Title
High-throughput generation of transmissible antivirals

Author
Notton, Timothy

Publication Date
2017

Peer reviewed|Thesis/dissertation

High-throughput generation of transmissible antivirals

by

Timothy Notton

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Abstract

High-throughput generation of transmissible antivirals

by

Timothy Notton

Doctor of Philosophy in Bioengineering

University of California, San Francisco

Leor Weinberger, Chair

Effective population-level control of viral infections faces significant challenges including: how to therapeutically target the highest-risk populations, circumvent behavioral barriers, and overcome pathogen persistence and resistance mechanisms. A potential solution to overcome these barriers is the use of transmissible antivirals such as defective interfering particles (DIPs) or recently-proposed therapeutic interfering particles (TIPs). These transmissible antivirals are molecular parasites and transmit by 'piggybacking' on wildtype viral replication. By competing effectively for pools of common goods produced by the wildtype virus, DIPs/TIPs can interfere with wildtype virus replication and reduce viral loads in patients. As obligate parasites, TIPs would transmit via the same risk factors and transmission routes as wildtype viruses, automatically reaching high-risk populations, and thereby substantially limiting viral transmission even in resource-poor settings.

At present, methods to generate such transmissible antivirals are *ad hoc* and rely on either expert knowledge to rationally design transmissible antivirals or a laborious and often lengthy process of prospecting for rare, spontaneously-occurring subgenomic mutants. Thus, while deletion mutants of human viruses are desired for use as viral vectors, live-attenuated vaccines, and transmissible antivirals, modern technologies to generate them at scale are not available.

We introduce a new tool to overcome this barrier: a high-throughput method to generate diverse libraries of barcoded viral deletion mutants ($> 10^5$ unique mutants) at modest expense in a period of fewer than 5 days. The method is scalable and cyclical: viral strains with multiple deletions can be obtained by iterating the process. As proof of concept, we demonstrate the construction and screening of libraries of $> 23,000$ barcoded deletion mutants of HIV-1 and $> 90,000$ barcoded deletion mutants of Zika virus (ZIKV). Through repeated *in vitro* passage and deep sequencing of the pooled viral mutants, we are able to comprehensively map the *cis*-acting elements of HIV-1 and ZIKV at single base resolution. Moreover, we are able to track the prevalence of each barcoded deletion mutant through *in vitro* passage, and identify a subset of deletion mutants that are efficiently mobilized and amplified by the wildtype virus.

For HIV-1, our results recapitulate empirical reports of *cis*-acting elements in the literature. We identify four *cis*-acting regions in the HIV-1 genome which could not be complemented *in trans*: (1) $5'$ LTR through the matrix domain of Gag, (2) cPPT/CTS, (3) RRE through SA7, (4) PPT through $3'$ LTR. Thus the minimal proviral size of an HIV-1 vector with two intact LTRs is approximately 2.6 kbp.

For ZIKV, we identify two *cis*-acting regions: (1) $5'$ UTR though C, (2) NS2A through the $3'$ UTR. We show that deletions which induce frameshift of the common open reading frame do not persist, in agreement with a basic mechanism of flaviviral replication. These results suggest a model where Pr, M, E, and NS1 can be provided *in trans*, but not C, NS2A/B, NS3, NS4A/B, and NS5.

Finally, we use the information garnered in the HIV-1 screen to construct a library of transmissible antivirals. Use a modular cloning strategy, we assemble a combinatorial library of multiply-deleted mutants that are composed of a subset of adaptive HIV deletion mutants. We find that mutants with deletions of the accessory region *vif*–*vpu*, when combined with

gag/pol deletions, interfere with wildtype virus replication and transmit efficiently in single round studies. These results suggest a model of interference where transcriptional asymmetry allows this subset of deletion mutants to compete effectively for a common pool of capsid proteins provided by the wildtype virus.

Taken in whole, we show that we have developed a framework for generating transmissible antivirals from first principles. The method is of general use in virology, where the technology can be used to generate live-attenuated vaccines, viral vectors, and replicons, as well as to understand fundamental principles of viral replication and genetics in diverse viral systems. It is of particular interest in emerging viral infections, where therapies must be quickly generated and deployed.

To Mom & Dad

Who gave me my first two microscopes.

Acknowledgments

I'm indebted to many.

Foremost, to Ashley, for her love and gracious sacrifices to accommodate me. For the rest of the SC family who insisted on their names in print, here you are: K.I., S.P., S., F.B., V., L.H.&J., G., G.J., F., C.B., S.V., B.B.(†), C.(†).

To the Notton family—Mom, Dad, Michael, Suzie, Ryan, Kaitlyn, Grant, Kari, and James—for their support and endurance.

To Victoria Saykally (Chapters 2, 4, 5, 6) and Cassandra Thompson (Chapter 6), for excellent technical assistance.

To all people who have worked on TIPs in the Weinberger Lab (in rough chronological order): Brandon Razooky, Vincent Metzger, Kate Franz, Igor Rouzine, Lisa Bishop, Josep Sardanyés, Colleen Scheitrum, Jac Luna, Grayson Kochi, Renee Ram, Moses Xiu, Anand Pai, Jonathan Klein, Luke Rast, Seung-Yong Jung, Elizabeth Tanner, Anjali Gopal, Victoria Saykally, Cassandra Thompson, Justin Jenkins, Matthew Dalton, and Joshua Glazier.

To current members of the Weinberger Lab not previously mentioned: Cynthia Bolovan-Fritts, Hye-in Son, Noam Vardi, Zoë Steier, Maike Hansen, Ravi Desai, Sonali Chaturvedi.

To Freeman Lan, who shared his work with transposons at the Bioengineering retreat in 2013, and sparked the idea for this work.

To Marielle Cavrois and Gilad Doitsh, who generously shared their knowledge of working with HIV in the BSL-3. To JJ Miranda, for several helpful conversations.

To three core facilities for their support: Gladstone BSL-3 Core, Gladstone Flow Cytometry Core, UCSF Center for Advanced Technology. To the funders of and generous donors to the NIH AIDS Reagent Program.

To Kim Osborn, for administrative assistance.

To Stephen Beverley, Kai Zhang, and Lon-Fye (George) Lye at WUSTL, for providing a strong foundation in science.

To Nick Abello and Nick Vasquez for pouring over 1000 plates and their assistance in the BSL-3.

To UCSF librarians, for assistance with research and absolution of late fees. To Lisa DiGiorgio-Haag, for providing outstanding care throughout my time at UCSF.

To Winnie Wen and Maike Hansen for sharing desk and bench space, often by commandeering. To Dan Gibson for many instructional papers.

To Michael McManus and John Dueber, for sound advice and critique. To Adam Abate and Shawn Douglas, for advice and for serving on my qualifying exam.

To Leor, for his constant support and from whom I have learned much, not least of all to seek out and champion good ideas.

To all those left unmentioned here, but who nonetheless made this work possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> "Biologically active defective viral particles have been described with increasing frequency in animal virus systems, suggesting that they may be more common than had been suspected. . . . the defective virus has been overlooked as a possible important determinant in the outcome of natural infections."
>
> –Alice Huang & David Baltimore (1970) in [1]

1.1 The need for transmissible and evolvable therapies

Viruses are dynamic pathogens – they are moving targets. Their movement is foremost physical, in their spread between hosts and through populations. During the most recent influenza A pandemic in 2009–2010, an H1N1 strain spread from Mexico to 72 other countries within four months, and throughout the world within 18 months [2].

The ability of a virus to spread is encapsulated by R_0, its basic reproductive ratio [3]. By definition, R_0 is the average number of secondary infections that a single infected host will generate, provided the remaining population is susceptible. If $R_0 > 1$, the virus will spread; if $R_0 < 1$, the infection will subside. Thus, R_0 is a threshold value for transmissibility. In the regime where $R_0 > 1$, a virus will spread exponentially during the early stages of an epidemic, as shown in Figure 1.1. For most human viruses, R_0 is estimated to be in the range of 1–20 [3].

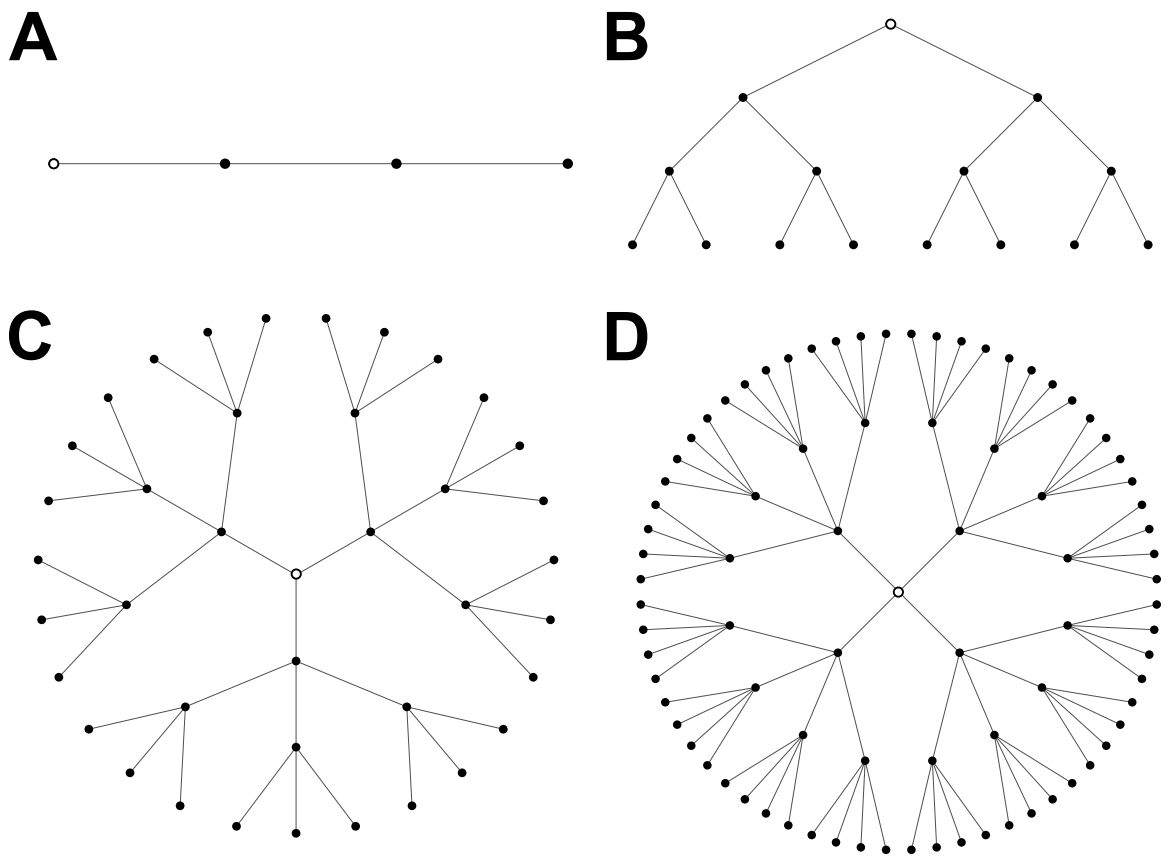


Figure 1.1: Infections spread exponentially for $R_0 > 1$. Four-generation transmission graphs demonstrate exponential spread of virus when R_0 exceeds 1. (A) $R_0 = 1$ (B) $R_0 = 2$, (C) $R_0 = 3$, (D) $R_0 = 4$. Each vertex (circle), corresponds to a single infected person. The unfilled vertex corresponds to the index patient. Edges of the graph indicate transmission of virus from one person to another.

Viruses almost move in an abstract sense. Due to their error-prone replication and ability to resort and recombine, viral populations can traverse vast sequence space inaccessible to their hosts. During human immunodeficiency virus type 1 (HIV-1) infections, the combination of a high mutation rate (3×10^{-5} substitutions/base/replication event) and large intrapatient population sizes (10^7–10^8 newly-infected cells/day) allow the viral swarm to exhaustively sample sequence space [4–7]. This complicates therapies that can be subverted by mutation of only a few points in the genome, as these mutations will occur with high

probability. Ineffectiveness of a previously efficacious therapy due to a viral phenotype is known as *resistance* and is a significant issue when treating infectious disease.

Importantly, the best tools for treating viral infections at a population-level are inherently *static* therapies, deployed against a *dynamic* pathogen. Static therapies do not transmit ($R_0 < 1$) and cannot evolve, as their form is fixed by design. These static therapies (small molecule antivirals, vaccines, antibodies) are critical for contemporary control of population-wide infectious disease, but suffer universally from the same fundamental limitations when treating dynamic pathogens. Static therapies have difficulty surmounting three barriers to effective disease control: pathogen resistance, heterogeneous transmission networks, and human behavior.

Viral resistance

The HIV-1 pandemic is a potent reminder of the relentless evolution of viral pathogens and the consequent emergence of resistance. Over 19 drugs are available to treat HIV-1 infections, and adherence to standard-of-care antiretroviral therapy (ART) can effectively control life-long infection and effectively prevent transmission [8]. Yet, drug-resistant HIV-1 strains have been described for many of these single-agent pharmaceuticals [8]. To avoid ART failure, standard-of-care ART requires treatment with multiple independent HIV-1 drugs, as fully resistant mutants emerge within months on monotherapies [5]. Combination therapies are effective if treatment is available and adhered to, but require lifelong treatment. Indeed, testing for resistance at the onset of ART or during loss of viral load control is a key component of HIV-1 care [8].

Resistance extends beyond small molecules. Over the last decade, there has been significant interest in developing therapies based upon broadly neutralizing antibodies that would be effective against many viral strains. While efficacious during *in vitro* testing, in clinical

trials (such as VRC-01) control over viral load is maintained for only a few months before resistant strains emerge and viral rebound occurs [9].

Heterogeneous transmission networks (superspreaders)

Transmission networks of infectious disease are often heterogeneous. For HIV-1 transmission, sexual contact drives new infections, and sexual contacts follow a power-law form, where some individuals, are more highly connected than others [10]. Although they are the minority per capita, these highly connected individuals, superspreaders, are recognized as a significant driving force of new infections [11]. Thus, therapeutic interventions will be less effective unless superspreaders individuals can be linked to care.

Treatment of HIV-negative (pre-exposure prophylaxis) and HIV-positive patients with ART can effectively prevent infection or reduce transmission [8, 12, 13]. Paradoxically, those with the highest risk of transmission may be unable or unwilling to seek care (e.g., commercial sex workers and their clients, intravenous drug users) due to socioeconomic factors, illicit activity, and disenfranchisement [14], as initiating ART or adherence to treatment is not-compulsory ("opt-in", not "opt-out").

Human behavior and environmental factors

As perception of risk changes, adherence to therapies and behavior may also vary. This phenomenon is known as *behavioral disinhibition*, an increase in risky behavior as the perception of risk decreases. For infectious disease control, this can manifest itself as an increase in activities that can promote transmission of infection (unprotected sex, non-adherence to chemotherapy).

For viral diseases that are transmitted by vectors, such as Dengue virus (DENV) and Zika virus (ZIKV) (both by mosquitoes), effective control of the virus requires control of

both the human and vector infections. In addition, the sylvatic cycle of certain viruses (e.g., DENV), where the virus can be propagated and maintained in hosts outside of humans, can complicate effective viral control [15].

Therapeutic interfering particles are transmissible antivirals

Static therapies are the best-available therapies against viruses, but are fundamentally flawed against dynamic viral pathogens. Static therapies cannot autonomously follow a virus across geography or through sequence space, but remain fixed in form, motionless against moving pathogens. Once resistance to a drug emerges, a likely event with monotherapies, efficacy of the treatment is greatly reduced. Static therapies must be deliberately deployed and enjoy few indirect benefits outside the area of their deployment (except for herd immunity).

Recently, we and others have proposed the use of engineered molecular parasites of viruses to treat population-wide infectious disease [16–19]. Such therapies, termed therapeutic interfering particles (TIPs) (related to naturally occurring defective interfering particles (DIPs)), are a class of virus that can interfere with the replication of a wildtype virus and piggyback on its replication to transmit autonomously through populations. Importantly, TIPs would function as transmissible antivirals (with $R_0 > 1$) and would serve as dynamic, transmissible therapies that can complement existing static therapies [17, 19]. I define and introduce DIPs/TIPs in the later sections of this chapter.

At present, therapeutic interfering particles are currently only theoretical. This dissertation bridges this fundamental gap between theory and practice: detailing a practical approach for generating and evaluating TIPs for many viral systems beginning from first principles. In a reduction to practice, I show how potential transmissible antivirals can efficiently generated for two clinically important epidemic viruses: HIV-1 (a retrovirus) and ZIKV (a flavivirus). The dissertation is structured as follows:

In the remainder of Chapter 1, I introduce the theory of therapeutic interfering particles and define their salient qualities.

In Chapter 2, I develop a high-throughput molecular biology method for generating diverse libraries of viral deletion mutants that can be used as feedstock for a TIP generation pipeline.

In Chapter 3, I construct a mathematical framework for analyzing deep sequencing data of the barcoded deletion libraries, and a methodology for identifying *cis*- and *trans*-acting elements of viruses, crucial information for evaluating which portions of the viral genome should be included in a DIP or TIP.

In Chapter 4 and Chapter 5, I use the techniques developed in previous chapters to screen libraries of HIV-1 and ZIKV deletion mutants, and identify key *cis*-acting and *trans*-acting elements of both viruses as well as deletion mutants that can serve as prototype TIPs.

In Chapter 6, I demonstrate how prototype HIV-1 transmissible antivirals can be synthesized using information from the screen of the HIV-1 deletion library. These prototype TIPs both interfere with HIV-1 replication and can be mobilized by wildtype HIV-1 (are transmissible).

1.2 *Omnis viri e cellula*

At its most essential, a virus is an infectious agent that can only replicate inside living cells [20]. All viruses are *absolutely dependent* on their cellular host for protein synthesis: all known viruses lack ribosomes[1]. Viruses have limited metabolism and depend on their cellular host for supply of high-energy metabolic precursors (ATP, nucleotide pools) and membrane metabolism (if enveloped). This translation dependence differentiates them from other par-

[1] All known viruses lack ribosomes, but some large DNA viruses have genomes encoding significant aspects of translation machinery, including all twenty necessary amino acid-tRNA synthetases [21].

asites, even though some viral genomes exceed the genome sizes of bacteria (mimiviruses vs. mycoplasma) [21, 22].

The atom of viral reproduction is the viral particle or virion. Virions exist in a dazzling variety of shapes and sizes[2] but all are comprised of a polynucleotide genome (DNA, RNA, or both) coated by a protein shell referred to as a *capsid*. If enveloped, the capsid is coated in one or more phospholipid bilayers, known as the envelope. Herpesviruses contain an additional layer of material between the capsid and the envelope known as the tegument. Components of the virion are sometimes referred to as structural proteins. Representative virions for two enveloped human RNA viruses, HIV and ZIKV are shown in Figure 1.2.

When a virion is taken in by a permissive and susceptible cell, the initiating virion is ultimately destroyed. Thus, to infect other susceptible cells, additional virions must be generated. This requires, at a minimum, replication of the viral genome and production of structural proteins to form a complete virion. Upon replication of the genome and production of structural proteins, new virions can be formed by a process of encapsidation (coating genome in capsid), and egress of the virion to complete the lifecycle.

1.3 Viral markets and the emergence of cheaters

We now approach viral infections from an economic perspective. When a virus infects a cell, a new cellular economy is created, where host- and virus-derived goods and services are exchanged in a market. Here, viral genomes are consumers and macromolecules are the marketable goods. Goods may be encoded in the viral genome (virus-derived) or they may be encoded in the genome of the host (host-derived). The composition of this economy varies

[2]The smallest virus known to infect human cells is porcine circovirus (1.3 kbp genome, 17 nm capsid [23]). Among the largest are herpesviruses (human cytomegalovirus: $\approx$ 240 kbp) and poxviruses (variola & vaccinia: $\approx$ 190 kbp) [20].

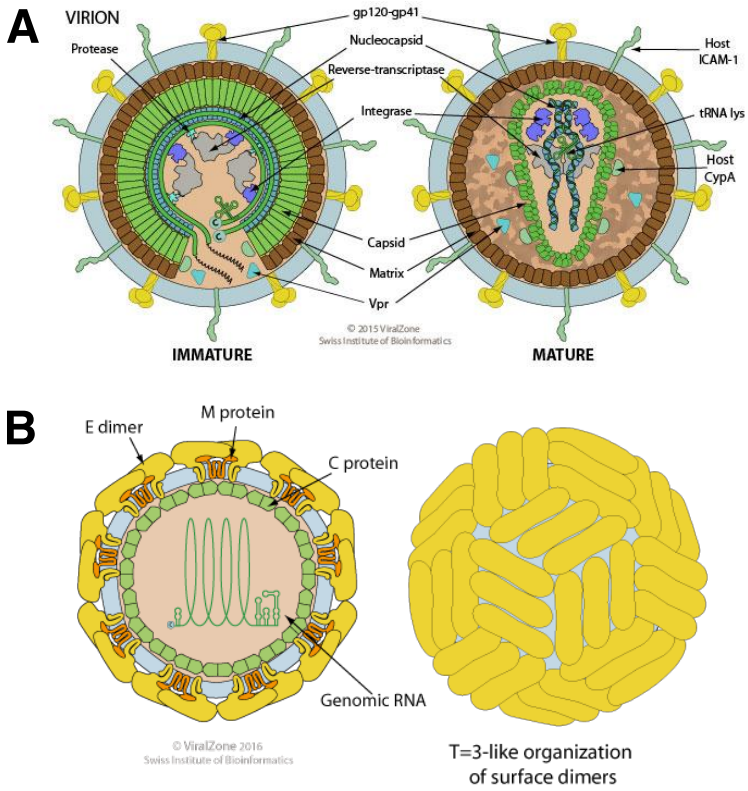


Figure 1.2: Structures of HIV-1 and ZIKV viral particles (A) Virion of HIV-1. The HIV-1 virion is comprised of two genomic RNAs, Gag, Pol, Env, host cell tRNAs, and Vpr (B) Virion of ZIKV. Composed of one copy of gRNA, C protein, E dimer, and M protein. The above images were obtained from ViralZone (`www.expasy.org/viralzone`) and used with permission from the SIB Swiss Institute of Bioinformatics under a Creative Commons Attribution-NonCommercial 4.0 International License.

	excludable	**non-excludable**
rival	private goods	common goods
non-rival	club goods	public goods

Table 1.1: Classes of goods

significantly among viruses/hosts, but all goods can be lumped into four categories based upon continuum qualities of rivalry and excludability (see Table 1.1).

A good is *rival* if consumption of the good by one consumer prevents simultaneous

class	product	virus	rationale
public	Tat	HIV-1	Tat supply $\gg$ demand during active rep. [24, 25]
common	Env	HIV-1	envelope protein, pseuodtyping possible
public	NS1	DENV/ZIKV	secreted immune defense, genome replication
private	NS5	DENV/ZIKV	cannot be *trans*-complemented
common	Gag	HIV-1	capsid protein, lentiviral vectors possible

Table 1.2: Example replication goods of human viruses

consumption by another consumer. Viral structural proteins (components of the virion) are rival as they can be consumed by only a single genome. If goods are plentiful relative to the number of viral genomes, they may be nonrival in effect, as the marginal cost of each good is low.

A good is *excludable* if access to the good is restricted to the viral genome which encoded it ("pay-to-play"). For some RNA viruses (Zika virus), encapsidation is linked to translation, therefore, capsid proteins are excludable goods in this scenario. In other RNA viruses, capsid proteins are available to any viral genome competent for encapsidation; here, capsid proteins are non-excludable.

Where goods fall on the rivalry and excludability axes determine the four general classes of goods in Table 1.1: public goods, club goods, private goods, and common goods.

Public goods are non-rival and non-excludable: all viral genomes in the cell benefit from their production and no genomes can be excluded. Thus there is competition for these, but they be present in such a great amount that stealing most of them will not effect other viral genomes. Examples include products that counteract innate immunity, modulation of host defenses, certain transactivators (HIV-1 Tat)— see Table 1.2 for examples.

Private goods are rival and excludable. During flaviviral infections, dramatic membrane alterations are observed within the host cell, including formation of 70–100 nm membranous structures that are the sites of viral RNA replication [26, 27]. These structures exclude

genomes outside their boundaries from accessing the internal goods. Some RNA viruses, such as HIV-2, link translation of capsid proteins to encapsidation of the genome, thus enforcing excludability [28, 29]. Club goods, like public goods, are non-rival, but have artificial scarcity, where only 'members' are allowed access to the nonrival goods.

Common goods are rival and non-excludable—all viral genomes can access a finite common pool of goods. Thus, there are elements of scarcity and competition. For example, HIV-1 genomes compete for a common pool of viral capsid proteins in the cytoplasm of infected cells [28, 29].

One useful abstraction from genetics is to think of viral genomes as being composed of *cis*-acting elements and *trans*-acting elements. Regions that are *trans*-acting encode for common or public goods. If they are deleted, the mutant can be rescued by providing the good *in trans*. Genomic regions that are *cis*-acting, if deleted, cannot be rescued *in trans*. Thus, *cis*-acting regions encode private goods or club goods, as they are excludable.

Common goods are susceptible to exploitation: certain viral genomes may access these goods without contributing to the common pool. This is known as the *free-rider problem* in economics and *cheating* in biology. These individuals which take from, but do not contribute the common pool of goods are known by a variety of pejoratives: cheaters, mooches, free-riders, and parasites. We will call them by another name in virology: defective interfering particles or DIPs.

1.4 Defective Interfering Particles (DIPs)

Defective interfering particles (DIPs) are a special class of virus. As their name implies, they are loss-of-function viral mutants that are defective in self-replication. DIPs have homology to a functional virus, termed a helper virus, and are generated from this wildtype virus by inactivation of one more *trans*-elements [20]. If they are the only viral species present, DIPs

are unable to self-replicate upon infecting a host cell, due to missing *trans*-acting elements. However, when a cell is co-infected with the helper wildtype virus and a DIP, the DIP can mobilized and infect new target cells, as the helper virus supplies the missing elements in *trans* (*trans*-complementation). Importantly, this symbiotic relationship between DIP and helper virus occurs at the expense of the helper virus—DIPs interfere with helper virus reproduction by acquiring common goods produced by the wildtype virus. In this way, they function as viral superparasites and can have significant influence on the behavior of the helper virus.

In virology, the ratio of viral infections per cell is Poisson distributed and is known as the *multiplicity of infection* (MOI) [20]. At an MOI of 1, each cell is infected, on average, by one virus, although some cells were remain uninfected, while other cells will be infected with more than one virus (superinfected). Infections at high MOI (MOI $\gg$ 1) favor the propagation of DIPs, as each cell has a high probability of receiving at least one helper virus to supply the necessary common goods. Conversely, infections at low MOI (MOI$\ll$ 1) do not favor the propagation of DIPs, as cells are rarely simultaneously infected with both a DIP and a helper virus.

The first reported defective interfering particles (DIPs) were observed during high MOI passage of influenza A in eggs [30]. When eggs were infected at high MOI, titers of the wildtype infectious virus greatly decreased, a phenomenon known as *interference*. von Magnus extended the knowledge into homologous and heterologous interference, and DIPs were identified as being derived from the wildtype virus [31].

During repeated passage of DIPs and helper virus at high MOI, titers of the DIP and helper virus undergo periodic "boom and bust" oscillations known as the von Magnus effect, a phenomenon akin to predator/prey and Lotka-Volterra dynamics [20]. When DIP levels are high, interference with helper virus is enhanced, and levels of the helper virus decrease. This

decrease in helper virus, leads to a drop in DIP levels (from lack of *trans*-complementation), which reduces interference with the helper virus replication, leading to rebound of helper virus. The rise in helper virus titers increases *trans*-complementation of DIPs, leading to amplification of DIPs in the culture, which begin to interfere strongly with the helper virus — the cycle continues *ad infinitum*. This phenomenon has been observed during production of the contemporary influenza A vaccine [32, 33].

Importantly, the intimate and parasitic relationship between a DIP and its cognate helper virus can lead to coevolution of the two species. In experiments with vesicular stomatitis virus, emergence of a DIP was followed by amplification of a DIP-resistant helper virus (Sdi^{-} phenotype) [34]. This in turn, was accompanied by amplification of DIP mutants that could again successfully parasitize the once DIP-resistant helper virus [34].

Advances in sequencing technology have allowed tracking of the emergence of DIPs at fine resolution [35], and we anticipate the expanded use of next-generation sequencing methodologies in tracking coevolution of DIPs/TIPs in future experiments.

1.5 Design principles of effective TIPs

In order for TIPs/DIPs to function at the population-level scale they must satisfy at least two basic requirements: TIPs must be mobilized by the wildtype virus via co-infection (transmissibility), TIPs must outcompete the wildtype virus in co-infected cells (interference). The combination of these effects leads to spread of TIPs by the wildtype virus and reduction of wildtype viral loads. For example, in the case of HIV-1, a ten-fold decrease in viral load reduces the probability of transmission by at least 2.5-fold [36].

In Figure 1.3, we show how a hypothetical TIP against HIV-1 satisfying the transmissibility and interference requirements might behave at three different scales: single-cell, tissue level, and host population scale. Our contention is that the effects of a TIP at the single-

cell scale can translate into large epidemiological effects at the population scale and reduce prevalence of viral disease [19].

What are the design requirements of an effective TIP? In order to be mobilized by the wildtype helper virus, an effective TIP *must* retain all necessary *cis*-acting elements required for efficient replication. Loss of any *cis*-acting element precludes efficient mobilization by definition. In addition, TIPs must retain all necessary *cis*-acting elements to compete for *trans*-acting elements (common-goods).

Furthermore, TIPs must lack at least one *trans*-acting element, which prevents them from self-replicating and establishes a *competition* for the common goods (*trans*-element encoded products) supplied by the wildtype virus. In contrast to other proposed antiviral biologicals, TIPs do not "poison" the pool of common goods, but rather compete for them. Other proposed therapies, which include *trans*-dominant mutants [37], antisense payloads [38], and shRNA [39], do not favor transmission of the therapeutic by design.

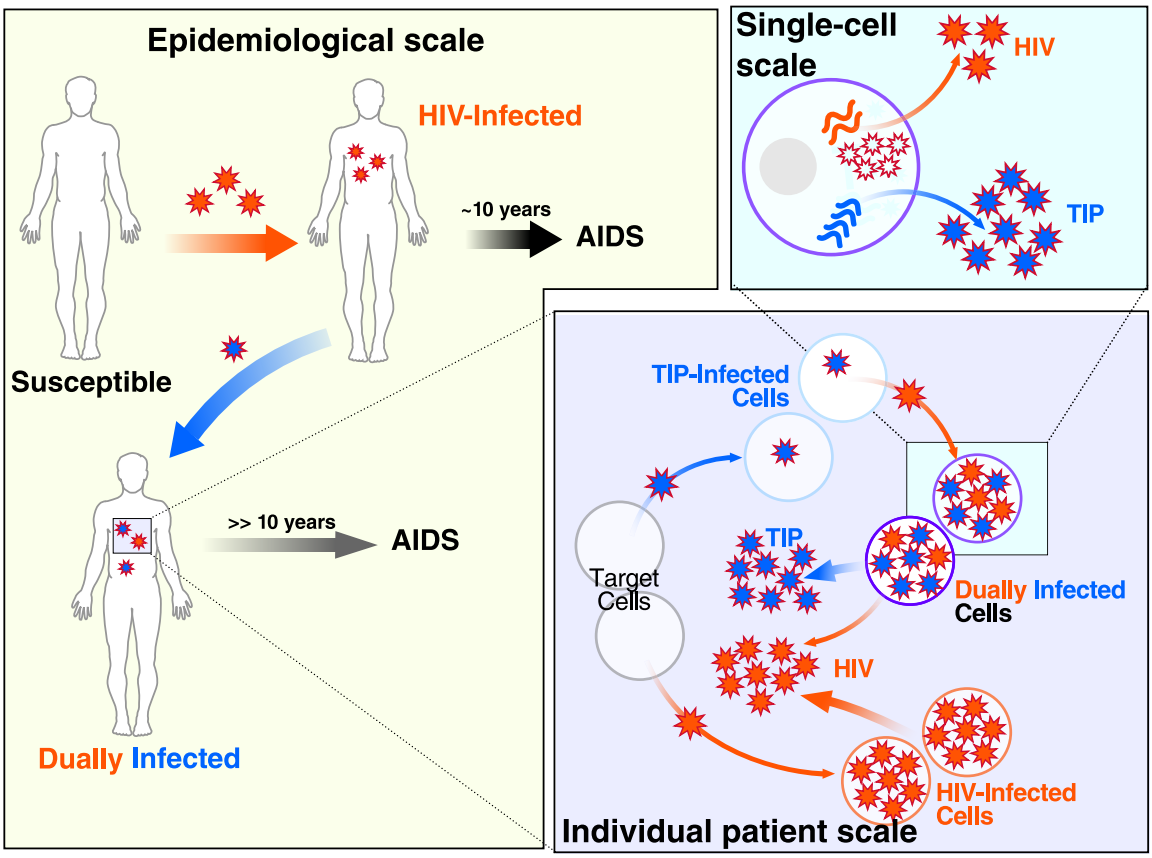


Figure 1.3: Model of TIP replication and spread from single-cell to population level. TIPs 'piggyback' to impact disease spread from the single cell scale to the epidemiological scale. At the single cell scale (top right), dually infected cells would produce both HIV and TIP genomes. TIP genomes, subgenomic mutants of wild type virus, would reach higher concentrations than HIV in dually infected cells and would stoichiometrically outcompete wildtype virus for common goods required by HIV for replication and packaging, and would thus interfere with HIV replication. At the individual patient level (lower right), dually-infected cells would produce more TIPs than HIV. The new TIPs would then infect additional target cells, pre-empting these cells to produce TIPs preferentially once they become infected by HIV. In this way, TIPs would reduce HIV viral load and would generate a TIP viral load in the patient. At the epidemiological scale (left), susceptible individuals would acquire HIV, would become HIV-infected, and would ultimately progress to AIDS in 10 years (higher viral loads lead to faster progression to AIDS, and lower viral loads delay progression). Owing to reduced HIV viral load, individuals dually infected with HIV and TIPs would transmit less HIV, and instead transmit TIPs, and would progress more slowly to AIDS.

Chapter 2

High-throughput generation of deletion libraries

2.1 Introduction

Generation of deletion mutants by serial high MOI passage (Section 1.4) is a straightforward but laborious endeavor. Deletion mutants arise at low frequency and remain rare unless a deletion confers increased fitness relative to the wildtype virus. To reach detectable levels by classic techniques (prevalence $> 1\%$) the selection coefficient need be large or the number of serial passages great. Thus, there is significant variability between virus systems in how frequently spontaneous deletion mutants arise and the selection advantage of these spontaneously-arising mutants.

For example, during *in vitro* passage of vesicular stomatitis virus, generation of transmissible deletion mutants is facile, and they emerge often within 5–10 high MOI passages (5–10 days) [40]. In contrast, one hundred high MOI passages of Rous Sarcoma virus (RSV) at a passage rate of 3–4 days per passage (over 1 year) lead to production of several deletion

mutants [41].

Thus, after the advent of recombinant DNA methods in the 1970s, investigators have sought ways to generate defined mutants and random deletions at an appreciable frequency. Indeed, generating random deletions using reverse genetic system is a common method in virology, and we survey the evolution of methods below.

Existing methods to create viral deletion libraries

Creation of deletion mutants using viral reverse genetic systems paralleled the development of molecular biology techniques and the acquisition of new infectious DNA clones of human viruses.

In the first reported example, in 1974, Lai and Nathans partially-digested purified SV40 DNA with *Hin*dIII restriction endonuclease, purified fragments of varying size, and reintroduced the linear DNA into a host cell for repair and complementation by temperature sensitive SV40 helper virus [42]. In 1975, Carbon, Shenk, Berg, report a method to produce small deletions in SV40 DNA by linearizing the DNA with a restriction endonuclease, briefly treating the ends with λ 5′-exonuclease to create 25-30 b 3′ overhangs, then introducing this DNA into cells for repair and virus production [43]. In 1976, they extended the method to initiate deletions at random locations by performing an incomplete digestion with DNAseI in the presence of Mn^{2+} [44]. Shenk also teaches that use of heteroduplex and S1 endonuclease treatment can increase the size of deletions [45]. In 1978, simple digestion of Ad5 DNA and religation with T4 DNA ligase was performed to generate deletion and insertion mutants [46].

More complex methods to allow targeted deletions soon became available. In 1980, Hearing and Shenk adopted a D-loop mutagenesis system from Green and Tibbetts ([47]) and Shortle *et al.* to investigate Adenovirus [48]. Here a D-loop of DNA was constructed by

catalyzing the hybridization of a ≈ 300 b ssDNA fragment by *E. coli* recA, then nicked by S1 endonuclease. Gaps were created by extending the nick with *M. luteus* DNA polymerase I, then the DNA linearize by digestion with S1 endonuclease. Ends were repaired with DNA polymerase I, then ligated by T4 DNA Polymerase, with deletions sizes of from 3 bp to 25 bp.

To increase the size of deletions, the use of an exonuclease combined with an endonuclease/exonuclease to cleave the trailing overhang can provide double-stranded truncation. Two example systems include BAL-31 and Exonuclease III / Mung Bean Nuclease [49, 50]. Kirkegaard and Nielsen developed a system using T4 DNA Polymerase and *only* one dNTP species to create random deletions in poliovirus, but are limited to a short deletions [51]. More modern methods to induce large deletions include [52].

In contrast to methods that introduce deletions, methods that introduce insertions can also be used to generate loss-of-function viral mutants. A common method is the use of bacterial transposon systems to insert a transposon cassette into viral genomes to disrupt *trans*-acting and *cis*-acting elements by separating protein domains or introducing missense and nonsense mutations [53–57]. If employed as the sole method of mutagenesis, the transposon is subject to excision by the virus during replication, causing reversion to a wildtype genotype. Some of these methods offer the capacity to remove the majority of the transposon sequence by introduction of restriction sites into the termini of the transposon, yet still leave a scar in the viral genome, and are prevented from performing the deletion process several times, due to the introduction of the restriction site into transposed DNA [54, 55, 58].

2.2 A method to produce tagged random deletion libraries

We sought to develop a method that would satisfy eight requirements, a method that: (1) can produce libraries of at least 10^5 tagged deletion mutants within 5 days, (2) can be used for many viral systems (generalizable), (3) uses widely-available reagents, (4) is inexpensive (< \$1 USD / mutant), (5) is amendable to manual and automated execution, (6) offers control over deletion size, (7) can be iterated to generate multiply deleted strains, and (8) has the ability to genetically tag mutants for next-generation sequencing to improve detection and lower costs [59]. As no existing methods could satisfy all of the above, we developed a new method.

The new technology is a high-throughput, molecular biology method which used cycles of *in vitro* transposition and exonuclease digestion to generate random deletions in circular DNA. An overview of the method is shown in Figure 2.1 on page 19 with a detailed schematic in Figure 2.2 on page 21. The method allows control over the size of random deletions and optionally tags each member of the diverse deletion library with a molecular barcode to facilitate analysis by deep sequencing. The examples provided below demonstrate the generation of a library of tagged viral mutants. Performing multiple iterations of the cycle allows for the generation of multiply-deleted viral strains.

2.3 Materials and Methods

Plasmids

pNL4-3 is a molecular clone of HIV-1 subtype B [60]. pNL4-3 (AIDS Reagent Program #114) was a kind gift of Malcom Martin.

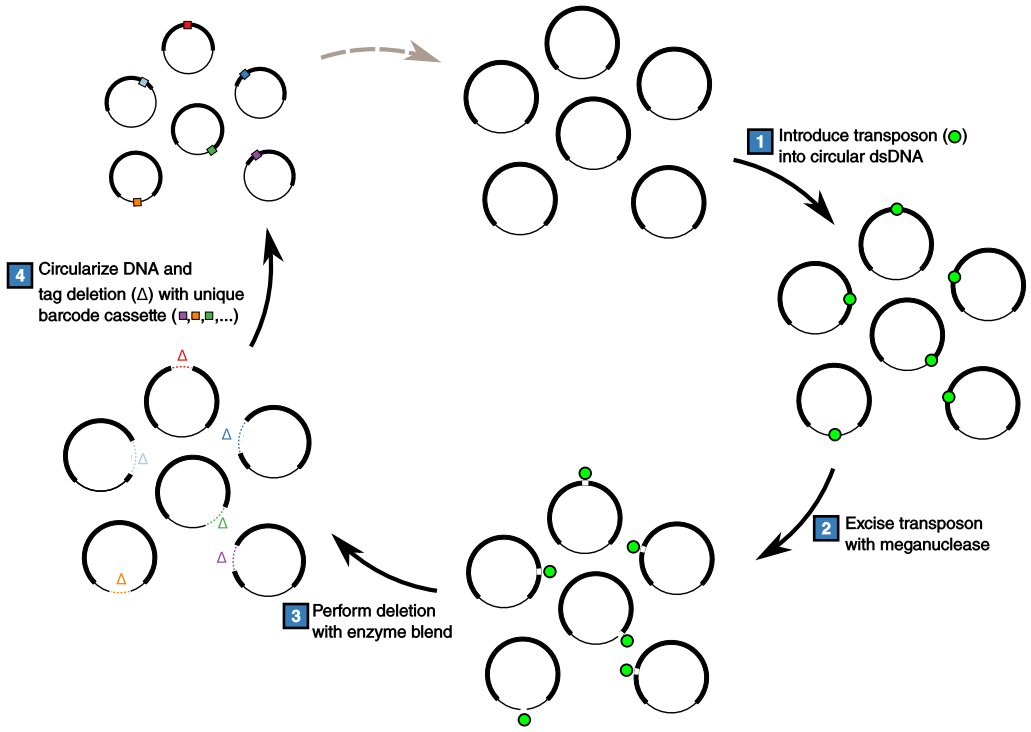


Figure 2.1: The Deletion Cycle: A one-pot method to created collections of barcoded deletion mutants. The starting material is a collection of plasmids. (1) Transposon cassettes harboring unique restriction sites are inserted into plasmids via *in vitro* transposition, one transposon per plasmid. Bacteria are transformed with the pool and outgrowth in selective media generates a pure population of insertion mutants. (2) Transposons are excised and the plasmids linearized by digesting the insertion library with a meganuclease (I-*Sce*I or I-*Ceu*I) via the unique restriction sites encoded in the transposon. (3) Deletions are performed by chewback from both DNA termini by simultaneous treatment with $\mathrm{RecJ_f}$ (a $5' \rightarrow 3'$ exonuclease) and T4 DNA Polymerase without dNTP (a $3' \rightarrow 5'$ exonuclease). Mean deletion size is modulated by adjusting the incubation temperature and duration of chewback. (4) The chewed termini are end-repaired, dA-tailed, then joined by ligation to a T-tailed 60 b barcode cassette. Each mutant is tagged with a statistically unique 20 b barcode. As chewback removes the inserted transposon cassette and meganuclease recognition sites, the process can be iterated to create multiply deleted and tagged molecules.

Two molecular clones of ZIKV, strain MR-766, were a generous gift from Matthew Evans [61]. Two versions were available: a wildtype clone (pMR766(+) and a mutant with a `GDD`→`GNN` mutation in NS5, which lacks a functional RNA Dependent RNA Polymerase, which we call pMR766(−).

Reagent sourcing

All enzymes were obtained from New England Biolabs (Billercia, MA, USA) unless indicated otherwise. All chemicals were obtained from Sigma-Aldrich (St. Louis, MO, USA), unless indicated otherwise. DNA oligonucleotides and synthetic dsDNA were obtained from Integrated DNA Technologies (Coralville, IA, USA).

Enzymatic reactions were conducted in disposable polypropylene plasticware. Reaction temperatures were controlled by placing the reaction vessel in either a thermocycler block (for volumes less than 200 μl) or a forced-air incubator (for volumes exceeding 200 μl).

Construction of transposon cassettes

Synthetic transposon cassettes were developed, based upon the naturally occurring Tn5 transposon and hyperactive *in vitro* transposition activity [63, 64]. Schematics of the transposon cassettes are shown in Figure 2.3 on page 44.

Transposon cassettes contained the following general structure, read from 5′ to 3′ in the open reading frame of the antibiotic-resistance marker: Tn5 mosaic-end, I-*Sce*I recognition site, triple stop codon, I-*Ceu*I recognition site, promoter, antibiotic resistance gene, terminator, *Bam*HI recognition site, I-*Sce*I recognition site, triple stop codon, I-*Ceu*I recognition site, Tn5 mosaic end. Promoter and terminator modules were adapted from [65, 66].

Transposon cassettes were ordered in 3 pieces as synthetic dsDNA (<500 bp) (gBlocks, Integrated DNA Technologies) and cloned by Gibson Assembly into pUC19 (linearized at

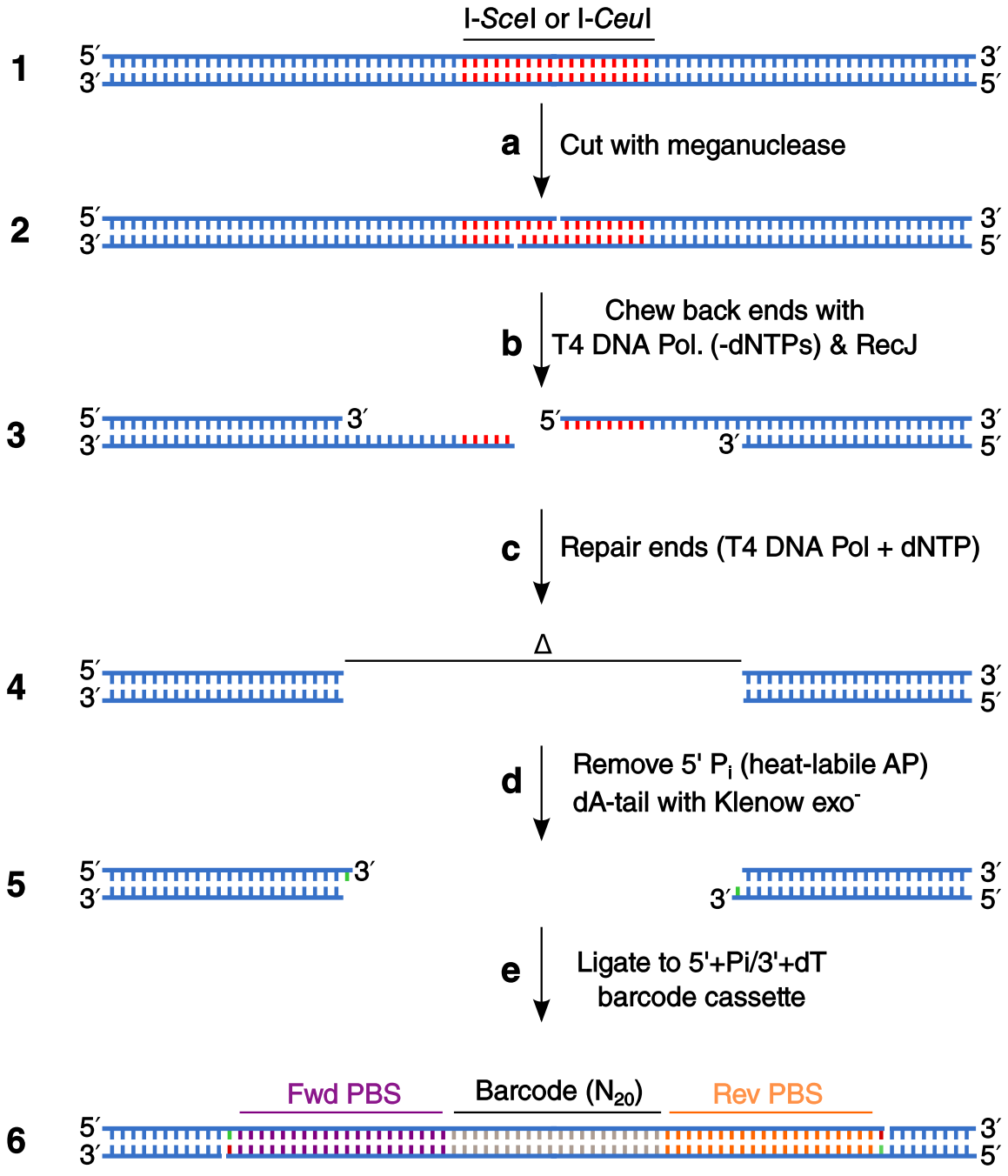


Figure 2.2: Deletion cycle details (caption on on next page)

Figure 2.2: (Caption from previous page) Details of chewback and end repair steps of the deletion chewback. The detail shown is the meganuclease recognition site encoded at each end of the transposon cassette.**(a)** Circular DNA (**1**) is linearized by digestion by treatment with a meganuclease (I-*Sce*I or I-*Ceu*I), which cleaves at recognition sites encoded on the inserted transposon. This creates linear DNA with 4 base 3′ overhangs (**2**). **(b)** Deletions are created by simultaneous treatment with two exonucleases that are active in a common buffer system. T4 DNA Polymerase has predominantly $3' \rightarrow 5'$ exonuclease activity on dsDNA when incubated with magnesium cations and no dNTPs. $\mathrm{RecJ_f}$ has $5' \rightarrow 3'$ exonuclease activity on 5′ overhangs of at least 7 bp, and increased exonuclease activity in the presence of single-stranded binding protein (SSB) [62]. Treatment with this trio of proteins creates a population of truncated deletion mutants with ragged ends (**3**). **(c)** Ragged DNA ends are blunted by treatment with T4 DNA Polymerase in the presence of dNTPs at 20 °C for a brief period to create a population of blunt-ended linear DNA fragments (**4**). **(d)** DNA ends are prepared for barcode cassette ligation by 5′ dephosphorylation with a heat-labile alklaline phosphatase, and addition of a single 3′-dA overhang with an exonuclease-deficient mutant of the Klenow Fragment of *E. coli* DNA Polymerase I. **(e)** The treated deletion mutants (**5**) are ligated to a barcode cassette with single 3′-dT overhangs and 5′ phosphoryl groups to create barcoded circular DNAs with 2 nicks separated by 60 bp (**6**). The nicks can be repaired or the hemiligated DNAs can be directly electroporated into *E. coli*.

the *Bam*HI site) [67]. For the chloramphenicol-resistance transposon cassette (TN5MC), chloramphenicol-resistant pTN5MC transformants were selected on LB supplemented with 100 μg/ml carbenicllin and 33 μg/ml chloramphenicol at 37 °C. For the kanamycin-resistance transposon cassette (TN5MK), kanamycin-resistant pTN5MK transformants were selected on LB supplemented with 100 μg/ml carbenicllin and 50 μg/ml kanamycin.

