of retroviral rna. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. (plainview, n.y., cold spring harbor laboratory press). pages 121–160.

[Temin, 1963] Temin, H. M. (1963). The effects of actinomycin d on growth of rous sarcoma virus in vitro. *Virology*, 20(577).

[Temin, 1964] Temin, H. M. (1964). Nature of the provirus of rous sarcoma. *National Cancer Institute Monographs*, 17:557.

[Temin et al., 1980] Temin, H. M., Keshet, E., and Weller, S. K. (1980). Correlation of transient accumulation of linear unintegrated viral dna and transient cell killing by avian leukosis and reticuloendotheliosis viruses. *Cold Spring Harb Symp Quant Biol*, 44 Pt 2,:773–8.

[Temin and Mizutani, 1970] Temin, H. M. and Mizutani, S. (1970). Rna-dependent dna polymerase in virions of rous sarcoma virus. *Nature*, 226(5252):1211–3.

[Temin and RUBIN, 1958] Temin, H. M. and RUBIN, H. (1958). Characteristics of an assay for rous sarcoma virus and rous sarcoma cells in tissue culture. *Virology*, 6(3):669–88.

[Tsichlis and Coffin, 1980] Tsichlis, P. N. and Coffin, J. M. (1980). Recombinants between endogenous and exogenous avian tumor viruses: role of the c region and other portions of the genome in the control of replication and transformation. *J Virol*, 33(1):238–49.

[Tsichlis et al., 1980] Tsichlis, P. N., Conklin, K. F., and Coffin, J. M. (1980). Mutant and recombinant avian retroviruses with extended host range. *Proc Natl Acad Sci USA*, 77(1):536–40.

[Vogt and Ishizaki, 1965] Vogt, P. K. and Ishizaki, R. (1965). Reciprocal patterns of genetic resistance to avian tumor viruses in two lines of chickens. *Virology*, 26(4):664–72.

[Vogt, 1997] Vogt, V. M. (1997). Retroviral virions and genomes. in retroviruses, j.m. coffin, s.h. hughes, and h. varmus, eds. (plainview, n.y., cold spring harbor laboratory press). pages 27–69.

[Voronin et al., 2009] Voronin, Y., Holte, S., Overbaugh, J., and Emerman, M. (2009). Genetic drift of hiv populations in culture. *PLoS Genet*, 5(3):e1000431.

[Wang et al., 2007] Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M., and Shafer, R. W. (2007). Characterization of mutation spectra with ultra-deep pyrosequencing: application to hiv-1 drug resistance. *Genome Research*, 17(8):1195–201.

[Wei et al., 1995] Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., and Hahn, B. H. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(6510):117–22.

[Weiss et al., 1971] Weiss, R. A., Friis, R. R., Katz, E., and Vogt, P. K. (1971). Induction of avian tumor viruses in normal cells by physical and chemical carcinogens. *Virology*, 46(3):920–38.

[Weller et al., 1980] Weller, S., Joy, A., and Temin, H. (1980). Correlation between cell killing and massive second-round superinfection by members of some subgroups of avian leukosis virus. *J Virol*, 33(1):494–506.

[Wilson et al., 1981] Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 a resolution. *Nature*, 289(5796):366–73.

[Wong-Staal et al., 1983] Wong-Staal, F., Hahn, B., Manzari, V., Colombini, S., Franchini, G., Gelmann, E. P., and Gallo, R. C. (1983). A survey of human leukaemias for sequences of a human retrovirus. *Nature*, 302(5909):626–8.

[Wright, 1931] Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97–159. (0) Journal Article United States.

[Xu and Boeke, 1987] Xu, H. and Boeke, J. D. (1987). High-frequency deletion between homologous sequences during retrotransposition of ty elements in saccharomyces cerevisiae. *Proc Natl Acad Sci USA*, 84(23):8553–7.

[Yamaguchi and Gojobori, 1997] Yamaguchi, Y. and Gojobori, T. (1997). Evolutionary mechanisms and population dynamics of the third variable envelope region of hiv within single hosts. *Proc Natl Acad Sci USA*, 94(4):1264–9.

[Yarchoan et al., 1989] Yarchoan, R., Mitsuya, H., Thomas, R. V., Pluda, J. M., Hartman, N. R., Perno, C. F., Marczyk, K. S., Allain, J. P., Johns, D. G., and Broder, S. (1989). In vivo activity against hiv and favorable toxicity profile of 2',3'-dideoxyinosine. *Science (New York, NY)*, 245(4916):412–5.

[Yoshinaka et al., 1985] Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proc Natl Acad Sci USA*, 82(6):1618–22.

[Young et al., 1993] Young, J. A., Bates, P., and Varmus, H. E. (1993). Isolation of a chicken gene that confers susceptibility to infection by subgroup a avian leukosis and sarcoma viruses. *J Virol*, 67(4):1811–6.

[Yuste et al., 1999] Yuste, E., Sánchez-Palomino, S., Casado, C., Domingo, E., and López-Galíndez, C. (1999). Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J Virol*, 73(4):2745–51.

[Zingler et al., 1995] Zingler, K., a Bélanger, C., Peters, R., Agard, E., and Young, J. A. (1995). Identification and characterization of the viral interaction determinant of the subgroup a avian leukosis virus receptor. *J Virol*, 69(7):4261–6.