Determination of minimal conditions for chewback reactions

Chewback reactions were conducted in NEB Buffer 2.1 (New England Biolabs), the composition of which at 1× concentration is:

50 mM NaCl
10 mM Tris-HCl
10 mM $MgCl_2$
10 μg/μl BSA

pH 7.9 @ 25 °C.

λ-*Hind*III digest, T4 DNA Polymerase (3 U/μl), RecJ$_f$ (30 U/μl), and ET SSB (500 ng/μl), were obtained from New England Biolabs. Template DNA, λ-*Hind*III (New England Biolabs, #N3012S), was prepared for chewback by heating to 60 °C for 3 min and immediately cooling on wet ice prior to addition to the chewback reaction to separate annealed cohesive *cos* ends.

A standard 50 μl chewback reaction was prepared on wet ice in a 0.2 ml PCR tube by combining:

30.3 μl of dH_2O
5.0 μl of 10× NEB 2.1
10.0 μl of λ-*Hind*III digest (500 ng/μl)
1.7 μl of T4 DNA Polymerase (3 U/μl)
0.5 μl of RecJ$_f$ (30 U/μl)
0.5 μl of ET SSB (500 ng/μl)

The 50 μl reaction was immediately transferred from wet ice to a thermocycler (Bio-Rad) set to a block temperature of 37 °C and a heated lid temperature of 50 °C , and incubated at 37 °C for 30 minutes to effect the double-strand chewback. After 30 minutes, 1 μl of 10 mM dNTPs was added (200 μM final dNTP concentration) and the reaction mixed and returned to 37 °C for 11 min to allow T4 DNA Polymerase to fill in recessed ends. After 11 minutes of fill-in, the reaction was halted by adding 2 μl of (500 mM EDTA, pH 8.0) to bring the EDTA concentration to 20 mM.

For the various dropout reactions depicted in Figure 2.7 and Figure 2.8, dH_2O was substituted for enzyme solutions.

Determination of chewback rate

An ≈ 4.3 kbp dsDNA template was obtained by purifying the 4361 bp fragment of λ-*Hind*III digest. Ten μg of λ-*Hind*III digest (New England Biolabs, #N3012S) were run out on a 0.8% low melting point agarose/TAE gel (Lonza SeaPlaque GTG agarose) and the gel stained for 20 minutes at 25 °C with SYBR Safe (Thermo Fisher Scientific) diluted to 1× concentration in TAE. DNA bands were visualized by placing the gel atop a blue light transilluminator (Lonza) and viewing the illuminated gel through UVEX S0360X blue light blocking safety glasses (Honeywell). Gel slices corresponding to the 4361 bp fragment were excised with a clean single-edge safety razor blade and transferred to tared microcentrifuge tubes. DNA was recovered by adding 0.1 gel volumes of 10× β-agarase I reaction buffer (New England Biolabs), melting gel slices briefly at 65 °C, cooling to 42 °C , and immediately adding 1 U of β-agarase I per 100 μl of molten gel (New England Biolabs). The mixture was incubated at 42 °C for 60 min to release DNA bound in the agarose matrix. DNA was precipitated from the digested fraction of the agarase reaction by adding 0.1 volumes of 3 M sodium acetate, pH 5.4, and 2 reaction volumes of 2-propanol. After mixing, the reaction was spun at 20000 × g for 15 min at 25 °C, and the supernatant aspirated. The DNA pellet was washed once with 900 μl of 70% ethanol, allowed to air dry briefly, then dissolved in (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0).

200 μl of 2× dNTP buffer were prepared by combining:

40 μl of 10× NEB 2.1 (2× final conc.)
8 μl of 10 mM dNTP (400 μM final conc.)
152 μl of dH_2O

and stored on wet ice.

150 μl of Stop Buffer was prepared by combining:

10 μl of 500 mM EDTA, pH 8.0 (33 mM final conc.)

140 μl of dH_2O

and stored at bench temperature (24 °C).

A 50 μl chew-back reaction was prepared on wet ice in a 0.2 ml PCR tube by combining:

30.3 μl of dH_2O
5 μl of 10× NEB 2.1
12 μl of 4.3 kbp dsDNA from above ($\approx$ 25 ng/μl)
1.7 μl of T4 DNA Polymerase (3 U/μl)
0.5 μl of $RecJ_f$ (30 U/μl)
0.5 μl of ET SSB (500 ng/μl)

The 50 μl reaction was immediately transferred to a thermocycler set to a block temperature of 37 °C and a lid temperature of 95 °C. At (0, 5,10, 15, 20, 25, 30, 40, 50, 60, 70, 80 min) post-transfer to 37 °C, a 4 μl aliquot was removed and combined with 4 μl of 2× dNTP buffer. These 8 μl reactions were incubated at 37 °C for 11 min to allow T4 DNA Polymerase to fill in the single-stranded tails that remain uncleaved by $RecJ_f$. After 11 min of fill-in, the reaction was halted by adding 12 μl of Stop Buffer to bring the EDTA concentration to 20 mM. Reactions were incubated on wet ice after addition of 12 μl of Stop Buffer.

The concentration of dsDNA was determined by a fluorimetric method (PicoGreen, Thermo Fisher Scientific). Five μl of each reaction was added to 95 μl of (10 mM Tris-HCl, pH 7.5; 1 mM EDTA, pH 7.5). To this, 100 μl of a PicoGreen working stock (diluted to 1/200× in (10 mM Tris-HCl, pH 7.5; 1 mM EDTA, pH 7.5) were added in an Opti-F black-walled plate and read with an Enspire plate reader (Perkin Elmer) with 480 nm excitation and 520 nm emission filter, and fluorescence compared to a λ DNA standard. All reactions were performed in triplicate.

Chewback rates at 37 °C were calculated by fitting the decay in dsDNA (fluorescence signal) at various timepoints to a linear regression model with the freely-available R statistical

software. Chewback rates were determined to be ≈ 60 bp/min for 0–20 min and ≈ 50 bp/min for 0–80 min (Figure 2.9).

Production of transposon insertion library

Linear transposon cassettes were constructed by PCR from pTN5MK plasmid template. Oligos oTN5-F (`/5Phos/CTGTCTCTTATACACATCTGCGGCCGC`) and oTN5-R (`/5Phos/CTGTCTCTTATACACATCTTTAATTAATTCGCTACC`) were ordered with covalent 5′-phosphorylation modification and standard desalting purification from Integrated DNA Technologies. A 300 µl PCR master mix was prepared by mixing the following at bench temperature (24 °C):

60 µl of 5× Q5 Polymerase Buffer (New England Biolabs)
171 µl of dH_2O
6 µl of 10 mM dNTP
30 µl of 6 µM oTN5-F
30 µl of 6 µM oTN5-R
6 µl of pTN5MK (1 ng/µl)
3 µl of HotStart Q5 Polymerase (2 U/µl) (New England Biolabs)

The 300 µl PCR master mix was briefly mixed, then distributed as 6×50 µl aliquots in 0.2 ml PCR tubes. PCR was performed using the following program in a thermocycler with heated lid (105 °C):

1 cycle of:
    98 °C for 30 s
15 cycles of:
    98 °C for 10 s
    68 °C for 20 s
    72 °C for 50 s
1 cycle of:
    72 °C for 300 s
    10 °C for ∞

Post-thermocycling, the $6 \times 50\,\mu$l reactions were pooled and linear transposon DNA purified with the Zymo DCC-5 Kit (Zymo Research) across two DCC-5 silica columns per the manufacturer's instructions. The DNA was eluted from each column by adding $12\,\mu$l of (10 mM Tris-Cl pH 8.0; 0.1 mM EDTA, pH 8.0), and the two eluted fractions were pooled to obtain $\approx 20\,\mu$l of purified transposon product. The concentration of linear DNA was determined to be $75\,\text{ng}/\mu$l by absorbance at 260 nm using a Nanodrop spectrophotometer (Thermo Fisher Scientific).

The ≈ 1.4 kbp linear transposon DNA was gel-purified by running out all $20\,\mu$l in a 2 cm wide well in a 0.8% agarose/TAE gel. Post-run, the gel was stained with 1× SYBR Safe (Thermo Fisher Scientific) in 1× TAE and a gel fragments corresponding to the 1.4 kbp linear transposon fragment excised upon illumination with a blue light transilluminator (Lonza) and viewed through UVEX S0360X blue light blocking safety glasses (Honeywell). DNA was recovered from the gel slice by adding 3 gel volumes of Buffer QG (Qiagen), and melting the gel slice by incubation at 37 °C with frequent mixing. The liquid gel mixture was applied to a DCC-5 silica column (Zymo Research) in repeated $600\,\mu$l volumes, interspersed by spinning the column at $10000 \times g$ for 40 s and discarding the flowthrough. The DCC-5 column was washed twice by adding $600\,\mu$l of Wash Buffer (Zymo Research), spinning the column at

10000×g for 60 s, and discarding the flow-though. The column was carefully transferred to a 1.5 ml DNA LoBind tube (Eppendorf) and 30 μl of (10 mM Tris-HCl pH 8.0; 0.1 mM EDTA, pH 8.0) applied to the column bed and incubated for 1 min. DNA was eluted by spinning the column for 1 min at $10000 \times g$. The concentration of linear DNA was found to be $\approx 30\,\mathrm{ng}/\mu\mathrm{l}$ by determining the absorbance at 260 nm using a Nanodrop spectrophotometer (Thermo Fisher Scientific).

in vitro transposition

An *in vitro* transposition reaction was performed using recombinant EZ-Tn5 transposase (Epicentre). The enzyme was supplied by the manufacturer at a concentration of $1\,\mathrm{U}/\mu\mathrm{l}$ in a storage buffer consisting of 50% glycerol, 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, and 0.1% Triton X-100). A reaction buffer (10× EZ-Tn5 reaction buffer) was also provided by the manufacturer (Epicentre) with the following 10× composition: 0.50 M Tris-acetate (pH 7.5), 1.5 M potassium acetate, 100 mM magnesium acetate, and 40 mM spermidine.

A 10 μl *in vitro* transposition reaction was assembled by combining the following in a 0.2 ml PCR tube:

2.0 μl of pNL4-3 ($100\,\mathrm{ng}/\mu\mathrm{l}$) [14825 bp] [21.83 fmol]
2.0 μl of TN5MK gel-purified transposon ($10\,\mathrm{ng}/\mu\mathrm{l}$) [1434 bp] [22 fmol]
1.0 μl of 10× EZ-Tn5 reaction buffer
4.0 μl of dH_2O
1.0 μl of EZ-Tn5 transposase ($1\,\mathrm{U}/\mu\mathrm{l}$)

The reaction was mixed and the 0.2 ml PCR tube transferred to a thermocycler set to a block temperature of 37 °C and a heated lid temperature of 50 dC for a 2 h incubation. After a 2 h incubation, 1 μl of 1% (m/v) SDS solution was added and the reaction heated to 70 °C

for 10 min to halt the reaction.

After equilibrating to room temperature, the entire volume of the reaction was pipetted upon a 0.025 μm membrane (13 mm outer diameter) (Millipore, #VSWP01300), floating on 25 ml of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0). Drop-dialysis was allowed to proceed for 1 h to remove inhibitory salts from the reaction mixture. After 1 h the reaction volume was recovered and placed in a DNA LoBind tube (Eppendorf) on wet ice.

Transformation and harvest of insertion library

1 μl of the reaction mixture ($\approx$ 10%) was added to 40 μl of ice-cold electrocompetent *E. coli* (strain DH10B), and introduced into a chilled 0.1 cm sterile electroporation cuvette (Bio-Rad) placed on wet ice. Bacteria were electroporated with a Gene Pulser II electroporation system (Bio-Rad) with pulse settings of 1.7 kV, 25 μF, 200 Ω. Immediately post-electroporation, 960 μl of SOC (Thermo Fisher) were added and the cell mixture transferred to a sterile 15 ml polypropylene culture tube, and allowed to recover for 90 min at 30 °C, 250 rpm shaking. After a 90 min recovery, 10 μl (1%) of the reaction was plated on a 10 cm Petri plate containing solid media comprised of Lysogeny Broth (Miller) supplemented with 1% agar, 100 μg/ml carbenicillin, and 50 μg/ml kanamycin (LBA+Carb_{100}+Kan_{50}) for estimation of total library size. The remaining 99% of the transformation volume was plated across 6×10 cm LBA+Carb_{100}+Kan_{50} plates. After incubation for 24 h at 32 °C , approximately 5000-10000 CFU were obtained across the six library plates. The bacteria were scraped from the agar surface using a sterile cell spreader after adding 5 ml of LB to each plate and recovered by centrifugation at $4000 \times g$ for 15 min. The bacterial pellet was resuspended in 6 ml of LB supplemented with 7% (v/v) DMSO and stored in 1.0 ml cryovials at -80 °C for future use.

Production of a barcoded deletion library from an insertion library

One vial of the frozen insertion library (TN5MK in pNL4-3) was thawed and used to inoculate 500 ml of LB-Miller (LB) supplemented with 100 μg/ml carbenicillin and 50 μg/ml kanamycin (LB+Carb_{100}+Kan_{50}) in a 2.8 L Fernbach flask. The culture was grown at 30 °C and 250 rpm in a shaking incubator (New Brunswick Scientific) until it reached an OD_{600} of 1.0, whereupon a dry bacterial pellet was obtained by centrifugation and frozen at −80 °C. Supercoiled plasmid DNA was obtained from the frozen bacterial pellet with a DNA Maxiprep Kit (Qiagen) per the manufacturer's instructions, and resuspended in (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0) at a concentration of > 200 ng/μl.

Linearization by meganuclease treatment

Next, inserted transposons were excised from the plasmid insertion library to create linear DNA molecules by treatment with meganuclease I-*Sce*I (New England Biolabs) in 1× CutSmart Buffer (New England Biolabs), whose 1× composition as specified by the manufacturer is: 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, 100 μg/ml BSA, pH 7.9 at 25 °C. Approximately 50 μg of insertion library DNA were digested in a 500 μl reaction by mixing the following in DNA LoBind tube (Eppendorf):

105 μl of TN5MK in pNL4-3 (485 ng/μl)
50 μl of 10× CutSmart Buffer (New England Biolabs)
20 μl of I-*Sce*I (5 U/μl)
335 μl of dH_2O

The reaction was incubated for 8 h at 37 °C, with brief mixing by inversion performed every 2 h. After 8 h, the reaction was cooled to bench temperature (24 °C), then extracted once with 500 μl of 25:24:1 phenol:chloroform:isoamyl alcohol equilibrated with (10 mM Tris-HCl,

pH 8.0; 1 mM EDTA, pH 8.0) (Thermo Fisher Scientific) followed by a second extraction with 500 μl of pure chloroform (Sigma-Aldrich). The upper aqueous layer was transferred to a new DNA LoBind tube, and 1.0 μl (25 μg) of co-precipitating GenElute Linear Polyacrylamide (Sigma-Aldrich) was added and the solution mixed to homogeneity.

Digested DNA was precipitated from the aqueous phase by $MgCl_2$/PEG-8000 precipitation. The $\approx$ 500 μl were adjusted to a final concentration of 12.5% (m/v) PEG-8000 and 20 mM $MgCl_2$ by adding 14 μl of 1 M $MgCl_2$ and 168 μl of 50% (m/v) PEG-8000. The reaction was inverted and flicked to mix, then spun at 20000 $\times$ g for 60 min in a refrigerated microcentrifuge (Eppendorf) at 25 °C to pellet the precipitated DNA. After centrifugation, the pellet was difficult to visualize. The supernatant was removed and discarded and 900 μl of freshly-prepared 70% ethanol were added and the tube contents mixed by inverting the tube and flicking. Upon addition of 70% ethanol, the DNA pellet became readily visible. The tube was spun at 20000 $\times$ g for 2 min to collect the pellet and the supernatant aspirated and discarded. An additional 900 μl of 70% ethanol were added to wash the pellet, and the tube spun again at 20000 $\times$ g for 2 min to collect the pellet. All supernatant was carefully removed and the pellet dried briefly at room temperature (5 min) until no visible liquid remained. The DNA was solubilized by adding 60 μl of (10 mM Tris-HCl , pH 8.0; 0.1 mM EDTA, pH 8.0), incubating the tube at 42 °C for 20 min and mixed by flicking the tube. The concentration of linear DNA was found to be $\approx$ 750 ng/μl by a fluorimetric assay (Quant-iT PicoGreen dsDNA Assay Kit from Thermo Fisher Scientific). DNA was stored at 4 °C until future use.

Chewback and end-repair of linearized library

Before the chewback reaction occurred, substrate DNA was heated to 60 °C for 3 min and immediately placed on wet ice to separate DNA aggregates. Four standard chewback reac-

tions (reactions **R5**, **R10**, **R15**, **R20**) were prepared on wet ice in four separate 0.2 ml PCR tubes. Each reaction was prepared by combining the following in a separate 0.2 ml PCR tube:

5.0 μl of 10× NEB 2.1
36.3 μl of dH_2O
6.0 μl of linearized pNL4-3/TN5MK insertion lib. (750 ng/μl)
1.7 μl of T4 DNA Polymerase (3 U/μl)
0.5 μl of $RecJ_f$ (30 U/μl)
0.5 μl of ET SSB (500 ng/μl)

All four 50 μl reactions were immediately transferred from wet ice to a thermocycler (Bio-Rad) set to a block temperature of 37 °C, a heated lid temperature of 50 °C, then incubated at 37 °C for a duration of 5–20 minutes to effect the double-strand chewback (**R5**: 5 min, **R10**: 10 min, **R15**: 15 min, **R20**: 20 min). At the appropriate time, the indicated reaction was removed from 37 °C incubation and 1 μl of 10 mM dNTPs were added (200 μM final dNTP concentration). The reaction was mixed and returned to 37 °C for 11 min to allow T4 DNA Polymerase to fill in recessed ends. After 11 min of fill in, the reaction was halted by adding 2 μl of (500 mM EDTA, pH 8.0) to adjust the EDTA concentration to 20 mM and placed on wet ice.

All four 50 μl chewback reactions (**R5**, **R10**, **R15**, **R20**) were pooled (200 μl volume) and then extracted once with 200 μl of 25:24:1 phenol:chloroform:isoamyl alcohol with TE (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0) (Thermo Fisher Scientific) followed by a second extraction with 500 μl phenol:chloroform:isoamyl alcohol equilibrated with TE, pH 8.0 (Thermo Fisher Scientific). The $\approx$ 200 μl upper aqueous layer was transferred to a new DNA LoBind tube (Eppendorf), and desalted by running 2 × 100 μl through separate Sephacryl gel filtration columns (Microspin S-400 HR columns (GE Lifesciences)).

The $2 \times 100\mu$l flowthrough fractions were pooled and the linear DNA blunt-ended by NEBNext End Repair Reaction Module (New England Biolabs). The composition of the 1× End Repair Reaction buffer is specified by the manufacturer as: 50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 0.4 mM dATP, 0.4 mM dCTP, 0.4 mM dGTP, 0.4 mM dTTP, pH 7.5 at 25 °C). The enzyme mix contains a blend of T4 Polynucleotide Kinase (10 U/μl) and T4 DNA Polymerase (3 U/μl). A 400 μl end-repair reaction was prepared by combining the following in a DNA LoBind tube (Eppendorf) on wet ice:

200 μl of linearized deletion library ($\approx$ 20 μg total)
40 μl of 10× End Repair Buffer (NEB)
140 μl of dH_2O
20 μl of EndRepair Enzyme Mix (NEB)

The 400 μl reaction was distributed as $2 \times 200\,\mu$l aliquots in 0.2 ml PCR tubes and incubated for 30 min in a thermocycler (Bio-Rad) with a block temperature set to 20 °C .

After the 30 min incubation, the $2 \times 200\,\mu$l fractions were pooled in a single DNA LoBind tube (Eppendorf). The 400 μl pool was extracted once with 400 μl of 25:24:1 phenol:chloroform:isoamyl alcohol equilibrated with TE, pH 8.0, (Thermo Fisher Scientific), and once with 400 μl of pure chloroform (Sigma). The upper aqueous phase was transferred to a new DNA LoBind tube, and 1.0 μl (25 μg) of co-precipitating GenElute Linear Polyacrylamide (Sigma) was added and the solution mixed to homogeneity.

DNA was precipitated by $MgCl_2$/PEG-8000 precipitation. The remaining 400 μl volume of aqueous phase was adjusted to a final concentration of 12.5% (m/v) PEG-8000 and 20 mM $MgCl_2$ by adding 13 μl of 1 M $MgCl_2$ and 135 μl of 50% (m/v) PEG-8000. The reaction was inverted and flicked to mix, then spun at $20000 \times g$ for 60 min in a refrigerated microcentrifuge (Eppendorf) at 25 °C to pellet the precipitated DNA. After centrifugation, the pellet was translucent and difficult to visualize. The supernatant was removed and discarded and

900 μl of freshly-prepared 70% ethanol were added and the tube mixed. Upon addition of 70% ethanol, the DNA pellet became white, opaque, and was readily visible. The tube was spun at 20000 $\times$ g for 2 min to collect the pellet and the supernatant aspirated and discarded. An additional 900 μl of 70% ethanol were added to wash the pellet, and the tube spun at 20000 $\times$ g for 2 min to collect the pellet. All supernatant was carefully removed and the pellet dried briefly at room temperature (5 min) until no visible liquid remained. The DNA was solubilized by adding 60 μl of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0), heating to 42 °C for 20 min, and mixed by flicking the tube. The concentration of linear DNA was found to be $\approx$ 200 ng/μl by determining the absorbance at 260 nm using a Nanodrop spectrophotometer (Thermo Fisher Scientific). DNA was stored at 4 °C until future use.

Addition of 3′-dA overhang to backbone

A 3′-dA overhang was added to the purified blunt-end barcode cassette with a 3′ $\rightarrow$ 5′ exonuclease-deficient Klenow Fragment of *E. coli* DNA Polymerase I (New England Biolabs). Linear, end-repaired library DNA from above was heated for 3 min at 60 °C, then immediately transferred to wet ice to cool. A 100 μl 3′-dA tailing reaction was prepared by mixing the following in a 0.2 ml PCR tube:

50 μl of end-repaired chewed-back library ($\approx$ 200 ng/μl)
10 μl of 10$\times$ NEB Buffer 2
32 μl of dH_2O
2 μl of 10 mM dTTP
6 μl of Klenow Fragment (exo-) (5 U/μl).

The 100 μl reaction was incubated for 1 h at 37 °C in a thermocycler (Bio-Rad) with a block temperature set to 37 °C and a heated lid temperature of 50 °C. After the 1 h incubation, the enzyme was heat-inactivated by incubation at 70 °C for 20 min. The reaction was

allowed to cool, and then 10 μl of 10× Antarctic Phosphatase Reaction Buffer (New England Biolabs) (1× composition: 50 mM Bis-Tris-Propane HCl, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, pH 6.0 at 25 °C) were added to the tube and the contents mixed and pulsed down. Two μl of Antarctic Phosphatase (5 U/μl) (NEB) were added, the reaction mixed, then incubated for 1 h at 37 °C . The enzyme was heat-inactivated by heating the reaction to 70 °C for 5 min.

Next, the 5′-dephosphorylated, >8 kbp 3′-dA tailed vector was purified from <8 kbp pieces of DNA (including the excised transposon cassette) by gel purification. Twenty μl of 6× Gel Loading Dye, Blue (New England Biolabs) (1× composition: 2.5% (m/v) Ficoll-400, 11 mM EDTA, 3.3 Tris-HCl, 0.017% (m/v) SDS, 0.015% (m/v) bromophenol blue, pH 8.0 at 25 °C) were added to 100 μl of 5′-dephosphorylated, 3′-dA tailed vector, and 12 μl of the mixture loaded in 10 wells of a 0.8% low melting point agarose/TAE gel (Lonza SeaPlaque GTG agarose) and run alongside a DNA ladder (Quick Load 1 kb extend ladder, NEB). The gel was run at 85 V for 90 min, until the bromophenol blue bands migrated to the middle of the gel, then stained for 30 min with SYBR Safe (gel stained for 20 min at 25 °C with SYBR Safe (Thermo Fisher Scientific) diluted to 1× concentration in 1× TAE.

DNA bands were visualized by placing the gel atop a blue light transilluminator (Lonza) and viewing the illuminated gel through UVEX S0360X blue light blocking safety glasses (Honeywell). Gel slices corresponding to fragments of size between $\approx$8–15 kbp were excised with a clean single-edge safety razor blade and transferred to tared microcentrifuge tubes. DNA was recovered by adding 0.1 gel volumes of 10× β-agarase I reaction buffer (New England Biolabs), melting gel slices briefly at 65 °C, cooling to 42 °C, and immediately adding 1 U of β-agarase I per 100 μl of molten gel (New England Biolabs). The mixture was incubated at 42 °C for 60 min to release DNA bound in the agarose matrix.

DNA was precipitated from the digested fraction of the β-agarase I reaction by adding 0.1 volumes of 3 M sodium acetate, pH 5.4, and 2 reaction volumes of 2-propanol. After

mixing, the reaction was spun at $20000 \times g$ for 15 min at 25 °C, and the supernatant aspirated. The DNA pellet was washed twice with 900 μl of 70% ethanol, allowed to air dry briefly, then dissolved in 40 μl of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0), and allowed to solubilize overnight at 4 °C. The concentration of vector DNA was found to be ≈ 150 ng/μl by determining the absorbance at 260 nm using a Nanodrop spectrophotometer (Thermo Fisher Scientific). A similar concentration (145 ng/μl) was determined by a fluorimetric assay (Quant-iT PicoGreen dsDNA Assay Kit from Thermo Fisher Scientific).

Preparation of barcode cassette inserts

We designed a barcode cassette using a sequence that would be heterologous to most human viruses. The left (`L`) and right (`R`) common regions are derived from Tobacco Mosaic Virus (TMV, GenBank Accession `X68110.1`) nucleotides of (+) strand 1992–2011 and (−) strand 2127–2108. PCR using these primer will amplify a 136 bp product from 1992–2127. Thus, even if TMV is present, the product size that will be amplified is greater than 60 bp and can be filtered out by size-selection during downstream Illumina library preparation.

Blunt-end, 5′-phosphorylated, 60 bp barcode cassettes were prepared by PCR from ss-DNA template. Oligonucleotides oBC20v1-F (`/5Phos/CCGTCCATGAAGGGTTCGAT`) and oBC20v1-R (`/5Phos/ACGAATCTGCCGTTGCCATA`) were ordered with covalent 5′-phosphorylation modification and standard desalting purification from Integrated DNA Technologies. Oligonucleotide pool oBC20-T:

```
CCGTCCATGAAGGGTTCGATNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT
```

where `N` indicates (`A`,`C`,`G`,`T`) was ordered with machine-mixed bases and standard desalting purification from Integrated DNA Technologies.

Two 1.1 ml aliquots of PCR master mix was prepared by mixing the following at bench temperature (24 °C) in two DNA LoBind Tubes (Eppendorf):

550 μl of 2× Q5 HotStart HiFi Master Mix (New England Biolabs)
385 μl of dH_2O
27.5 μl of 20 μM oBC20-F (fwd oligo)
27.5 μl of 20 μM oBC20-R (rev oligo)
110.0 μl of 100 nM oBC20-T (template)

The 2×1.1 ml PCR master mixes were briefly mixed, then distributed as 40×50 μl aliquots in 0.2 ml PCR tubes. PCR was performed using the following program in a thermocycler (Bio-Rad) with heated lid (105 °C):

1 cycle of:
    98 °C for 30 s
5 cycles of:
    98 °C for 10 s
    65 °C for 75 s
1 cycle of:
    65 °C for 300 s
    10 °C for ∞

Post-thermocycling, the 40 × 50 μl PCR reactions were pooled and the barcode cassette purified across ten DCC-5 columns using a DNA Clean and Concentrator 5 kit (Zymo Research) per the manufacturer's instructions. DNA was eluted from each of the ten silica DCC-5 columns with 15 μl of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0) and the five eluted fractions pooled to obtain $\approx$ 50 μl of purified blunt-end barcode cassette. The concentration of blunt-end barcode cassette DNA was determined to be $\approx$ 26 ng/μl by a fluorimetric assay (Quant-iT PicoGreen dsDNA Assay Kit from Thermo Fisher Scientific).

The blunt-end barcode cassette was stored in a DNA LoBind tube (Eppendorf) at −80 °C for future use.

A 3′-dT overhang was added to the purified blunt-end barcode cassette with a 3′ → 5′ exonuclease-deficient Klenow Fragment of *E. coli* DNA Polymerase I (New England Biolabs). A 300 µl reaction was prepared by mixing the following in a DNA LoBind tube (Eppendorf):

100 µl of blunt-end purified 60 bp barcode cassette (26 ng/µl)
30 µl of 10× NEB Buffer 2
146 µl of dH_2O
6 µl of 10 mM dTTP
18 µl of Klenow Fragment (exo−) (5 U/µl).

The 300 µl reaction was distributed as 3 × 100 µl aliquots in 0.2 ml PCR tubes and incubated for 3 h in a thermocycler (Bio-Rad) with a block temperature set to 37 °C and a heated lid temperature of 50 °C.

Post-incubation, the 3 × 100 µl PCR reactions were pooled and the 3′-dT barcode cassette purified with a single DCC-5 silica column using a DNA Clean and Concentrator 5 kit (Zymo Research) per the manufacturer's instructions. DNA was eluted with 80 µl of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0). The concentration of the 3′-dT blunt-end barcode cassette DNA was determined to be ≈ 11 ng/µl by a fluorimetric assay (PicoGreen from Thermo Fisher Scientific). The 3′-dT barcode cassette was stored in 15 µl aliquots in DNA LoBind tubes (Eppendorf) at −80 °C for future use.

Ligation of barcode cassettes and chewed vector

Next, 3′-dT-tailed barcode cassettes were ligated into 3′-dA-tailed vector and the DNA circularized using a large amount of T4 DNA Ligase in a PEG-6000 containing buffer (Quick Ligation Buffer, New England Biolabs). The composition of the 1× Quick Ligation Buffer as

specified by the manufacturer is: 66 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 7.5% (m/v) PEG-6000, pH 7.6 at 25 °C . A 200 µl ligation reaction was prepared in a DNA LoBind tube by combining the following:

7.0 µl of 3′-dA-tailed vector pNL4-3 (150 ng/µl)
10.5 µl of 3′-dT-tailed barcode (11 ng/µl)
82.5 µl of dH_2O
100.0 µl of 2× Quick Ligation Buffer
10 µl of Quick Ligase (T4 DNA Ligase at 2000 U/µl).

The reaction was incubated at bench temperature (24 °C) for 2 h to perform the ligation at a 30:1 insert:vector molar ratio. The reaction was halted by adding 8 µl of (500 mM EDTA, pH 8.0) and mixing. Next, 10 µl of Proteinase K (800 U/ml) (New England Biolabs) were added, the reaction mixed, then incubated for 30 min at 37 °C to cleave bound T4 DNA Ligase from the DNA.

Sealing of nicks in hemiligated DNA

Nicked DNA was sealed by sequential treatment with T4 Polynucleotide Kinase (T4 PNK) and *Taq* DNA Ligase. Hemiligated DNA was 5′-phosphorylated with T4 PNK in T4 DNA Ligase Reaction Buffer (New England Biolabs). The 1× composition of T4 DNA Ligase Reaction Buffer as specified by the manufacturer is: 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH 7.5 at 25°C. During the 30 min Proteinase K treatment, 65 µl of T4 PNK (New England Biolabs) Master Mix was prepared by combining the following:

6.5 µl of 10× T4 DNA Ligase Reaction Buffer
57.5 µl of dH_2O
1.0 µl of T4 PNK (10 U/µl).

After 30 min, the reaction was purified with 1.8 reaction volumes (360 μl) of AMPure XP beads (Beckman-Coulter Genomics) per the manufacturer's direction, and eluted from the paramagnetic beads with 65 μl of the prepared T4 DNA Ligase master mix. The eluate was incubated at 37 °C for 60 min to phosphorylate DNA at the nicked sites.

After a 60 min incubation, the nicks were sealed by treatment with *Taq* DNA Ligase (New England Biolabs) in 1× *Taq* DNA Ligase Reaction Buffer (1× composition: 20 mM Tris-HCl, 25 mM potassium acetate, 10 mM magnesium acetate, 1 mM NAD^+, 10 mM DTT, 0.1% (v/v) Triton X-100, pH 7.6 at 25 °C). A 100 μl reaction was prepared by combining the following in a 0.2 ml PCR tube:

60.0 μl of T4 DNA Ligase Reaction from above
10.0 μl of 10× *Taq* DNA Ligase Reaction Buffer
26.0 μl of dH_2O
4.0 μl of *Taq* DNA Ligase (40 U/μl).

The reaction was then introduced into a thermocycler (Bio-Rad) with block temperature of 50 °C and heated lid temperature of 75 °C for 15 min to seal nicks. Ligated DNA was purified from the reaction mixture with 1.8 reaction volumes (180 μl) of AMPure XP beads per the manufacturer's instructions (Beckman Coulter Genomics), and eluted with 42 μl of (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0). The DNA was stored at 4 °C for future use.

Library transformation and outgrowth

Fifteen μl of the 40 μl of purified ligation ($\approx$ 30%) were electroporated into electrocompetent *E. coli* (strain DH10B) in 15 separate transformation reactions. In each transformation, 1 μl of the library was added to 40 μl of ice-cold electrocompetent cell slurry, mixed, and introduced into a chilled 0.1 cm sterile cuvette (Bio-Rad) placed on wet ice. Bacteria were

electroporated with a Gene Pulser II electroporation system (Bio-Rad) with pulse settings of 1.7 kV, 25 μF, 200 Ω (time constant of $\approx$5 msec). Immediately post-electroporation, 960 μl of SOC (Thermo Fisher) were added and the cell mixture transferred to a sterile 15 ml polypropylene conical tube, and allowed to recover for 90 min at 30 °C.

After a 90 min recovery, the 15×1.0 ml recovered transformations were pooled and mixed. One hundred μl of the 15 ml pool was added to 900 μl of SOB, vortexed to mix, then 100 μl of this dilution (corresponding to 1/1500 of the library pool) plated on each of 6×10 cm plates containing solid media comprised of LB-Miller supplemented with 1% agar and 100 μg/ml carbenicillin (LBMA+Carb_{100}). The remaining 14.9 ml of the transformation was used to inoculate 500 ml of SOB+Carb_{100} in a 2.8 L Fernbach flask and grown at 30 °C in a shaking incubator at 250 rpm for 24 h, until an OD_{600} of 3.0–4.0 was reached. The culture was chilled on wet ice and 450 ml of the 500 ml culture were harvested by centrifugation and used to prepare dry bacterial pellets for plasmid DNA isolation. The remaining 50 ml was spun down and resuspended in 14 ml of LB-Miller. To this 14 ml volume, 1.05 ml of pure DMSO (Corning) was added to a final concentration of 7% (v/v) DMSO. The cell slurry was aliquoted to 15 sterile cryovials (Corning) and stored at -80 °C as seed stocks for the future library preparations.

The six plates had CFU counts of {69, 68, 73, 63, 79, 71}, avg of 70.5, which led to estimate of plated library size as 106,000 CFU ($70.5 \cdot 1500$).

Genotyping of plasmid libraries

NGS libraries were prepared for paired-end sequencing on the Illumina HiSeq/MiSeq platforms. Separate sequencing libraries were prepared with a Nextera XT Kit (Illumina) from 1 ng of pNL4-3 insertion library and 1 ng of pNL4-3 deletion library. Transposon insertion and PCR enrichment were performed per the manufacturer's instructions, but the subli-

braries were pooled and size-selected by running out on a 1.5% agarose gel, staining with 1× SYBR Safe (Thermo Fisher), and excising a gel fragment corresponding to DNA of size range of 350–500 bp. DNA was purified from the gel slice using Qiagen Buffer QG, Buffer PE (Qiagen), and DCC-5 columns (Zymo Research). The sublibraries were pooled and sequenced on a single lane of a HiSeq4000 (Illumina), using 2×125 b reads at the Center for Advanced Technology at University of California, San Francisco.

Transposon insertion locations were computed by filtering for high-quality reads containing an exact match of either mosaic end sequence of TN5MK, then extracting flanking regions to build an insertion map.

A lookup table matching deletion loci to barcode sequence was determined by 1) searching reads for the forward oBC20-F (`CTGTCTCTTATACACATCTGCGGCCGC`) and reverse (oBC20-R (`CTGTCTCTTATACACATCTGCGGCCGC`) common barcode sequences and extracting the intermediate 20 b 2) assembling a list of barcode sequences 3) assigning flanking regions to each barcoded deletion using custom Python software.

Creation of ZIKV deletion libraries

Some modifications were made with respect to the HIV library. The structure of the barcode cassette was changed slightly, such that the fwd oligo oBC20v2-F (`/5Phos/CCGTCCATGAAGCGTTCGAT`) and rev oligo oBC20v2-R (`/5Phos/ACGAATCTGCCGTTGCCATA`) and the corresponding template oligo was:

`CCGTCCATGAAGCGTTCGATNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT`.

These sequences were modified to change a triple repeat in the forward barcode read (`GGG`), to avoid problems with sequencing of low-diversity libraries on the Illumina HiSeq4000.

2.4 Construction of an HIV-1 random deletion library

We sought to develop a method that would allow for the production of large, diverse libraries of barcoded viral deletion mutants in order to increase the sample space of potential transmissible antivirals and identify key *cis*-acting elements.

The proposed method is a deletion cycle which combines three technologies to perform deletions: *in vitro* transposition to seed clones with a common and unique restriction sites, dual-exonuclease chewback to create truncations, and molecular barcodes so that the population of mutants could be tagged and tracked over time by deep sequencing of the barcode cassettes. Thus, rare mutants could be followed over time with resolution unavailable to standard sequencing of pooled viral nucleic acids [59]. We consider these three aspects in turn: *in vitro* transposition, exonuclease digestion, and molecular barcoding.

Development of transposon cassettes

We designed a pair of synthetic transposon cassettes (TN5MK and TN5MC) that are compatible with the well-characterized hyperactive Tn5 transposase [63, 64, 68]. Schematics of both cassettes are shown in Figure 2.3. The hyperactive Tn5 transposase is commercially available (EZ-Tn5 Transposase, Epicentre), but can also be purified in-house [69].

Each transposon cassette contains an antibiotic-resistance marker so that plasmids harboring the transposon can be selected with antibiotics (TN5MK: kanamycin, TN5MC: chloramphenicol). The pair of cassettes are otherwise identical: we generated a pair to give a choice of selection in the event that the plasmid of interest has a preexisting kanamycin or chloramphenicol resistance marker.

Each transposon also contains a pair of meganuclease recognition sites at each end (Figure 2.4). Integration of the transposon cassette into a target DNA molecule introduces unique restriction sites for meganucleases I-*Sce*I and I-*Ceu*I, which have long recognition

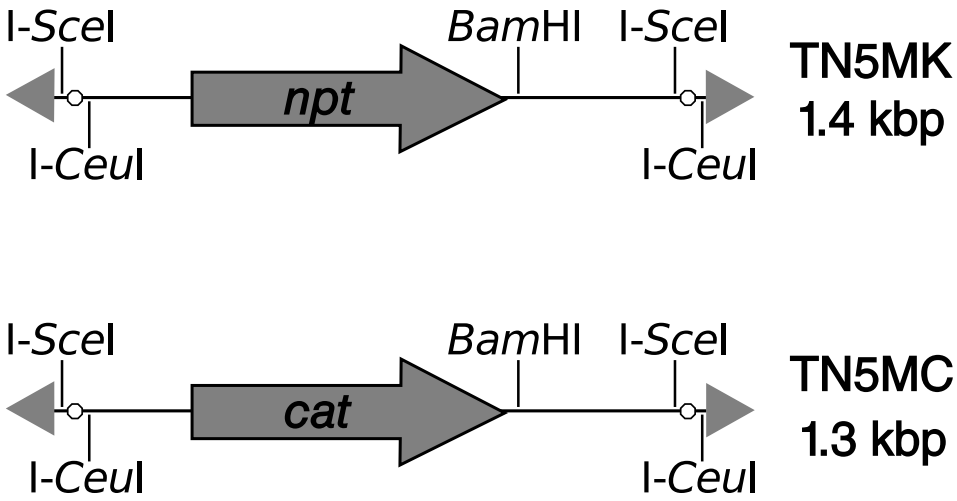


Figure 2.3: Meganuclease transposons TN5MK and TN5MC. Each synthetic transposon is comprised of an antibiotic resistance cassette (K: kanamycin-resistance encoded by neomycin phosphotransferase I (*npt*); C: chloramphenicol-resistance encoded by chloramphenicol acetyl transferase (*cat*)), flanked by meganuclease restriction sites for I-*Sce*I and I-*Ceu*I and Tn5 mosaic ends (gray triangles). Each transposon also contains a internal unique *Bam*HI recognition site. Two versions were developed to offer a choice in antibiotic selection in the event that the target plasmid has a preexisting *npt* or *cat* marker.

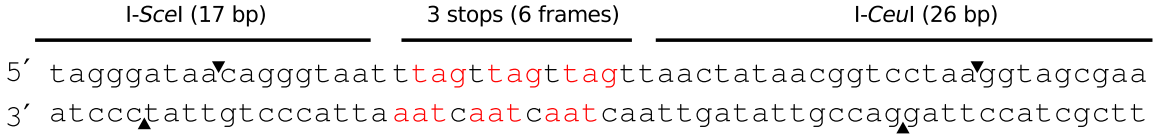


Figure 2.4: Detail of meganuclease restriction sites for I-*Sce*I and I-*Ceu*I encoded in transposons TN5MK and TN5MC.

sites (17 or 26 bp, respectively). The length of the recognition site confers specificity, and these meganuclease are infrequent or non-cutters in most genomes, including the human genome. As such, we anticipate that these cassettes can be used without modification in many systems.

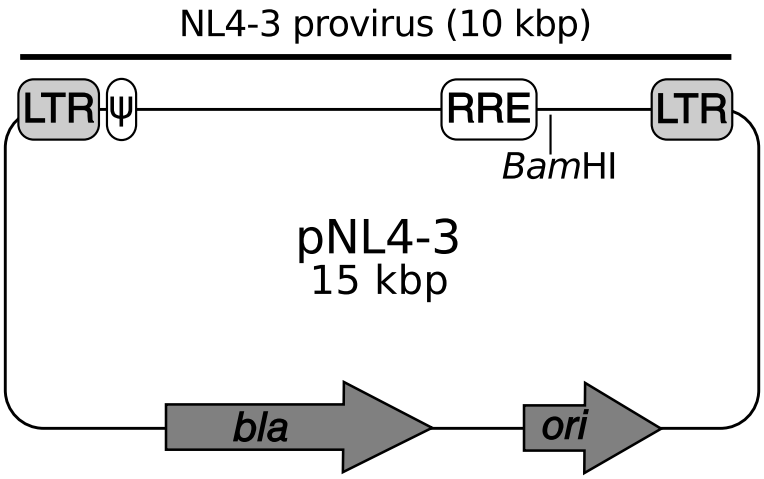


Figure 2.5: pNL4-3, an HIV-1 molecular clone, is a 14825 bp plasmid harboring the 9709 bp NL4-3 provirus (HIV-1 subtype B). NL4-3 is a chimera of two viruses (*N*Y5 and *L*AV). The plasmid was generated from bacteriophage λ libraries of human cells infected with NY5 and LAV and cloned into pUC18 at the *Pvu*I site. Approximately 2.7 kbp of the plasmid is human genomic DNA, which flanks the 5′ and 3′ long terminal repeats (LTRs). *bla*: β-lactamase (ampicillin/carbenicillin resistance), *ori*: origin of replication.

Production of an insertion library

We first validated the synthetic meganuclease transposons by creating an insertion library of the classic HIV-1 molecular clone, pNL4-3, a 15 kbp plasmid harboring the 9709 bp NL4-3 provirus (Figure 2.5). Using *in vitro* transposition, we integrated TN5MK randomly into pNL4-3 at a ratio of one TN5MK transposon per pNL4-3 plasmid, then transformed electrocompetent *E. coli* with the reaction mixture and selected for TN5MK integrants by dual selection with kanamycin and carbenicillin. After an overnight incubation, we observed a total of $\approx$ 10,000 CFU on pNL4-3 insertion library plates (nominal integration rate of 1 integration per 1.5 bp), while no colonies were observed on the Tn5 transposase negative control. Hereafter, we refer to the TN5MK in pNL4-3 insertion library as pNL4-3$^+$ for simplicity.

After purification of the transposed DNA, the result is a polyclonal population of pNL4-3$^+$,

where each plasmid contains a different single transposon insertion, i.e., a population of transposon-inserted circular target DNAs. We confirmed integration of the transposon cassettes in the plasmids by digesting the insertion library (pNL4-3$^+$) with I-*Sce*I or I-*Ceu*I (Figure 2.10 on page 52). Digestion with either meganuclease excises the 1.4 kbp TN5MK cassette with a concomitant gel shift of the vector DNA due to loss of supercoiling.

We next characterized pNL4-3$^+$ by deep-sequencing the plasmid library. After quality-filtering, we searched high-quality reads for Tn5 mosaic end sequences and developed a map of insertion sites across the genome. A map of TN5MK insertions within pNL4-3 is shown below in Figure 2.6.

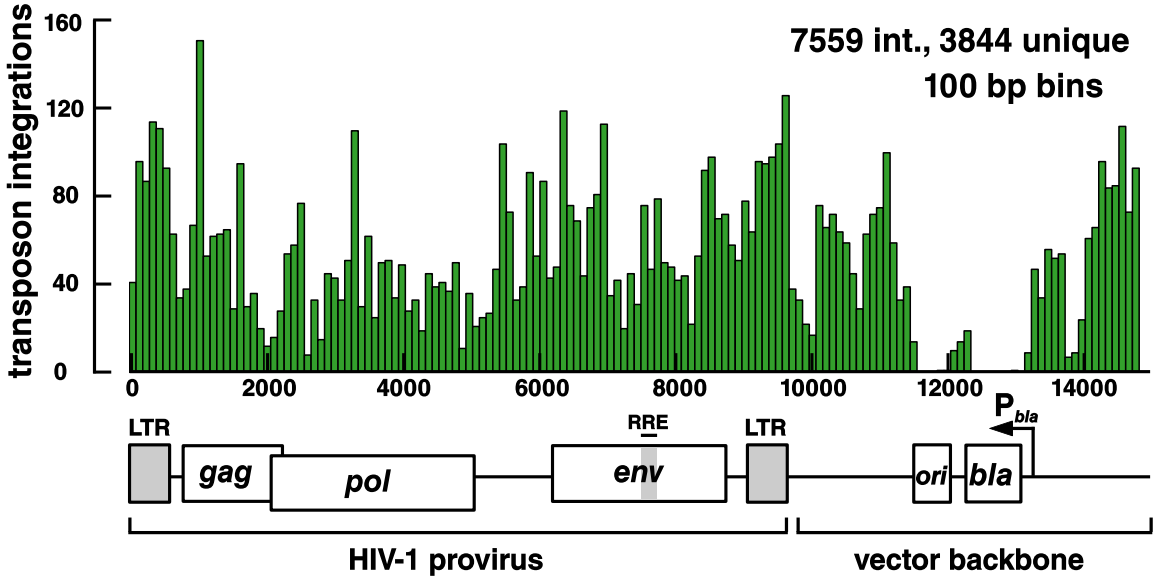


Figure 2.6: Location of TN5MK insertions in pNL4-3$^+$ reveal integration of the transposon throughout the genome. We observed 7559 integrations at 3844 unique locations. Note the lack of insertions in *ori* and *bla*, which were selected against during outgrowth as they prevent proper replication of the plasmid in *E. coli*.

TN5MK was integrated broadly throughout pNL4-3 at a diversity of sites, including 3844 unique locations. Due to the near identity of the HIV-1 LTRs, we were unable to determine if some integrations were present in the 5′ or 3′ LTR. Across the HIV-1 genome, transposons

were integrated at high density and without large gaps between neighboring integrations—there was at least one transposon integrated every 100 bp. In the plasmid backbone, there were two integration gaps: one in origin of replication, *ori*, and the other in the ampicillin/carbenicillin resistance marker, β-lactamase, as both are required for propagation of the plasmid in *E. coli*. This provides a natural control for sequencing, as insertion mutants in these two regions are selected against during growth in culture.

Exonuclease chewback and modulation of deletion size

Many of the existing methods to create truncations of dsDNA utilized enzymes that were incompatible in the same buffer system, notably Exonuclease III and Mung Bean Nuclease. Both enzymes require careful adjustment of template amounts, salt concentration, temperature, and duration of incubation to favor exonuclease activity over low levels of inherent endonuclease activity.

We elected to use a pair of enzymes with complementary single-stranded exonuclease activities on double-stranded DNA, and could utilize a common buffer system. Deletions are created by simultaneous treatment with two exonucleases. T4 DNA Polymerase has predominantly $3' \rightarrow 5'$ exonuclease activity on dsDNA when incubated with magnesium cations and no dNTPs. $\mathrm{RecJ_f}$ has $5' \rightarrow 3'$ exonuclease activity on $5'$ overhangs of at least 7 bp, and increased exonuclease activity in the presence of single-stranded binding protein (SSB) [62]. Treatment with this trio of proteins creates a population of truncated deletion mutants with ragged ends.

We verified minimal chewback conditions using λ-*Hind*III digest as template. λ-*Hind*III consists of 8 DNA fragments ranging from 125 bp to 20 kbp, and shortening of any of these fragments from chewback will result in lower apparent molecular weight and increased migration in an agarose gel. In Figure 2.7 and Figure 2.8, during a 30 min chewback reaction (no

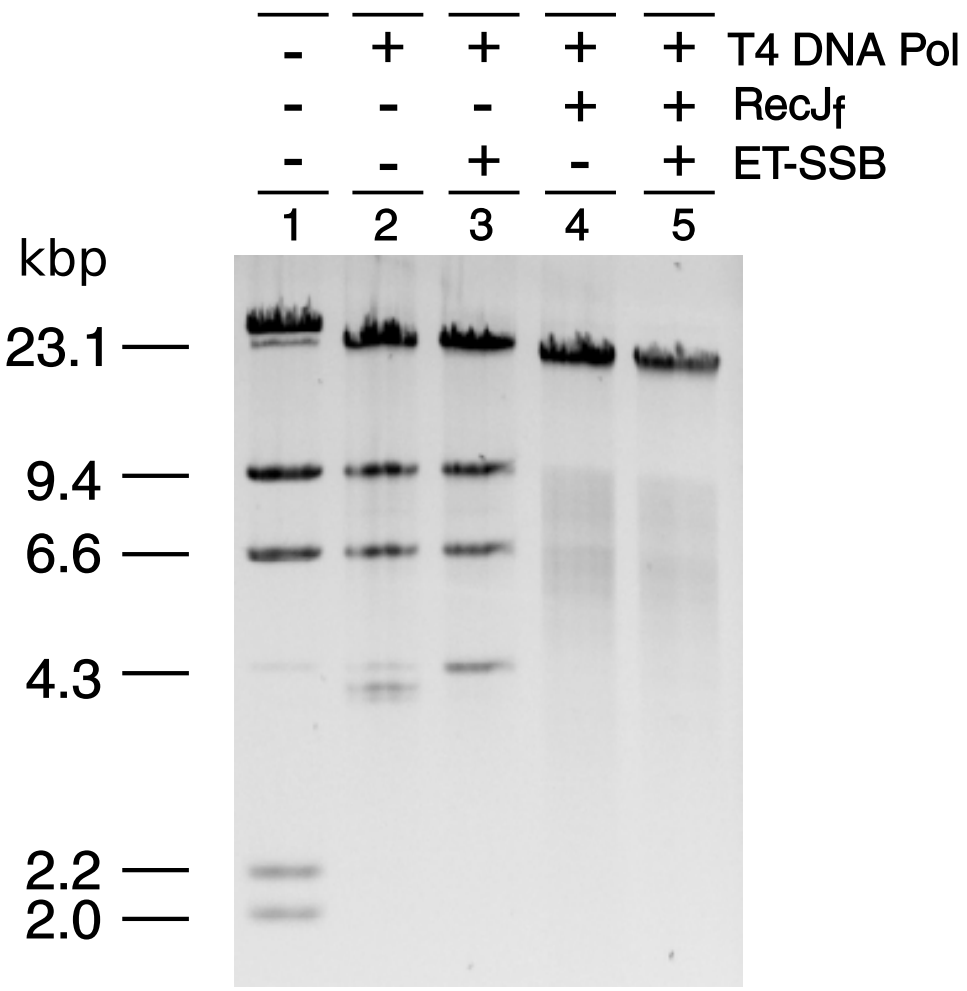


Figure 2.7: Minimal conditions to perform a chewback reaction. λ-*Hind*III digest was treated with a combination of enzymes and without dNTPs for 30 minutes, then dNTPs were added to fill-in the ends at 37 °C. For the various dropout reactions, dH_2O was substituted for enzyme solutions. Lane 1: 0 μl enzyme, Lane 2: 1.7 μl T4 DNA Pol, Lane 3: 1.7 μl T4 DNA Pol, 0.5 μl ET SSB, Lane 4: 1.7 μl T4 DNA Pol, 0.5 μl $RecJ_f$, Lane 5: 1.7 μl T4 DNA Pol, 0.5 μl $RecJ_f$, 0.5 μl ET SSB.

dNTP) and 10 min fill-in (200 μM dNTP), we only observe a decrease in molecular weight when both T4 DNA Polymerase and $RecJ_f$ were present. Addition of SSB protein appeared to slightly increase the chewback efficiency (compare lanes 4 to lane 5), and helped resolve formation of structures between *cos* sites (compare lanes 2,3).

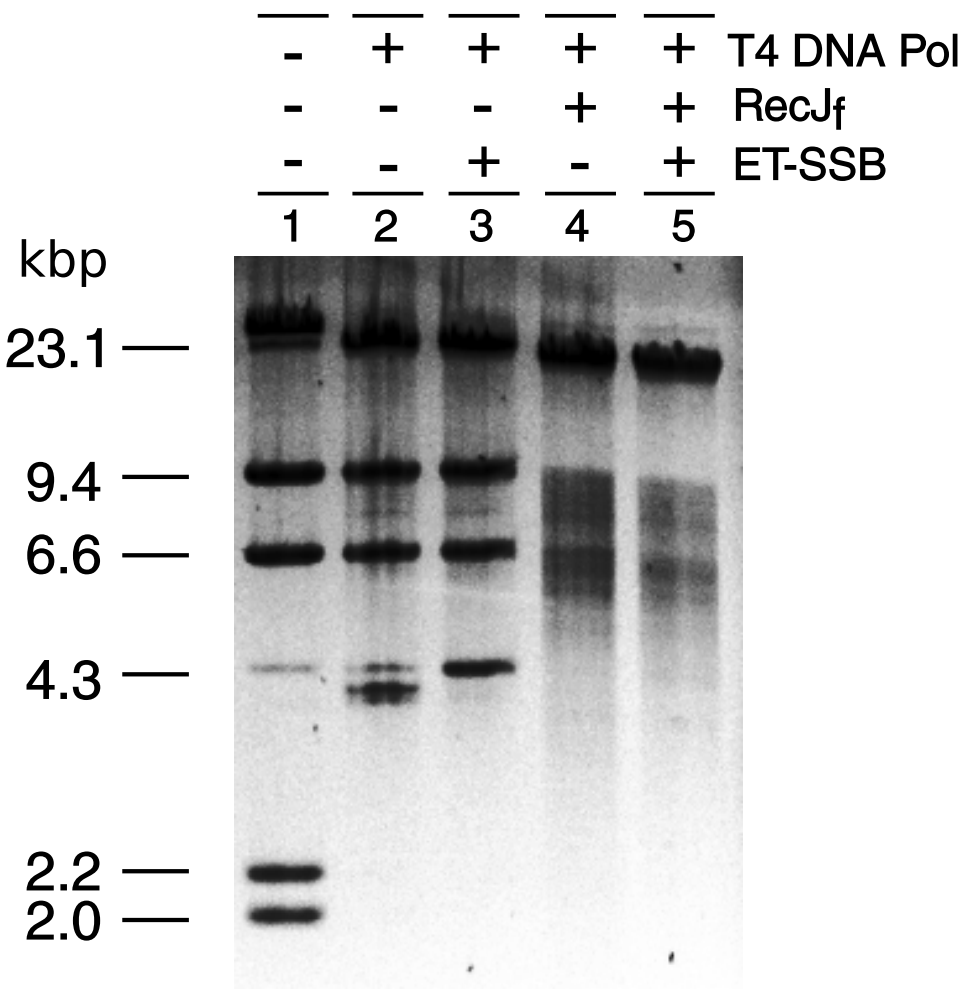


Figure 2.8: Minimal conditions to perform chewback reaction, overexposed. The gel is the same as in Figure 2.7, but with increased exposure time and contrast to visualize low abundance DNA.

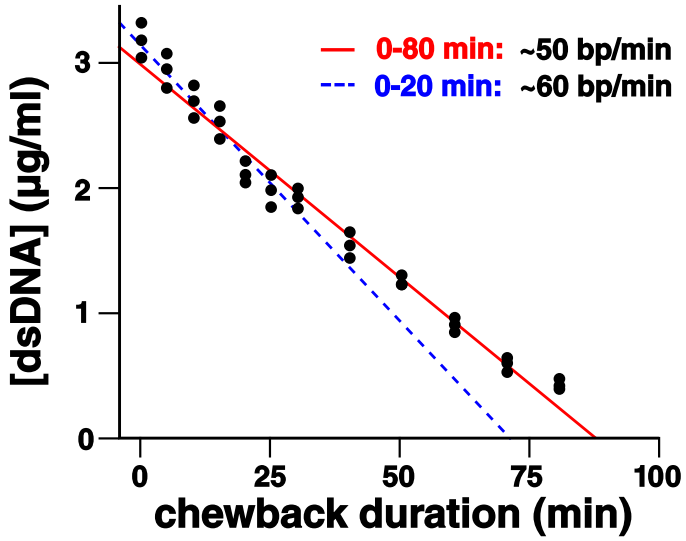


Figure 2.9: Dual-end chewback rate is 50-60 bp/min. The chewback rate was determined by treating a 4 kbp fragment of dsDNA with the specified conditions of T4 DNA Polymerase, $\mathrm{RecJ_f}$, and ET SSB for increasing amounts of time, then halting chewback by adding dNTP to 200 μM to allow T4 DNA Polymerase to fill in 5′ ends for 10 minutes. Reactions were halted by adding EDTA to 20 mM. Reactions were performed in triplicate. DNA concentrations were established by measuring fluorescence of PicoGreen with a plate reader and comparing to a λ DNA standard.

We next determined the chewback rate at 37 °C by a dsDNA fluorometric method, using a pure 4 kbp fragment of the λ-*Hind*III digest as template. As the ends were progressively shortened by chewback, the fluorescence signal of a dsDNA-specific dye (PicoGreen) decreased proportionally. The measured double-end chewback rate, as determined by linear regression, was approximately 50–60 bp/min, as shown in Figure 2.9.

An HIV-1 deletion library

Now confident that we could construct a transposon insertion library and had a functioning method for performing dsDNA chewback, we combined the two methods to create a deletion library of pNL4-3.

TN5MK transposons were excised from pNL4-3$^+$ by digestion with I-*Sce*I to generate a population of linear DNAs. A chewback reaction was setup using the linearized DNA as template and 25% of the reaction was removed after 5, 10, 15, and 20 minutes and ragged ends allowed to fill-in by adding dNTPs for 10 min. Thus, using the chewback rate we determined previously (50–60 bp/min), we expected 25% of the library to have a mean deletion sizes of 300 bp (5 min chew), 600 bp (10 min chew), 900 bp (15 min), 1200 bp (20 min). The four sublibraries were pooled, end-repaired, dA-tailed/dephosphorylated to prevent background ligation, then recircularized by ligation to a 3′ dT-tailed barcode cassette a 3′-dT-tailed barcode cassette drawn from a pool of random barcode cassettes (a 60 bp cassette including 20 bp random barcode flanked by 20 bp primer binding sites). In this way, each deletion mutant was tagged with a unique barcode cassette with high probability. The hemiligations were nick sealed and electrocompetent *E. coli* transformed with a portion of the ligation and expanded in liquid culture. For convenience, we refer to the tagged deletion library as pNL4-3Δ_1.

By plating small sample of the library (0.1%) before outgrowth, we estimated a plated size of 110,00 CFU (99% CI of 93,000–120,000 CFU). In a control reaction, where barcode cassettes were not included in the ligation, fewer than 10% of the number of colonies on the actual were observed, indicating low levels of background ligation.

To estimate the tagging efficiency of pNL4-3Δ_1, we performed colony PCR on the plated library with primers specific to the common regions of the barcode cassette. Of the 24 colonies picked, 23 were positive for the barcode cassette, giving an estimated tagging efficiency of

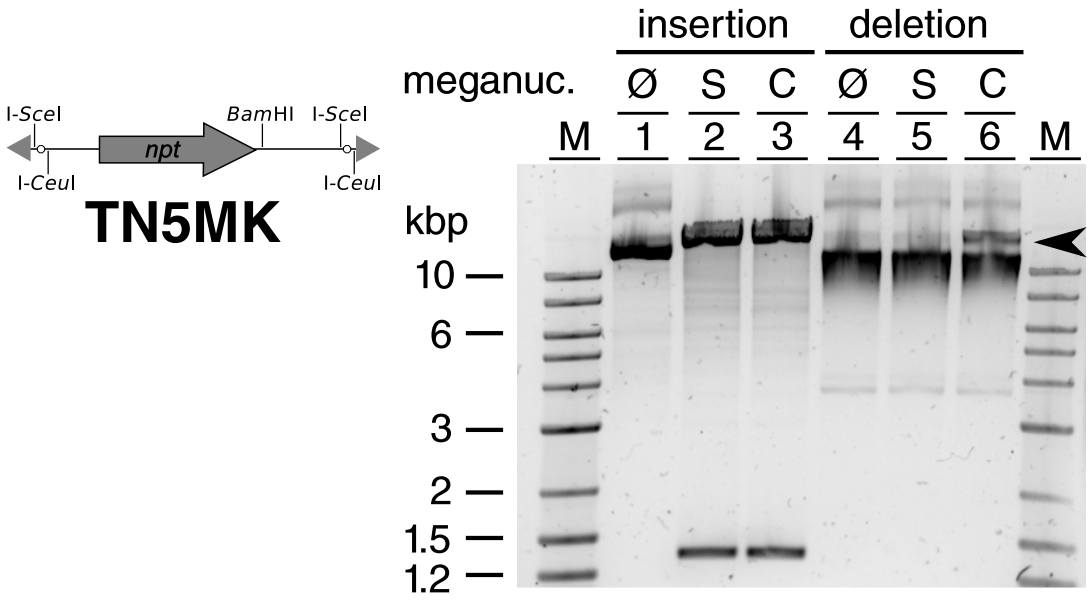


Figure 2.10: Loss of meganuclease sites in pNL4-3Δ_1. The pNL4-3$^+$ (insertion library) and pNL4-3Δ_1 (deletion library) were either undigested (∅) or cut with a single meganuclease (I-*Sce*I (S) or I-*Ceu*I (C)). Digestion of the insertion library shows release of the 1.4 kbp TN5MK transposon and upward shift of the supercoiled library (lanes 1-3). Treatment of the deletion libraries under the same condition shows that all I-*Sce*I sites have been lost, but some I-*Ceu*I sites remain, indicating incomplete chewback of the transposon flanks (indicated by the arrowhead in lane 6).

96% (95% CI of 78–99%).

We confirmed that we had created a tagged deletion library using several orthogonal methods. First, we assessed if the TN5MK termini left after excision with I-*Sce*I had been completely removed during chewback, by digesting pNL4-3$^+$ and pNL4-3Δ_1 with I-*Sce*I and I-*Ceu*I (Figure 2.10). The transposon insertion library (pNL4-3$^+$) could be cleaved by I-*Sce*I and I-*Ceu*I as predicted. In contrast, the deletion library (pNL4-3Δ_1) was insensitive to cleavage by I-*Sce*I, and only partially sensitive to cleavage by I-*Ceu*I. The residual I-*Ceu*I sensitivity is likely due to small deletions, where chewback had not progressed far enough to remove the 30 bp of genetic material left after excising the transposon.

The parental pNL4-3 plasmid harbors a unique *Bam*HI restriction site. Upon integration of TN5MK, a second *Bam*HI site is introduced as TN5MK contains an internal *Bam*HI site (see inset in Figure 2.11. Thus, all members of pNL4-3$^+$ have two *Bam*HI sites unless TN5MK is integrated within the pre-existing *Bam*HI site.

Treatment of pNL4-3$^+$ with *Bam*HI will cleave the plasmid into two pieces with a combined length of ≈ 16.3 kbp (14.9 kbp plasmid and 1.4 kbp TN5MK transposon), as there are two cut sites: one within the transposon and one in the plasmid backbone. Thus, we expect to observe a broad distribution in fragment sizes with gaps corresponding to regions where the TN5MK transposon could not stably integrate and persist, such as the *ori* and *bla*.

After excision of the TN5MK transposon, only one *Bam*HI restriction site remains (situated within *env*). If the *Bam*HI site is intact, treatment with *Bam*HI will linearize the plasmid. In the deletion library, we expect some deletions to span this site and delete the *Bam*HI locus, rendering the plasmids insensitive to *Bam*HI digestion. After treatment of the pNL4-3Δ_1 deletion library with *Bam*HI digest and RecBCD (which digests linear DNA to single nucleotides), we show that there is a population of clones that are resistant to *Bam*HI/RecBCD, indicating that *Bam*HI has been deleted (Figure 2.11). In contrast, similar treatment of the insertion library shows that all mutants are sensitive to *Bam*HI/RecBCD.

Furthermore, in Figure 2.11, we see that the deletion library exhibits a range of subgenomic sizes (diffuse spread of supercoiled DNA under 10 kbp). In the insertion library, we observe a range of integration sites, as exhibited by the evident smear due to the difference in length between the preexisting *Bam*HI site and the *Bam*HI site introduced by integration of the TN5MK transposon cassette.

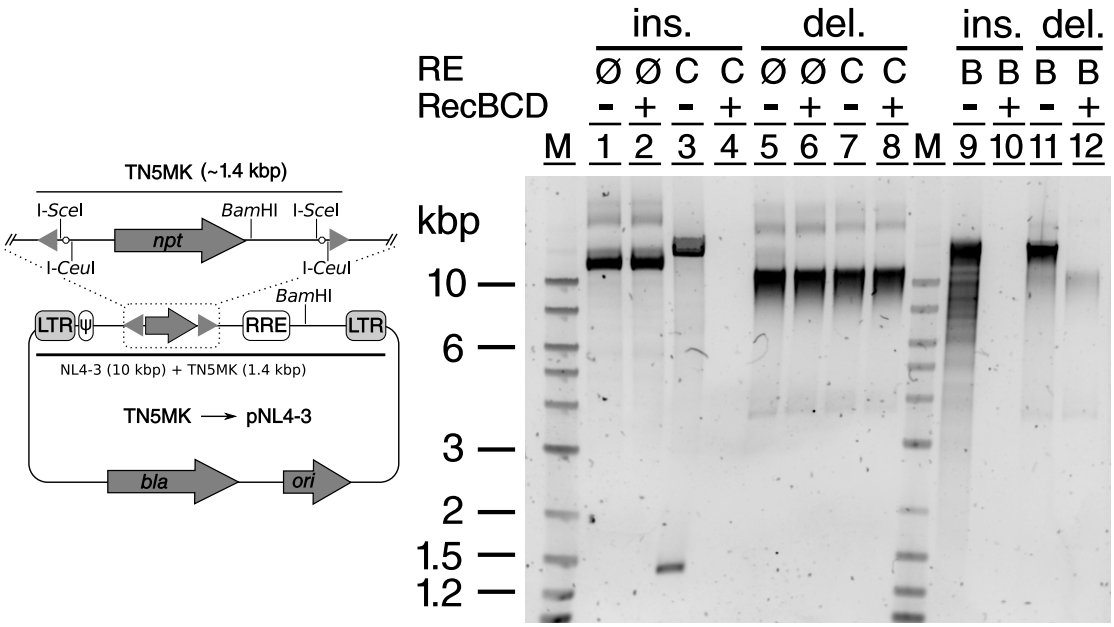


Figure 2.11: Loss of *Bam*HI sites in the deletion library pNL4-3Δ_1. The pNL4-3 plasmid contains a *Bam*HI site at base 8465, between RRE and the end of *env*. The pNL4-3 insertion library and pNL4-3 deletion library were not digested (∅) or cut with I-*Ceu*I (C) or *Bam*HI (B) and subjected to binary treatment with RecBCD, which digests linear DNA to completion. The undigested insertion (lanes 1,2) and deletion libraries (lanes 5,6) are not sensitive to RecBCD, indicating they are circular DNA. After I-*Ceu*I digestion, the transposon insertion library is sensitive to RecBCD (lanes 3,4) (has I-*Ceu*I sites) while the deletion library is insensitive (loss of transposon) (lanes 7,8). Digestion with *Bam*HI shows a smear in the insertion library (lane 9), where the sizes differ depending on where the TN5MK integrates relative to the *Bam*HI site. Treatment with RecBCD shows no DNA remaining (lane 10). However, in the deletion library, we see a range of sizes peaking around 15 kbp (lane 11), and a population is resistant to *Bam*HI/RecBCD (lane 12), indicating deletion of the preexisting *Bam*HI site.

NGS sequencing of pNL4-3 library

We next performed next-generation sequencing to build a mapping between barcode cassette and deletion locus. The plasmid library was fragmented and deep-sequenced (2×125b reads on HiSeq4000), and the results used to link barcode sequences to deletion sites. The result was a lookup table mapping the set of barcodes ($B = \{b_1, b_2, b_3, \dots\}$ to deletion loci ($D = \{d_1, d_2, d_3, \dots\}$).

After deep sequencing of the library of the (estimated plated size of $\approx 10^5$ CFU), we could uniquely map 23,851 barcode cassettes to deletions, yielding the tagged diversity of the library as 23,851 mutants (see Table 2.1).

description	pNL4-3Δ_1
parental plasmid size (C)	14,825
genome length (v_0 length)	9709 bp
plated size (CFU)	110,000
sequencing library	LTN3A-01
read type	2×125 bp
number of PF clusters	148,418,934
reads with barcode	4,356,618 (2.9%)
number of barcodes	116,427
number of unique barcodes	25,095
number of mappable deletions	23,851
mean deletion size	1,420 bp
median deletion size	1,110 bp
stdev. of deletion size	712 bp

Table 2.1: Summary of pNL4-3Δ_1

Plotting the deletion depth profile of pNL4-3Δ_1 in Figure 2.12, we see fairly even coverage over the plasmid library, with a peak at *env*, and a region of zero deletion depth at *ori*/*bla*. Deletion of these regions on the plasmid backbone comprises the ability of the plasmid to be

propagated effectively in *E. coli*. The peak centered at the N-terminus of *env* likely reflects increased growth of deletion mutants with deletion in the signal-peptide of *env*, a region known to be toxic in bacteria. Across the NL4-3 proviral genome, each base is covered by several hundred deletion mutants.

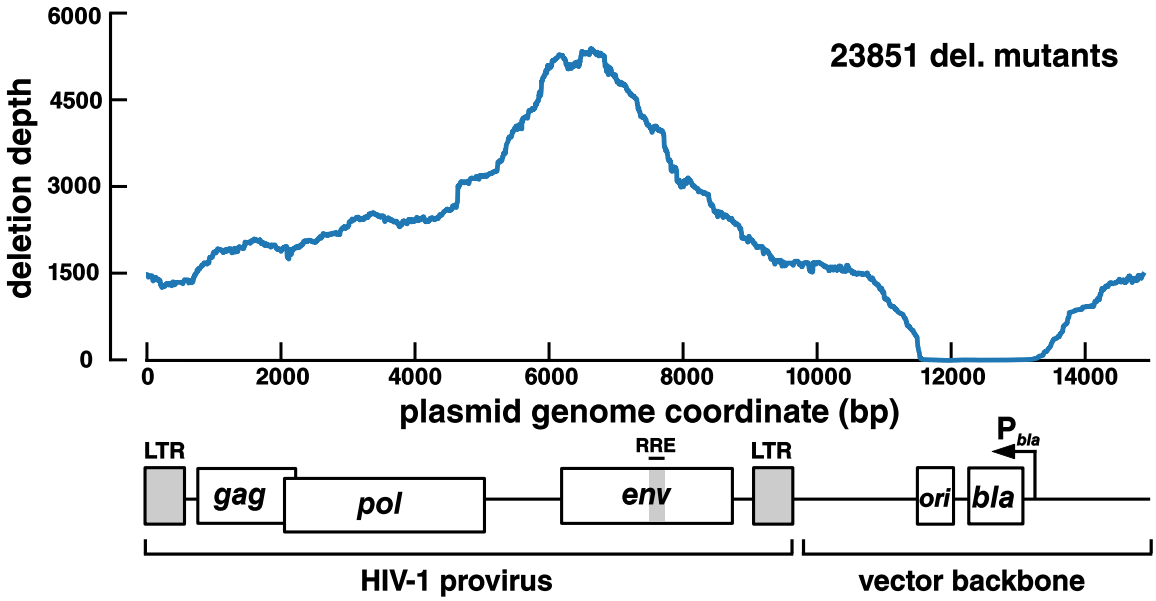


Figure 2.12: Deletion Depth Profile reveals saturating coverage of the HIV-1 genome. Calculation of the deletion depth profile of the pNL4-3Δ_1 genome indicates that each base is covered by hundreds to thousands of deletion mutants. Regions where deletion depth is large indicate that the deletions are tolerated in this region. Note the two regions where deletions are not tolerated in the plasmid backbone: *ori*, the origin of replication and *bla*, β-lactamase, the ampicillin/carbenicillin resistance marker

The library was composed of a range of deletion sizes, shown in histogram format in Figure 2.13, with a median deletion size of ≈ 1.1 kbp with a minimum deletion of 30 bp and a maximum deletion size of > 6 kbp. The skewing of the library (with the long right tail), may be due to mechanical shearing of DNA during some of the cleanup steps.

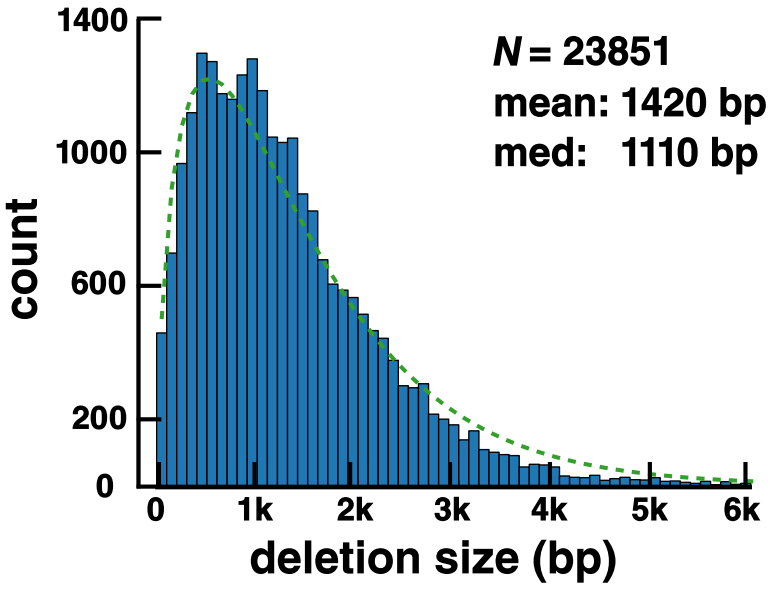


Figure 2.13: pNL4-3Δ_1 is composed of 23,851 tagged mutants with a range of deletion sizes. The right-skewed histogram of deletion sizes in pNL4-3Δ_1, with bins of 100 bp (shown in blue) is well-fit by a Γ distribution (green, broken-line). The mean deletion size of the 23,851 tagged mutants is 1420 bp, with a median deletion size of 1100 bp.

2.5 Construction of ZIKV random deletion libraries

Encouraged by our success with creating an HIV-1 deletion library, we next prepared deletion libraries of an unrelated virus, Zika virus (ZIKV), which had an available molecular clone. ZIKV is a flavivirus and has a (+)-stranded, ssRNA genome of approximately 11 kb, and replicates predominantly in the cytoplasm of infected cells. We elected to use a recently available molecular clone of MR-766, the classic 1947 Ugandan strain of ZIKV [61].

In this molecular clone, the ZIKV genome is encoded as a cDNA, and infectious virus generated by transfecting a permissive cell line (293T) with the plasmid. This is possible due to the presence the CMV IE promoter at the 5′ end and a hepatitis D ribozyme at the 3′ end, which cleaves itself out of the genome post-transcription to generate an authentic 3′ end. In addition, there is a synthetic intron located within NS1 to alleviate stability issues during

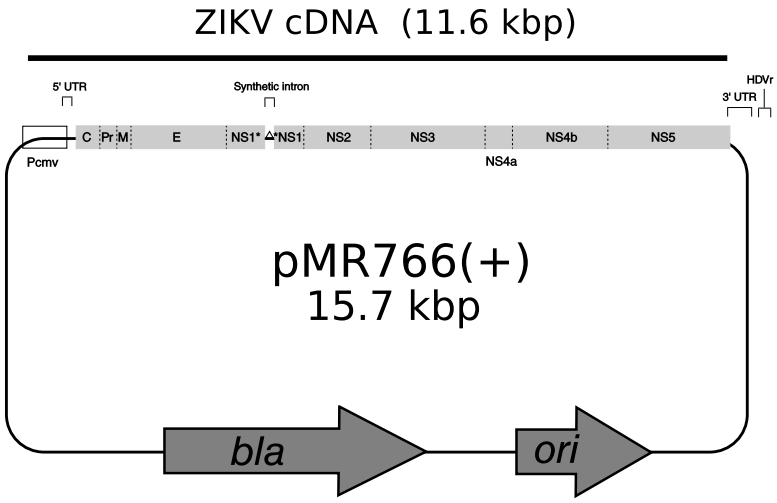


Figure 2.14: pMR766(+), a Zika virus molecular clone. In pMR766(+), the MR-766 Zika virus genome is encoded as a cDNA driven by the CMV IE2 promoter. At the 3′ end of the genome, a self-cleaving Hepatitis Delta ribozyme allows for creation of an authentic 3′ end post-transcription. An intron sequence is present within NS1 to allow maintenance in bacteria, but is spliced out during transcription in host cells.

growth in bacteria, due to the presence of a cryptic promoter in NS1 [61]. The synthetic intron is removed during splicing of the ZIKV transcript in the nucleus, reconstituting the genuine genome. Thus, the plasmid is only required to generate the initial viral RNA genome to seed further replication in the cytoplasm.

Two versions of the ZIKV MR766 molecular clone were available: Pol(+) MR766, which we refer to as pMR766(+), and Pol(−) MR766, which we call pMR766(−). A schematic of both plasmids is shown in Figure 2.14. The (+) version is the clone of the wildtype virus while the (−) version is replication-defective mutants due to a `GDD`→`GNN` residue substitution in the active site of NS5, the essential RNA dependent RNA polymerase. Thus, the MR766(−) virus is not replication-competent, but it is possible that it may be rescued by providing NS5 *in trans*.

We generated transposon insertion libraries of pMR766(+) and pMR766(−) by integrating TN5MK via *in vitro* transposition and recovered $\approx 10^3$ insertion mutants for each after

plating. Digestion of the polyclonal insertion libraries (pMR766(+)$^+$ and pMR766(−)$^+$) with I-*Sce*I and I-*Ceu*I revealed the presence of integrated transposons (Figure 2.15). We note one difference from HIV-1 insertion library in Figure 2.11: the presence of numerous subgenomic (< 15 kbp) DNAs. It is unclear if these were created by double integrations of TN5MK or from instability of the pMR766(+) and pMR766(−) plasmids.

We excised transposons from pMR766(+)$^+$ and pMR766(−)$^+$ by digestion with I-*Sce*I, then performed two chewback reactions for each insertion library. In chewback 1, we pulled samples at 5,10,15,20,25 minutes (mean deletion sizes of 300–1800 bp), and in chewback 2, we pulled samples at 30,40,50,60 minutes (mean deletion size of 1500–3600 bp). Thus member of chewback 1 (5-25 min chewback) will have an average smaller deletion than those in chewback 2 (30-60 min chewback). After chewback the truncated DNAs were end-repaired, 3′-dA-tailed/dephosphorylated, and ligated to 3′-dT-tailed/phosphorylated barcode cassettes, then electroporated into *E. coli* and outgrown in liquid culture.

Plating a sample of each library before outgrowth allowed us to estimate library sizes of 10^5 CFU. A control ligation, where barcodes were not included, generated fewer than 10% of the genuine ligations. At this point, we have generated a total of four pMR766 tagged deletion libraries— pMR766(+)Δ_1, pMR766(+)Δ_2, pMR766(−)Δ_1,pMR766(−)Δ_2 —each genetic background has a “short” and “long” chewback duration, therefore different deletion sizes.

We established the tagging efficiency of each library by picking a number of single colonies and performing PCR for the barcode cassette and the ZIKV C region. Typically, 5 of 6 colonies contained a barcode cassette ($\approx$ 80% tagged).

We next characterized libraries by treatment with restriction endonucleases, using the same approach as the HIV-1 library. In Figure 2.15, we see that the undigested libraries (supercoiled) show a range of apparent sizes that fall between 2–10 kbp on the linear DNA

ladder, whereas the insertion libraries are predominantly a single size that migrates at or above 10 kbp.

pMR766(+) and pMR766(−) harbor a unique restriction site for *Kpn*I. Digestion with *Kpn*I will linearize the plasmid unless the site is deleted. Digestion with *Kpn*I shows a range of sizes < 15 kbp for the deletion libraries, but a single mass for insertion libraries. The discrete bands at 2 and 1.5 kbp evident in the deletion libraries were not sensitive to digestion with *Kpn*I, I-*Sce*I, or I-*Ceu*I.

NGS genotyping of the ZIKV deletion libraries

After whole plasmid sequencing of the four pMR766 series deletion libraries, we observed tagged and genotyped diversities that ranged between 10^3 and 5×10^4 tagged deletion mutants (Table 2.2). The rationale for the low diversity in the short-chewback libraries (Δ_1) versus the long-chewback (Δ_2), is that the short-chewback reactions failed to chew past the transposon cassette, rendering the mutation unable to be located properly with the short read length (125 b).

In addition, we observed a deletion size distribution that was different than HIV (Figure 2.16). The distribution was clearly bimodal with peaks at small and large deletion sizes. Increasing the length of the chewback (duration of chewback) shifted the lower peak, but not the upper peak. One reason for this is that clones that have lost the ZIKV cDNA insert have have a much greater replication advantage in outgrowth in *E. coli* than the wildtype genome.

Plotting the deletion depth profile of pMR766(+)Δ_2 in Figure 2.17, we see fairly even coverage over the ZIKV genome, with a peak at NS1, and a region of zero deletion depth at the flanking region of the genome, which corresponds to the plasmid backbone, harboring *ori*/*bla*. Deletion of these regions on the plasmid backbone comprises the ability of the

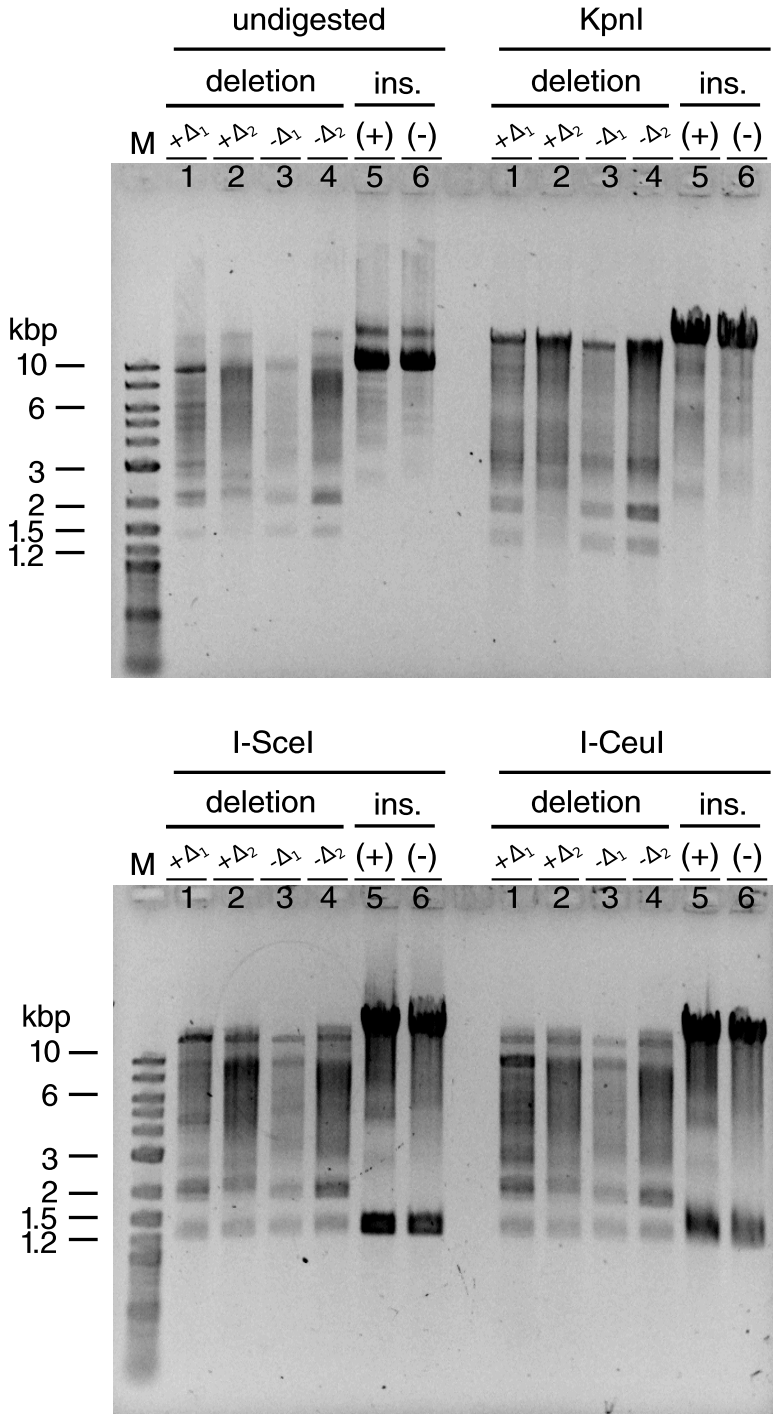


Figure 2.15: Diagnostic gels of ZIKV (pMR766) insertion/deletion libraries

plasmid to be propagated effectively in *E. coli*. The peak centered at NS1 likely reflects increased growth of deletion mutants with deletion in the NS1 which is known to have cryptic promoter activity and cause reduced growth in *E. coli* [61, 70]. Across the ZIKV cDNA genome, each base is covered by several hundred deletion mutants.

description	pMR766(+)Δ_1	pMR766(+)Δ_2	pMR766(−)Δ_1	pMR766(−)Δ_2
parental plasmid size (C)	15,739	15,739	15,739	15,739
genome length (v_0 length)	9709 bp	9709	9709 bp	9709 bp
plated size (CFU)	$\approx$100,000	$\approx$100,000	$\approx$100,000	$\approx$100,000
sequencing library	LTN5A-01	LTN5A-02	LTN5A-03	LTN5A-04
read type	2×125 bp	2×125 bp	2×125 bp	2×125 bp
number of PF clusters	106,081,700	92,750,026	117,525,278	67,106,817
reads with barcode	3,672,013 (1.7%)	4,467,985 (2.4%)	4,355,269 (1.9%)	3,050,618 (2.3%)
number of barcodes	116,427	152,214	53,954	121,152
number of unique barcodes	1,645	48,277	4,324	48,277
number of mappable deletions	1,608	41,553	4,292	47,455
mean deletion size	7,274 bp	5,747 bp	6,140 bp	6,110 bp
median deletion size	8,224 bp	5,408 bp	6,787 bp	5,724 bp
stdev. of deletion size	3,668 bp	3,441 bp	4,182 bp	3,025 bp

Table 2.2: Summary of pMR766Δ series libraries

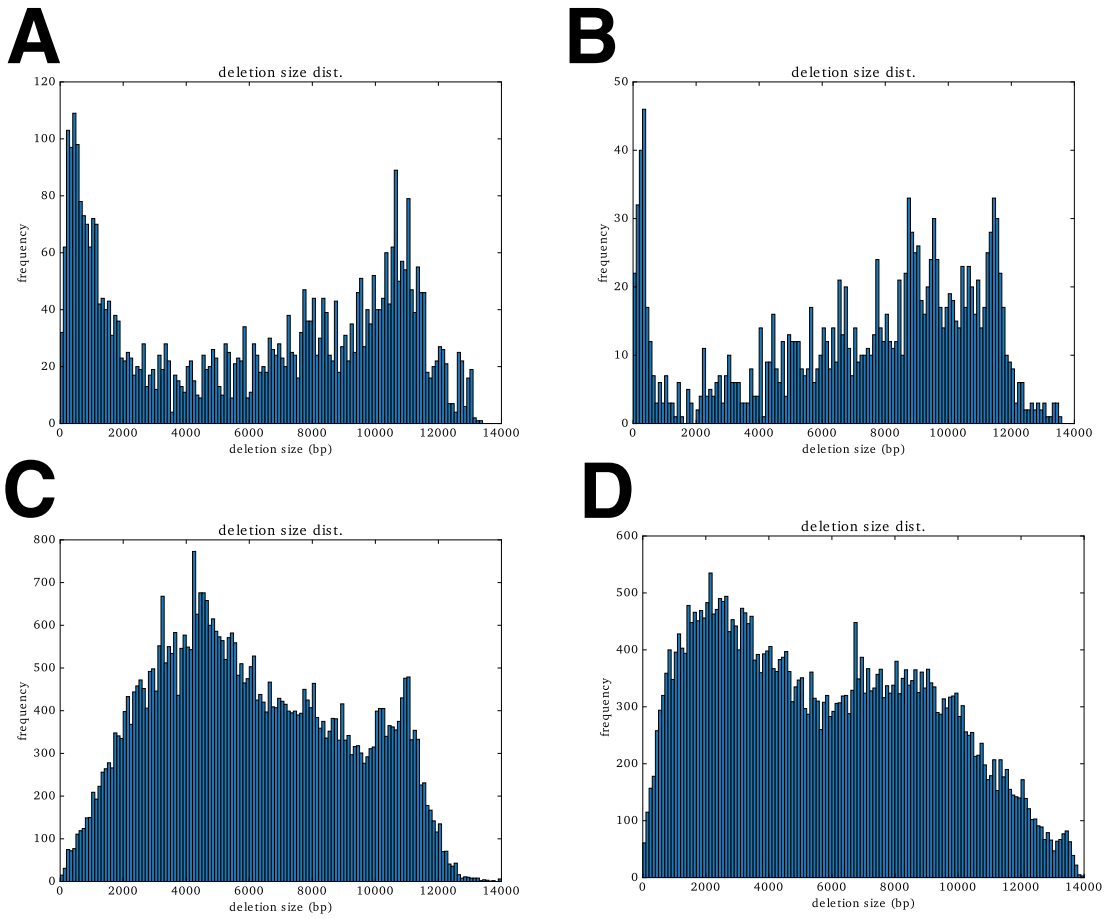


Figure 2.16: pMR766 deletion libraries have a bimodal distribution of deletion sizes. Histograms of deletion sizes are shown for the four ZIKV deletion libraries: two short-chewback (A) pMR766$(-)\Delta_1$, (B) pMR766$(+)\Delta_2$; and two long-chewback (C) pMR766$(-)\Delta_2$, (D) pMR766$(+)\Delta_2$.

2.6 Discussion

The products of this chapter are 5 random deletion libraries: one HIV-1 deletion library and four Zika virus deletion libraries. Each library consists of 1×10^3–5×10^4 barcoded deletion mutants. From both deep sequencing and classical diagnostic methods, we are able

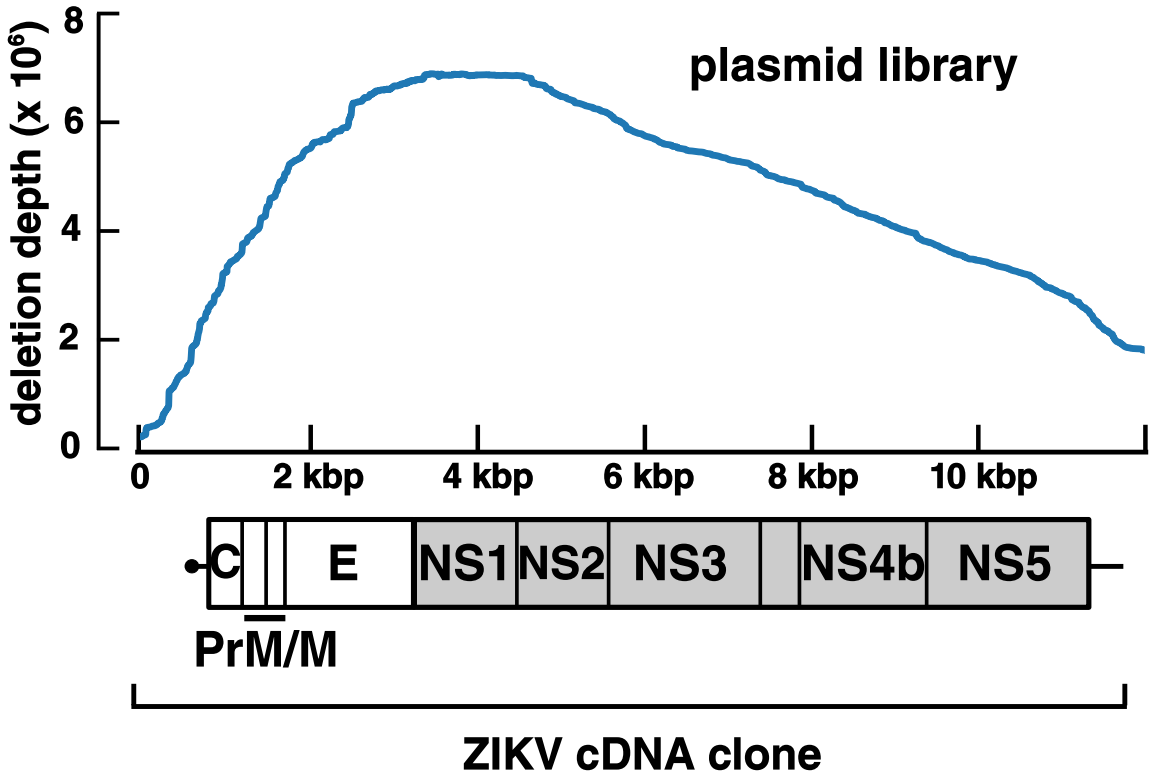


Figure 2.17: pMR766(+)Δ_2 has comprehensive coverage across the ZIKV genome. The deletion depth profile indicates that each base in the ZIKV genome is covered by several hundred deletion mutants.

to show that the libraries are genuine deletion libraries: these tagged mutants are subgenomic mutants of the wildtype plasmid.

In the introduction to this chapter, we set out a list of eight requirements that the library generation method should meet. We now briefly review these and confirm that we have satisfied each requirement.

Diverse libraries within 5 days: Although all libraries consisted of $< 10^5$ tagged deletion mutants, we plated only 10%–50% of the ligation mixture, yielding a true size of $> 10^5$ tagged deletion mutants. We successfully generated libraries within five days of work. Day 1, insert transposons and transform, Day 2, harvest insertion library and outgrow, Day 3,

maxiprep DNA, excise transposons, perform chewback, end-repair Day 4, dA-tail, ligate barcode, transform, Day 5: harvest library.

Generalizability: The deletion library does not depend on the presence of any restriction sites within the plasmid, only the absence of either meganuclease site: I-*Sce*I or I-*Ceu*I. These meganuclease sites are rare as their recognition sites are large (17 or 24 bp), and are not found within the human (3×10^9 bp haploid) genome, and their existence is unlikely in most viral genomes, which are much smaller ($< 10^6$ bp). If a plasmid has chloramphenicol or kanamycin resistance, the current system will not work, but a new compatible transposon cassette can be easily generated, or the molecular clone resistance marker replaced.

At present, we have only tested the system on plasmids ranging in size from 2.7 kbp to $\approx$ 16 kbp. The system may work for larger plasmids (up to $\approx$ 50 kbp), with reduced efficiency. For large viruses encoded on BACs and YACs, the method will likely not work on the full-length plasmid due to well-known inefficiencies with transformation and technical issues with manipulating large DNAs. One possible solution would be to divide the DNA into modules that could be introduced into the large plasmid using methods such as REXER [71] or homologous recombination in yeast [72].

Availability of reagents/equipment: All reagents are available "off-the-shelf from major laboratory suppliers. The method requires only basic equipment (micropipettors, thermocycler, incubator, gel box, computer) that can be found in any laboratory performing basic molecular biology. The transposon cassettes can be provided or synthesized easily.

Marginal cost: Generating a tagged deletion library uses $\approx$\$200 USD of consumable reagents to generate the tagged library and $\approx$\$500 USD of sequencing costs, and we achieved diversities of 1×10^3–5×10^4 mutants, yielding a marginal cost of <\$ 1 USD per genotyped mutant. If we had transformed more of the original library or enriched for transposon cassettes, then the marginal cost could be further reduced. In addition, we made no effort

to enrich for barcodes cassettes during plasmid library genotyping, which can greatly reduce sequencing costs, as only about 3% of reads harbored a barcode.

Manual and automated execution: In this chapter, all manipulations were performed manually, but all steps are amendable to use of liquid handling systems. Most manipulations are enzymatic, and these treatments and subsequent cleanup steps can be performed in a multi-well plate format.

Control of deletion size: By altering the duration of chewback time, the average size of deletions can be controlled, as double-end chewback proceeds at approximately 50–60 bp/min. To make larger or smaller deletions in the same time period, it may be also possible to adjust chewback temperature to increase or decrease the rate of chewback.

Iteration to generate multiply deleted strains: During chewback, this method effectively removes the inserted transposon cassette during meganuclease treatment and removes flanking meganuclease sites during chewback. Thus, the process can be repeated many times with the same transposon cassette and a different unique barcode cassette to generate multiply-deleted strains.

Barcoding of mutants: Each mutant is barcoded with a 60 bp barcode cassette, which consists of random 20 bp barcode flanked by 20 bp common forward and reverse primer binding regions. As we will show in later chapters, Illumina sequencing libraries can be easily prepared by PCR of the barcode cassettes.

Future improvements

Can random libraries be made larger or more cheaply? The one-pot method create extremely diverse, libraries *in vitro*, but transformation of *E. coli* limits the diversity of the library. By transformation of *E. coli* with a greater proportion of the library or performing many transformations in parallel, we anticipate that this could be scaled up at least 100×, so

libraries of 10^6–10^7 tagged mutants could be reached easily. If outgrowth in *E. coli* could be avoided, by the careful use of amplification methods such as rolling circle amplification, then extremely diverse libraries could be achieved. Use of other amplification methods may also reduce artifacts due to growth in *E. coli*, such as selection against potentially "toxic" or unstable sequences in viral genomes.

An additional area of improvement is during genotyping of the plasmid library. During genotyping, we made no effort we made no effort to enrich for barcode cassettes during library preparation. Thus most of the reads were wasted (less than 5% of reads contained a barcode cassette). If we selected for these reads by a variety of available methods, then we could improve efficiency of genotyping and reduce costs. Suitable methods would include enrichment of barcode cassettes by bait-capture hybridization methods or inverse PCR ([22]).

For viruses with much larger genomes, such as herpes viruses (encoded on 250 kbp BACs), a few of the steps are likely too inefficient to produce diverse libraries [73]. Transformation of bacteria with high molecular weight DNA is inefficient and large genomes are easily damaged by shearing during physical manipulations required for cleanup. One potential solution is to divide the large genomes into smaller pieces that can be manipulated more easily, mutated separately, and then reassembled using suitable methods such as REXER [71].

Chapter 3

Random deletion library screening theory

3.1 Introduction

In Chapter 2, I described a method to generate tagged libraries of deletion mutants derived from a single plasmid, a molecular clone. This chapter introduces a general theory behind making such collections and demonstrates how *cis*-acting elements and other functional features of viral genomes can be identified through controlled *in vitro* or *in vivo* passage of these deletion libraries.

3.2 Definitions

Nomenclature used in this chapter is summarized in Table 3.2 on page 72. We begin with a plasmid, p_0, which we represent as a sequence of C letters from the four letter DNA alphabet $\Omega = \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$. The plasmid, p_0, encodes a contiguous subsequence of L letters corresponding to a complete viral genome, v_0. The remaining $C - L$ letters of p_0 that

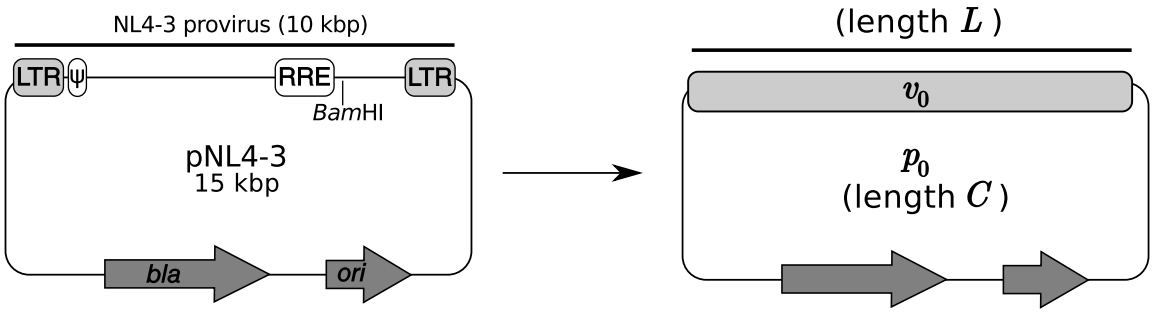


Figure 3.1: Abstraction of an infectious DNA clone plasmid as a sequence p_0 of length C, encoding a viral genome v_0 of length L.

do not encode the viral genome comprise the plasmid backbone. The backbone region encodes sequences and gene products necessary for the maintenance of the plasmid in a microbial host, typically *E. coli*. In pNL4-3, two of the sequences encoded on the backbone are the pUC19 origin of replication, *ori*, and a gene encoding for β-lactamase, which confers ampicillin/carbenicillin resistance to the bacterium harboring the plasmid. Abstraction of the pNL4-3 plasmid in this fashion is demonstrated in Figure 3.1.

When we generate a deletion library, we create a set of k plasmid deletion mutants, $\mathcal{P} = \{p_1, p_2, \ldots, p_k\}$, where each mutant in $\mathcal{P}$ is a derivative of the parental, or wildtype (WT), plasmid p_0. This is accomplished by randomly removing a contiguous subsequence from p_0 and replacing the deleted subsequence with a barcode cassette, a new fixed length sequence of length b. In general, the barcode cassette sequence is shorter than the deleted region. Thus, each mutant in $\mathcal{P}$ harbors a single deletion (relative to p_0) and is tagged at the deletion locus with a barcode cassette drawn randomly from a large set of B potential barcode cassettes, which we call $\mathcal{B}$. We stipulate that $B \gg k$, so that the preponderance of mutant plasmids in $\mathcal{P}$ are tagged with a unique barcode.

As the viral genome is a subsequence of p_0 ($L < C$), not all plasmid deletion mutants in $\mathcal{P}$ are guaranteed to harbor deletions in the parental viral genome v_0—a subset of mutants

will retain the wildtype v_0 sequence, but harbor a truncation of the plasmid backbone. If the backbone deletion does not inactivate elements required for plasmid maintenance in the microbial host, then these mutants will exist in the final library. For the molecular clones of HIV-1 and ZIKV used in this dissertation, $L/C \cong 2/3$, so the majority of mutants will contain a deletion in v_0. As $L/C \to 1$, the efficiency of introducing deletions in v_0 approaches 1, such that all plasmid deletion mutants will harbor a deletion in v_0.

In particular we are interested in $\mathcal{V} \subseteq \mathcal{P}$, the subset of deletion mutants in $\mathcal{P}$ that harbor deletions in v_0, the parental viral genome. We will refer to this subset of m mutants as $\mathcal{V}$, the set of viral deletion mutants ($\mathcal{V} = \{v_1, v_2, \ldots, v_m\}$).

3.3 The universe of possible deletion libraries

We can conveniently describe each deletion mutant as a sparse bitvector of length L, where a `1` component indicates that the base has been deleted and `0` component indicates that the base has been retained. For a hypothetical virus with a small genome ($L = 20$), the wildtype genotype vector has the representation:

$$\mathbf{m}^\intercal = \texttt{0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0},$$

and a deletion mutant with a 5 bp internal deletion can be written as

$$\mathbf{m}^\intercal = \texttt{0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0}.$$

For convenience, we do not model the b bp insertion induced by the tagging of each deletion mutant with a b bp barcode cassette.

variable	definition
p_0	sequence of WT/parental molecular clone plasmid
C	length of WT/parental virus plasmid
L	length of WT/parental viral genome (v_0) encoded on p_0
b	fixed length of each barcode cassette
$\mathcal{B}$	set of potential barcode cassettes
B	size (cardinality) of set of potential barcode cassettes ($\|\mathcal{B}\|$)
$\mathcal{P}$	set of plasmid deletion mutants ($\mathcal{P} = \{p_1, p_2, \ldots, p_k\}$)
k	size (cardinality) of set of plasmid deletion mutants ($\|\mathcal{P}\|$)
$\mathcal{V}$	set of viral mutants ($\mathcal{V} = \{v_1, v_2, \ldots, v_m\}$)
m	size (cardinality) of set of viral mutants ($\|\mathcal{V}\|$)
v_0	WT/parental virus
v_i	viral deletion mutant i
f_i	frequency of barcode i
μ	viral mutation rate
N	viral population size
$\mathbf{m_i}$	genotype bitvector of deletion mutant i (dimensions $L \times 1$)
$\mathbf{M}$	genotype binary matrix of deletion mutant library (dimensions $L \times m$)
$\mathbf{c}$	count vector of deletion mutant library (dimensions $m \times 1$)
$\mathbf{d}$	deletion depth profile ($\mathbf{d} = \mathbf{Mc}$) (dimensions $L \times 1$)
$D(L)$	number of possible contiguous subgenomic deletion mutants of length L

Table 3.1: Variable definitions

The set of subgenomic mutants

As discussed in Chapter 2, one universe of possible DIPs is the set of viral sequences that are strictly deletion mutants of the wildtype virus, the set of subgenomic mutants of v_0. How large is the set of subgenomic mutants of v_0?

Using our bitvector notation, the size of the set of all possible bitvectors of length L is simply 2^L. Each of the L bases of v_0 can represented by either a `0` or a `1`, so there are 2^L potential combinations.

In this set of 2^L bitvectors of length L, one of the members is the wildtype virus, which

is a 0-vector of size L (no deletions). The remaining $2^L - 1$ bitvectors represent subgenomic mutants, as they have a deletion at a least one base (one or more 1s). One of these $2^L - 1$ subgenomic mutants is fully deleted, it is a 1-vector of length L, and has no functionality, as it does encode for any letters.

Thus, for a viral genome of length L, there are a maximum of $2^L - 1$ possible subgenomic deletion mutants[1]. For HIV-1 and ZIKV, $L \approxeq 10^4$, so there are a maximum of $(2^{10000} - 1) \approxeq 10^{301}$ possible deletion mutants.

Subgenomic mutants by single contiguous deletions

Suppose we restrict the fashion in which subgenomic mutants are created to reflect the class of subgenomic mutants created in the random deletion libraries specified in Chapter 2. Now, we only permit generation of subgenomic mutants which harbor a single contiguous deletion with respect to the parental/WT sequence. What is the size of this constrained set of mutants?

Let $D(L)$ be the number of subgenomic mutants of a sequence of length L, where a contiguous subsequence of length 1 to $L - 1$ are removed. If $L = 1$, then no mutants are possible, this is the base case. If $L = 2$, then $D(L) = 2$, as either the first base or the second base will be deleted. If $L = 3$, then we have 3 possible one-bp deletions, and 2 possible

[1]The maximum of $2^L - 1$ subgenomic is an upper bound on the number of unique subgenomic mutants. Subgenomic mutants that are represented by unique bitvectors may yield the same primary acid sequence, as the DNA/RNA alphabet is composed of only 4 letters. As an example, suppose there is a virus with $v_0 =$ AATTC. Then subgenomic mutants 10100 and 01010 are degenerate as both yield the same primary sequence: ATC

two-bp deletions, so $D(3) = 5$. Thus in general, we have

$$\begin{aligned} D(L) = \sum_{i=1}^{L-1} i &= 1 + 2 + \cdots + (L-1) \\ &= \frac{(L)(L-1)}{2} = \frac{L^2 - L}{2} \\ &\approxeq L^2/2. \end{aligned} \tag{3.1}$$

Which for large L is about $L^2/2$ with $1/L$ relative error. For HIV-1 and ZIKV, $L \approxeq 10^4$ so $N = 5 \cdot 10^7$. Thus having libraries of 10^5 mutants, we can sample 0.2% of the potential subgenomic mutants generated by single contiguous deletion of parental HIV-1 and ZIKV genomes.

3.4 Determining the length of barcode cassettes

In Chapter 2, we proposed that tagging each deletion mutant with a barcode cassette (a molecular barcode) would aid in the screening of large diverse libraries of random deletion mutants by greatly reducing the amount of sequencing required in downstream experiments.

After an initial genotyping of the plasmid library by whole plasmid sequencing, which allows construction of a mapping between barcode cassettes to deletion loci, no further whole genome sequencing would be required. Instead, sequencing libraries would be prepared by PCR of the barcode cassette sequences by their common primer binding sites and the genotype of each mutant inferred by consulting the previously computed mapping between barcode cassette and deletion locus.

The utility of this method is dependent on whether molecular tags are unique. During library construction, each mutant should be tagged with a unique barcode cassette with a probability near 1, so that the barcode cassette serves as a unique molecular tag during

genotyping. During passage of the libraries, the molecular barcodes must be of sufficient length so that they can acquire mutations and still be distinguished from similar tags (which may also acquire mutations).

We now ask investigate the unique tagging property. Given a library of m mutants tagged with barcodes drawn with equal probability from a set of size B, what proportion of the m mutants are uniquely tagged?

The viral mutant library is a set of m viral mutants obtained by randomly deleting a portion of the viral genome and then genetically tagging the deletion locus with a barcode cassette. The barcodes cassettes are 60-bp dsDNA oligonucleotides composed of a string of 20 random bases, $(\mathtt{N})_{20}$, flanked by common 20-bp primer binding sites. We assume that each barcode is drawn with equal probability from a pool of B random barcode cassettes. To maximize the proportion of mutants tagged with unique barcodes, we stipulate that $B \gg m$, such that most drawn barcodes are drawn only once.

We model the tagging of m mutants from a pool of B barcodes as drawing m random integers from a discrete uniform distribution with range $[1, B]$. Here, each integer from 1 to B represents a unique barcode and is drawn with equal probability $1/B$.

Let $P(A)$ be the probability that at least two of the m mutants are tagged with the same barcode. The probability of the complementary event, that all barcodes are unique, is then $P(A') = 1 - P(A)$. $P(A')$ can be calculated as follows. If $m > B$, then $P(A') = 0$ by the pigeonhole principle, as at least two mutants will share the same barcode.

What is $P(A')$ for $m \leq B$? The probability of a drawing a unique barcode on the first draw is $(B/B = 1)$, as there are B available barcodes of the B possible barcodes. The probability of drawing a unique barcode on the second draw that does not match the first barcode is $(B-1)/B$, as only $B-1$ of the B barcodes remain. Following this pattern, the probability on the third draw will be $(B-2)/B$, the fourth draw $(B-3)/B$ and the m-th

draw $(B-m+1)/B$. These events are independent so $P(A')$ is simply their product:

$$\begin{aligned} P(A') &= \underbrace{\frac{B}{B} \times \frac{(B-1)}{B} \times \frac{(B-2)}{B} \times \cdots \times \frac{(B-m+1)}{B}}_{m \text{ factors}} \\ &= \frac{(B)(B-1)(B-2)\cdots(B-m+1)}{B^m} \\ &= \frac{B!}{B^m(B-m)!} = \frac{{}_BP_m}{B^m}, \quad \text{for } 0 < m \leq B. \end{aligned} \tag{3.2}$$

where ${}_BP_m = B!/(B-m)!$, the number of permutations of m elements drawn from B elements.

In general, if a set of size B is sampled m times, then the expected number of unique samples, $u(m;B)$, is exactly:

$$u(m;B) = m\left(1 - \frac{1}{B}\right)^{m-1}. \tag{3.3}$$

The probability that each sample or drawing a non-unique sample is $1/B$, so the probability of drawing a unique sample is the complement, just $(1-1/B)$. For $B \gg 1$ we can use the Taylor series approximation of $(1-x)^n \approxeq 1-nx$ for $0 < x \ll 1$, which gives:

$$u(m;B) = m\left(1 - \frac{m-1}{B}\right). \tag{3.4}$$

For $m \gg 1$, a good approximation (with error of $\approx m/B$) is:

$$u(m;B) \approxeq m\left(1 - \frac{m}{B}\right) = m - \frac{m^2}{B}. \tag{3.5}$$

Thus, the proportion of non-unique barcodes, p_n is simply $p_n = 1 - u(m,B)/m$. I've tabulated p_n for library sizes of 10^5, 10^6, and 10^7, with barcode pool sizes of 10–20 random bases ($4^{10} \leq B \leq 4^{20}$) below.

m	B	p_n
10^5	4^{10}	$9.1 \cdot 10^{-2}$
10^5	4^{12}	$5.9 \cdot 10^{-3}$
10^5	4^{14}	$3.7 \cdot 10^{-4}$
10^5	4^{16}	$2.3 \cdot 10^{-5}$
10^5	4^{18}	$1.5 \cdot 10^{-6}$
10^5	4^{20}	$9.1 \cdot 10^{-8}$
10^6	4^{10}	$6.1 \cdot 10^{-1}$
10^6	4^{12}	$5.8 \cdot 10^{-2}$
10^6	4^{14}	$3.7 \cdot 10^{-3}$
10^6	4^{16}	$2.3 \cdot 10^{-4}$
10^6	4^{18}	$1.5 \cdot 10^{-5}$
10^6	4^{20}	$9.1 \cdot 10^{-7}$
10^7	4^{10}	≈ 1.0
10^7	4^{12}	$4.5 \cdot 10^{-1}$
10^7	4^{14}	$3.7 \cdot 10^{-2}$
10^7	4^{16}	$2.3 \cdot 10^{-3}$
10^7	4^{18}	$1.5 \cdot 10^{-4}$
10^7	4^{20}	$9.1 \cdot 10^{-6}$

For one million mutants (10^6) tagged with random 20 bp barcodes, roughly 15 are expected to have non-identical barcodes ($15 = (10^6)(1.5 \cdot 10^{-5})$). After acquiring up to 6 point mutations ($20-6 = 14$ bp unique), only about 1 in 300 ($p_n = 3.7 \cdot 10^{-3}$) mutants are expected to share a barcode.

3.5 Frequency of background mutations

Construction of random deletion libraries requires amplification of DNA, either by culturing bacteria harboring a plasmid molecular clone or by enzymatic amplification of template DNA (PCR or ϕ29 rolling circle amplification). As each amplification process has a known error-rate, we expect that off-target mutations will occur (those that are outside the deliberately introduced deletions loci). Thus, some barcoded deletion mutants will also harbor secondary mutations acquired during plasmid amplification that we will not account for during con-

struction of the barcode to deletion locus mapping. How common are off-target mutations introduced during amplification in bacteria?

We begin with a single parental/wildtype viral species, v_0, whose DNA genome is of length L. The v_0 DNA genome can be propagated in bacteria on a circular piece of dsDNA termed a plasmid. Preparations of plasmid DNA at these stage have low diversity, essentially the background mutation rate in the host bacterial strain.

To show this, we first consider that during a DNA prep, approximately x copies of the plasmid are replicated from a single template with a mutation rate μ_b. The expected number of errors due to this amplification is:

$$\text{errors} = \mu_b x L.$$

For our example, $\mu_b = 10^{-8}$, $x = 10^{12}$, and $L = 10^4$, so we expect about 10^8 errors in the total mass of DNA. For the deletion library, we tag approximately 10^5 of these sequences, so the expected number of errors in this pool is just

$$\text{errors} = \frac{10^5}{10^{12}} 10^8 = 10. \tag{3.6}$$

Roughly 1 in 10,000 deletion mutants will have an off-target point mutation.

3.6 Identification of *cis*-acting elements

We describe each deletion mutant as a sparse bitvector of length C, where `1`'s indicate that the base has been deleted and `0`'s indicate that the base has been retained. For a short virus,

where $C = 20$, the wildtype vector has the representation:

$$\mathbf{m}^\intercal = 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0,$$

and a mutant with a 5 bp internal deletion has the representation

$$\mathbf{m}^\intercal = 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0.$$

To calculate the deletion depth at each timepoint, we simply add the mutant vectors:

$\mathbf{m_1}^\intercal$	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0
$\mathbf{m_2}^\intercal$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
$\mathbf{m_3}^\intercal$	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
$\mathbf{m_4}^\intercal$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
$\mathbf{d}^\intercal$	0	0	0	0	0	1	1	1	1	2	3	2	2	2	2	1	0	1	1	1

In general, deep sequencing the barcoded libraries, allows computation of a count c_i for each mutant i in M:

$$\mathbf{d} = \sum_{i=1}^{M} c_i \mathbf{m}_i$$

This could also be represented as a multiplication between a count vector, $\mathbf{c}$, with dimensions $M \times 1$ and a mutant matrix, $\mathbf{M}$, a $(0, 1)$-matrix of dimensions $L \times M$:

$$\mathbf{d} = \mathbf{M}\mathbf{c}. \tag{3.7}$$

Example: mapping the *cis*-acting elements of HIV-1

We now show how to interpret data obtained by passage of random deletion libraries to identify *cis*-acting elements of viruses using Equation 3.7.

For a virus, we use a toy-model of HIV-1. Assume that we can create a random deletion libraries of HIV-1, in which the deletion sizes of tagged deletion mutants are normally

distributed with mean λ and standard deviation σ ($\mathcal{D} \sim \mathcal{N}(\lambda, \sigma)$).

We first create a random deletion library by combining four sublibraries of 15,000 random deletion mutants:

$$\mathcal{D} \sim (\mathcal{N}_1(\lambda_1, \sigma_1) + \mathcal{N}_2(\lambda_2, \sigma_2) + \mathcal{N}_3(\lambda_3, \sigma_3) + \mathcal{N}_4(\lambda_4, \sigma_4)), \tag{3.8}$$

where $(\lambda_1, \sigma_1) = (250, 50)$, $(\lambda_2, \sigma_2) = (500, 200)$, $(\lambda_3, \sigma_3) = (750, 200)$, and $(\lambda_4, \sigma_4) = (1000, 300)$. These values are first-order estimates of expected chewback sizes after 5,10,15,20 minutes of exonuclease chewback as specified in Chapter 2.

Thus, the library consists of 60,000 deletion mutants. In this example, deletions are made in a plasmid molecular clone of 15 kbp ($C = 1.5 \cdot 10^4$) and the length of the provirus is exactly 10 kbp ($L = 10^4$). The starting point of each deletion is drawn from a uniform distribution of the plasmid genome. Representative deletions are shown below in Figure 3.2. To simulate barcode sequencing, we draw $6 \cdot 10^5$ samples (with replacement) from the deletion library (10× oversampling). During plasmid sequencing, we stipulate that the probability of drawing each deletion mutant is identical. Sample deletion depth profiles of the plasmid were calculated using equation 3.7 and are displayed as red curves in Figure 3.3 and Figure 3.4.

Now we impose some selection on the library to simulate differences in replication. We designate several regions of the virus to be *cis*-acting elements: CAE1 from 1–1190, CAE2 from 4781–4898, CAE3 from 7710–8861, CAE4 from 9064–9083, and CAE5 from 9612–9617. These intervals correspond roughly to the 5′LTR–Ψ, cPPT/CTS, RRE, PPT, and 3′ LTR, respectively. Mutants which harbor deletions of *cis*-acting elements are sampled with probability of 0.1× the probability of mutants in which all *cis*-acting elements are intact. This is the equivalent to a 10× decrease in replicative fitness during a single round of replication.

In Figure 3.3, we set the *cis*-acting elements to be CAE1, CAE3, and CAE5. In Figure 3.4,

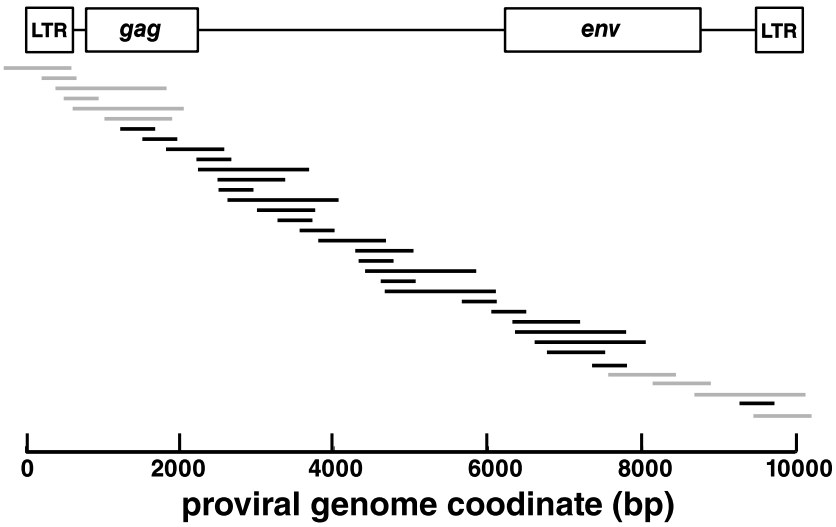


Figure 3.2: Example deletions mapped to the HIV-1 provirus. The extent of each deletion mutant is designated by a horizontal bar, and mutants sorted by increasing leftmost genome coordinate. Deletions in gray overlap *cis*-acting elements. Deletions in black do not overlap *cis*-acting elements.

we set the *cis*-acting elements to be CAE1–CAE5. Deletion depth profiles for these libraries (after a single round of replication) are shown as blue curves in Figure 3.3 and Figure 3.4.

In both simulated experiments, we are able to recover the approximate location of *cis*-acting elements by comparing the deletion depth profiles of the plasmid library (red curves) to the deletion depth profiles of the library after 1 round of replication (blue curves). The *cis*-acting elements can be located by noting the location of the minima of the deletion depth profiles.

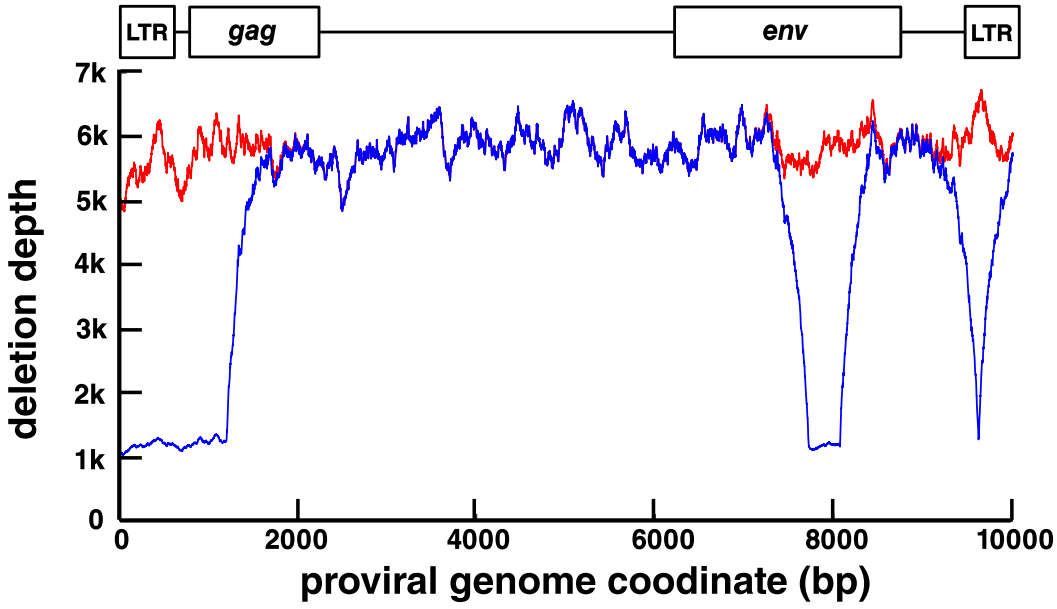


Figure 3.3: Simulated *cis*-acting element identification, 3 CAE. Deletion depth (vertical axis) for a simulated deletion library is plotted against genome coordinate (horizontal axis). Deletion depth profiles for a plasmid library (red) and a library after 1 round of selection (blue) are displayed. Viruses harboring deletions that intersect *cis*-acting elements incur a 10-fold decrease in replicative fitness. The specified *cis*-acting elements are CAE1 from 1–1190, CAE3 from 7710–8861, and CAE5 from 9612–9617.

3.7 Sequencing analysis of deletion libraries

Every barcoded deletion mutant is tagged with one or more random 60-bp DNA barcode cassettes:

$$5'\text{–}\texttt{CCGTCCATGAAGGGTTCGATNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT}\text{–}3'$$

where `N` is `A`,`C`,`G`, or `T`. Each barcode cassette is composed of a unique 20-bp random barcode, B, flanked by 20-bp left (L) and right (R) common sequences that are shared by all barcodes. Thus, each barcode cassette has the structure LBR. As barcode cassettes are inserted into each plasmid, we must also consider the context of the barcode flanking

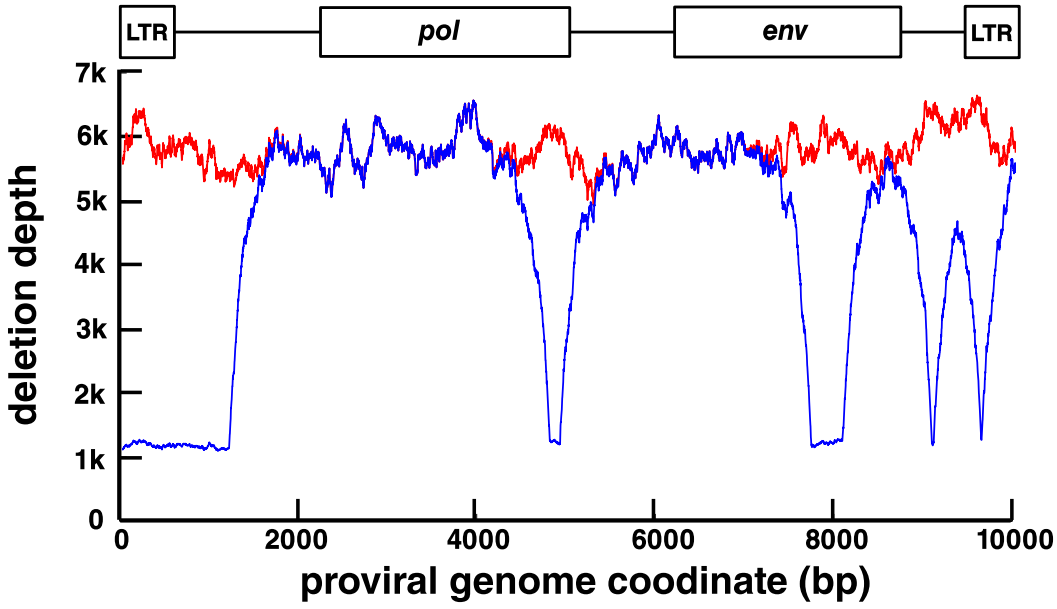


Figure 3.4: Simulated *cis*-acting element identification, 5 CAE. Deletion depth (vertical axis) for a simulated deletion library is plotted against genome coordinate (horizontal axis). Deletion depth profiles for a plasmid library (red) and a library after 1 round of selection (blue) are displayed. Viruses harboring deletions that intersect *cis*-acting elements incur a 10-fold decrease in replicative fitness. The specified *cis*-acting elements are CAE1 from 1–1190, CAE2 from 4781–4898, CAE3 from 7710–8861, CAE4 from 9064–9083, and CAE5 from 9612–9617.

sequences (plasmid sequences), which we represent as $\mathsf{F_L}$ and $\mathsf{F_R}$, where $\mathsf{F_L}$ are the bases to the left of L ($5'$ of L) and $\mathsf{F_R}$ are the bases to the right of R ($3'$ of R).

In unenriched populations, reads containing a complete barcode cassette sequence (LBR) are rare as the length of LBR ($b = 60$) is much less than the length of the host plasmid ($C > 4000$). The frequency of sequencing a complete barcode cassette is $O(b/C)$, where b is the length of the barcode cassette and C is the size of the wildtype/parental plasmid. For the pNL4-3 plasmid, this ratio is 60/15000 or 0.4%.

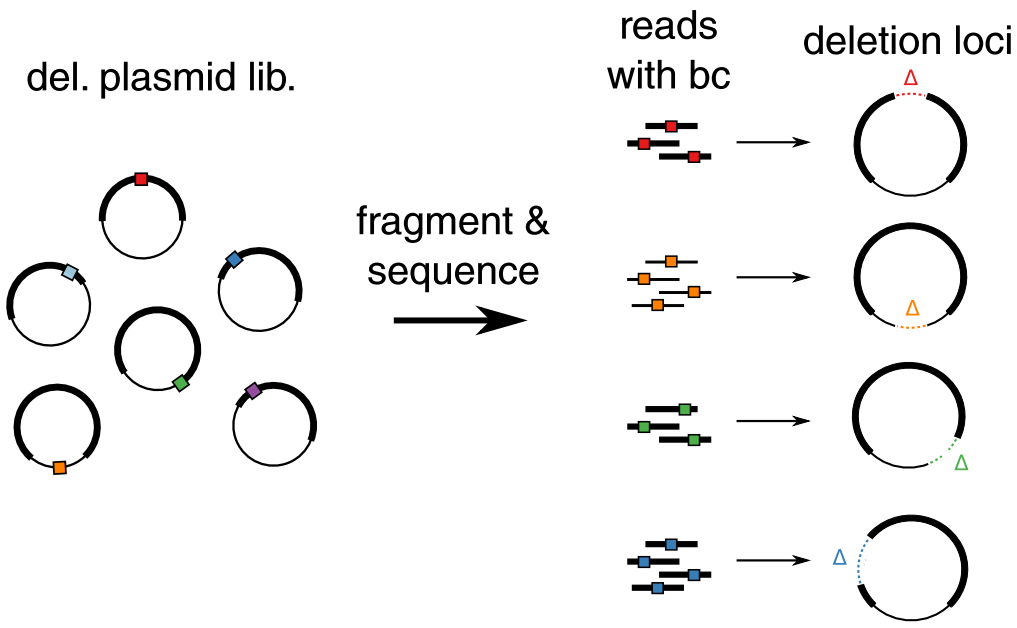


Figure 3.5: Genotyping of plasmid random deletion libraries. A pool of deletion mutant plasmids are fragmented and sequenced on a next-generation sequencing platform to generate a series of reads. Most reads do not harbor barcode cassettes but a small fraction do. Barcode cassettes are grouped by common barcode sequences than mapped to the reference genome to identify the correction deletion locus with high probability. Details in text below.

Mapping deletion loci to barcode cassettes

After creation of a plasmid random deletion library, whole genome sequencing of the plasmid library is used to construct a mapping between barcode cassettes and deletion loci (Figure 3.5). This initial genotyping step allows for efficient sequencing in downstream procedures, as only barcode cassettes need to to be sequenced and the barcode-to-deletion mapping consulted to identify which deletion mutants persist throughout passage. If barcode cassettes were not used, whole genome sequencing would be required to identify persistent deletion mutants.

Barcodes were mapped to deletion loci using paired-end sequencing data obtain from Illumina NGS (HiSeq4000, 2×125 b reads). Custom `python` software (`rdl-seq`) performed

the following procedure:

1. Filter all high-quality reads for those containing L or R or the reverse complement of L or the reverse complement of R (inclusive or).

2. Create a hash table/dictionary `flanks`. For filtered read from above, search for substrings matching the complete pattern LBR. For each match, create 2-tuple of (F_L,F_R). Retrieve `flanks[B]` array. Append 2-tuple to array and assign appended array to `flanks[B]`.

3. Create hash table/dictionary `consensus`. For each barcode index in `flanks`, build consensus F_L and F_R sequences using all values in array indexed by B. Store new consensus values under `consensus[B]`.

4. Create hash table/dictionary `dloci`. For each barcode index in `consensus`, align consensus F_L and F_R sequences to reference genome by Boyer-Moore-Horspool algorithm (default string matching in `python3.4`) and assign deletion locus. Store deletion locus as ordered pair in `dloci[B]`.

During downstream processing of barcodes the `dloci` hash table is consulted to map barcode to deletion loci.

Chapter 4

Screen for HIV-1 therapeutic interfering particles

4.1 Introduction

Human immunodeficiency virus type 1 is a lentivirus of family *Retroviridae*. Due to at least three introductions from chimpanzees (*Pan troglodytes*) since the 1920s ([74, 75]), more than 60 million people have been infected between 1920 and present, leading to over 30 million deaths [20]. HIV-1 productively infects a variety of human tissue types, predominantly $CD4^+$ T cells, macrophages, and dendritic cells [20]. HIV-1 spreads through tissues via cell-to-cell transfer of virions through a structure known as virological synapse but also through cell-free transfer [76, 77].

The preponderance of new infections of human hosts occur during sex, as semen, pre-ejaculate, vaginal fluid, and blood all carry significant viremia [20]. More minor routes of transmission include perinatal transmission, breastfeeding, blood transfusions, and reuse of needles during intravenous drug use [20].

HIV-1 infection is chronic, and if untreated, lead to progressive failure of the immune system over a period of 6 months–15 years, typically leading to death from opportunistic infections unchecked by a failing immune system [20].

The ≈ 9 kb (+)-sense ssRNA genome encodes nine genes: 3 genes encoding structural proteins that form the virion (*gag*, *pol*, *env*) and 6 genes encoding non-structural proteins (*vif*, *vpr*, *vpu*, *tat*, *rev*, *nef*) that are required for regulation of HIV-1 gene expression and for modulation of host innate and adaptive immunity [78–81]. A genome diagram of prototypical HIV-1 strain is shown in Figure 4.1 on page 89. Most HIV-1 proteins can be complemented *in trans* (Vif [82],Nef [80],Vpr, Rev, Tat [25], Env [83]), which has led to the development of lentiviral vectors that are subgenomic mutants of HIV-1 [83–85]. In lentiviral vector systems, the *trans*-acting elements that have been deleted in the viral vector are provided *in trans* by co-transfection of packaging plasmids into a permissive cell line, such as 293T.

Upon infection of a susceptible cell, the HIV-1 genome is converted into dsDNA form by HIV-1 reverse transcriptase and integrated into the host genome by HIV-1 integrase (both domains are encoded in *pol*) as an ≈ 10 kbp proviral genome. Under certain conditions, where levels of host transcription factors (NF-$\kappa\beta$) and a viral transactivator (Tat) are low, viral genes are not expressed, and the virus is referred to be in *latent*. Under activating conditions, where sufficient transcription factors and Tat are available, the HIV-1 genome is transcribed as a singe RNA from a common promoter in the 5′ LTR, then alternatively spliced into as many as 109 unique transcripts [86]. Only the full-length, unspliced genome is competent for encapsidation, and requires export from the nucleus by binding of a *cis*-acting element, the Rev Responsive Element (RRE), to Rev, which leads to export of the unspliced genome via the Crm1 pathway [20].

A number of *cis*-acting elements have been putatively identified in the literature and are reported in Table 4.1. Regions at the 5′ UTR of the virus, including stem loops 1–4 are crucial

for binding of the unspliced viral genome to the viral nucleocapsid protein. Other genomic regions such as PBS, cPPT/CTS, cPPT, are necessary for efficient reverse transcription and integration.

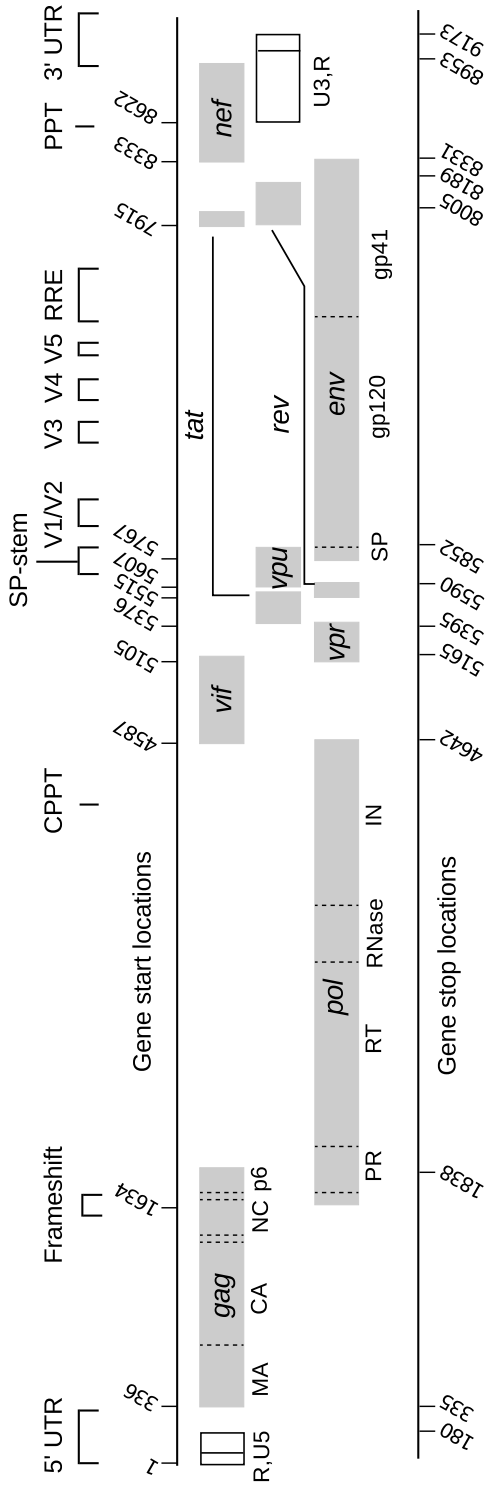


Figure 4.1: The RNA genome of HIV-1. The nine protein coding genes of HIV-1 are depicted as gray rectangles, with protein domains annotated underneath each rectangle and protease cleavage sites shown as vertical dashed lines. Two of the genes are bi-exonic (*tat* and *rev*) while the remainder are mono-exonic. The LTR at the genome termini are shown as unfilled rectangles. Regions of known secondary structure and potential *cis*-acting elements are annotated above gene start locations and include: the 5′ untranslated region *gag-pol* ribosomal frameshift, central polypurine tract/central termination sequence, Env signal-peptide stem, V1, V2, V3, V4, V5 loops of gp120; the Rev Response Element; the polypurine tract; and the 3′ UTR. Adapted from Figure 1a/1b of [87].

	HXB2CG provirus[a]				
element	start	stop	length	description	ref
U3	1	455	456	unique 3′ region	
R	456	551	96	repeat region	
U5	552	634	83	unique 5′ region	
Lys-tRNA	636	653	18	Lys-3 tRNA PBS	
ISRE	655	673	19	Interferon-Stimulated Response Element	
pkg?	674	690	17	pkg signal RNA secondary structure	
SL1	691	734	44	dimerization initiation site, DIS	[88–93]
SL2	736	754	19	major splice donor, MSD	[88, 91–93]
SL3	766	779	14	NC-binding domain, Ψ	[88, 90–93]
SL4	790	810	21	unknown	[88, 91–93]
ARP	790	5096	4307	A-rich *pol* sequences (reverse transcription)	[94]
RFS	2086	2091	6	Ribosomal slip site	
GRPE[b]	2022	2188	167	Gag-Pol ribosomal packaging element	[95]
cPPT/CTS	4781	4898	118	central polypurine tract / central termination sequence	[96, 97]
RRE	7710	8061	353	Rev Responsive Element	
PPT	9069	9093	25	polypurine tract	
U3	9086	9540	456	unique 3′ region	
R	9541	9636	96	repeat region	
U5	9637	9719	83	unique 5′ region	

[a]HXB2CG reference sequence can be obtained from NCBI GenBank (`K03455.1`)
[b]The role of GRPE in encapsidation of HIV gRNA is controversial [98].

Table 4.1: Putative *cis*-acting elements of HIV-1 in the literature

Conditionally replicating HIV-1 vectors

Although deletion mutants are known to spontaneously arise during HIV-1 infections, both *in vivo* and *in vitro* (see [99]), we are unaware of any any naturally occurring defective interfering particles for HIV-1 that arise during *in vivo* infections. During *in vitro* studies, there are reports of deletion mutants that emerge during HIV-1 infections to interfere with HIV-1 replication, but without evidence of transmission between cells [100].

After the development of lentiviral vectors, a number of groups developed therapeutics derived from subgenomic HIV-1 mutants (often lentiviral vectors) that were conditionally replicating HIV-1 strains (examples in Table 4.2. These mutants lack at least one essential *trans*-acting element, and were reported to interfere with wildtype virus replication during *in vitro* studies. Most encoded a payload designed to inhibit replication of wildtype HIV-1. Sample payloads included ribozyme sequences designed to cleave transcripts of wildtype helper virus, shRNA hairpins to downregulate wildtype expression, dominant negative capsid proteins, and antisense *env* transcripts (see Table 4.2). These viruses can be distinguished from TIPs, as they cannot be transmitted efficiently ($R_0 < 1$), thus their influence is limited to the lineage of transduced cells. Unless introduced and maintained in the entire population of target cells (above the "herd immunity threshold), they have limited therapeutic benefit. In short, these therapies achieve disease dependence but "poison the well" of common pool goods, effectively starving the supply of common goods required for mobilization of both helper virus and DIP.

group	year	name	payload	reference
I. Chen	1999	DAEGFP	EGFP	[101]
*	*	DAt1ruEGFP	EGFP	*
*	*	HR′CMVEGFP	PcmvEGFP	*
*	*	SRαLEGFP	PcmvEGFP	*
B. Dropulić	1999	crHIV-1.1	1 × ribozyme	[102]
*	*	crHIV-1.11	2 × ribozymes	*
*	*	crHIV-1.12	2 × ribozymes	*
*	*	crHIV-1.111	3 × ribozymes	*
L. Naldini	1999	HIV vector	none	[103]
J. Rosenblatt	2001	DHIV	none	[104]
*	*	DHIV-RRE$^-$	none	*
K. Morris	2005	HIV2	none	[105]
C. June	2006	VRX496	antisense-env	[38]
L. Weinberger	2011	SC	none	[106]
*	*	SC1216	mCherry	*
*	*	Ld2mCh	d2mCherry	*
*	2013	TIP-SR2	EF1A-mTagBFP2	[107]
*	*	TIP-SR2-D1	EF1A-mTagBFP2	*
*	*	TIP-SR2-ΔEF1A	mTagBFP2	*
*	*	TIP-SR2-ΔEF1A-ΔmTagBFP2	none	*

Table 4.2: Examples of conditionally-replicating HIV

Mapping of *cis*- and *trans*-acting elements

HIV-1 mutagenesis studies are common, point mutations have been scanned extensively [108], stop codons introduced into all points in the genome except for tat [109], and recombinant deletion mutants have also been generated and studied [110, 111].

In this chapter, we extend these mutagenesis studies using the pNL4-3 deletion library prepared in Chapter 2 to screen the HIV-1 genome for *cis*- and *trans*-acting elements, and identify deletion mutants that are not replication-competent, but can be mobilized by the wildtype virus. Thus from first principles, we can identify *cis*- and *trans*-acting elements of viruses and use these data as a base of information to build transmissible antivirals.

4.2 Materials and Methods

Tissue culture

293T (synonyms: HEK 239T, 293tsA1609neo) were obtained from the American Type Culture Collection (ATCC, # CRL-3216). Adherent 293T were propagated in **D10** media, comprised of Dulbecco's Modified Eagle's Medium (DMEM; Corning, #10-013-CV) supplemented to a final concentration of: 10% (v/v) heat-inactivated fetal bovine serum (FBS) (Corning, #35-011-CV), 100 IU/ml penicillin & 100 μg/ml streptomycin (Corning, #35-002-CI), 25 mM HEPES (Thermo Fisher, #15630080), and 2 mM L-alanyl-L-glutamine (Corning, #25-015-CI). The cell line was cultivated in 15-cm polystyrene dishes in a humidified incubator at 37 °C with 5% CO_2. Subcultures were prepared by removing media, washing once with DPBS-CMF, treating with 0.25% Trypsin/2.21 mM EDTA in HBSS for 2–5 min at 37 °C, then quenched by adding an equal volume of **D10**.

MT-4 [112] were obtained through the NIH AIDS Reagent Program, (Catalog #120) Division of AIDS, NIAID, NIH as a generous gift from Dr. Douglas Richman. Cells were

propagated in **R10** media, comprised of Roswell Park Memorial Institute (RPMI-1640; Corning, #10-040-CV) supplemented to a final concentration of: 10% (v/v) heat-inactivated fetal bovine serum (FBS) (Corning, #35-011-CV), 100 IU/ml penicillin & 100 μg/ml streptomycin (Corning, #35-002-CI), 10 mM HEPES (Thermo Fisher, #15630080), and 2 mM L-alanyl-L-glutamine (Corning, #25-015-CI). The cell line was cultivated in polystyrene T flasks in a humidified incubator at 37 °C with 5% CO_2. Once established, cultures were maintained a density of $2 \cdot 10^5$–$2 \cdot 10^6$ cells per ml . Subcultures were prepared by diluting cultures with fresh **R10** media to about $5 \cdot 10^5$ cells/ml every 2–3 days.

Cell cultures were not tested for mycoplasma or verified for authenticity. Lines were maintained for less than 30 days in continuous culture before a new seed culture was used (maximum of 30 divisions / 10 passages).

Production of viral stocks by transfection

Virus pool was obtained by co-transfection of 293T with pNL4-3 (WT virus) and the pNL4-3 deletion library. On the day of transfection, a suspension of 293T was obtained by trypsinization of subconfluent 15 cm plates of 293T and brought into single cell suspension by gentle passage through a 40 μm mesh filter (Corning). A cell count was obtained with an automated Coulter cell counter (Moxi, ORFLO), and cells were diluted to a concentration of $5 \cdot 10^5$ cells/ml in **D10**. Thirty-six ml of this suspension ($1.8 \cdot 10^7$ cells) were added to $3 \times$ T175 flasks. Next, 18 μg of pNL4-3, 18 μg of pNL4-3Δ_1., and 108 μl of 1 μg/μl polyethyleneimine (Polysciences, PEI) were added to serum-free DMEM supplemented with 25 mM HEPES and the volume brought to 3.6 ml with extra serum-free DMEM with 25 mM HEPES, incubated at bench temperature (24 °C) for 15 min, then added to the upright flasks. The upright flasks were gently rocked, then lowered to a horizontal position in a 37 °C/5% CO_2 incubator. Media was replaced after an overnight incubation (16–20 h), and virus was harvested at 40-48

hours post-transfection by passing through 0.45 μm sterile filters (Millipore). Virus stocks were concentrated by ultracentrifugation. Twenty-five ml of clarified supernatant were underlaid with a 6% (m/v) iodixanol in DPBS-CMF solution in SW28 ultracentrifuge tubes, then adjusted to a 38.6 ml final volume with additional clarified supernatant. The tubes were spun for 90 min at 20000 rpm in an SW28 rotor at 4 °C. Supernatant was decanted and the invisible viral pellets resuspended in pure heat-inactivated FBS and store frozen in single-use aliquots at −80 °C.

Titration of viral stocks

HIV-1 stocks were titrated by infecting cultures of MT-4 in 6-well plates with viral inocula and scoring for HIV p24-producing cells at 24 hours post-infection. Briefly, 100 μl of inoculum was added to 1.0 ml of **R10** containing $2 \cdot 10^6$ MT-4, mixed briefly, then incubated for 4 hours at 37 °C. After four hours, an additional 1.0 ml of **R10** was added and the infection allowed to proceed for an additional 20 hours (a single-round of replication). At 24 hours post-infection, cultures were fixed by adding 0.1 volumes of 20% formaldehyde (tousimis) (final concentration 2.0%) and incubated for at least 1 hour at 4 °C.

Formaldehyde-fixed cells were permeabilized by treatment with 75% ice-cold methanol for 10 minutes, then stained with a phycoerythrin-labelled monoclonal antibody (KC57-RD1,BD) in DPBS-CMF + 2% FBS + 2 mM EDTA + 0.1% IGEPAL-CA630 for 30 min before washing once in stain buffer. At least 50,000 live cells were counted by flow cytometry on a FACS Calibur DxP8. Gates were drawn based upon stained naive cell population.

High-MOI passage scheme

On day 0 (0 hpi), $2 \cdot 10^6$ MT-4 were infected at an MOI of 5-20 with the virus pool containing WT HIV-1 and tagged HIV-1 deletion mutants for 4 hours in a volume of 2 ml, then trans-

ferred to a T25 flask containing 10 ml of MT-4 at a concentration of 10^6 cells/ml. On day 2 (40 hpi), the 12 ml of culture was transferred to a T175 flask containing 60 ml of MT-4 in R10 at a concentration of 10^6 cells/ml. On day 3 (70-72 hpi), supernatant from the MT-4 was clarified by centrifugation and 0.45 μm filtration, then concentrated by ultracentrifugation as described above. This cycle corresponds to 3 rounds of HIV-1 replication (completed on day 1, day 2, day 3). The cycle was repeated a total of four times (12 passages / rounds of replication) to select for deletion mutants that could be efficiently mobilized by HIV-1 in high MOI passage and retained all necessary *cis*-acting elements. The passage scheme was conducted with 3 biological replicates.

Preparation of sequencing libraries

Viral RNA was isolated from frozen aliquots of the concentrated virus pool at various time points (passage 0, passage 3, passage 6, passage 9, passage 12) using a QIAmp Viral RNA Mini Kit (Qiagen) per the manufacturer's instructions with two exceptions: 1) carrier RNA was replaced with 5 μg of linear polyacrylamide (Sigma) per isolation 2) $5 \cdot 10^6$ copies of bacteriophage MS2 RNA (Roche) were spiked in per isolation.

Total cellular RNA from 293T was isolated using Trizol (Life Technologies) from cell pellets obtained at the time of viral harvest. A poly(A) fraction, representing mRNA, was isolated by annealing total RNA to magnetic $d(T)_{25}$ beads to pull down polyadenylated transcripts using a commercial kit (NEBNext Poly(A) mRNA magnetic isolation module).

Purified RNA was reverse-transcribed with Superscript III (Thermo Fisher) and Random Primer Mix (New England Biolabs). cDNA was used as template in real-time qPCR to quantitate barcode cassette concentrations with oligonucleotides oBC20v1-F (`CCGTCCATGAAGGGTTCGAT`) and oBC20v1-R (`ACGAATCTGCCGTTGCCATA`) and compared to a standard curve prepared with dilutions of a barcode standard, oBC20v1-T:

CCGTCCATGAAGGGTTCGATNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT,

in 10 μl reactions with Fast SYBR Green Master Mix (Thermo Fisher). Levels of total HIV RNA were estimated by levels of HIV *pol* with oligos oNL43pol-F (GAGACAGGGCAAGAAACAGC) and oNL43pol-R (AACAGGCGGCCTTAACTGTA). Samples were normalized for recovery by determining levels of MS2 RNA recovered by oligos oMS2-F (TCCTGCTCAACTTCCTGTCGAG) and oMS2-R (CAGGTCAAACCTCCTAGGAATG) (sequences from [113]). Samples were not DNaseI-treated before reverse transcription, but levels of background DNA were acceptably low (-RT controls had barcode levels of <1/1000x of +RT reactions).

Illumina sequencing libraries were prepared by a modification of a method specified in [114] and detailed in Figure 4.2. Barcode cassettes were amplified from cDNA from above using a minimum number of cycles (typically 12-18) to prevent overamplification (post log-phase PCR) as evidenced by the RT-qPCR data from above. Illumina adaptors were added by two rounds of PCR (5 cycles each), to add phasing adaptors, random barcodes, and multiplexing barcodes. Sublibraries were size-selected on 5% TBE polyacrylamide gels and pooled for sequencing.

Figure 4.2: Preparation of Illumina sequencing libraries from PCR of barcode cassettes, version 1. Three consecutive PCR reactions were used to add adapter sequences compatible with the TruSeq sequencing system. Introduction of molecular barcodes (orange) and multiplexing barcodes (green), allowed for high-throughput quantification of the barcoded libraries.

20–30 sublibraries were sequenced on two lanes of a HiSeq4000 (Illumina) (spiked with 25% PhiX), using a single 1×50 b reads at the Center for Advanced Technology at University of California, San Francisco.

Barcodes were tallied using custom Python software and matched to deletion loci using the lookup table prepared previously to calculate deletion depth.

4.3 Results

In order to function as an effective TIP, a viral mutant must retain all necessary *cis*-acting elements required for replication so that it may be complemented by the wildtype virus. To identify *cis*-acting elements of HIV-1, we conducted a genetic screen by serial, high-MOI passage of NL4-3 and the NL4-3Δ_1 deletion library prepared in Chapter 2.

In a high MOI infection, *trans* factors are provided by co-infection with replication-competent virus. Therefore, genome regions that can tolerate deletion (as measured by enrichment of specific barcodes) correspond to *trans*-acting elements while regions that are intolerant of deletion correspond to *cis*-acting elements. Here we sought to functionally characterize the collection of diverse deletion mutants in order to map *cis*-acting elements of HIV-1. Mutants which persisted through multiple passages are either (a) replication-competent (b) able to be *trans*-complemented by the wildtype virus.

As discussed in Chapter 3, we had two technical considerations when planning the experiment. To achieve efficient *trans*-complementation, cells should be infected at a high MOI, so that on average, each cell is infected with more than one virus, and preferably at least one copy of the wildtype virus to supply *trans* factors. For MOI > 5, the majority of cells should be receive at least one copy of the wildtype genome to supply the missing *trans* factors.

In a second consideration, we wished to gain meaningful data from the experiment and avoid drawing conclusions from noisy data. Thus, to keep strong selective pressure and avoid

drift, we limited the diversity of the library to be fewer than the number of infected cells, so that most of the library would be sampled multiple times during infection. In addition, we also setup 3 biological replicates (flasks K, L, M) that were seeded from a common pool of virus. Thus, properly behaving mutants should be reliably enriched in all three flasks. We also developed a parallel control, where only the wildtype virus would be introduced (flasks A, B, C). In this wildtype only arm (flasks A,B,C), we expect to observe no tagged deletion mutants (barcode cassette). The experiment is diagrammed in Figure 4.3.

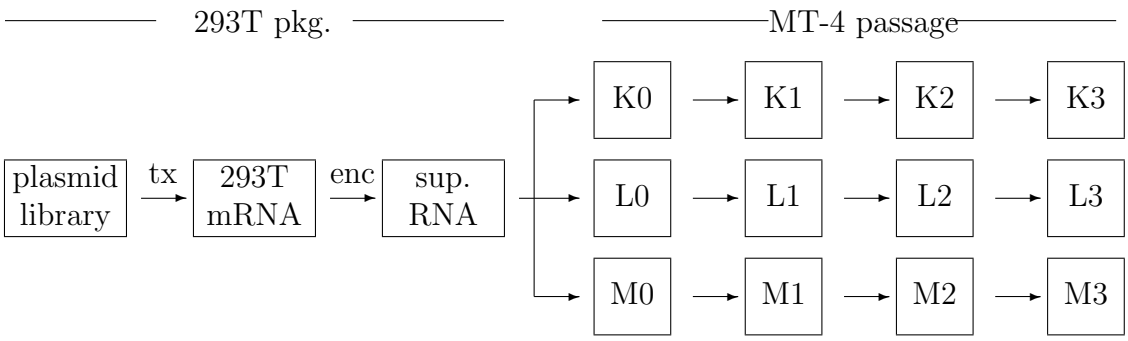


Figure 4.3: Block design of high-MOI passage of NL4-3Δ_1 passage. 293T were co-transfected with equal masses of pNL4-3 and pNL4-3Δ_1 to generate a pool of infectious virus containing both wildtype virus and deletion mutants. The common virus pool was divided and used to infect MT-4 in triplicate (K,L,M). Using a combination of cell-mediated and cell-free transfers, cell-free virus was harvested after 1 week (3 passages: K0/L0/M0), 2 weeks (6 passages (K1/L1/M1)), 3 weeks (9 passages (K2/L2/M2), and 4 weeks (12 passages (K3/L3/M3).). In addition, samples of mRNA from 293T (tx: transcription) and cell-free supernatant (enc: encapsidation) were saved. At the same time, a pool of virus containing only NL4-3 (no deletion mutants) was passaged identically (Flasks A,B,C: not shown).

High MOI passage conditions is achieved

A virus pool containing wildtype HIV-1 (NL4-3) and deletion mutants (NL4-3Δ_1) was prepared by co-transfection of 293T with equal masses of the pNL4-3Δ_1 library and the parental

wildtype plasmid (pNL4-3). The virus-containing supernatant was collected, clarified by 0.45 μm filtration, then concentrated by ultracentrifugation at 48 hpi. The concentrated stock was used to infect MT-4 at high multiplicity of infection (> 5).

Virus stocks were titrated by infecting MT-4 (details in Materials and Methods) and scoring for cells producing HIV Gag (p24) by flow cytometry. This includes all wildtype viruses and also those mutants that are still competent to produce p24. An example titration is shown in Figure 4.4 on page 102. Values are reported as IU/ml, where an IU (infectious unit) is the amount of virus needed to produce a $p24^+$ MT-4 cell at 24 hpi. As shown in Figure 4.4, titers of concentrated stocks averaged $5 \cdot 10^7$ IU/ml.

Now that we had confirmed initial viral stocks were of sufficient titer to provide good coverage of the library ($2 \cdot 10^6$ cells infected at MOI of 5 yields 10^7 infections), we set out to confirm that high MOI would be maintained in subsequent passages.

In this scheme, shown in Figure 4.5, virus can be transmitted by cell-to-cell and cell-free transfer. We confirmed that high MOI could be maintained by staining samples of cells for HIV p24 production throughout the passage. A representative example shown in Figure 4.6) shows that $> 99\%$ of cells were $p24^+$ at various points in the transfer, indicating a high-MOI passage. These conditions was maintained throughout the passage scheme. We can then safely anticipate that the scheme will select for two phenotypes of virus: (a) replication-competent viruses and (b) replication-defective viruses that are efficiently *trans*-complemented by wildtype virus (mobilized).

Barcodes persist throughout the passage

Viral RNA was purified from concentrated viral stocks and cell pellets at various points during the high MOI passage. During purification of viral RNA, $5{\cdot}10^6$ copies of a heterologous sequence (bacteriophage MS2) were spiked in to each purification to normalize recoveries and

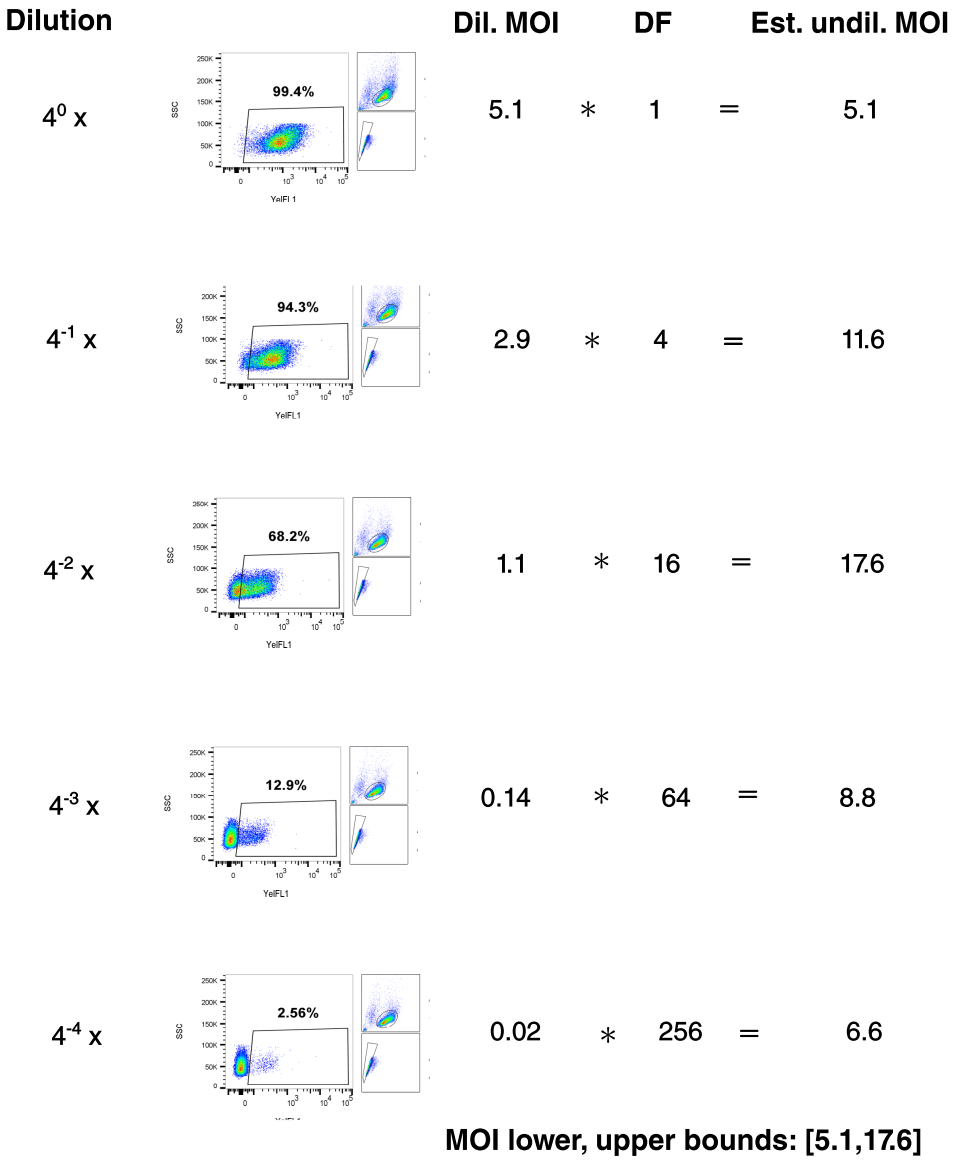


Figure 4.4: An example titration of HIV-1 stocks on MT-4. $2{\cdot}10^6$ MT-4 were infected with 100 μl of serial dilutions of an HIV-1 stock, and stained for HIV-1 Gag production (p24) at 24 hours post-infection, then assessed by flow cytometry. (DF: dilution factor, dil. MOI: MOI of dilution, est.: estimated. For each dilution, the largest flow cytometry dot plots depict side scatter (SSC) vs EGFP and the gating used to establish which cells are EGFP-positive; the small upper right plot depicts live cell gating (forward scatter (SSC) v SSC); the small lower right plot shows singlet gating (forward scatter width vs forward scatter area).

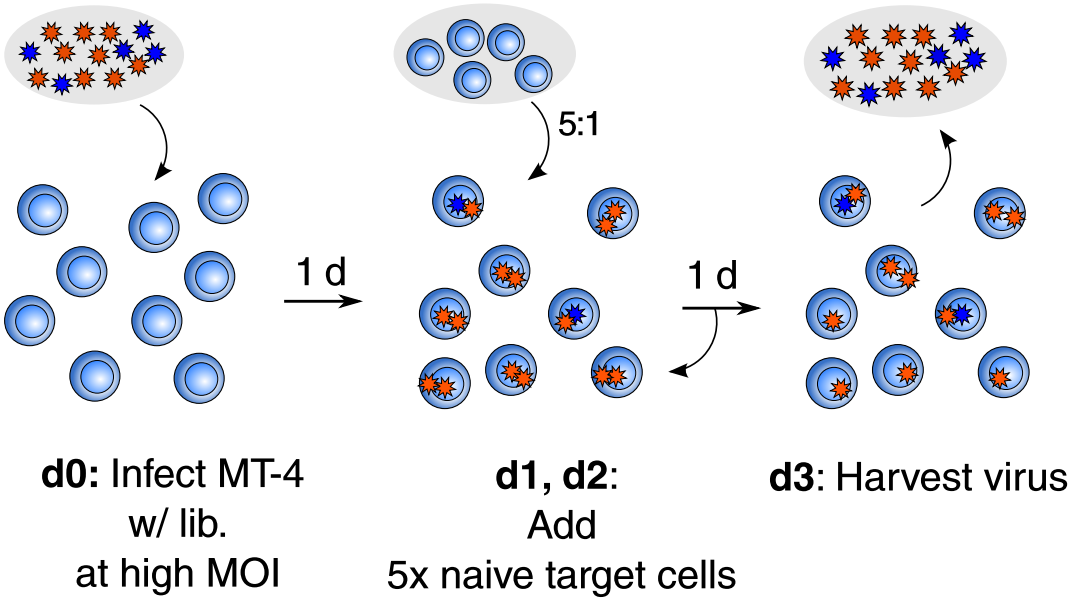


Figure 4.5: HIV-1 high MOI passage screen. On day 0, $2 \cdot 10^6$ MT-4 (blue double discs) are infected at high MOI with a pool of virus (HIV-1) containing both wildtype (red stars) and deletion mutants (blue stars). At day 1, 10^7 additional naïve MT-4 are added and the volume expanded to 12 ml. On day 2, $6 \cdot 10^7$ additional naïve MT-4 were added and the volume expanded to 72 ml. On day 3, cell-free supernatant was harvested and virus purified by ultracentrifugation.

to serve as a control

Levels of barcode cassette (BC) and HIV-1 *pol* (POL) were determined by RT-qPCR. Genomes containing a tagged deletion will harbor BC and genomes that retain the wildtype *pol* sequence (all wildtype HIV and mutants that do not have a deletion in *pol*). Thus BC signal correlates with the concentration of tagged deletion mutants and POL as a measure of total viral genome concentration (provided that *pol* deletion mutants are in the minority). RT-qPCR data for supernatant concentrations are listed in Table 4.3 as quantification cycle (C_q) values. Higher values of C_q indicate reduced abundance of a template molecule; lower values of C_q indicate increase abundance and the scale is logarithmic in base 2, with a ΔC_q of 1 corresponding to a $\approx$ 2-fold difference and ΔC_q of 3 corresponding to an $\approx$ 8-

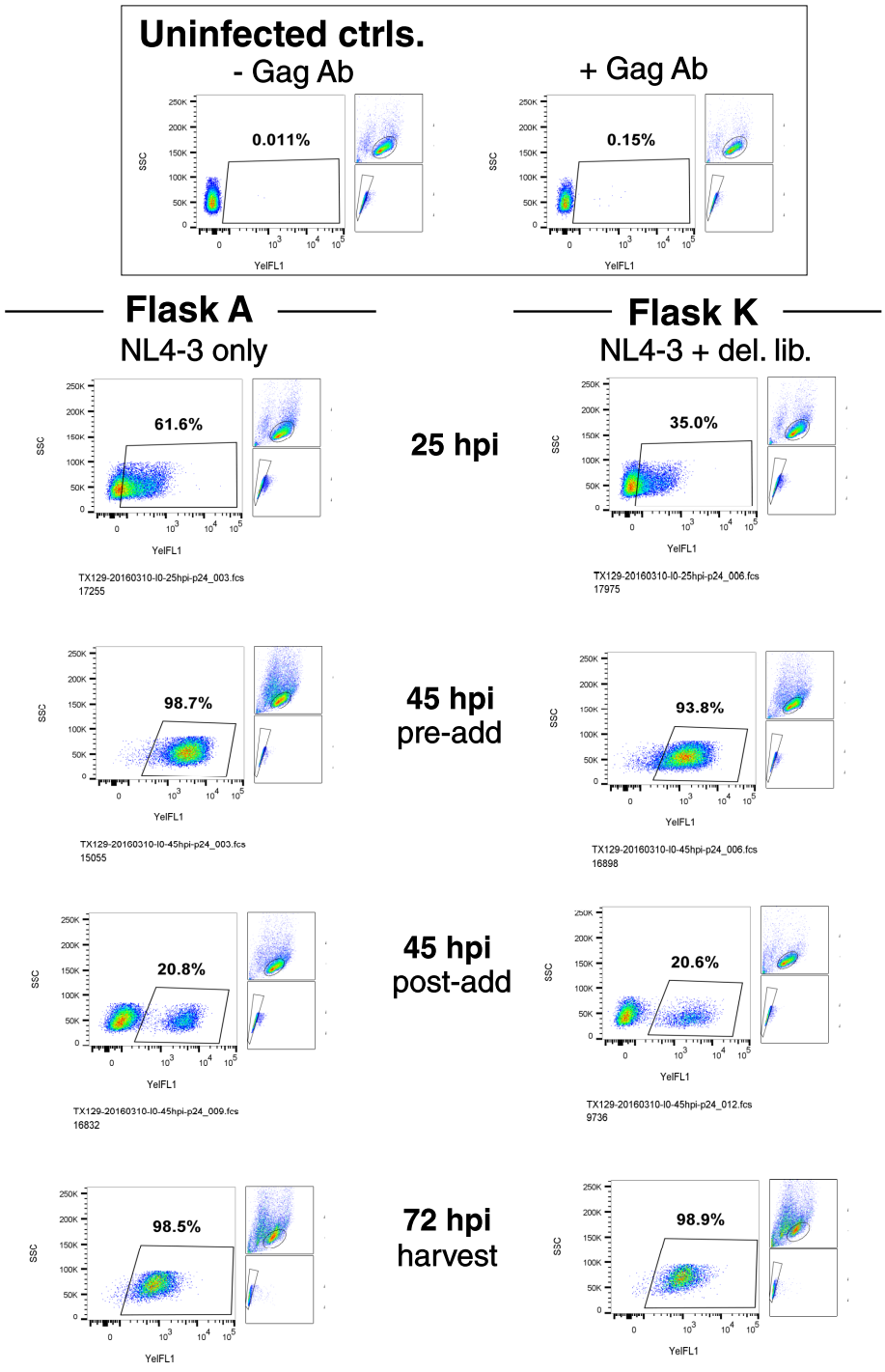


Figure 4.6: High MOI is maintained throughout a week of passage. Cells that are positive for HIV-1 capsid protein are indicated in the polygonal gate and reported as percentage of the population. For each sample, the largest flow cytometry dot plots depict side scatter (SSC) vs EGFP and the gating used to establish which cells are EGFP-positive; the small upper right plot depicts live cell gating (forward scatter (SSC) v SSC); the small lower right plot shows singlet gating (forward scatter width vs forward scatter area).

fold difference. Importantly, the BC primers form a stable primer dimer structure that is amplified even without the presence of template (no-template controls: NTC reactions). The limit of detection for BC is somewhere in the interval of C_q 29–35. True primer dimers and true barcode samples can be differentiated by a melt-curve after cycle 40 of qPCR.

HIV *pol* was detected in the cultures containing WT virus, but was only detectable at background levels in media obtained from uninfected MT-4. BC was detectable in supernatant from cultures containing the deletion library (KLM), but only at background levels in the WT only culture (C). In samples where reverse transcriptase was not added, (−RT), levels of barcode cassette and *pol* were > 10 C_q above the +RT controls, indicating that DNA contamination was negligible (RNA predominates by $\approx 1000\times$ over cDNA). The source of DNA is likely cell fragments and membrane blebs from dying cells. Another possible source of contamination may also be due to reverse transcription products from intravirion reverse transcription, a phenomenon observed in retroviruses. As levels of DNA were low compared to RNA, we did not perform additional DNAseI treatment before downstream processing.

Throughout the passage, levels of BC and POL remained relatively constant from passage to passage and between the three triplicate flasks (K/L/M). Thus the barcoded deletion mutants were stably maintained. We did not observe any differences in total HIV copy number from flasks with and without the NL4-3Δ_1 library, indicating no significant amounts of interference from mutants lacking *pol*.

Intracellular RNA

We also isolated intracellular RNA from 293T that were transfected with pNL4-3 (WT HIV) and the pNL4-3Δ_1 deletion library and performed RT-qPCR. Total RNA was isolated with TRI-Reagent, and a poly(A) fraction was obtained by annealing to paramagnetic poly-dT beads. The total RNA fraction contained a considerable amount of DNA (from transfection

			$C_q \pm$ STDEV ($n = 3$)		
sample	flask	Δ	BC	POL	MS2
293T	ABC	−	34.8 ± 2.2	13.9 ± 0.6	30.8 ± 0.2
293T	KLM	+	16.9 ± 0.2	13.1 ± 0.3	29.1 ± 0.8
passage 3	C	−	34.8 ± 0.1	11.0 ± 0.1	29.6 ± 0.3
passage 3	K	+	16.8 ± 0.2	11.9 ± 0.1	30.7 ± 0.3
passage 3	L	+	15.8 ± 0.8	11.5 ± 0.1	30.7 ± 0.2
passage 3	M	+	16.5 ± 0.2	11.2 ± 0.1	30.6 ± 0.1
passage 6	C	−	34.7 ± 0.3	11.7 ± 0.1	30.3 ± 0.2
passage 6	K	+	16.7 ± 0.4	10.9 ± 0.1	29.7 ± 0.4
passage 6	L	+	17.7 ± 0.2	11.8 ± 0.1	30.1 ± 0.4
passage 6	M	+	17.5 ± 0.3	12.1 ± 0.1	30.3 ± 0.1
passage 9	C	−	35.2 ± 0.8	12.0 ± 0.1	29.1 ± 0.4
passage 9	K	+	17.7 ± 0.1	11.4 ± 0.1	29.5 ± 0.2
passage 9	L	+	18.1 ± 0.2	11.9 ± 0.1	29.5 ± 0.2
passage 9	M	+	17.7 ± 0.1	11.5 ± 0.1	29.8 ± 0.1
passage 12	C	−	35.3 ± 0.3	13.6 ± 0.1	32.7 ± 0.2
passage 12	K	+	18.6 ± 0.2	11.4 ± 0.3	30.6 ± 0.1
passage 12	L	+	17.3 ± 0.2	10.1 ± 0.1	29.7 ± 0.4
passage 12	M	+	17.6 ± 0.1	10.6 ± 0.1	30.0 ± 0.4
media only		−	> 40	37 ± 1	30.6 ± 0.9
80 ng MS2 RNA		−	39.6 ± 0.2	35.2 ± 0.6	11.6 ± 0.1
NTC		−	> 40	37 ± 2.1	36.9 ± 0.3
293T pool, −RT		+	24.7 ± 0.2	21.3 ± 0.2	34 ± 1
p3 pool, −RT		+	29.2 ± 0.2	24.9 ± 0.1	34 ± 1.6
p6 pool, −RT		+	30.1 ± 0.3	24.8 ± 0.3	35.8 ± 0.9
p9 pool, −RT		+	31.2 ± 0.1	25.1 ± 0.4	35.6 ± 0.3
p12 pool, −RT		+	31.2 ± 0.5	24.3 ± 0.1	35.9 ± 0.3

Δ: −/+ NL4-3Δ_1 library
NTC: no template control
−RT: no reverse transcriptase control

Table 4.3: Barcode cassettes are maintained throughout 12 high MOI passages. RT-qPCR from viral RNA purified from supernatant.

				$C_q \pm$ STDEV ($n = 3$)		
sample	flask	WT	Δ	BC	POL	MS2
total	naïve	−	−	34.6 ± 0.7	29.9 ± 0.1	36.9 ± 0.2
total	ABC	+	−	34.6 ± 0.4	13.6 ± 0.1	37.6 ± 0.7
total	KLM	+	+	17.2 ± 0.3	14.0 ± 0.1	37.2 ± 0.2
poly(A)	naïve	−	−	33.6 ± 0.7	28.9 ± 0.4	35.7 ± 0.4
poly(A)	ABC	+	−	34.3 ± 0.5	14.9 ± 0.1	35.8 ± 0.2
poly(A)	KLM	+	+	18.1 ± 0.1	15.6 ± 0.2	37.2 ± 0.7
total	pooled, −RT	+	+	23.6 ± 0.5	19.8 ± 0.2	36.7 ± 0.6
poly(A)	pooled, −RT	+	+	30.4 ± 0.3	32.3 ± 0.9	36.0 ± 0.1
MS2	—	−	−	36.1 ± 1.4	36.5 ± 0.5	28.4 ± 0.2
NTC	—	−	−	> 40	34 ± 5	37.0 ± 0.8

WT: −/+ NL4-3
Δ: −/+ NL4-3Δ_1 library
NTC: no template control
−RT: no reverse transcriptase control

Table 4.4: Barcode cassettes are present in poly(A) RNA in transfected 293T. RT-qPCR from transfected 293T.

plasmids), but the poly(A) fractionation effectively removed most of this. The poly(A) fraction corresponds to poly(A) mRNA, which requires complete transcription. RT-qPCR results are listed in Table 4.4. As with the supernatant data in Table 4.3, BC was only detected when the deletion library was included, and BC and POL were not detectable in naïve 293T.

Barcode distributions are robust

Using the C_q values from Table 4.3 and Table 4.4 we prepared Illumina sequencing libraries to characterize which barcodes were persisting throughout the high MOI passage. Barcode cassettes were amplified exponentially for a minimum number of cycles, then multiplexing indices and Illumina TruSeq adapters were added by two additional rounds of PCR. From

RT-qPCR, we established that the number of template molecules for each PCR reaction was $> 10^7$.

Using the barcode-to-deletion mapping for the 23,851 mappable mutants in pNL4-3Δ_1 (prepared in Chapter 2), the count of each barcode cassette (with a known deletion locus), was tabulated, and used to compute the prevalence, f_i of each barcode cassette i in each sample.

How robust is the selection procedure? Are the barcodes enriched in one arm of the experiment also enriched in the others? How predictive is knowledge of barcode prevalence in one flask vs. the others?

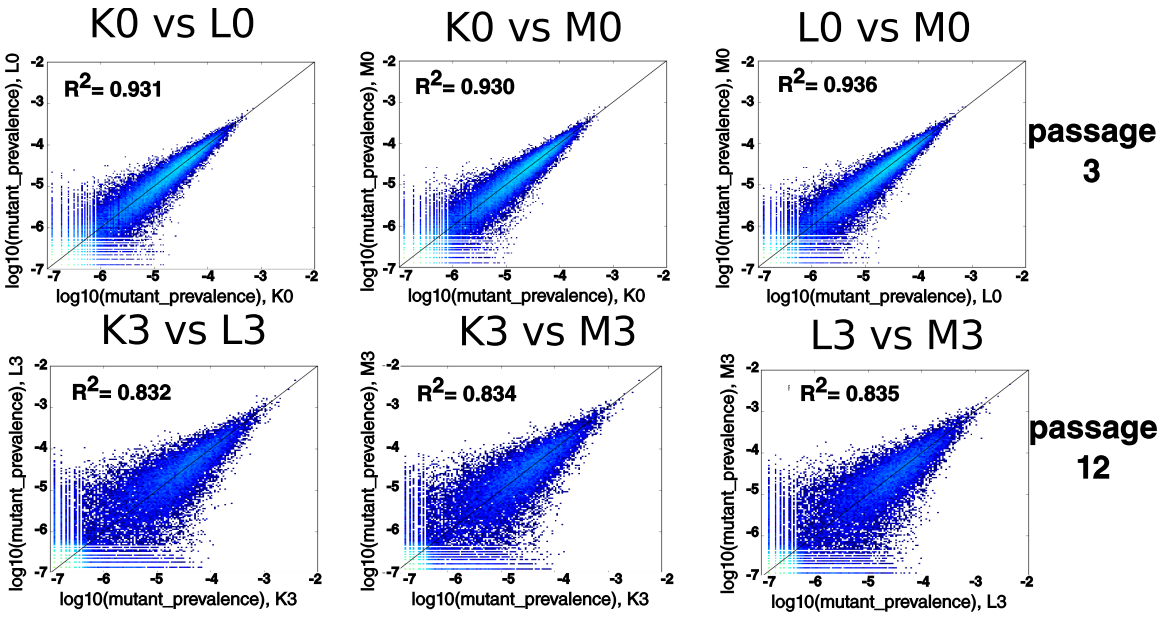


Figure 4.7: Biological replicates are robust Pairwise dot plots show R^2 values of 0.83-0.93 at passage 3 and passage 12 passages.

To answer these questions, we constructed pairwise correlation plots of all barcodes in triplicate flasks (K,L,M) at passage 3 and passage 12 (Figure 4.7). Coefficients of determination (R^2) at early timepoints (passage 3) were on average greater than at passage 12

($R^2 \approxeq 0.93$) and at later timepoints dipped slightly to ($R^2 \approxeq 0.83$). Thus, there was strong concordance between the triplicate infections.

Identification of deleterious and adaptive mutations

During the high MOI passage, we select for two phenotypes of deletion mutants that persist throughout passage. Persistent mutants are either replication-competent (are self-mobilized) or are replication-defective but can be efficiently *trans*-complemented by the wildtype virus.

Using deep sequencing, we are able to tabulate the prevalence of each mutant in the total population of barcoded mutants, and compute prevalence trajectories throughout the passage. If a particular mutation becomes more prevalent over time, we can conclude that it is adaptive: it confers a fitness advantage (fitness is greater than average fitness at that point). If a particular mutant or mutation becomes less prevalent, than the mutation is deleterious (fitness decrease relative to average fitness). If prevalence remains constant, than the mutation is neutral. Once a mutant falls below a prevalence of $1/N$ where N is the population size (10^5 to 10^6 in this experiment), it likely to be lost due to drift [6]. This corresponds to $\log_{10}$(prevalence) of -5.0 to -6.0.

Of the 23,851 mappable NL4-3 deletion mutants, only 4390 (18%) were extant in all three replicate flasks by passage 12—the remaining 19,461 (82%) were extinct in at least one of the three replicates.

We computed prevalence trajectories for these 4390 extant mutations and calculated the slope in prevalence versus passage number by linear regression, as shown below in Figure 4.8.

After fitting linear regression lines to these 4390 mutants, we observed that 1390 (32%) increased in prevalence through every passage, indicating that deletion mutants harboring these deletions were transmitting better than the average member of the barcoded population (Figure 4.9). The remaining 3000 mutants remained steady or decreased in prevalence

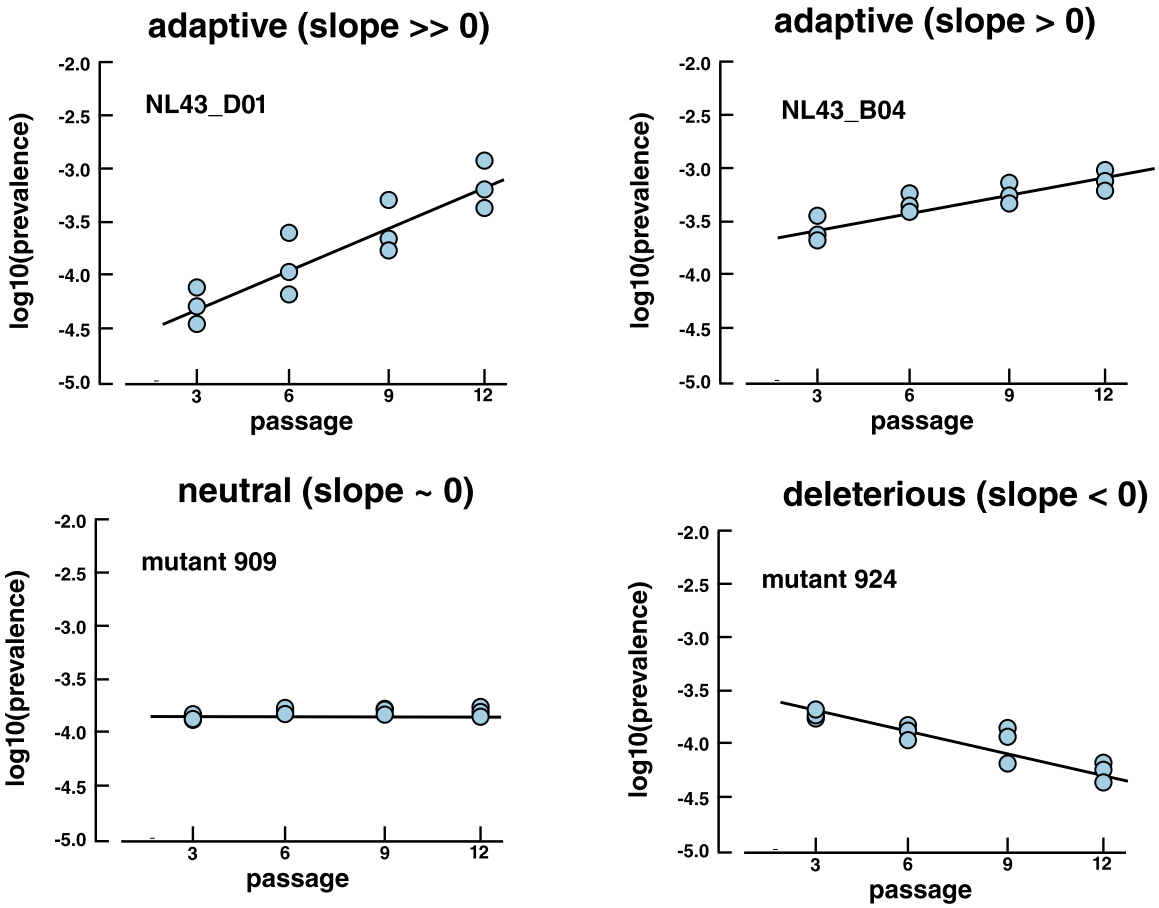


Figure 4.8: Representative mutation trajectories during high MOI passage. Some deletions were adaptive (upper panels: NL43_D01 and NL43_B04), while others were netural (mutant 909) or deleterious (mutant 924). Data points correspond to the triplicate flasks (K,L,M) at each passage. Prevalence is in reference to the total barcode cassette pool (tagged mutants).

through every passage. As barcode levels were relative constant to total HIV (as shown previously in RT-qPCR data), we conclude that these 1390 were transmissible and could spread through the population as fast or faster than the wildtype virus. Thus, these 1390 deletions represent a potential collection of deletion mutants that are transmissible ($R_0 > 1$) under these conditions of high MOI passage in MT-4 as they are expanding in the population.

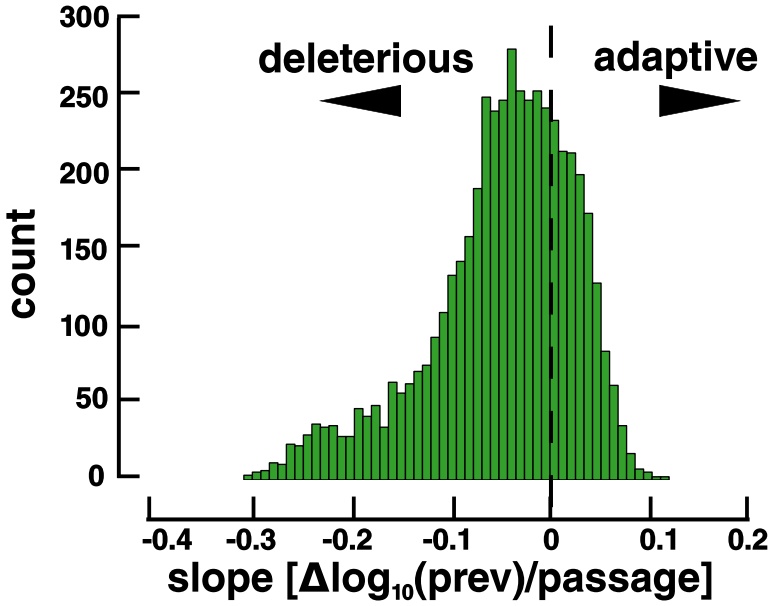


Figure 4.9: Distribution of fitness in deletion mutants that are not extinct by passage 12. Of the 4390 mutants that are not extinct, 1390 (30%) are increasing in prevalence through every passage. The dashed vertical line marks the neutral fitness boundary (slope of 0).

Identification of *cis*-acting elements

Using deep sequencing counts of each barcode cassette (vector $\mathbf{c}$) and the barcode-to-genotype mapping prepared in Chapter 2 (matrix $\mathbf{M}$), we were able to compute a deletion depth profile for the HIV-1 genome at each timepoint using the formula ($\mathbf{d} = \mathbf{Mc}$) (see Equation 3.7 on page 79 of Chapter 3 for details). Regions where the deletion depth is large

correspond to genomic intervals that are tolerant to deletion (*trans*-acting elements encoding common goods). Intervals where the deletion depth is low or tends to 0 correspond to *cis*-acting elements, and cannot be complemented *in trans*. Deletion depth is presented in terms of raw units (unnormalized), and depends on the sequencing depth of each library.

As a baseline, we determined the deletion depth profile of the pNL4-3Δ_1 deletion library across the pNL4-3 provirus, shown below in Figure 4.10.

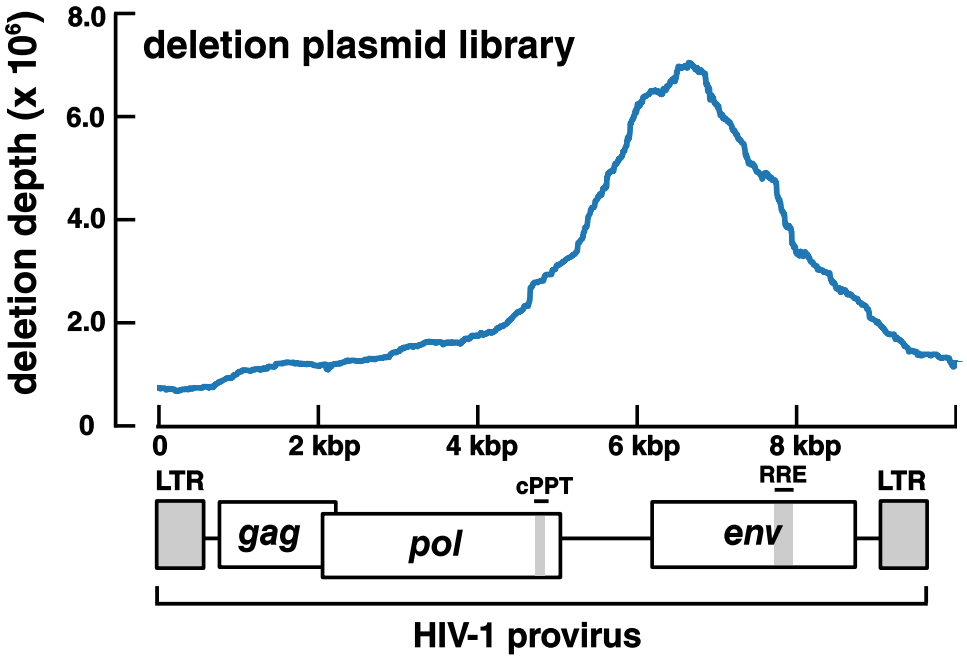


Figure 4.10: Deletion depth profile of the pNL4-3Δ_1 plasmid deletion library. Note the peak centered on the signal peptide of *env* and coverage of the entire HIV-1 genome.

The deletion depth across the genome is nonzero and flat across 0–4 kbp before rising to a peak centered at the N-terminus of *env*, and falling towards the 3′ end of the genome. Bias in deletion depth at this stage corresponds to differences in growth rate in the *E. coli* host harboring each plasmid. Faster growing bacteria will cause their harbored plasmid to be overrepresented in the library. The signal peptide of *env* and sequences at the N-terminus

are known to be toxic to bacteria, therefore bacteria harboring *env* deletion mutant plasmids are likely to outgrow bacteria harboring plasmids that retain *env*.

cis-acting elements required for transcription

The poly(A) fraction of RNA isolated from 293T co-transfected with the pNL4-3Δ_1 deletion library and pNL4-3 represents mRNA. Barcodes found in this fraction represent genomes that have been successfully transcribed from the HIV-1 provirus. Areas of low deletion depth (compared to the plasmid library profile) correspond to *cis*-acting elements required for transcription in 293T. A deletion depth profile for this poly(A) RNA fraction is shown below in Figure 4.11.

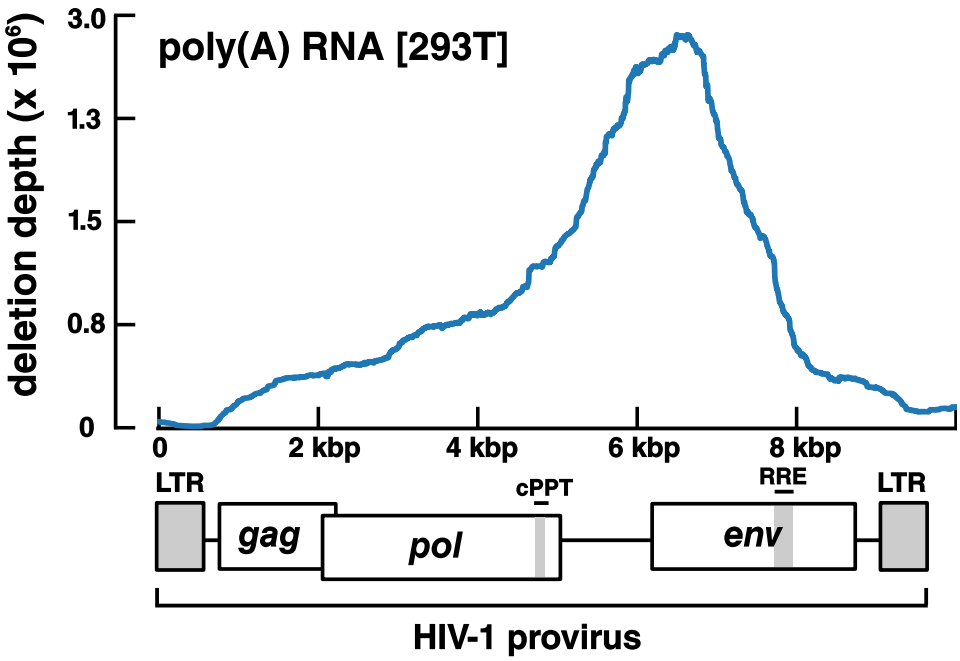


Figure 4.11: The 5′ LTR and 5′ UTR are required for efficient transcription. The deletion depth profile of poly(A) RNA from transfected 293T is shown. Note the region of near-zero deletion depth at the 5′ end of the genome.

The deletion depth profile of the poly(A) fraction has the same basic profile as the plasmid

deletion library with two notable differences. Deletions at the 5′ end of the genome (spanning the 5′ LTR through SL1–SL4) strongly inhibit transcription, as reflected by the low deletion depth. This regions encodes the LTR promoter and regions necessary for splicing and efficient transcription (TAR loop). Thus deletions here are expected with current knowledge of HIV (see Table 4.1 on page 90).

At the 3′ end of the genome, we observe a decrease after the RRE, falling towards the 3′ LTR with respect to the plasmid deletion library profile in Figure 4.10.

cis-acting elements required for encapsidation and egress

We next built a deletion depth profile of RNAs that were present in the virus-containing fraction of supernatant from transfected 293T. Barcodes found in this fraction have tagged viral genomes that could be transcribed, exported from the nucleus, and packaged into virions (encapsidated), and exported from the cell (egressed).

As shown in Figure 4.12, the deletion depth profile is strikingly different from the transcription of plasmid library profiles, and shows two key areas of low deletion depth, which correspond to cis-acting elements that are required for transcription and encapsidation. At the 5′ end of the genome, we observe an area of zero deletion depth that begins at the 5′ LTR and continues through the 5′ UTR, (SL1,SL2,SL3,SL4) to the start codon of *gag*. This region often referred to as Ψ, the lentiviral packaging element, and is present in most lentiviral vectors. Structural studies have identified this region to be a minimal packaging element [115].

At the 3′ end of the genome, there is a region of zero deletion depth that maps exactly to the Rev Responsive Element (RRE), a region of secondary structure that binds HIV-1 Rev to export incompletely spliced RNAs from the nucleus via the Crm1 pathway [116]. We can conclude that the RRE is important for the steps of nuclear export and encapsidation, but

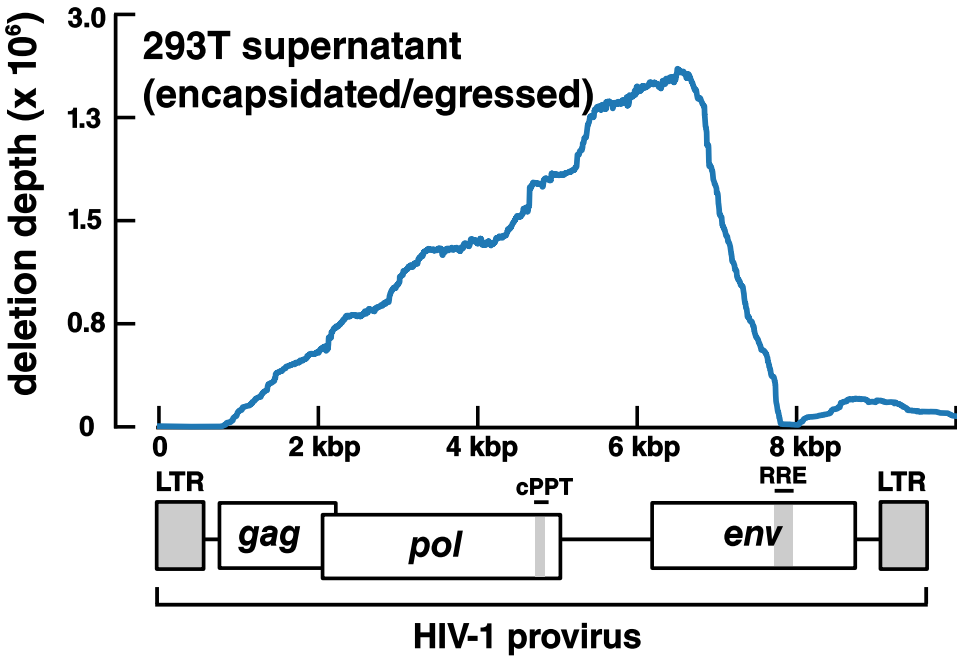


Figure 4.12: The 5′ UTR and the HIV-1 RRE are required for export and encapsidation. A deletion depth profile built from tagged viral deletion mutants in the supernatant of transfected 293T. Key regions are the 5′ end of the genome (5′LTR, 5′ UTR, and the RRE).

we cannot differentiate between nuclear export or encapsidation as the potential block, as we did not fractionate nuclear and cytoplasmic RNAs.

These two regions (5′ LTR/UTR and RRE) appear to be the only elements required for transcription and encapsidation of the viral genome - all other genome intervals can tolerate some amount of deletion. This is in contrast to literature reports of the GRPE (a region at the frameshift of *gag/pol*, that is putatively necessary for genome encapsidation [95]. It is however, consistent with a model of Gag binding to Ψ and the RRE of HIV-1 [117, 118].

Identification of *cis*-acting elements through high MOI passage

The pool of NL4-3Δ_1 deletion mutants was next passaged 12 times at high multiplicity-of-infection (MOI) in a permissive T-cell line, MT-4. Samples were collected every 3 passages and deletion depth profiles built by deep sequencing of the barcoded deletion mutants. Deletion depth profiles at passage 3, passage 6, and passage 12 are shown in Figure 4.13, Figure 4.14, and Figure 4.13.

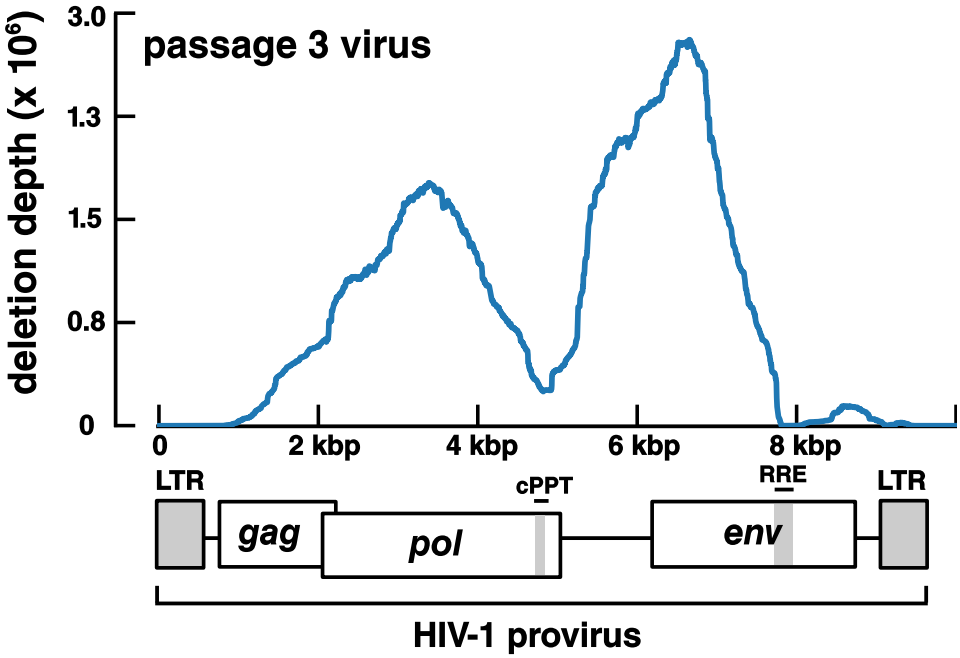


Figure 4.13: Deletion depth profile after 3 high MOI passages in MT-4. Key *cis*-acting regions are the 5′ end of the genome (5′ LTR through Gag Matrix (p17)), cPPT/CTS, RRE, and the 3′ end of the genome (PPT-3′ LTR).

At passage 3, the deletion depth profile plotted in Figure 4.13 diverges notably from the the 293T supernatant profile. We observe three key differences: (1) a valley that appears with a minimum centered above the cPPT/CTS (2) a shift of the region of zero-deletion depth at the 5′ end of the genome, which now encompasses the 5′ LTR, 3′ UTR, and the first three hundred bases of *gag*, (3) a widening and 3′ shift of the valley situated above the

RRE.

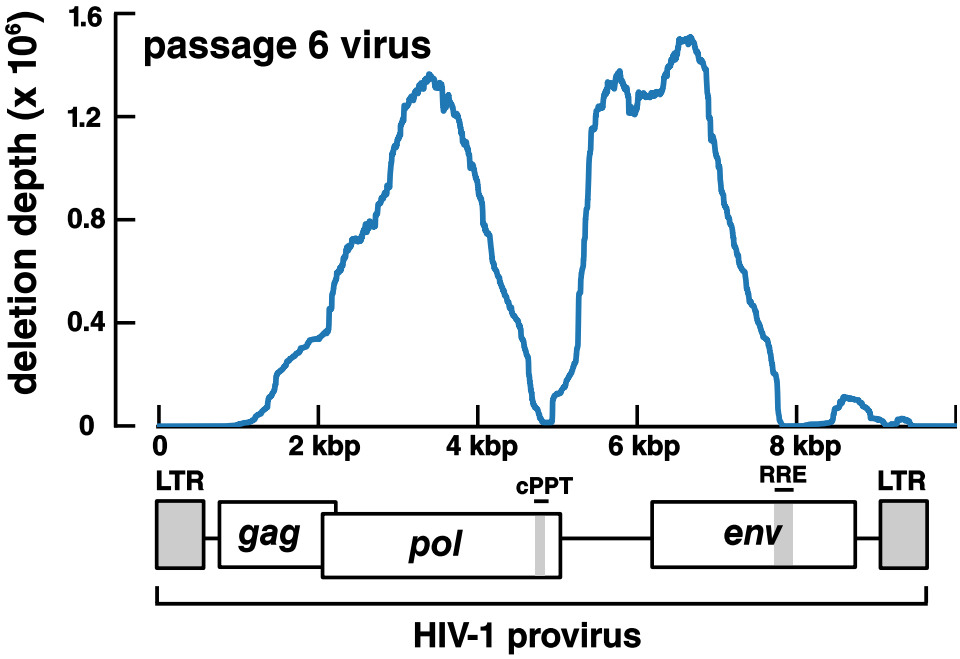


Figure 4.14: Deletion depth profile after 6 high MOI passages in MT-4. Key *cis*-acting regions are the 5′ end of the genome (5′ LTR through Gag Matrix (p17)), cPPT/CTS, RRE, and the 3′ end of the genome (PPT-3′ LTR).

After 6 *in vitro* passages (Figure 4.14), these features become more pronounced where each valley flattens to a deletion depth of zero. The 5′ shoulder of the peak situated above *env* also appears to increase in height.

After 12 high MOI *in vitro* passages, areas of low deletion depth (indicating intolerance to deletion) have flattened and reached a deletion depth of approximately zero. This actual deletion depth is non-zero (values of 10–300), and could reflect cross-talk between libraries during demultiplexing on the Illumina platform in addition to persistent mutants.

In Figure 4.15, we note three regions of the HIV-1 genome that are tolerant to deletion, indicating that they encode elements that can be complemented efficiently in *trans*. The deletion depth profile at passage 12 is annotated in Figure 4.16. We observe two *trans*-

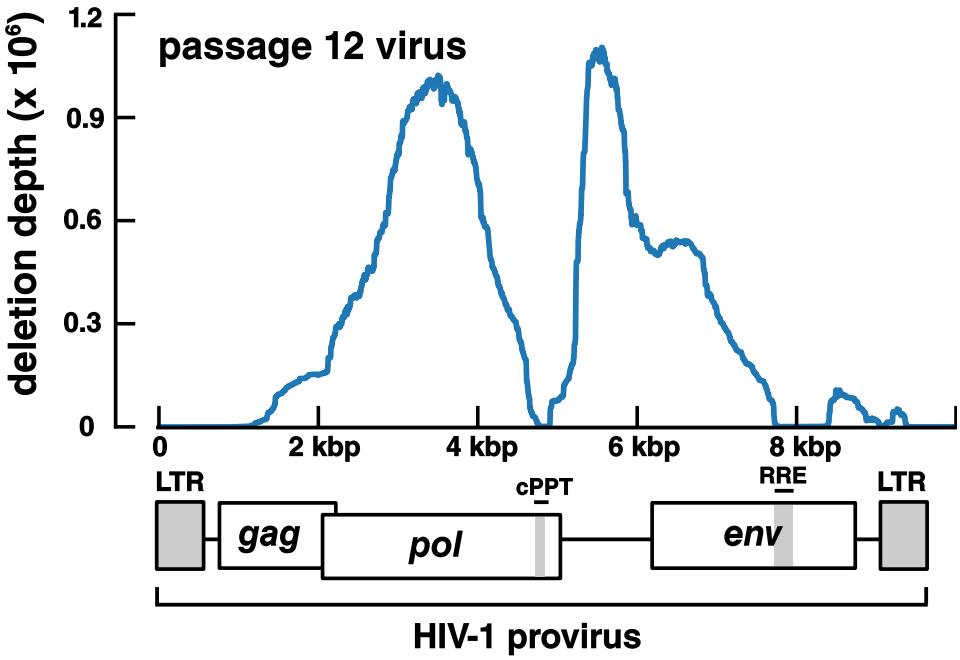


Figure 4.15: Deletion depth profile after 12 high MOI passages in MT-4. Key *cis*-acting regions are the 5′ end of the genome (5′ LTR through Gag Matrix (p17)), cPPT/CTS, RRE, and the 3′ end of the genome (PPT-3′ LTR).

acting element "mountains" and one *trans*-acting "hill", with peaks at the center of *pol* (*trans*-acting element 1: TAE1), in the accessory gene tract of HIV (*vif*–*vpu*) (TAE2), and in the 3′ end of *env*. The deletion depth profile in TAE1 (*gag*-*pol*) has a small shoulder and an inflection point corresponding to the start codon of *pol*. The deletion depth profile of TAE2 (post cPPT to pre-RRE) has a shoulder corresponding to the start codon of *env*. The last *trans*-acting element (TAE3) begins after splice acceptor 7 (SA7) which marks the second exon of *tat* and *rev*, and ends at the PPT. The last hill after TAE3 (within the U3 region of the 3′ LTR) may be an artifact due to the mechanism of HIV-1 genome replication, where the U3 region of the 3′ LTR is copied to the 5′ end of the DNA genome during reverse transcription. Thus, mutants with apparent deletions in U3 may persist

The deletion depth profile after passage 12 also contains four regions of low or zero

deletion depth, indicating that these genomic regions cannot tolerate deletion and cannot be complemented efficiently *in trans*: these regions are putative *cis*-acting elements (CAE).

CAE1 is 1115 bp in length and maps to nucleotides 1–1114 of NL4-3 proviral genome. This 5′ end of genome encompasses the 5′ LTR, stem loops 1–4, and the first 325 b of *gag*, which maps to the Gag MA (p17). This region is included in lentiviral vectors and a minimal packaging element, Ψ ([115]), has been mapped to this region. Why deletions in Gag MA are not tolerated remains unknown, although others have speculated that there is a relationship between encapsidation and translation of the HIV-1 genome [28, 29].

CAE2 is 126 bp in length and maps to nucleotides 4779–4905 of NL4-3, at the 3′ end of *env*. This region maps exactly to the cPPT/CTS, a region previously identified as being necessary for reverse-transcription and integration of HIV-1 [96, 97]. This region is conserved in lentiviruses but not retroviruses [96].

CAE3 is 671 bp in length and maps to nucleotides 7710–8381 of NL4-3, at the 3′ end of *env*. It begins exactly at the RRE, and ends precisely at splice acceptor 7 (SA7). This splice acceptor is used for splicing of several multiply spliced HIV-1 transcripts, including (*vpr*, *tat*, *rev*, *nef*)[86]. This region, too, is included in many lentiviral vectors.

CAE4 is 684 bp in length and maps to nucleotides 9025–9709 of NL4-3. This region spans the PPT, necessary for reverse transcription in HIV-1 and the 3′ LTR [20]. Identifying this region as a *cis*-acting element is uncontroversial.

4.4 Discussion

cis-acting elements of HIV-1

In this chapter, we conducted a genetic screen to map the *cis*- and *trans*-acting elements of HIV-1. In the screen, 293T were transfected with the NL4-3Δ_1 barcoded deletion library

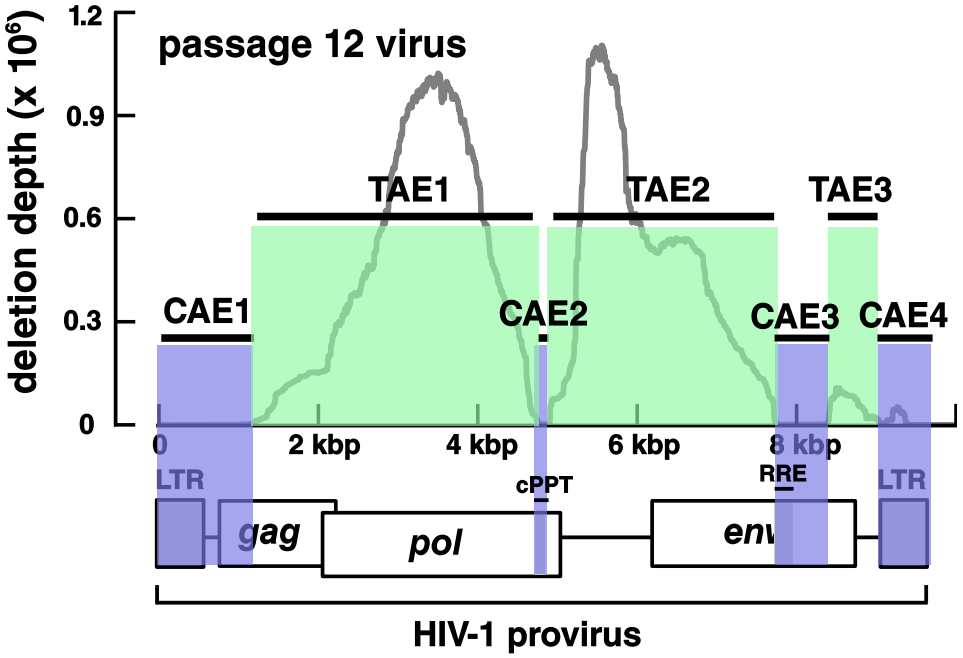


Figure 4.16: A model of HIV-1 *cis*- and *trans*-acting elements: The HIV-1 genome is composed of 4 *cis*-acting elements, CAE1–CAE4, (highlighted in blue) and 3 *trans*-acting elements, TAE1–TAE3, (highlighted in green).

.

and wildtype NL4-3 to produce a pool of infectious virus. The virus pool was then passaged 12 times at high MOI (5–20) in a T-cell line (MT-4).

During high MOI infection, target cells are infected with more than one virus with high probability, allowing replication-defective viruses that retain all necessary *cis*-acting elements to persist through the passage by *trans*-complementation with common goods produced by functional wildtype virus. Thus, this screen selects two viral phenotypes: (a) replication-competent viruses (b) replication-defective viruses that retain all necessary *cis*-acting elements and can be mobilized effectively by wildtype HIV-1.

Sequencing of the barcodes at critical points in the infection cycle (transcription, encapsidation/egress, passage 3, passage 6, passage 12) and referencing the barcode/deletion lookup table allowed for the identification of genomic regions where deletions were tolerated in each

step of the HIV-1 replication cycle (*trans*-acting elements) or were intolerant to deletion (*cis*-acting elements).

Using the high-resolution deletion depth profiles, we clearly identified four distinct *cis*-acting elements, regions that that could not be complemented *in trans*. These regions are annotated in Figure 4.16: **CAE1:** nt 1–1144 of NL43, spanning the 5′ LTR, stem loops 1–4, and the first 325 b of *gag*; **CAE2:** nt 4779–4905 of NL4-3, the cPPT/CTS; **CAE3:** nt 7710–8381 of NL4-3, which begins exactly at the RRE and ends precisely at splice acceptor 7 (SA7); **CAE4:** nt 9025–9709 of NL4-3, which covers the PPT and the 3′ LTR.

Based upon these results, a minimal HIV-1 therapeutic interfering particle (in proviral form) might be obtained by concatenation of the sequences of CAE1 (1115 bp), CAE2 (126 bp), CAE3 (671 bp), CAE4 (684 bp), which sum to a total length of 2596 bp. This of course, is under the naïve assumption that the regions can act independently and the multiple deletions do not demonstrate epistasis.

Despite previous claims that the GRPE is important for encapsidation ([95]) we show that deletions of this region do not effect mobilization, in agreement with Nikolaitchik and Hu ([98]).

There are a few potential weaknesses to our this study. We conducted the screen using a single molecular clone of HIV-1 and a single clonal cell line. It is possible that *cis*-acting elements can vary between viral strains and between cell lines and tissue types. Conducting the screen in a tissue explants (PBMC or HLAC cultures) may reveal different results.

We were also unable to monitor recombination between viruses ([119]), which could produce viral strains that have acquired more than one deletion, and create linkage effects. Similarly, we did not perform sequencing outside of the barcode cassette and could not monitor the appearance of additional mutations. We point out that the dot plots in Figure 4.7, show strong correlation between replicates, showing the our observed selection was

deterministic, not in the region of drift.

Generation of HIV-1 DIPs *in vivo*

In contrast to other viral systems, *in vivo* defective interfering particles have not been reported for HIV-1. Why might this be?

We can first eliminate one potential explanation, that HIV-1 single deletion mutants do not arise with appreciable frequency. Small insertions/deletions in HIV-1 genomes are commonly observed during sequencing of *in vivo* HIV-1 infections in patients [7]. Moreover, large single deletion mutants are often recovered as integrants in resting T-cells (as replication-defective proviruses) [99, 120–122]. Thus, spontaneous single deletion mutants do arise during *in vivo* HIV-1 infections.

Even though mutants with single deletions are spontaneously generated, the wildtype virus may require multiple deletions to be converted to an HIV-1 DIP. From the Figure 4.16 model of HIV-1 *cis*- and *trans*-acting elements (page 120), we see that the *trans*-acting elements (TAE1, TAE2, TAE3), which we expect to be deleted in a DIP, are interleaved with *cis*-acting elements (CAE2 and CAE3). Inactivation of all three TAE by a single, contiguous deletion would also remove two *cis*-acting elements, rendering the deletion mutant unable to be mobilized efficiently. DIP-inducing deletions cannot span *cis*-acting elements.

Thus, generation of an HIV-1 DIP would require (a) simultaneous inactivation of TAE1, TAE2, and TAE3 (a rare event) or (b) stepwise acquisition of deletions/mutations in TAE1, TAE2, and TAE3 while preserving all four *cis*-acting elements. Retention of the activities of *nef* (in TAE4), and *vpu* (in TAE3), would downregulate CD4 and prevent superinfection by the helper virus. Sufficient intracellular levels of any of the HIV-1 *trans*-acting elements may be enough to trigger destruction of the infected cell. The complexity of the deletions required may be why HIV-1 DIPs have not be observed in natural HIV-1 infections.

Future direction: low MOI screens

In a more stringent screen, the HIV-1 deletion library, pNL4-3Δ_1, could be passaged in permissive cells at low MOI (MOI of < 1), instead of high MOI as performed in this chapter. An example scheme is depicted in Figure 4.17 on page 124.

Target cells are transduced/infected at low MOI with a pool of virus containing tagged deletion mutants and wildtype helper virus, then allowed to recover in the presence of a drug (or neutralizing antibody) to restrict viral replication to a single round. During the recovery period, cells infected with wildtype virus or cytopathic deletion mutants will be killed, while uninfected cells and cells infected with non-cytopathic deletion mutants will survive. After the recovery period, the population of cells is infected at high MOI with the wildtype virus to mobilize deletion mutants. In this fashion, we can select for deletion mutants which do not kill their host cell but also retain all *cis*-acting elements required for transmission.

The low MOI screen is more stringent than a high MOI screen performed in this chapter, as it selects for mutants that are non-cytopathic but can be mobilized. The high MOI screen selects for mutants that are replication-competent or can be mobilized by the wildtype virus, but does not select for or against cell killing. Successive passages (every 24 hours), were performed before cell death occurs in HIV-1 infections, typically 24-72 hours post-infection [123].

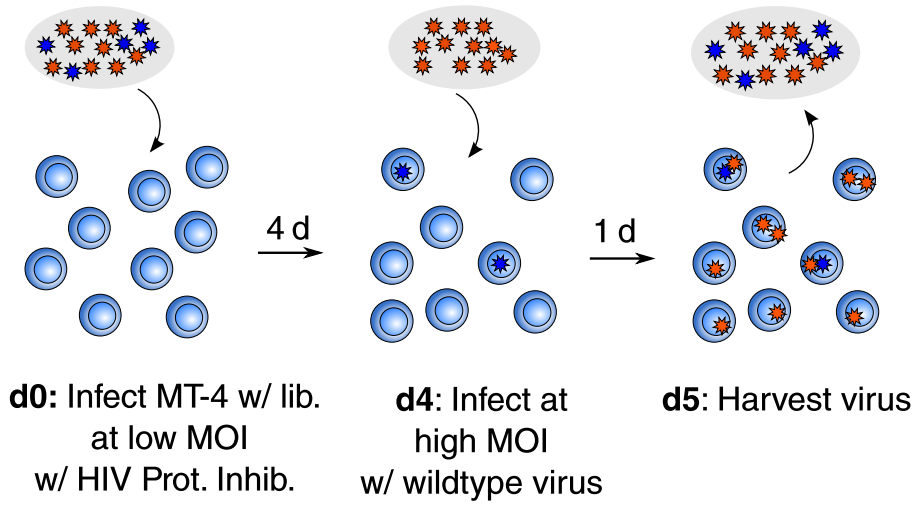


Figure 4.17: Low MOI screen for noncytopathic HIV-1 DIPs On day 0, MT-4 (blue double discs) are infected at low MOI with a pool of virus (HIV-1) containing both wildtype (red stars) and deletion mutants (blue stars). This infection occurs in the presence of an HIV-1 protease inhibitor, such as Darunavir, which restricts viral replication to a single round. Cells infected with cytopathic HIV-1 (such as the wildtype helper virus) are killed between 0 and 4 days post-infection. Uninfected cells and cells infected with deletion mutants that do not cause cell death are enriched. At 4 days post-infection, the entire population of cells is infected at high MOI with wildtype helper virus (red stars), to achieve efficient *trans*-complementation of deletion mutants. On day 5, the virus-containing supernatant is harvested, now enriched for deletion mutants that are non-cytopathic and retain all *cis*-acting elements. The virus pool can be used to infect cells at low MOI and the process iterated.

Chapter 5

Screen for ZIKV therapeutic interfering particles

5.1 Introduction

Zika virus (ZIKV) is a flavivirus (member of *Flaviviridae*) and has a (+)-sense, single-stranded RNA genome. The ≈ 11 kb RNA genome encodes a single polypeptide that is cleaved by viral- and cellular-encoded proteases into 10 proteins. Four of these proteins (C, Pr, M, E) are structural and form the virion. The remaining six are non-structural (NS2A, NS2B, NS3, NS4A, NS4B, NS5) and are important for viral replication and immune evasion [58]. In infected cells, ZIKV replicates predominantly in the cytoplasm, specifically, in membranous compartments formed from the endoplasmic reticulum [27, 124].

Zika virus is capable of replication in mammalian hosts, notably humans, and in mosquitoes of the genus *Aedes*, in particular *Aedes aegypti* and *Aedes albopictus*. Human-mosquito transmission occurs when a female mosquito takes in a blood meal. Human-human transmission can occur through perinatal transmission (mother to child), sex (infectious ZIKV can be

found in semen), and significant exchange of infected tissues, such as during blood transfusions.

Zika virus has been observed since the 1940s (initial report in 1947), and remained endemic in Africa over the last 60 years [61]. Within the last 5 years, a series of transmission events allowed introduction from French Polynesia, and to South, Central, and North America [125, 126]. At present, there are no vaccines or drugs available to treat ZIKV infections.

ZIKV is closely related to Dengue fever virus (DENV), which is also spread by the bite of *Aedes* mosquitoes. Although no natural DIPs have been reported for ZIKV, there is precedent for naturally-occurring Dengue DIPs. In natural infections, there is a report of a defective virus with a stop-codon in the envelope (E) protein that spread hundreds of kilometers by transmission between mosquitoes and humans, presumably mobilized by the wildtype virus during co-infection [127]. During *in vitro* replication experiments, mutants with deletions of the entire $\approx$ 10 kb open reading frame (retaining only the $5'$ and $3'$ UTRs) can arise [115, 128].

A common tool of RNA viruses are *replicons*: viruses that are deficient in one or more structural proteins, so they cannot form infectious viral particles, but retain the necessary replication machinery to replicate their genome in permissive host cells. Several varieties are available for DENV and other flaviviruses, such as Hepatitis C virus [129].

A variety of structural information about ZIKV has become recently available, including high resolution structural and functional information about intact virions ([130]) and individual proteins: envelope ([131]), NS1 ([132, 133]), and NS5 ([134]).

In this chapter, we map the *cis*-acting elements of ZIKV at fine resolution, by passaging the pMR766Δ deletion libraries prepared in Chapter 2 at a high multiplicity of infection in Vero cells. Moreover, we identify several mutants that are efficiently *trans*-complemented by the wildtype virus, and can potentially serve as prototype TIPs.

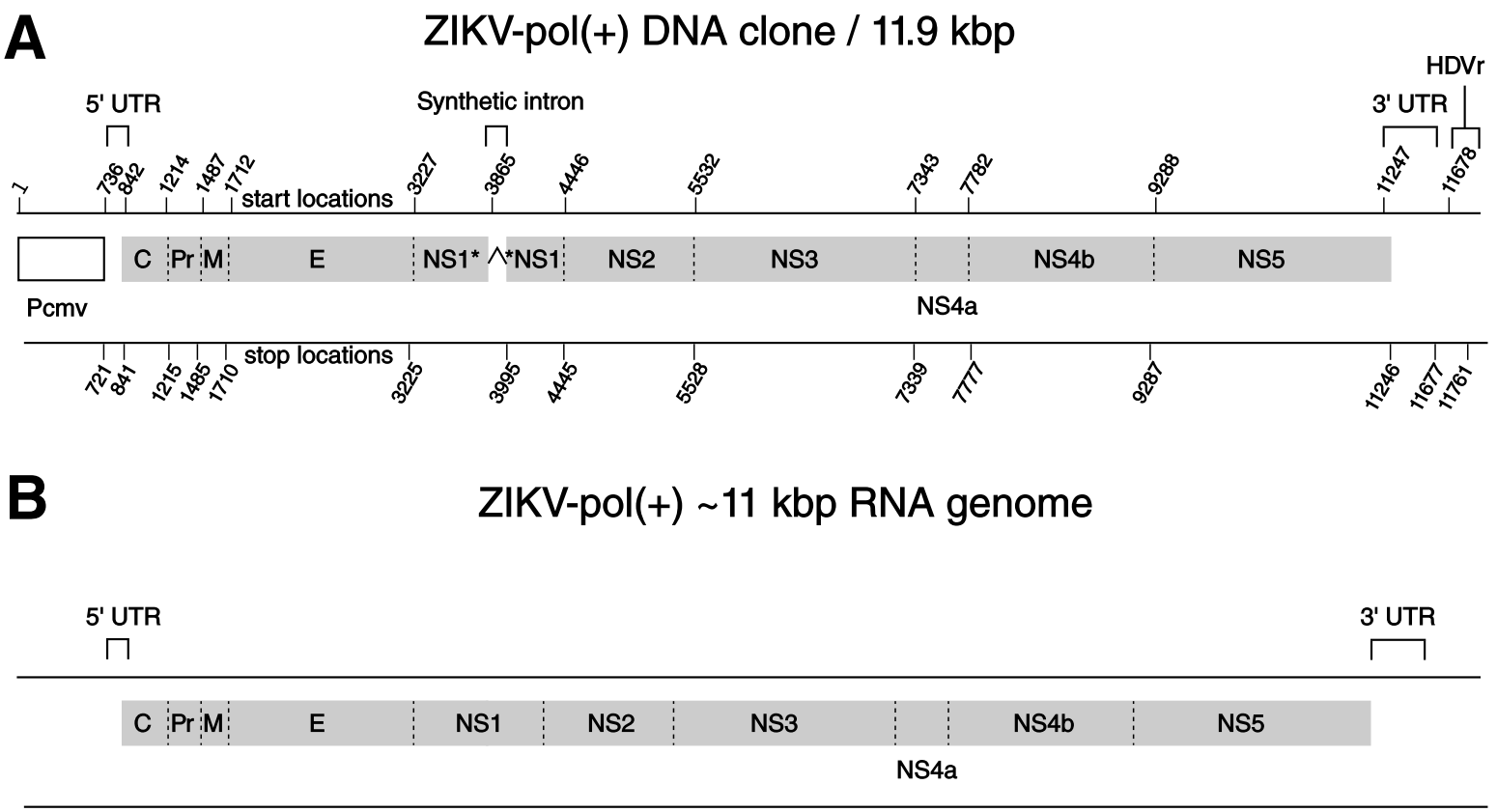


Figure 5.1: pMR766, a ZIKV molecular clone. (A) Depiction of the pMR766 plasmid from [61]. Upon transfection of permissive cells, the MR766 genome is transcribed from the CMV IE2 promoter at the 5′ end of the genome. At the 3′ end, transcription is terminated by a Hepatitis Delta virus ribozyme (HDVr), which cleaves itself out of the transcript to generate an authentic 3′ end. A synthetic intron is situated in NS1 to enhance the stability of genome during bacteria replication. (B) After transcription and post-transcriptional splicing, the MR-766 genome is recovered, now without introns and the hepatitis Delta ribozyme.

5.2 Materials and Methods

Tissue culture

293T (synonyms: HEK 239T, 293tsA1609neo) were obtained from the American Type Culture Collection (ATCC, # CRL-3216). Vero cells (African green monkey [*Cercopithecus aethiops*] kidney cells) were also obtained from the ATCC (# CCL-81). Both cell lines were propagated in **D10** media as described in Section 4.2 on page 93. Cells were cultivated in polystyrene dishes in a humidified incubator at 37 °C with 5% CO_2. Subcultures were prepared by removing media, washing once with DPBS-CMF, treating with 0.25% Trypsin/2.21 mM EDTA in HBSS for 2–5 min at 37 °C, then quenched by adding an equal volume of **D10**. Cell cultures were not tested for mycoplasma or verified for authenticity. Lines were maintained for less than 30 days in continuous culture before a new seed culture was used (maximum of 30 divisions / 10 passages).

Production of viral stocks by transfection

A molecular clone of ZIKV strain MR766 was a generous gift of Matthew Evans [61]. The molecular clone described in Chapter 2, which we term pMR766(+), is a ZIKV cDNA driven by a CMV IE2 promoter and terminated by a Hepatitis Delta Ribozyme sequence to generate an authentic 5′ end. A β-globin intron is inserted into NS1 but spliced out post-transcription to generate an authentic genome. The interruption is necessary for growing the molecular clone in bacteria, as a cryptic promoter within NS1 makes a non-interrupted NS1 unstable in bacteria [70]. A virus pool was obtained by co-transfection of 293T with pMR766(+) and one of the four MR-766 deletion libraries prepared in Chapter 2: pMR766(+)Δ_1, pMR766(+)Δ_2, pMR766(−)Δ_1, or pMR766(−)Δ_2.

On the day of transfection, a suspension of 293T was obtained by trypsinization of sub-

confluent 15 cm plates of 293T and brought into single cell suspension by gentle passage through a 40 μm mesh filter (Corning). A cell count was obtained with an automated Coulter cell counter (Moxi, ORFLO), and cells were diluted to a concentration of $5 \cdot 10^5$ cells/ml in **D10**. Thirty-six ml of this suspension ($1.8 \cdot 10^7$ cells) were added to 15-cm polystyrene dishes. Next, 18 μg of pMR766(+), 18 μg of pMR766 deletion library, and 108 μl of 1 μg/μl polyethyleneimine (Polysciences, linear PEI, 25 kDa) were added to serum-free DMEM supplemented with 25 mM HEPES and the volume brought to 3.6 ml with extra serum-free DMEM with 25 mM HEPES, incubated at bench temperature (24 °C) for 15 min, then added to 15-cm dishes. The dishes were gently rocked to mix, then placed in a 37 °C/5% CO_2 incubator. PEI-containing media was replaced after an overnight incubation (16–20 h) with fresh **D10**, and virus was harvested at 72 hours post-transfection by brief centrifugation, then passage through 0.45 μm sterile filters (Millipore).

Concentration of viral stocks

Virus stocks were concentrated by ultrafiltration. Clarified supernatant was added to a 100 kDa MWCO filtration device in 20 ml aliquots. The device was spun at $1200 \times g$ for 20–30 min until the concentrate volume was less than 1 ml. The flowthrough fraction was removed and an additional 20 ml were added to the upper reservoir and the process repeated. Generally, clarified supernatant was concentrated 20–40× (20–40 ml original to 1 ml of concentrate).

Concentrated stocks were adjusted to 20% (v/v) FBS and 10 mM HEPES (to reduce loss in infectivity from freeze-thawing) then frozen in single-use aliquots at −80 °C.

Titration of viral stocks

ZIKV stocks were titrated by plaque assay [135]. On the day before infection, Vero cells were seeded in 6-well or 12-plates at approximate 50% confluency. On the day of infection

(0 dpi), serial 10-fold dilutions of sample stocks were prepared by dilution in DMEM supplemented with 3% (v/v) heat-inactivated FBS. The media from each well of the infection plate was gently removed and replaced with 100 μl (12-well plates) or 200 μl (6-well plates) of serially-diluted inoculum. The plate was gently rocked and returned to the incubator for a period of 1 hour, with gently rocking applied every 15 minutes. After one hour of adsorption, the inocula were removed and the cultures overlaid with a viscous solution of 1% (w/v) carboxymethylcellulose in DMEM-F12. This solution was prepared by mixing equal volumes of 2% (m/v) carboxymethylcellulose (Sigma #C4888) in water and 2× DMEM-F12 supplemented with pen/strep and 8% heat-inactivated FBS.were added. The infection plates were returned to the incubator and left undisturbed for 5 days. At 5 dpi, the wells were fixed by addition of 0.1 volume of 20% formaldehyde and mixed gently for 1 hour. The supernatant was carefully removed and the culture stained with a solution of 1% crystal violet in 20% ethanol for 15 minutes. The well was destained by rinsing with deionized water. Plaques were 1-2 mm in diameter and could be visualized as clear circular patches on the stained purple monolayer.

High-MOI passage scheme

On day 0 (0 hpi), Vero cells were infected at an MOI of 16–30 with the virus pool containing WT ZIKV and tagged ZIKV deletion libraries. The inoculum was applied in a low volume in a 6-well plate for 1 hour, then removed. Supernatant was collected at 1, 2, 3 dpi and saved for future experiments. This corresponds to one passage.

Virus from passage 1 was titrated by plaque assay and used to infect Vero cells for an additional passage. The passage scheme was conducted with 2 biological replicates.

Preparation of sequencing libraries

Viral RNA was isolated from frozen aliquots of the concentrated virus pool at various time points (293T, passage 1, passage 2, using a QIAmp Viral RNA Mini Kit (Qiagen) per the manufacturer's instructions with one exception: 1) carrier RNA was $5 \cdot 10^6$ copies of bacteriophage MS2 RNA (Roche) were spiked in per isolation.

Purified RNA was reverse-transcribed with MuLV-R (New England Biolabs) and Random Primer Mix (New England Biolabs). cDNA was used as template in real-time qPCR to quantitate barcode cassette concentrations with oligonucleotides oBC20v2-F (`CCGTCCATGAAGCGTTCGAT`) and oBC20v2-R (`ACGAATCTGCCGTTGCCATA`) and compared to a standard curve prepared with dilutions of a deletion mutant plasmid containing both a barcode cassette and the ZIKV C region. oBC20v1-T:

`CCGTCCATGAAGGGTTCGATNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT`,

in 10 μl reactions with Fast SYBR Green Master Mix (Thermo Fisher). Levels of total ZIKV RNA were estimated with oligos specific to the Capsid (C) protein of MR766: fwd oligo oMR766-C-F (`TTGGTCATGATACTGCTGATTGC`) and rev oligo oMR766-C-R(`CCTCCATGTTCCAAGACAAC`). Samples were normalized for recovery by determining levels of MS2 RNA recovered by oligos oMS2-F(`TCCTGCTCAACTTCCTGTCGAG`) and oMS2-R (`CAGGTCAAACCTCCTAGGAATG`) (sequences from [113]). Samples were DNAseI-treated before reverse transcription, and levels of background DNA were acceptably low (-RT controls had barcode levels of <1/1000x of +RT reactions).

Illumina sequencing libraries were prepared by a modification of a method specified in [114] and detailed in Figure 5.2. Barcode cassettes were amplified from cDNA from above using a minimum number of cycles (typically 14-30) to prevent overamplification (post log-phase PCR) as evidenced by the RT-qPCR data from above. Illumina adaptors were added

by two rounds of PCR (5 cycles each), to add phasing adaptors, random barcodes, and multiplexing barcodes. Sublibraries were size-selected on 5% TBE polyacrylamide gels and pooled for sequencing.

Figure 5.2: Preparation of Illumina sequencing libraries from PCR of barcode cassettes, version 2. Three consecutive PCR reactions were used to add adapter sequences compatible with the TruSeq sequencing system. Introduction of molecular barcodes (orange) and multiplexing barcodes (green), allowed for high-throughput quantification of the barcoded libraries.

5.3 Results

Virus pools containing wildtype ZIKV (strain MR-766) and MR-766 deletion mutants (MR766(+)Δ_1, MR766(+)Δ_2, MR766(−)Δ_1, and MR766(−)Δ_2) was prepared by co-transfection of 293T with equal masses of the each deletion library prepared in Chapter 02 and the wildtype molecular clone (pMR766(+)). The virus-containing supernatant was collected after 72 hours, clarified by 0.45 μm filtration, then concentrated by ultrafiltration. The concentrated stock was titrated by plaque assay and used to infect Vero cells at high multiplicity of infection (MOI of > 16). The virus pool was passaged twice in Vero cells (P1,P2) and fractions of the virus pool saved at each passage. The passage was performed in duplicate alongside a control infection that only contained wildtype virus (no deletion libraries). We tested a range of MOI (1–30) and also varied harvest timing (1,2,3 dpi) to optimize yields between 1–30.

Viral RNA was purified from concentrated viral stocks and cell pellets at various points during the high MOI passage. During purification of viral RNA, $5 \cdot 10^6$ copies of a heterologous sequence (bacteriophage MS2) were spiked in to each supernatant sample to normalize recoveries and to serve as a control. For isolation of intracellular RNA, β-actin was used as a control template.

Barcodes are maintained throughout the passage experiment

Levels of barcode cassette (BC) and ZIKV Capsid (ZIKV-C) were determined by RT-qPCR. Genomes containing a tagged deletion will harbor BC and genomes that retain the wildtype capsid sequence (all wildtype ZIKV and mutants that do not have a deletion in Capsid). Thus BC signal correlates with the concentration of tagged deletion mutants and ZIKV-C as a measure of total viral genome concentration (provided that deletion mutants of ZIKV-C are in the minority).

We first isolated intracellular RNA from 293T that were transfected with pMR766(+) (WT ZIKV) and one of the four ZIKV deletion libraries: pMR766(+)Δ_1, pMR766(+)Δ_2, pMR766(−)Δ_1, and pMR766(−)Δ_2. Total RNA was isolated and treated with DNAseI before reverse transcription.

RT-qPCR results from 293T RNA are listed in below in Table 5.1. BC was only detected when a deletion library was included in the transfection, and BC and ZIKV-C were not detectable in naïve 293T. Levels of BC and ZIKV-C were comparable between the 4 deletion libraries. The BC primers form primer dimers, and have a SYBR Green C_q of ≈ 30. Authentic products are distinguishable from primer dimers from a melt curve or by running out on a 5% polyacrylamide gel.

		$C_q \pm$ STDEV ($n = 3$)		
DNA1	DNA2	BC	ZIKV-C	ACT_B
pMR766(+)	—	29.8 ± 0.4	11.5 ± 0.1	14.0 ± 0.2
pMR766(+)	pMR766(+)Δ_1	15.2 ± 0.1	11.6 ± 0.1	13.7 ± 0.2
pMR766(+)	pMR766(+)Δ_2	14.2 ± 0.1	11.7 ± 0.2	13.2 ± 0.3
pMR766(+)	pMR766(−)Δ_1	15.3 ± 0.1	11.4 ± 0.1	13.8 ± 0.1
pMR766(+)	pMR766(−)Δ_2	14.2 ± 0.1	11.3 ± 0.1	13.5 ± 0.1
pooled, -RT		26.3 ± 0.4	29.9 ± 0.7	$24 \pm NA$
pooled, +RT		15.2 ± 0.1	11.8 ± 0.1	13.8 ± 0.4
water, +RT		29.1 ± 0.4	33 ± 6	31 ± 10
NTC		28.7 ± 0.1	35.5 ± 0.7	34 ± 2.7

NTC: no template control
−/+RT: ± reverse transcriptase controls

Table 5.1: Deletion mutants and ZIKV genomes are highly expressed in transfected 293T. RT-qPCR was performed on total RNA collected from 293T co-transfected with the wildtype plasmid (pMR766(+)) and one of the four deletion libraries: pMR766Δ. Values are reported as cycles of quantification for 3 primer sets: tagged deletion mutants (BC), ZIKV genomes with intact capsid (C), and a beta-actin, a cellular housekeeping gene.

				$C_q \pm$ STDEV ($n = 3$)		
dpi	wt	Δ	MOI	BC	ZIKV-C	MS2
1	−	−	–	30.0 ± 0.3	32.6 ± 0.3	17.8 ± 0.2
1	+	−	1	30.0 ± 0.3	19.0 ± 0.1	17.7 ± 0.3
1	+	−	4	29.8 ± 0.2	17.3 ± 0.1	17.8 ± 0.1
1	+	−	16	29.6 ± 0.1	15.1 ± 0.1	16.7 ± 0.2
1	+	+	1	28.8 ± 0.3	17.8 ± 0.1	17.3 ± 0.2
1	+	+	4	27.6 ± 0.2	16.1 ± 0.1	16.9 ± 0.1
1	+	+	16	27.0 ± 0.1	15.9 ± 0.1	17.8 ± 2.1
1	+	+	30	26.3 ± 0.4	15.6 ± 0.1	17.5 ± 0.4
2	−	−	–	29.6 ± 0.2	37.2 ± 1.3	17.7 ± 0.4
2	+	−	1	30.1 ± 0.4	15.1 ± 0.1	18.7 ± 0.1
2	+	−	4	29.7 ± 0.3	13.9 ± 0.1	17.6 ± 0.3
2	+	−	16	29.8 ± 0.1	13.2 ± 0.1	16.6 ± 0.1
2	+	+	1	27.7 ± 0.6	14.7 ± 0.1	NA
2	+	+	4	26.0 ± 0.1	14.3 ± 0.2	17.7 ± 0.4
2	+	+	16	25.5 ± 0.2	14.1 ± 0.1	18.1 ± 0.1
2	+	+	30	24.9 ± 0.1	13.6 ± 0.1	16.9 ± 0.3
3	−	−	–	30.3 ± 0.9	> 40	17.7 ± 1.6
3	+	−	1	30.0 ± 0.5	14.5 ± 0.1	NA
3	+	−	4	30.0 ± 0.3	14.0 ± 0.1	NA
3	+	−	16	29.8 ± 0.1	14.3 ± 0.2	17.9 ± 0.2
3	+	+	1	26.4 ± 0.1	14 ± 0.1	17.6 ± 0.6
3	+	+	4	26.4 ± 0.2	14.6 ± 0.1	18.2 ± 0.2
3	+	+	16	25.4 ± 0.2	14.0 ± 0.1	17.7 ± 0.2
3	+	+	30	25.8 ± 0.4	14.5 ± 0.1	NA
pooled, -RT				29.6 ± 0.1	35.6 ± 0.2	37.2 ± 0.2
pooled, +RT				26.0 ± 0.1	14.4 ± 0.1	17.4 ± 0.2
water, + RT				29.5 ± 0.3	36 ± 2	35 ± 7
NTC				29.4 ± 0.3	37 ± 1.6	37.6 ± 0.8

Δ: −/+ pooled pMR766Δ libraries
NTC: no template control
−RT: no reverse transcriptase control

Table 5.2: Barcoded deletion mutants persist throughout high MOI passage. RT-qPCR from supernatant during P1 in Vero cells reveals presence of barcode cassettes (marking deletion mutants), especially at the highest MOI infections.

RT-qPCR results for supernatant samples (extracellular virus) are listed in Table 5.2 as quantification cycle (C_q) values. Viral RNA was isolated and treated with DNaseI to remove any contaminating DNA, and levels of BC and ZIKV-C were determined by RT-qPCR.

In mock-infected Vero, levels of BC and ZIKV-C were present at background levels. In infections with only the wildtype (MR766(+)) virus, levels of ZIKV-C were detectable at 1 dpi and continued to increase over 2–3 dpi, but BC levels were at background. In co-infection experiments (MR766(+) and MR766Δ pool), BC levels were detectable above background and continued to increase over 2–3 dpi. Levels of both BC and ZIKV-C were proportional to the infection MOI.

By performing qPCR with a plasmid standard that harbored exactly one copy of ZIKV-C and one copy of BC, we were able to compute a relation between the template copy number and the C_q of ZIKV-C and BC. We found the primer efficiencies were identical and the qPCR amplicons were of similar size. Therefore, if ZIKV-C and BC are present at the same concentration, then $\Delta C_q = 0$, where $C_q = C_q(\text{BC}) - C_q(\text{ZIKV-C})$. Thus, if $\Delta C_q = 0$, then all genomes are tagged (and therefore deletion mutants), and as ΔC_q increases, the fraction of genomes harboring a BC tag goes to 0.

We used this relation to calculate ΔC_q at early (293T transfection) and late passage timepoints (passage 1). In co-transfections (intracellular RNA), ΔC_q was in the range of 3–4, so roughly 1 in 8 to 1 in 16 of ZIKV genomes were tagged.

After passage 1, the ΔC_q widened to about 10–12, indicating that roughly 1:1000 to 1:10,000 genomes were tagged, implying a strong selective process (bottleneck) between transcription and encapsidation in 293T.

Identification of *cis*-acting elements

Using the C_q values from Table 5.1 and Table 5.2, we prepared Illumina sequencing libraries to characterize which barcodes persisted throughout high MOI passage of the deletion libraries in Vero cells. Barcode cassettes were amplified exponentially for a minimum number of cycles, then multiplexing indices and Illumina TruSeq adapters were added by two additional rounds of PCR. From RT-qPCR, we established that the number of template molecules for each PCR reaction was $> 10^5$. Using the barcode-to-deletion mapping for the $\approx 90,000$ mappable deletion mutants in the pooled ZIKV deletion libraries (prepared in Chapter 2), the count of each barcode cassette (with a known deletion locus), was tabulated, and used to compute the prevalence, f_i, of each barcode cassette i in each sample.

Using the methods detailed in Chapter 3 and Chapter 4, we constructed deletion depth profiles of the ZIKV library at several crucial points in the passage experiment: plasmid library, intracellular 293T, extracellular 293T, after passage 1, after passage 2. Persistent barcodes allowed us to identify which regions of the genome tolerate deletion (*trans*-acting elements) and which cannot (*cis*-acting elements).

We began by constructing a deletion depth profile of the plasmid deletion library, shown in Figure 5.3. Coverage across the ZIKV genome is relatively even (within 3×), with a single peak centered at NS1 and tails sloping gently downwards to each end of the genome.

There are two reasons for the shape of the profile. The average deletion size is ≈ 6 kbp (see Table 2.2 on page 63) and the plasmid backbone is situated at each end of the genome: deletion there are not tolerated during outgrowth of the plasmid pool in bacteria, thus the ends of the genome are somewhat underrepresented. To add to this, protein NS1 is known to harbor intrinsic promoter activity in bacteria and expression of ZIKV proteins $5'$ to this region is known to be toxic to bacteria [70]. Thus, bacteria which harbor plasmids with deletions in NS1 will tend to grow faster than their counterparts with an intact NS1, and

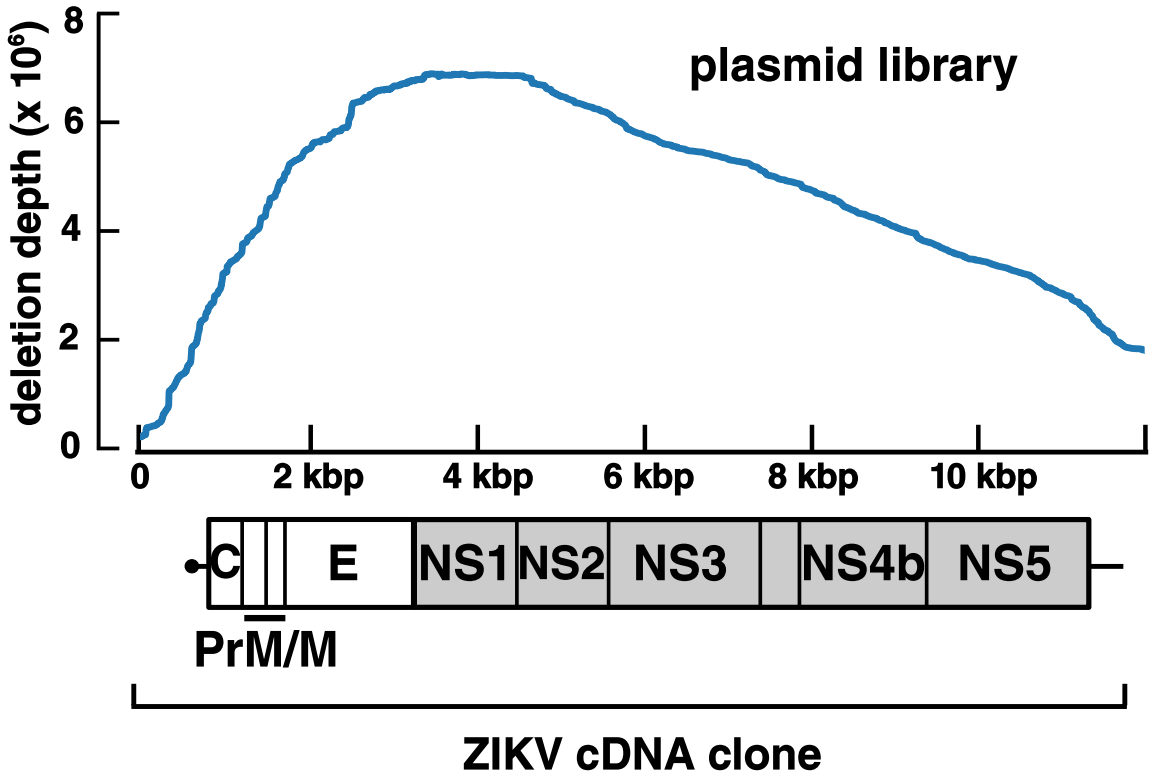


Figure 5.3: The ZIKV genome is well-represented in the pMR766(+)Δ_2 library, with some bias. Each base of the ZIKV genome is covered by several hundred different deletion mutants, with preference towards mutants that harbor deletions in NS1, which contains the cryptic promoter sequence, reflecting growth biases in the polyclonal bacterial library which harbors the plasmid library.

become overrepresented. As shown in Figure 5.3, deletions in NS1 are enriched for in the plasmid deletion library. Flavivirus NS1 is known to have cryptic promoter activity, and encodes a downstream polypeptide that is toxic in *E. coli* [61, 70]. Thus, plasmid deletion mutants that have a deletion in NS1 have a selective advantage over slower growing intact plasmids.

cis-acting elements required for transcription from a ZIKV molecular clone

A pool of infectious ZIKV containing both wildtype and deletion mutant genomes was prepared by transfecting 293T with pMR766(+) (wildtype) and pMR766 deletion libraries. During transfection, the ZIKV cDNA is transcribed from the CMV IE2 promoter to generate a poly(A) RNA, which is then self-cleaved by the encoded Hepatitis delta ribozyme at the 3′ end and 5′ capped by cellular enzymes.

We plot the deletion depth profile of the total RNA fraction isolated from co-transfected 293T in Figure 5.4. The basic profile is roughly similar to the plasmid library profile with several notable exceptions. At the 5′ end of the genome, we note that deletions of the internal CMV promoter (bases 1–721) inhibit transcription of the genome, a sensible result in this reverse genetics sytem, where the CMV promoter is required for production of the ZIKV RNA genome. At the 3′ end of the genome, deletions of the HDV ribozyme and poly(A) sequence are also inhibitory with respect to the plasmid library. Within the authentic ZIKV genome, we do not observe any regions of zero deletion depth, indicating that deletions here to affect transcription from the CMV promoter in the molecular clone.

Structural information from deletion libraries

After viewing the deletion depth profile in Figure 5.4, we noted two regions where the slope of the deletion depth profile changed dramatically from nearby regions, indicating that deletions of the area spanned had a significant effect on RNA levels. Region A is found in the ZIKV envelope protein, E, (nt 2460–2529) and Region B is at the N-terminus of NS5 (NS4 (8250–9229) and NS5 (9274–9853)).

In the literature, known flaviviral *cis*-acting elements are restricted to the 5′ UTR and

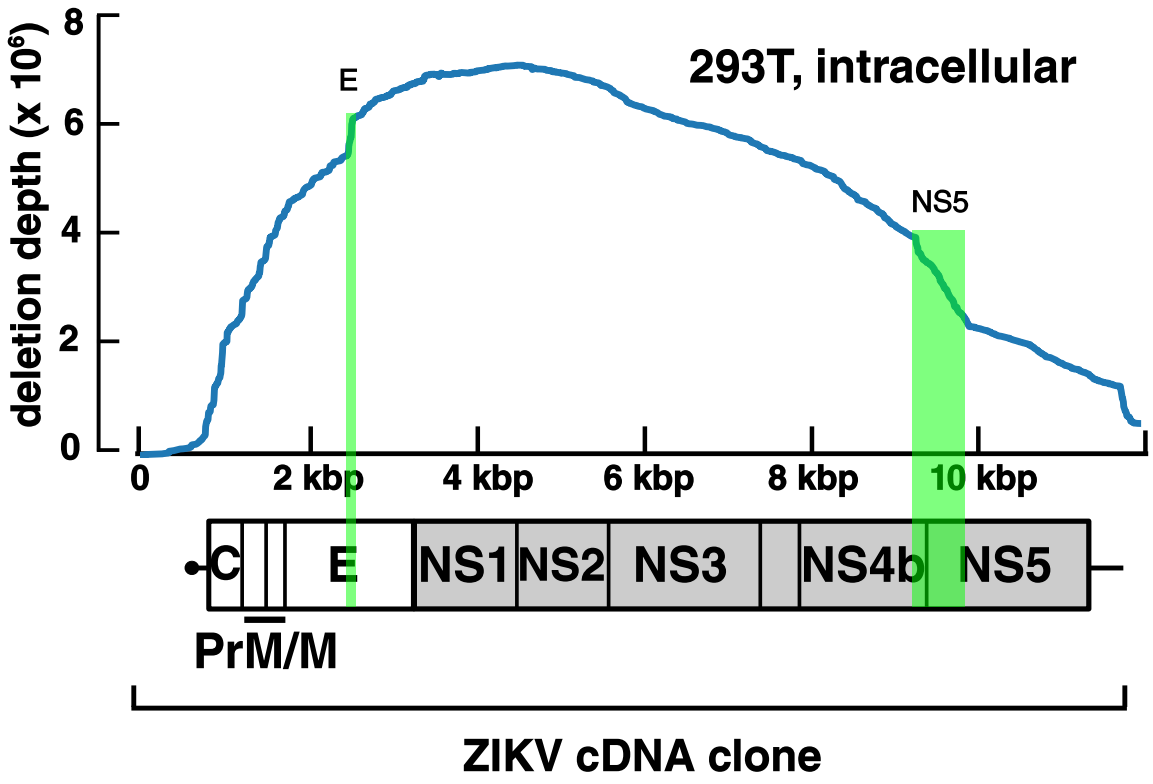


Figure 5.4: Only deletions in the 5′ and 3′ UTRs strongly affect ZIKV genome transcription from the plasmid molecular clone. Deletion depth profile of the pMR766(+)Δ_2 library obtained from deep sequencing of barcode cassettes isolated from intracellular RNA of 293T co-transfected with the wildtype virus plasmid and the pooled deletion library Deletions are well-tolerated apart from the CMV promoter / 5′ UTR and regions within the 3′ UTR (the HDVr site). The sudden rise/fall of the deletion depth profile in E and NS5 map to an α-helix in E ([131]) and the methyltransferase domain of NS5 ([134]), based on available crystal structures.

3′ UTR, and region A and region B fall within the interval coding sequence, well inside the UTRs.

Curious about the implications, we used available structures for ZIKV E protein and ZIKV NS5 to determine if these genomic regions clearly mapped to any known protein domains.

In [131], the authors report a crystal structure of ZIKV E (Protein Data Base accession number `5JHM`). Using this structure, we see that nt 2460–2529 correspond to amino acids 250–273 of ZIKV-E, which is an α-helix within the finger-like domain II of the dimerization interface. Deletions here have a positive effect on barcode survival by an unknown mechanism.

In [134], we see that NS5 is composed of two functional domains, an N-terminal methyltransferase (aa 1-265), an RNA-dependent RNA polymerase (aa 275-903), joined by a flexible linker from aa 265–274. Region B clearly maps to the N-terminal methyltransferase of NS5, and deletions here are inhibitory to RNA levels in the cell.

Frameshifting

The ZIKV coding sequence is encoded as single polypeptide, which is cleaved after translation by host- and virus-derived proteases into a series of peptides that correspond to the individual proteins. Thus if the combination of truncation during chewback and insertion of the 62 b molecular barcode yield a change in coding sequence size that is not a multiple of 3, a reading frame shift will occur. This has the effect of introducing missense and nonsense mutations downstream of the tagged deletion locus.

We examined the top 21 most prevalent barcodes after two passages and observed that 18 of the 21 did not cause a reading frame shift and 3 of the 21 caused a (+1) reading frame shift. If mutations occurred randomly, we would expect 7 to cause no-shift, 7 to cause +1,

and 7 to cause +2 frameshifts. Testing these two populations by a χ^2 test, we find that the difference is significant at the $\alpha = 0.001$ level ($p < 10^{-4}$, df= 2, $\chi^2 = 26.571$). The three mutations that appear to cause frameshift within our region of sequencing (within 100 bp of barcode loci) may have acquired insertions of deletions outside of the region of tagging that restored an intact reading frame.

cis-acting elements of ZIKV

We next performed two high MOI *in vitro* passages of the pooled MR766 deletion libraries in Vero cells. Deletion mutants which persisted through two MOI passages must be either self-replicating (replication competent) or can be efficiently *trans*-complemented by the wildtype helper virus.

Although the pool of viral deletion mutants was composed of four sublibraries, by the end of passage 2, the vast of the observed barcodes ($\approx 95\%$) were derived from the MR766(+)Δ_2 library, with the remaining amount ($\approx 5\%$ from the MR766(+)Δ_1 library). Barcodes derived from the RDRP deficient parent virus (MR766(−)Δ_1 and MR766(−)Δ_2) were infrequently observed ($< 1\%$) , despite the only differences between the MR766(+) and MR766(-) deletion libraries being slight (see Figure 2.16 and Table 2.2), with similar average deletion sizes and diversities.

One possible reason is that ZIKV genome replication cannot be rescued by supplying NS2-NS5 *in trans*. In the closely related Dengue virus, proteins NS2-NS5 form a replication complex that assembles within from membranous compartments formed from the endoplasmic reticulum [27]. Indeed, in studies of related flaviviruses (West Nile Virus and Yellow Fever Virus), we see that NS5 is not efficiently complemented *in trans*.

A representative deletion depth of profile of the MR766(+)Δ_2 library is shown in Figure 5.5, and is composed of roughly 200-300 of the 40k genomes ($< 1\%$). The deletion depth

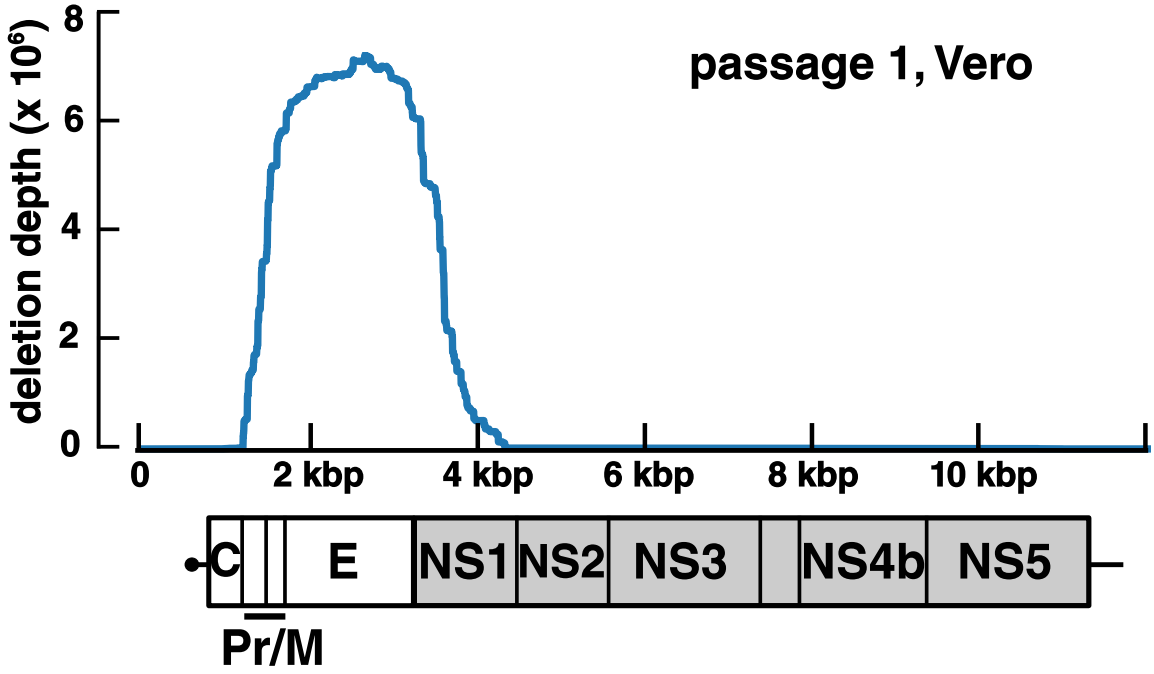


Figure 5.5: Only deletions in Pr–NS1 can be *trans*-complemented by wildtype ZIKV. After one passage in Vero cells, the deletion depth is only positive for a region spanning Pr, M, E, and NS1, suggesting this region behaves as a *trans*-acting element. Deletions of the regions 5′ and 3′ of this are region are not tolerated, suggesting that they are *cis*-acting elements and cannot be complemented *in trans*.

profile shows a single peak, centered at ZIKV-E, that slopes downward to a deletion-depth of zero beyond the 5′ border of PrM and the 3′ border of NS1. Thus, the genome region of PrM-NS1 can tolerate deletion (*trans*-acting element), but regions flaking this are *cis*-acting: (a) 5′ UTR and (b) NS2-3′ UTR.

Importantly, Pr, M, and E are 3 of the 4 structural proteins that comprise the viral particle (C is the last). Deletions of these proteins will prevent the formation of a complete, infectious virions, and render the mutant replication defective.

Using the barcoded library, we observed several mutants that increased 200–500× in prevalence within 2 passages over their counterparts. Some of these had 1-2 kbp deletions

that spanned C–NS1. Thus deletion mutants of Pr–NS1 can be complemented *in trans* and may serve as potential TIPs or replicons.

5.4 Discussion

In this chapter, we used deletion libraries of $> 90,000$ barcoded ZIKV deletion mutants to uncover key *cis*- and *trans*-acting elements of a ZIKV clone, based of an African isolate of ZIKV. After two high MOI passages, we profiled the sequencing library to reveal basic mechanisms of viral replication and identify key *cis*- and *trans*-acting elements.

We find that only a 3 kbp single genomic interval of MR766 can tolerate deletion and be efficiently complemented *in trans*: a region beginning exactly at Pr and ending precisely at the end of NS1. The remaining portions of the genome are functional *cis*-acting elements. We identify two *cis*-acting elements: (1) $5'$ UTR though C, (2) NS2 through the $3'$. A model built using these results is displayed in Figure 5.6.

Two additional pieces of evidence support the requirement for NS2–NS5. We show that deletions which induce frameshift of the common open reading frame do not persist, confirming a basic mechanism of flaviviral replication. In addition, Pol(−) mutants were not efficiently mobilized, confirming that NS5 cannot be efficiently complemented *in trans*. In vitro studies show that NS5 specifically binds to the $5'$ UTR in Dengue [136]. Taken together, these results suggests a model where Pr, M, E, NS1 can be provided *in trans*, but not C, NS2A, NS2B, NS3, NS4A, NS4B, and NS5. If translation could be initiated at more than one site (e.g., IRES), than frameshifting would be restricted to certain genomic interval.

These results are both agree and disagree with the existing flaviviral literature. Flaviviral replicons for Hepatitis C (HCV), Yellow Fever Virus (YFV), and Dengue Virus (DENV) exist and commonly have deletions in M and E, and often a portion of C [129]. In DENV, a close relative of ZIKV, there is a report of a mutant with a premature stop codon in E, and this

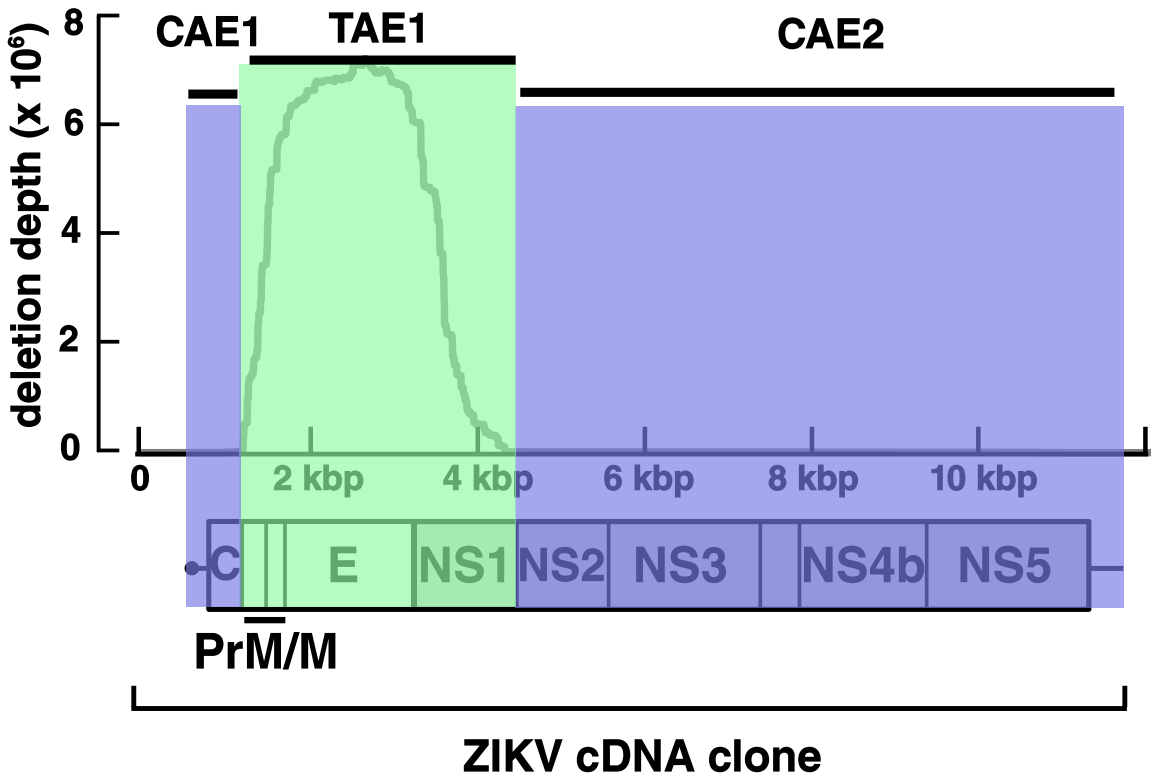


Figure 5.6: A model of *cis*- and *trans*-acting elements of ZIKV. The two *cis*-acting elements are highlighted in blue and cannot be complemented *in trans*. The *trans*-acting element is highlighted in green and can be complemented *in trans*. Deletions are tolerated within structural proteins Pr, M, E and nonstructural protein NS1. Structural protein C and nonstructural NS2–NS5 are intolerant to deletion.

mutant appears to be transmitted over hundreds of kilometers in mosquitoes and humans in Myanmar within a single rainy season [127].

There are reports in the literature of minimal DENV interfering particles comprised of the $5'UTR$ and $3'$ UTR of the virus, and likely appear in natural and *in vitro* infections as byproducts of viral genome replication when genome circularization occurs. Yet, we find no evidence for this, as the replication complex (NS2–NS5) must be maintained *in cis*. This is consistent with *trans*-complementation studies in other flaviviruses, in particular, Kunjin [137].

There are a few caveats in our study. We performed passage with a single viral species, in a single cell line. Natural ZIKV infections involve cross-species transmission between mammals and mosquitoes (*A. aegypti* and *A. albopictus*), and there is some evidence that *cis*-acting elements may differ between species [138]. Moreover, different ZIKV strains, although closely related, do have varying infection phenotypes, including variations in eliciting host immune response [139]. Thus, performing the high MOI experiment using different ZIKV strains, and a *Aedes* mosquito cell line may be necessary, such as C6/36. Animals studies may also be informative.

As a note in proof, the results obtained in this chapter are consistent with a recently published study of mutational tolerance in ZIKV, using the same strain of virus [58]. Here, the authors used a transposon system to introduce 5 aa insertions (in-frame) throughout the genome, and screened for mutants that were replication-competent. Insertions in NS2-NS5 were not tolerated, except at the regions proximal to the protein cleavage sites.

Chapter 6

Construction of prototype TIPs for HIV-1

6.1 Introduction

As detailed in Chapter 1, an ideal transmissible antiviral (a DIP/TIP) would possess the following qualities: (1) it would compete effectively for common goods provided by the wildtype helper virus during co-infection (interference) (2) it would be retain all necessary *cis*-acting elements required for efficient mobilization by the helper virus (mobilization, $R_0 > 1$), (3) it would be unable to transmit without co-infection with wildtype helper virus (no self-replication).

We reasoned that we would be able to use the results of the high MOI passage screen for HIV-1 *cis*-acting elements in Chapter 4 to develop transmissible antivirals against HIV-1. From the deletion depth profile, we were able to identify four *cis*-acting elements that were intolerant of deletion and could not be complemented *in trans* (see Figure 4.16 on page 120). Therefore, any HIV-1 potential transmissible antiviral *must retain* these regions

in its genome. Concatenation of these four *cis*-acting elements provides the minimum proviral size of an HIV-1 transmissible antiviral, about 2.6 kbp.

The high MOI passage of the pNL4-3Δ_1 deletion library also allowed an estimation of the fitness effect of each deletion (see Figure 4.8 and Figure 4.9). While most of the 23,851 deletions were deleterious (about 95%), 1390 mutants consistently increased in prevalence within the barcoded pool, suggesting these mutations had an adaptive effect (increase in R_0). A subset of the adaptive mutations increased in prevalence by as much as 1–2 $\log_{10}$ over 12 passages.

By generating viral mutants harboring one or multiple adaptive deletions, as identified in the high MOI screen, it may be possible to assemble a library of potential TIPs. Such viruses can be further optimized by point mutations and further deletion.

In this chapter, we generate a collection of 61 HIV-1 subgenomic deletion mutants that were reconstituted from adaptive deletions identified from the high MOI screen in Chapter 4. We demonstrate that a subset of these satisfy the basic requirements of a transmissible antiviral in single-round replication studies: they interfere with HIV-1 replication, reducing wildtype viral loads; they can be mobilized by the wildtype helper virus to new cells, and they do not self-replicate.

6.2 Materials and Methods

Cloning strategy

The pNL4-3 genome (an HIV-1 molecular clone) was subdivided into 7 blocks (A, B, C, D, E, F, G). Each block is comprised of a 400–2100 bp subsequence of pNL4-3 plasmid with flanking adapter sequences to allow assembly of full-length virus genomes by Gibson Assembly via flanking 40 bp sequences, as depicted in Figure 6.1.

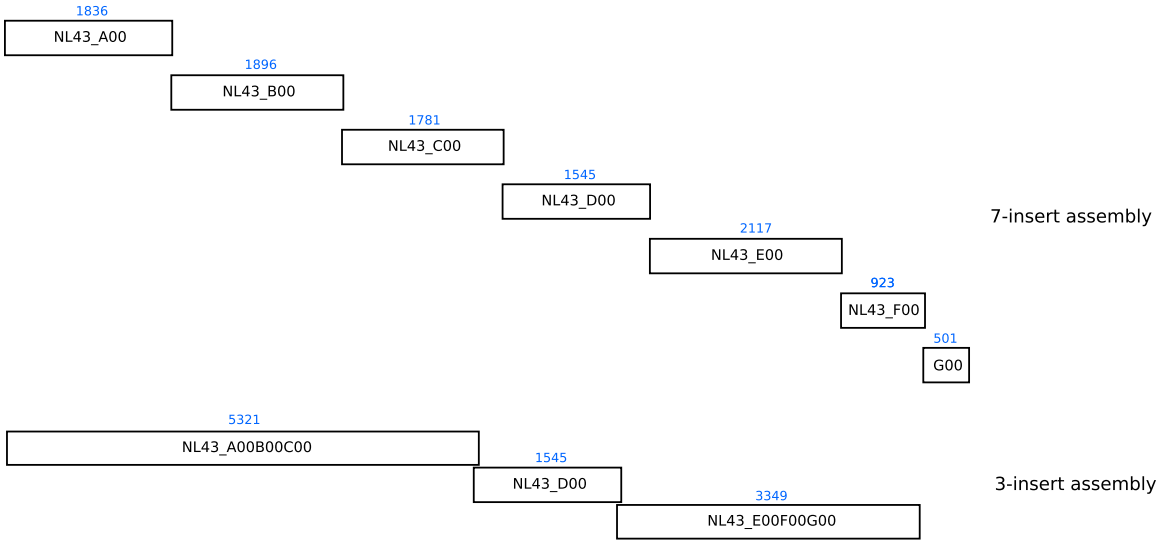


Figure 6.1: Assembly strategy to generate multiply deleted strains. New HIV-1 mutants can be created by 7-insert Gibson Assemblies (top) or 3-insert assemblies, using previously assembled ABC and EFG blocks.

Construction of wildtype single-block plasmids

All preparative PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs, #M0493L), hereafter referred to as **Q5**. All PCR was conducted at the 50 μl scale in 1× Q5 Reaction Buffer supplemented with single-thaw aliquots of dNTPs and with thermocycler parameters recommended by the **Q5** manufacturer's protocol. Oligonucleotide sequences are listed in Appendix B on page 195.

Linearized vectors were prepared by 15 cycles of PCR using 1 ng of pUC19 as template with **Q5**. pUC19Δ1 was obtained by PCR with oligos y470 and y471. pUC19Δ2 was obtained by PCR with oligos y472 and y473 that had been 5′ phosphorylated with T4 PNK (New England Biolabs). pUC19Bam was obtained by PCR with oligos y541 and y542. After the cycling was completed, each reaction was supplemented with 20 U of *Dpn*I (New England Biolabs), mixed, then incubated for 1 h at 37 °C to eliminate template DNA, then purified by a silica column cleanup with Zymo DCC-5 (Zymo Research, DCC-5).

name	size (bp)	template	fwd olig.	rev olig.
pUC19Δ1	2522	pUC19/*Bam*HI	y470	y471
pUC19Δ2	2384	pUC19/*Bam*HI	y472+P_i	y473+P_i
pUC19Bam	2684	pUC19/*Bam*HI	y541	y542
NL43_A00_inner	1788	pNL4-3/*Eco*RI	y474	y476+P_i
NL43_A00	1844	NL43_A00_inner	y475+P_i	y476+P_i
NL43_B00	1904	pNL4-3/*Eco*RI	y477+P_i	y478+P_i
NL43_C00_inner	1766	pNL4-3/*Eco*RI	y479+P_i	y480
NL43_C00	1789	NL43_C00_inner	y479+P_i	y481+P_i
NL43_D00	1633	pNL4-3/*Bam*HI	y482+P_i	y483+P_i
NL43_E00_inner	2102	pNL4-3/*Eco*RI	y484	y486+P_i
NL43_E00	2125	NL43_E00_inner	y485+P_i	y486+P_i
NL43_F00	931	pNL4-3/*Eco*RI	y487+P_i	y488+P_i
NL43_G00_inner	453	pNL4-3/*Eco*RI	y489+P_i	y490
NL43_G00	509	NL43_G00_inner	y489+P_i	y491+P_i

Table 6.1: PCR to construct wildtype NL4-3 blocks. HIV-1 blocks with wildtype NL4-3 sequence were constructed by one or two-step PCR from pNL4-3 template. Each row lists the name of the PCR amplicon, the expected size in bp, the precise template used (a forward slash indicates that the template was digested with a restriction enzyme), the forward PCR oligo, and the reverse PCR oligo. A P_i suffix indicates that the primer was 5′-phosphorylated (either during synthesis or by T4 PNK). Complete primer sequences are listed in Appendix B.

Three of the seven single-block inserts (NL43_B00, NL43_D00 and NL43_F00) were amplified by 15 cycles of PCR using 1 ng of *Eco*RI- or *Bam*HI-cut pNL4-3 as template and the oligos listed in Table 6.1. Post-thermocycling, each reaction was supplemented with 20 U of *Dpn*I (New England Biolabs), mixed, then incubated for 1 h at 37 °C to eliminate template DNA, then purified by a silica column cleanup with Zymo DCC-5 (Zymo Research, DCC-5).

The remaining four single-block inserts (NL43_A00, NL43_C00, NL43_E00, NL43_G00) were amplified by two rounds of PCR, using the oligos listed in Table 6.1. An "inner" PCR product was obtained by 15 cycles of PCR using 1 ng of *Eco*RI-cut pNL4-3 as template.

The final linear product by an additional 10 cycles of PCR, using 0.5 μl of the "inner" PCR product as template in a new 50 μl PCR reaction. Post-thermocycling, each reaction was supplemented with 20 U of *Dpn*I (New England Biolabs), mixed, then incubated for 1 h at 37 °C to eliminate template DNA, then purified by a silica column cleanup with Zymo DCC-5 (Zymo Research, DCC-5).

Single-block plasmids were constructed by TA-cloning. Linearized vectors (pUC19Δ1and pUC19Δ2) were 3′ dA-tailed by incubating 2 μg linear vector DNA, 200 μM dATP, 15 U of Klenow Fragment (exo$^-$) (New England Biolabs), in a 50 μl reaction in 1$\times$ NEB Buffer 2 for 2 hours at 37 °C. DNA was purified from the reaction by DCC-5 silica column cleanup (Zymo Research). The seven wildtype inserts (NL43_A00, NL43_B00,..., NL43_G00) were 3′ dT-tailed by incubating 2 μg insert DNA, 200 μM dTTP, 15 U of Klenow Fragment (exo$^-$) (New England Biolabs), in a 50 μl reaction in 1$\times$ NEB Buffer 2 for 2 hours at 37 °C. DNA was purified from the reaction by DCC-5 silica column cleanup (Zymo Research).

All ligations were performed using T4 DNA Ligase (New England Biolabs, Quick Ligation Kit, #M2200L). To construct pUC19Δ1-NL43_D00, 50 ng of 3′ dT-tailed NL43_D00 were ligated into 50 ng of 3′ dA-tailed pUC19Δ1 in a 20 μl reaction for 30 min at 25 °C.

To construct the remaining six single-block wildtype plasmids (pUC19Δ2-NL43_**x**, where **x** is one of {A00,B00,C00,E00,F00,G00}), 50 ng of 3′ dT-tailed NL43_**x** were ligated into 50 ng of 3′ dA-tailed pUC19Δ2 in a 20 μl reaction for 30 min at 25 °C.

Chemically competent DH10B *E. coli* were transformed with the ligation mixture, recovered for 1 hour at 37 °C, then plated on LB plates supplemented with 100 μg/ml carbenicillin and grown overnight at 37 °C. Single colonies were obtained and plasmids characterized by Sanger sequencing across the insert region and diagnostic restriction digests. All wildtype single-block plasmids are enumerated in Table 6.2.

Construction of mutant single-block plasmids

Deletion mutants pieces were ordered as synthetic dsDNA molecules (Integrated DNA Technologies, gBlocks). A total of 13 mutant blocks were ordered (7 B's (B01, B02, B03, B04, B05, B06, B07), 1 C (C01), 4 D's (D01, D02, D03, D04), and 1 F (F01)).

Mutant blocks NL43_D01, NL43_D02, NL43_D03, and NL43_D04 were cloned into linear PCR pUC19Δ1product by Gibson Assembly (Gibson Assembly HiFi 1-Step Kit, SGI-DNA) via the manufacturer's protocol. The remaining 9 mutant blocks (NL43_B01,...,NL43_F01) were cloned into linear PCR pUC19Δ2product by Gibson Assembly (Gibson Assembly HiFi 1-Step Kit, SGI-DNA). After a 1 hour incubation at 50 °C, the 10 μl Gibson Assembly reactions were diluted by adding 50 μl of dH_2O. 1 μl of the diluted mix was used to transform 40 μl of electrocompetent DH10B *E. coli*. The transformations were recovered for 1 hour at 37 °C, then plated on LB plates supplemented with 100 μg/ml carbenicillin and grown overnight at 37 °C. Single colonies were obtained and plasmids characterized by Sanger sequencing across the insert region and diagnostic restriction digests. All single-block plasmids are enumerated in Table 6.2.

Construction of 3-block plasmids

To construct plasmids harboring subassemblies of 3 blocks (ABC or EFG as shown in Figure 6.1), single blocks were liberated from their respective plasmids by *Not*Idigestion and gel-purified from plasmid backbone. To assemble ABC 3-block plasmids, a pool of liberated A,B,C blocks were combined with linear pUC19Δ1and incubated for 1 hour at 50 °C in Gibson Assembly master mix (Gibson Assembly HiFi 1-Step Kit, SGI-DNA). To assemble EFG 3-block plasmids, a pool of liberated E,F,G blocks were combined with linear pUC19Δ1and incubated for 1 hour at 50 °C in Gibson Assembly master mix (Gibson Assembly HiFi 1-Step Kit, SGI-DNA).

		post-digest		
name	size	insert	backbone	enzyme
pUC19Δ2-NL43_A00	4229	1836	2394	*Not*I
pUC19Δ2-NL43_B00	4290	1896	2394	*Not*I
pUC19Δ2-NL43_B01	3334	935	2394	*Not*I
pUC19Δ2-NL43_B02	2989	590	2394	*Not*I
pUC19Δ2-NL43_B03	2803	404	2394	*Not*I
pUC19Δ2-NL43_B04	3442	1043	2394	*Not*I
pUC19Δ2-NL43_B05	3525	1126	2394	*Not*I
pUC19Δ2-NL43_B06	3490	1091	2394	*Not*I
pUC19Δ2-NL43_B07	3554	1155	2394	*Not*I
pUC19Δ2-NL43_C00	4174	1781	2394	*Not*I
pUC19Δ2-NL43_C01	3250	857	2394	*Not*I
pUC19Δ1-NL43_D00	4076	1545	2531	*Asc*I
pUC19Δ1-NL43_D01	622	622	2531	*Asc*I
pUC19Δ1-NL43_D02	561	561	2531	*Asc*I
pUC19Δ1-NL43_D03	519	519	2531	*Asc*I
pUC19Δ1-NL43_D04	397	397	2531	*Asc*I
pUC19Δ2-NL43_E00	4510	2117	2394	*Not*I
pUC19Δ2-NL43_F00	3317	923	2394	*Not*I
pUC19Δ2-NL43_F01	3205	812	2394	*Not*I
pUC19Δ2-NL43_G00	2895	501	2394	*Not*I

Table 6.2: Single-block HIV plasmids were cloned as described above. Each row corresponds to a single-block HIV plasmid with columns describing the: name, name of the plasmid ("00" suffix is WT, all others are mutants); size, the size of the undigested plasmid in bp; insert, size of the linear block insert in bp after digestion of the complete plasmid with *Not*I or *Asc*I; backbone, size of the vector backbone in bp after digestion of the complete plasmid; enzyme, restriction endonuclease used to liberate block insert from backbone of plasmid.

After a 1 hour incubation at 50 °C, the 10 µl Gibson Assembly reactions were diluted by adding 50 µl of dH_2O. 1 µl of the diluted mix used to transform 40 µl of electrocompetent DH10B *E. coli*.

The transformations were recovered for 1 hour at 37 °C, then plated on LB plates supplemented with 100 µg/ml carbenicillin and grown overnight at 37 °C. Single colonies were obtained and plasmids characterized by Sanger sequencing across junction regions joined by Gibson Assembly and diagnostic restriction digests.

Construction of 7-block plasmids (full-length virus)

3-block inserts (ABC and EFG) were liberated from 3-block plasmids by digestion with *Asc*Iand gel-purified from the backbone. D blocks were liberated by *Asc*Idigestion of 1-block D plasmids. A three-insert Gibson assembly was performed by combining an ABC 3-block insert, a D-block insert, and a EFG 3-block insert with with pUC19Bam and incubated 1 hour at 50 °C in Gibson Assembly master mix (Gibson Assembly HiFi 1-Step Kit, SGI-DNA) per the manufacturer's instruction.

After a 1 hour incubation at 50 °C, the 10 µl Gibson Assembly reactions were diluted by adding 50 µl of dH_2O. 1 µl of the diluted mix used to transform 40 µl of electrocompetent DH10B *E. coli*.

The transformations were recovered for 90 min at 30 °C, then plated on LB plates supplemented with 25 µg/ml carbenicillin and grown overnight at 32 °C. Single colonies were obtained and plasmids characterized by Sanger sequencing across junction regions joined by Gibson Assembly and diagnostic restriction digests.

All 7-block plasmids (full-length virus) are listed in Table A.1 in Chapter A.

Production of virus and pseudovirus stocks

Virus pools were obtained by co-transfection of 293T with one or more plasmids. On the day of transfection, a suspension of 293T was obtained by trypsinization of subconfluent 15 cm plates of 293T and brought into single cell suspension by gentle passage through a 40 μm mesh filter (Corning). A cell count was obtained with an automated Coulter cell counter (Moxi, ORFLO), and cells were diluted to a concentration of $5 \cdot 10^5$ cells/ml in **D10**. Two ml of this suspension (10^6 cells) were added to each well of a series of 6-well polystyrene tissue culture plates.

Transfection complexes were prepared in 96-well polypropylene PCR plates under sterile conditions. A total of 2 μg of supercoiled plasmid DNA in (10 mM Tris-Cl, pH 8.0; 0.1 mM EDTA, pH 8.0) was added to each well and the volume brought to 100 μl by the addition of serum-free DMEM supplemented with 25 mM HEPES. Next, 106 μl of mixture comprised of 100:6 (v:v) of serum free-DMEM supplemented with 25 mM HEPES and 1 μg/μl polyethyleneimine in dH_2O (Polysciences, PEI) were added and the contents mixed by pipetting up and down 15$\times$ with the multichannel pipette volume set to 106 μl. The transfection mixture was incubated at room temperature (24 °C) for 15 min, then the contents of each well in the 96-well plate added to a corresponding well of 10^6 293T prepared above. The 6-well plates were gently rocked to distribute the transfection complexes, then placed in a humidified 37,°C incubator with a 95% air/5% CO_2 atmosphere.

At 16 h post-transfection, the culture media was aspirated and replaced with 2.5 ml of **D10**. At 43-48 hours post-transfection, the virus-containing media was harvested and clarified by passage through a 0.45 μm sterile filters (Millipore). The filtrate was immediately used to infect target cells.

Assay for replication competence

Virus stocks were prepared as described above, where the 2 μg mass of plasmid DNA in each well was comprised entirely of a single HIV-1 molecular clone or deletion mutant. An EGFP-tagged molecular clone (pNLENG1-IRES , which encodes NL4-3 with an EGFP::IRES::nef cassette in the *nef* locus was used as a positive control for replication competence [119]. We refer to this clone as NL43G for simplicity.

MT-4 [112], (a highly-permissive, highly-susceptible T cell line) were infected with the prepared viral stocks in a 96-well U-bottom plate format. All infections were performed in duplicate. In each well, 10^5 MT-4 in 150 μl of **R10** were mixed with 50 μl of virus stock by pipetting up and down 4×, then returned to a 37 °Cincubator. Cultures were split 1:10 with fresh **R10** media every 3–4 days and the wells assayed for infection visually (for cytopathic effect / cell lysis) at each split.

At 5 days post-infection, 90 μl of the cell mixture from each well was transferred to a 96-well opaque plastic plate (OptiPlate96-F, PerkinElmer) and mixed with 10 μl of Presto-Blue Cell Viability Reagent (ThermoFisher, #A13261). The plate was sealed with a gas-permeable adhesive seal (ThermoFisher, #AB0718) and incubated for 90 min at 37 °C. The reaction was quenched and virus inactivated by addition of 50 μl of 3% (m/v) SDS solution to each well. Fluorescence was read on an EnSpire Plate Reader (PerkinElmer) with the monochromator set to excitation/emission wavelengths of 560/590 nm.

Interference assay by co-transfection

Virus stocks were prepared as described above, but the 2 μg mass of plasmid DNA in each well was comprised of 1 μg of an single HIV-1 molecular clone or deletion mutant and 1 μg of pNLENG1-IRES.

MT-4 [112], (a highly-permissive, highly-susceptible T cell line) were infected with the

prepared viral stocks in a 96-well U-bottom plate format. All infections were performed in duplicate. In each well, 10^5 MT-4 in 150 μl of **R10** were mixed with 50 μl of virus stock by pipetting up and down 4×, then returned to a 37 °C incubator.

At 24 and 48 hours post-infection, cells were resuspended and a 50 μl (25% of the volume) removed for analysis by flow cytometry. The 50 μl sample was fixed by addition of 0.1 volumes of a 20% formaldehyde solution (tousimis, #1008A) and incubated for at least 1 hour at 4 °C before cytometry. Cells were scored for EGFP-expression by analyzing a portion of each sample on an HTFC IntelliCyt flow cytometer (488 nm excitation, 530/30 nm bandpass emission).

Assay for mobilization from 293T

Virus stocks were prepared as described above, but the 2 μg mass of plasmid DNA in each well was comprised of 500 ng of an single HIV-1 molecular clone or deletion mutant, 500 ng of a VSV-G pseudotyping plasmid (pMD.G from [83]), and 1000 ng of an 2nd generation lentivirus packaging plasmid (pCMVΔR8.91 from [84]).

MT-4 were transduced/infected with the prepared viral stocks in a 96-well U-bottom plate format. All infections were performed in duplicate. In each well, 10^5 MT-4 in 150 μl of **R10** were mixed with 50 μl of virus stock by pipetting up and down 4×, then returned to a 37 °C incubator. Cultures were split 1:10 with fresh **R10** media every 3–4 days and the wells assayed for infection visually (for cytopathic effect / cell lysis) at each split.

At 5 days post-infection, the transduced cells were washed 2× in 700 μl DPBS and genomic DNA isolated from the cell pellet using a commercial kit (Macherey-Nagel, NucleoSpin Blood).

PCR for block B (fwd: `GCATCCAGTGCATGCAGGGCCT`, rev: `CTGGCAGCACTATAGGCTGTACTGTCCA`) and block D (fwd: `GCTCCTCTGGAAAGGTGAAGGGGCAGTAG`, rev:`CTACTTCTTGTGGGTTGGGGTCTGTGGG`)

was performed on each gDNA sample to test for transduction/mobilization. PCR reactions were performed with OneTaq DNA Polymerase (New England Biolabs, #M0480L). Each 20 μl PCR reaction consisted of 2.0 μl template DNA, 0.4 μl 10 μM forward oligo, 0.4 μl 10 μM reverse oligo, 0.4 μl 10 mM dNTP, 4.0 μl 5$\times$ OneTaq DNA Polymerase Reaction Buffer, 0.1 μl OneTaq DNA Polymerase (5 U/μl), and 12.7 μl dH_2O. Thermocycling conditions (in mm:ss format) were (1 cycle of 95 °C for 5:00; 35 cycles of 94 °C for 0:15, 61 °C for 0:30, 68 °C for 1:30; 1 cycle of 68 °C for 5:00; 1 cycle of 10 °C for ∞).

6.3 Results

Construction of prototype TIPs of HIV-1

From the high MOI passage data in Chapter 4, we identified three *trans*-acting elements of the HIV-1 genome that were tolerant to deletion (see Figure 4.16). The three *trans*-acting elements are interleaved by four *cis*-acting elements elements that are intolerant to deletion. We reasoned that we could use this genome organization to our advantage when cloning multiply-deleted strains. As the *cis*-acting elements must be preserved, they could be used as fixed sequence "fasteners" to join variably-sized pieces of the *trans*-acting elements.

To develop a modular cloning strategy, we subdivided the 9709 bp genome into 7 blocks (A, B, C, D, E, F, G) that can be joined by Gibson Assembly to create a single provirus. By creating different version of blocks A through G, we could create a combinatorial library by building mutants composed of different versions of each common block to generate multiply deleted and mutated strains of HIV-1.

Using the high MOI data results, we selected a number of deletions that were reliably enriched (4–250$\times$) over 12 *in vitro* passages in all three replicates of the HIV-1 high MOI passage experiment (Figure 4.3). Additionally, each of the four *cis*-acting elements is encoded

by a single block (see Table 6.3).

		Δ span				
mutant	Δ (bp)	left	right	barcode	enrich. (K/L/M)×	genes **fully** or partially deleted
NL43_B01	1016	1636	2651	TTGAACCCATGGCCGGATCA	17/21/19×	*gag*/*pol*
NL43_B02	1361	1455	2815	TGGACCCGCCCGGACGCTCA	10/16/20×	*gag*/*pol*
NL43_B03	1547	1448	2994	CAGGCAGCTCAAGAAACACG	4/12/21×	*gag*/*pol*
NL43_B04	908	1469	2376	TCGACAGTGTAATGTCTGCA	27/16/19×	*gag*/*pol*
NL43_B05	825	1484	2308	GGGGGGCGGCGATTACTTGT	20/39/7×	*gag*/*pol*
NL43_B06	860	1469	2328	GAGCTGCAGGGCATGTTATA	23/20/32×	*gag*/*pol*
NL43_B07	796	1560	2355	GGCGACCCTCGGCCGTGGAC	26/16/16×	*gag*/*pol*
NL43_C01	987	3620	4606	CAGCTTGTTCAGATGCTGTA	30/36/43×	*pol*
NL43_D01	986	5073	6058	TCTTGAACAGCGCGGTCTGT	250/91/130×	*vif*/***vpr***/*tat*/*rev*
NL43_D02	986	5073	6058	D01 but without barcode	NA	*vif*/***vpr***/*tat*/*rev*
NL43_D03	1089	5071	6159	GATCGGTCGTCGCAGCGGTC	50/36/23×	*vif*/***vpr***/*tat*/*rev*/*vpu*
NL43_D04	1211	5041	6251	TATCTGTAGCCAACATTCGA	18/18/9×	*vif*/***vpr***/*tat*/*rev*/***vpu***/*env*
NL43_F01	174	9116	9289	CGTAAAGTGGGATAGTTTTT	17/17/14×	*nef*/U3

Table 6.3: Reconstructed single contiguous deletions from pNL4-3Δ_1. For each deletion, the following information are provided: block name; size of deletion in bp; location of deletion in the NL4-3 provirus genome; barcode sequence in integrated barcode cassette; enrichment from plasmid library to passage 12 virus in replicate K, L, and M; which genes are potentially inactivated by truncation. For example: NL43_B01 represents a 1016 bp deletion in *gag* and *pol* (bases 1636–2651 of pNL4-3/NL4-3 and was enriched 17-fold in replicate K, 21-fold in replicate L, and 19-fold in replicate M after 12 rounds of *in vitro* passage in MT-4.

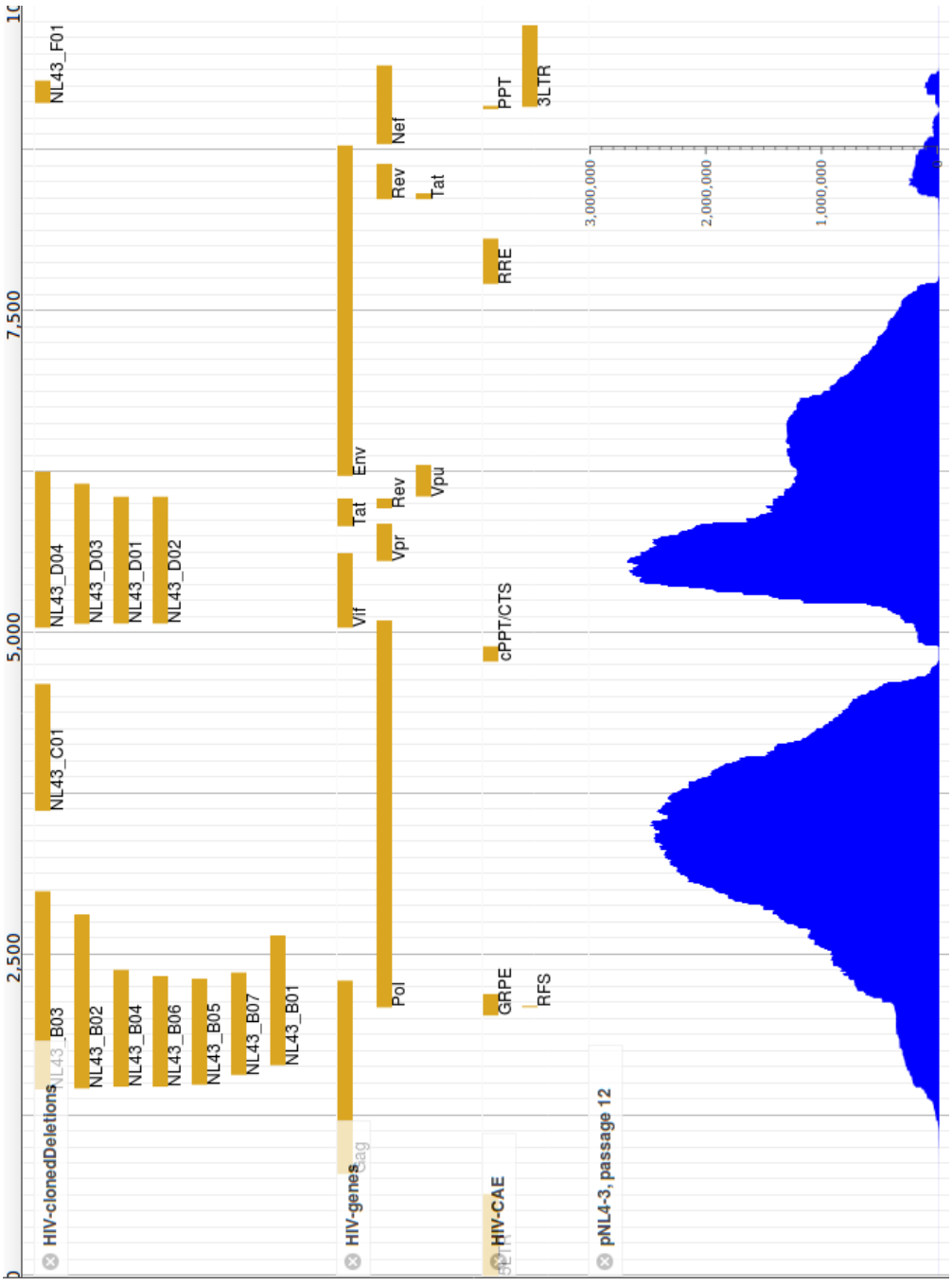


Figure 6.2: Reconstitution of adaptive single contiguous deletions of HIV-1. Information is displayed in four horizontal tracks, with horizontal numbering in bp. From top to bottom: **HIV-clonedDeletions** mapping of adaptive HIV-1 deletions to proviral genome; **HIV-genes** location of HIV-1 genes; **HIV-CAE** Putative HIV-1 *cis*-acting elements in the literature; **pNL4-3, passage 12**, deletion depth profile of the NL4-3Δ_1 library after twelve high MOI passages

In all, we selected 7 mutants from block B, 1 mutant from block C, 4 from block D, and 1 from block F. When combined with the wildtype piece from each block, we could create a total of 160 viruses (159 mutants and 1 wildtype virus). These blocks can be seen in Figure 6.2.

We use the following system to describe the genotype of each virus. For each block, "00" is the wildtype version, thus the wildtype version of NL4-3 can be written descriptively (yet verbosely) as NL43_A00B00C00D00E00F00G00. A mutant which is composed of mutant blocks B02, D01 and wildtype versions of the remaining blocks has the genotype of NL43_A00B02C00D01E00F01.

Using this combinatorial strategy, we construction 61 viruses (60 deletion mutants and 1 wildtype virus) using the assembly strategy demonstrated in Figure 6.1. A complete listing of the deletion mutants can be found in Appendix A.

Arrayed screen of prototype HIV-1 TIPs

We were interested in 3 properties of the deletion mutants. Are they replication-competent (can spread without wildtype virus)? Do they interfere with WT virus (by competing for common goods or another mechanism)? Can they be mobilized by the wildtype virus (efficiently *trans*-complemented)?

Supercoiled plasmid stocks of the mutant library were arrayed in 96-well plate format at a normalized DNA concentration (33 ng/μl). The collection of deletion mutants used the same vector backbone as the wildtype HIV-1 and had similar sizes (9–12 kbp mutants compared to 15 kbp wildtype). A related HIV-1 molecular clone (NLENG1-IRES), which we refer to as NL43G was used as positive control. NL43G is tagged with EGFP in the *nef* locus. Productively infected cells can be visualized by GFP fluorescence after approximately 20 hours of infection.

Most mutants are not replication competent

We first determined if the deletion mutants were replication-competent, by evaluating if they could infect and kill susceptible/permissive T cells. Viral stocks of deletion mutants were prepared by transfection of 293T with the arrayed library of deletion mutants and controls. At 48 hours post-transfection, the virus-containing supernatant was harvested, clarified by 0.45 μm filtration, and used to infect a highly permissive, highly susceptible, T cell line (MT-4). At 5 days post-infection, the MT-4 were assessed for viability visually and by a metabolic assay (PrestoBlue), which detects live, metabolically active cells that have a functional electron transport chain. Thus, wells infected with replication-competent virus will be mostly killed (based on live cell data), and have a low metabolic activity, which is detectable by low fluorescence signal in the plate reader assay.

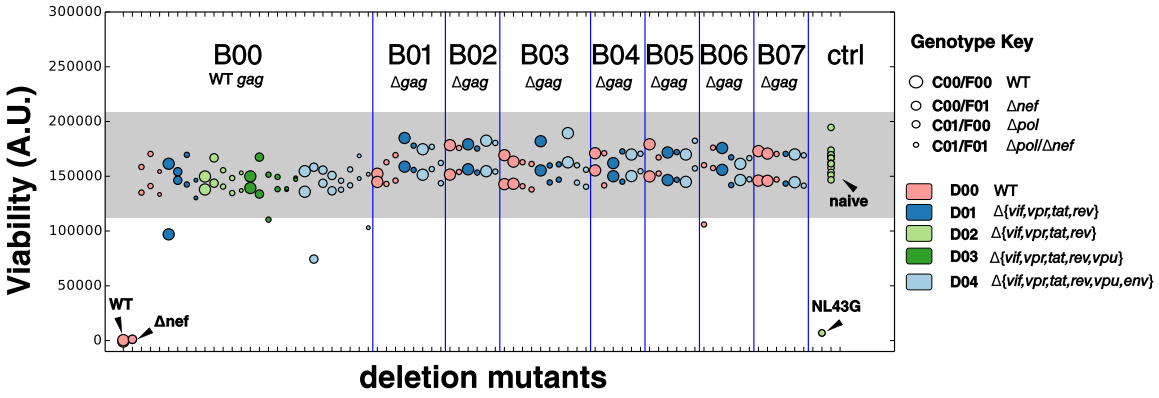


Figure 6.3: Most deletion mutants are not replication competent. MT-4 were infected with viral inocula produced by transfection of 293T with cloned HIV-1 mutants, and evaluated for viability by PrestoBlue assay at 5 days post-infection. Each clone was assessed by two independent experiments ($n = 2$). The shaded region represents viability (fluorescence) values within ±30% of uninfected cell mean (naïve). The genotype of each mutant can be read off using the key and annotations in the plot. The only clones to exhibit consist cell killing (replication) were the reconstructed wildtype (WT) and F01-only mutant (Δnef). The positive control (NL43G) exhibited the expected killing as well.

As shown in Figure 6.3, the only clones for which cell killing is evident are NL43G

(WT control), the A00B000C00D00E00F00G00 mutant (WT genotype) and a mutant with a single deletion in *nef*-deletion mutant (F01). This is consistent with literature reports, as *nef* is dispensable for replication of certain strains of HIV-1 in permissive cells lines, such as MT-4 [110]. Thus, this validates that the cloning strategy employed can reconstruct viruses with full wildtype activity.

No loss in cell viability was observed for the other 59 of 61 molecular clones, which contain deletions in one or more essential *trans*-acting elements. These cultures were outgrown and inspected visually for up to 14 days, but no cytopathic effect of loss of viability was observed in the remaining 59 mutants (data not shown).

Thus we have confirmed that these mutants are not replication competent and must be *trans*-complemented by wildtype virus to transmit.

Select deletion mutants interfere with wildtype virus

Next, we assessed the ability of each mutant to interfere with the replication of wildtype virus in a co-transfection experiment. 239T were co-transfected with a equal masses (1:1) of a mutant plasmid and a GFP-tagged WT molecular clone (NLENG1-IRES/NL43G). After 48 hours, the virus-containing supernatant was harvested, clarified by filtration, and used to infect a highly permissive/susceptible T cell line (MT-4). The cell population was scored for GFP+ cells by flow cytometry at 24 and 48 hpi. All mutants are untagged: the only source of GFP production is infection with wildtype NL43G virus.

If a mutant does not interfere with WT virus replication, than the presence or absence of the subgenomic mutant will not affect GFP+ levels. Conversely, if we observe a change in GFP+ compared to wildtype (NL43G only), than the mutant either enhances wildtype virus replication (higher levels of GFP+ cells) or interferes (decreases levels of GFP+ cells).

Results of the interference assay are shown in Figure 6.4, Figure 6.5, and Figure 6.6.

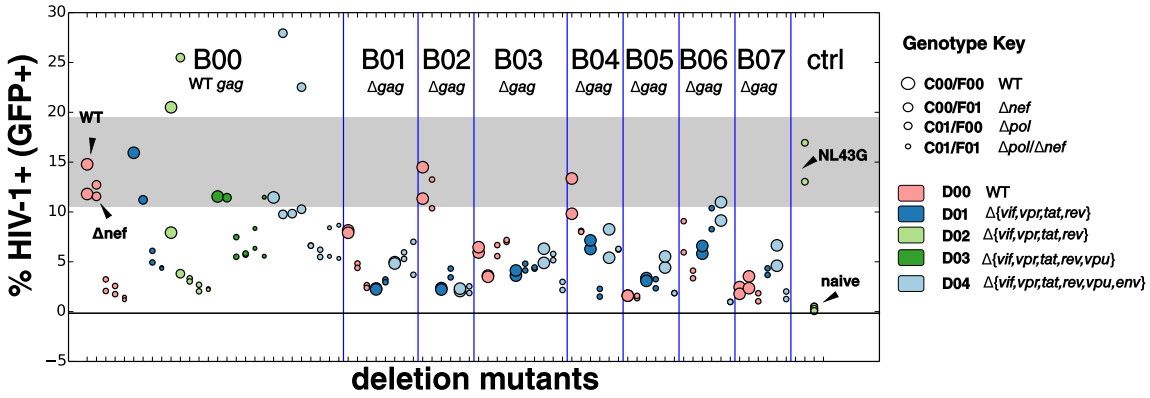


Figure 6.4: Select deletion mutants interfere with wildtype HIV-1 replication (24 h, −Darunavir). MT-4 were infected with viral inocula produced by transfection of 293T with cloned HIV-1 mutants and a replication competent, GFP-tagged virus (NL43G), then evaluated for wildtype infection (GFP fluorescent cells) 24 hours post infection (one round of replication). Each clone was assessed by two independent experiments ($n = 2$): both points are plotted above in a single column. The shaded region represents GFP+ values within ±30% of the mean of the NL43G only control. Clones with values below this region interfere with wildtype viral replication, clones with values above enhance wildtype virus replication. Uninfected cells (naïve) have a background GFP+ value of≈ 1%. The genotype of each mutant can be read off using the key and annotations in the plot. Replication-competent clones (annotated at WT and Δ*nef*) did not interfere with wildtype replication.

At 24 hpi (Figure 6.4), 13–17% of cells were GFP+ in the control NL43G infection, compared to < 1% in the uninfected control (naïve). Co-transfection with the reconstituted replication-competent HIV-1 clones had little effect on single-round wildtype HIV-1 NL43G replication (WT: 11–15% GFP+, Δ*nef*: 11-13% GFP+).

In contrast, many of the multiply deleted clones showed strong interference with wildtype HIV-1 replication, eliciting 0.5 $\log_{10}$–1.0 $\log_{10}$ reductions in infectious titer with respect to wildtype. The strongest interference effect was observed in clones harboring both deletions in *gag* and deletions in the accessory tract of the HIV-1 genome (*vif*–*vpu*). With no deletions in *gag* (B00), few of the clones exhibiting strong interference, except for those that also had deletions in *pol* and *nef*. We did not observe any consistent enhancement of HIV-1 infection

by any of the clones (average increase of >30%).

At 24 hpi, the presence (Figure 6.5) or absence (Figure 6.4) of an HIV-1 protease inhibitor (Darunavir) did not affect the percentage of infected cells, reflecting that 24 hpi was appropriate timing to conduct single round studies of HIV-1 replication.

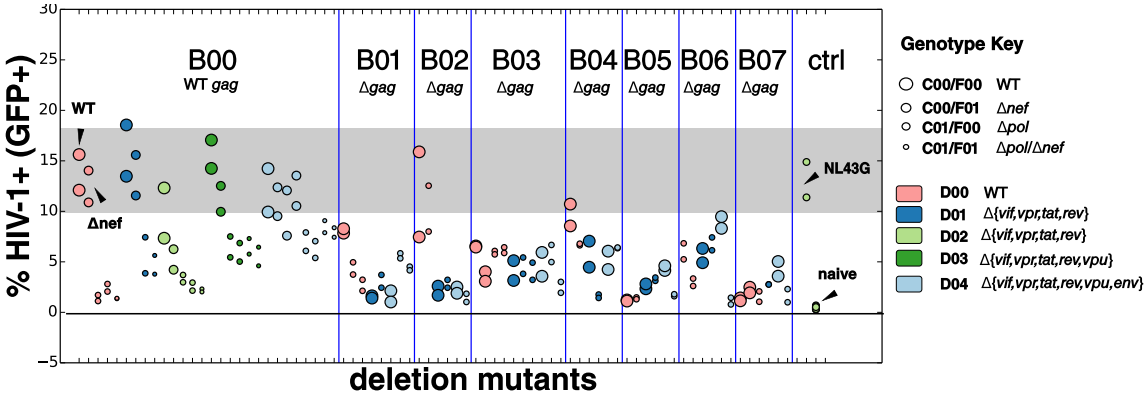


Figure 6.5: Select deletion mutants interfere with wildtype HIV-1 replication (24 h, +Darunavir). MT-4 were infected with viral inocula produced by transfection of 293T with cloned HIV-1 mutants and a replication competent, GFP-tagged virus (NL43G), then evaluated for wildtype infection (GFP fluorescent cells) 24 hours post infection. An HIV-1 protease inhibitor (Darunavir) was added at the time of infection to restrict transmission to a single round. Each clone was assessed by two independent experiments ($n = 2$): both points are plotted above in a single column. The shaded region represents GFP+ values within ±30% of the mean of the NL43G only control. Clones with values below this region interfere with wildtype viral replication, clones with values above enhance wildtype virus replication. Uninfected cells (naïve) have a background GFP+ value of≈ 1%. The genotype of each mutant can be read off using the key and annotations in the plot. Replication-competent clones (annotated as WT and Δ*nef*) did not interfere with wildtype replication.

By 48 hpi, (Figure 6.6), an additional round of HIV-1 replication will have occurred and interference effects will be easily detectable. As before, the strongest interference was observed in clones harboring both deletions in *gag* and deletions in the accessory tract of the HIV-1 genome (*vif*–*vpu*). With no deletions in *gag* (B00), few of the clones exhibit strong interference, except for those that also had deletions in *pol* and *nef*. We did not observe any consistent enhancement of HIV-1 infection by any of the clones (average increase of >30%).

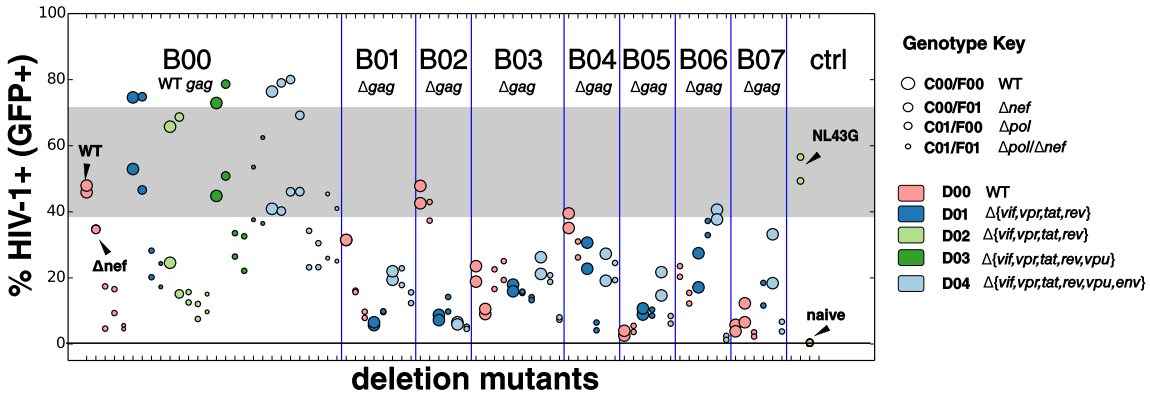


Figure 6.6: Select deletion mutants interfere with wildtype HIV-1 replication (48 h). MT-4 were infected with viral inocula produced by transfection of 293T with cloned HIV-1 mutants and a replication-competent, EGFP-tagged virus (NL43G), then evaluated for wildtype infection (GFP fluorescent cells) 48 hours post infection (one round of replication). Each clone was assessed by two independent experiments ($n = 2$): both points are plotted above in a single column. The shaded region represents GFP+ values within ±30% of the mean of the NL43G only control. Clones with values below this region interfere with wildtype viral replication, clones with values above enhance wildtype virus replication. Uninfected cells (naïve) have a background GFP+ value of≈ 1%. The genotype of each mutant can be read off using the key and annotations in the plot. Replication-competent clones (annotated as WT and Δnef) did not interfere with wildtype replication.

Mutants can be mobilized by providing WT *trans*-factors

We next determined if cloned mutants were transmissible: could they be encapsidated and transmitted by supply the missing *trans*-acting elements via transfection. A packaging cell line (239T) was co-transfected with the mutant viral genome plasmid and two additional plasmids: a packaging plasmid (pCMVRΔ8.91), which provides several HIV-1 proteins *in trans* (Gag, Pol, Tat, Rev), but is not replication-competent), and an envelope pseudotyping plasmid which provides a pan-tropic VSV-G envelope protein (pMD.G).

48 hours post-transfection of 293T, virus-containing supernatant was harvested, clarified, and used to transduce a T-cell line (MT-4), and the cells allowed to recover and outgrow for 5 days.

After 5 days, each transduced culture was assessed for transduction efficiency by performing PCR with primers specific to block B and block D, using DNA isolated from the transduced cells as template.

As shown in Figure 6.7, the appearance of clear amplicons at a length less than the WT Block B (1.9 kbp) or WT Block D (1.5 kbp), indicates that MT-4 were successfully transduced with the deletion mutant (indicating mobilization) and that the mutant did not lead to cell death (as occurred in some control reactions). For this reason, BTN503 (reconstituted wildtype) and BTN504 (Δnef) showed weak amplification, as the cells were destroyed by viral infection. Approximately 50% (31 of 61) of the deletion mutants showed clear mobilization from 293T from co-transfection.

6.4 Discussion

In this chapter, we use the information garnered during the high MOI screen of the NL4-3Δ_1 deletion library in Chapter 4 to construct and verify a collection of prototype TIPs for HIV-1. This was made possible by two critical pieces of genetic data were obtained in Chapter 4.

First, the high MOI passage experiment allowed us to delineate *cis*- and *trans*-acting elements of HIV-1. To create functional TIPs, we must preserve all *cis*-acting elements and remove or inactivate at least one *trans*-acting element to increase competition between TIP and wildtype helper virus.

Secondly, we were able to measure how specific deletions in *trans*-acting elements contributed to the ability of the deletion mutant to be mobilized by the wildtype virus with respect to its peers. Using these data, we identified a number of single deletions to incorporate into a TIP prototype.

Combining these data, we selected a number of "adaptive" deletions and generated a combinatorial library of HIV-1 deletion mutants, where each mutant harbors one or more

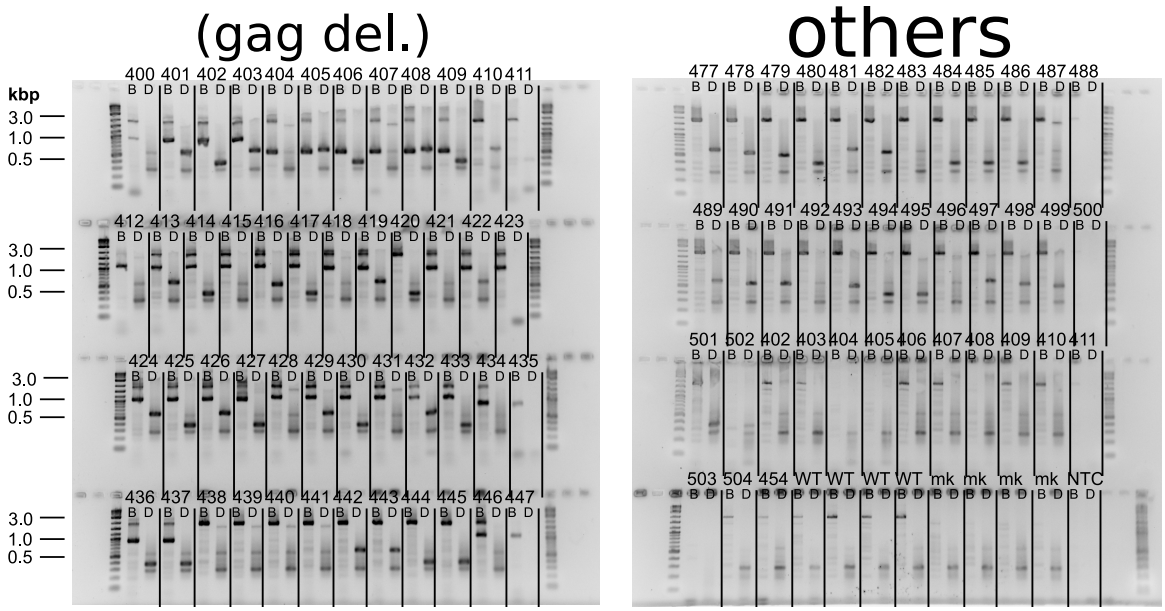


Figure 6.7: Deletion mutants can be mobilized by providing missing common goods *in trans*. Deletion mutant pseudovirus stocks were prepared by packaging HIV-1 mutant in 293T by co-transfection with a VSV-G envelope plasmid and pCMVRΔ8.91, which provides four HIV-1 proteins (Gag, Pol, Tat, Rev). After 5 days of recovery in the presence of Darunavir (an HIV-1 protease inhibitor), PCR for blocks B and D was performed on DNA isolated from the transduced cells. Each pair of lanes is marked with the BTN strain number in Appendix A (e.g. 427 is BTN427, 405 is BTN405), where B corresponds to the block B amplicon and D to the block D amplicon. The wildtype (undeleted) block B amplicon is 1.9 kbp and the WT block D amplicon is 1.5 kbp. A distinct, heavy band at size less than the WT amplicon indicates that the deletion mutant had been successfully mobilized and did not kill transduced cells. Control abbreviations are WT (infected with NL43G only), mk (mock-transfection, no deletion mutant genome was added to transfection), NTC (no template control, PCR with water as template).

barcoded deletions. A modular cloning strategy allowed simplified construction of a large number of strains is amendable to further scaling.

We next characterized the collection of cloned mutants, and tested each mutant or cloned WT virus for the properties of a TIP. As detailed in Chapter 1, an ideal transmissible antiviral (a DIP/TIP) would possess the following qualities: (1) it would compete effectively for common goods provided by the wildtype helper virus during co-infection (interference) (2) it would be retain all necessary *cis*-acting elements required for efficient mobilization by the helper virus (mobilization, $R_0 > 1$), (3) it would be unable to transmit without co-infection with wildtype helper virus (no self-replication).

We found that reconstituted WT and Δnef viruses were able to replicate and transmit (Figure 6.3), but do not interfere with wildtype HIV-1 in co-transfection studies (Figure 6.5). In contrast, deletion mutants were in general not replication-competent (59 of 60 in Figure 6.3) but *did interfere* with wildtype HIV-1 (Figure 6.4) and could be mobilized to T cells during single-round infection studies (Figure 6.7). Thus, these mutants have the desired qualities of TIPs and are transmissible antivirals.

In particular, we found that deletions of the accessory region of HIV-1 (*vif–vpu*), when combined with *gag/pol* deletions, allowed deletion mutants to interfere with wildtype virus replication and transmit efficiently in single round studies. This suggests a model of interference where transcriptional asymmetry allows these HIV-1 deletion mutants to compete effectively for a common pool of capsid proteins provided by the wildtype virus.

Further studies will be required to verify if these deletion mutants can function as TIPs. This study was confined to a single round of replication *in vitro*, using a single target cell type and single virus strain. Tests of mobilization between T cells and sustained control of viral load will be necessary to conduct in the future.

Taken in whole, this dissertation has developed a framework for generating transmissible

antivirals from first principles. The method is of general use in virology, where the technology can be used to generate live-attenuated vaccines, viral vectors, and replicons, as well as to understand fundamental principles of viral replication and genetics in diverse viral systems. It is of particular interest in emerging viral infections, where therapies must be quickly generated and deployed.

Bibliography

1. Huang, A. & Baltimore, D. Defective viral particles and viral disease processes. *Nature* **226,** 325–327 (1970).

2. Fineberg, H. V. Pandemic preparedness and response–lessons from the H1N1 influenza of 2009. *The New England Journal of Medicine* **370,** 1335–1342 (14 Apr. 2014).

3. Nowak, M. & May, R. M. *Virus dynamics: mathematical principles of immunology and virology* 256 pp. (Oxford University Press, UK, Nov. 11, 2000).

4. Rouzine, I. M. & Coffin, J. M. Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 10758–63 (Sept. 1999).

5. Rouzine, I. M., Rodrigo, A. & Coffin, J. M. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiology and Molecular Biology Reviews* **65,** 151–185 (1 Mar. 2001).

6. Rouzine, I. M. & Coffin, J. M. Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics* **170,** 7–18 (1 Apr. 13, 2017).

7. Zanini, F. *et al.* Population genomics of intrapatient HIV-1 evolution. *eLife* **4** (Dec. 2015).

8. Cohen, M. S. *et al.* Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *The New England Journal of Medicine* **375,** 830–839 (9 July 18, 2016).

9. Bar, K. J. *et al.* Effect of HIV Antibody VRC01 on Viral Rebound after Treatment Interruption. *The New England Journal of Medicine* **375,** 2037–2050 (21 Nov. 9, 2016).

10. Newman, M. E. Spread of epidemic disease on networks. *Physical Review E* **66,** 016128 (2002).

11. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438,** 355–359 (7066 Nov. 2005).

12. Grant, R. M. *et al.* Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *The New England journal of medicine* **363,** 2587–2599 (27 Dec. 2010).

13. Baeten, J. M. *et al.* Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *The New England journal of medicine* **367,** 399–410 (5 Aug. 2012).

14. Graw, F., Leitner, T. & Ribeiro, R. M. Agent-based and phylogenetic analyses reveal how HIV-1 moves between risk groups: injecting drug users sustain the heterosexual epidemic in Latvia. *Epidemics* **4,** 104–116 (2 June 2012).

15. Vasilakis, N., Cardosa, J., Hanley, K. A., Holmes, E. C. & Weaver, S. C. Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nature reviews. Microbiology* **9,** 532–541 (7 June 2011).

16. Weinberger, L. S., Schaffer, D. V. & Arkin, A. P. Theoretical Design of a Gene Therapy To Prevent AIDS but Not Human Immunodeficiency Virus Type 1 Infection. *Journal of Virology* **77,** 10028–10036 (2003).

17. Metzger, V. T., Lloyd-Smith, J. O. & Weinberger, L. S. Autonomous targeting of infectious superspreaders using engineered transmissible therapies. *PLoS Comput Biol* **7,** e1002015 (Mar. 2011).

18. Rouzine, I. M. & Weinberger, L. S. Design requirements for interfering particles to maintain coadaptive stability with HIV-1. *Journal of Virology* **87,** 2081–2093 (Feb. 2013).

19. Notton, T., Sardanyés, J., Weinberger, A. D. & Weinberger, L. S. The case for transmissible antivirals to control population-wide infectious disease. *Trends in Biotechnology* **32,** 400–405 (Aug. 2014).

20. Knipe, D. M. & Howley, P. M. *Fields Virology* 6th ed. 2664 pp. (Lippincott Williams&Wilki, July 25, 2013).

21. Schulz, F. *et al.* Giant viruses with an expanded complement of translation system components. *Science* **356,** 82–85 (Apr. 6, 2017).

22. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351,** aad6253 (6280 Mar. 2016).

23. Finsterbusch, T. & Mankertz, A. Porcine circoviruses–small but powerful. *Virus research* **143,** 177–183 (2 Aug. 2009).

24. Razooky, B. S. *The Role of Tat Positive Feedback in the Establishment of and Reactivation from HIV Latency* PhD thesis (University of California, San Francisco, 2014), 242.

25. Razooky, B. S., Pai, A., Aull, K., Rouzine, I. M. & Weinberger, L. S. A hardwired HIV latency program. *Cell* **160,** 990–1001 (5 Feb. 2015).

26. Gillespie, L. K., Hoenen, A., Morgan, G. & Mackenzie, J. M. The endoplasmic reticulum provides the membrane platform for biogenesis of the flavivirus replication complex. *Journal of virology* **84,** 10438–10447 (20 Oct. 2010).

27. Chatel-Chaix, L. *et al.* Dengue Virus Perturbs Mitochondrial Morphodynamics to Dampen Innate Immune Responses. *Cell host & microbe* **20,** 342–356 (3 Sept. 2016).

28. Butsch, M. & Boris-Lawrie, K. Translation is not required To generate virion precursor RNA in human immunodeficiency virus type 1-infected T cells. *Journal of virology* **74,** 11531–7 (Dec. 2000).

29. Butsch, M. & Boris-Lawrie, K. Destiny of unspliced retroviral RNA: ribosome and/or virion? *J Virol* **76,** 3089–3094 (Apr. 2002).

30. Henle, W. & Henle, G. Interference of Inactive Virus with the Propagation of Virus of Influenza. *Science* **98,** 87–89 (July 1943).

31. Von Magnus, P. Incomplete forms of influenza virus. *Advances in virus research* **2,** 59–79 (1954).

32. Frensing, T. *et al.* Continuous influenza virus production in cell culture shows a periodic accumulation of defective interfering particles. *PloS one* **8,** e72288 (Jan. 2013).

33. Frensing, T. Defective interfering viruses and their impact on vaccines and viral vectors. *Biotechnology journal* **10,** 681–689 (5 May 2015).

34. DePolo, N. J., Giachetti, C. & Holland, J. J. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. *Journal of Virology* **61,** 454–464 (2 Feb. 1987).

35. Timm, C., Akpinar, F. & Yin, J. Quantitative characterization of defective virus emergence by deep sequencing. *Journal of virology* **88,** 2623–2632 (5 Mar. 2014).

36. Cohen, M. S., Shaw, G. M., McMichael, A. J. & Haynes, B. F. Acute HIV-1 Infection. *The New England Journal of Medicine* **364,** 1943–1954 (20 May 19, 2011).

37. Tanner, E. J. *et al.* Dominant drug targets suppress the emergence of antiviral resistance. *eLife* **3** (Nov. 2014).

38. Levine, B. L. *et al.* Gene transfer in humans using a conditionally replicating lentiviral vector. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 17372–7 (Nov. 2006).

39. Leonard, J. N., Shah, P. S., Burnett, J. C. & Schaffer, D. V. HIV evades RNA interference directed at TAR by an indirect compensatory mechanism. *Cell Host & Microbe* **4,** 484–494 (Nov. 2008).

40. Thompson, K. A. S. & Yin, J. Population dynamics of an RNA virus and its defective interfering particles in passage cultures. *Virology journal* **7,** 257 (Sept. 2010).

41. Coffin, J. M., Tsichlis, P. N., Barker, C. S., Voynow, S. & Robinson, H. L. Variation in avian retrovirus genomes. *Ann NY Acad Sci* **354,** 410–425 (Nov. 1980).

42. Lai, C. J. & Nathans, D. Deletion mutants of simian virus 40 generated by enzymatic excision of DNA segments from the viral genome. *Journal of molecular biology* **89,** 179–193 (1 Oct. 1974).

43. Carbon, J., Shenk, T. E. & Berg, P. Biochemical procedure for production of small deletions in simian virus 40 DNA. *Proceedings of the National Academy of Sciences of the United States of America* **72,** 1392–1396 (4 Apr. 1975).

44. Shenk, T. E., Carbon, J. & Berg, P. Construction and analysis of viable deletion mutants of simian virus 40. *Journal of virology* **18,** 664–671 (2 May 1976).

45. Shenk, T. A biochemical method for increasing the size of deletion mutations in simian virus 40 DNA. *Journal of molecular biology* **113,** 503–515 (3 July 1977).

46. Jones, N. & Shenk, T. Isolation of deletion and substitution mutants of adenovirus type 5. *Cell* **13,** 181–188 (1 Jan. 1978).

47. Green, C. & Tibbetts, C. Targeted deletions of sequences from closed circular DNA. *Proceedings of the National Academy of Sciences of the United States of America* **77,** 2455–2459 (5 May 1980).

48. Hearing, P. & Shenk, T. Functional analysis of the nucleotide sequence surrounding the cap site for adenovirus type 5 region E1A messenger RNAs. *J Mol Biol* **167,** 809–822 (July 1983).

49. Wu, R. *et al.* Synchronous digestion of SV40 DNA by exonuclease III. *Biochemistry* **15,** 734–740 (4 Feb. 1976).

50. Henikoff, S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28,** 351–359 (June 1984).

51. Kirkegaard, K. & Nelsen, B. Conditional poliovirus mutants made by random deletion mutagenesis of infectious cDNA. *J Virol* **64,** 185–194 (Jan. 1990).

52. Krishnakumar, R. *et al.* Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. *Nucleic Acids Research* **42,** e111 (14 Aug. 2014).

53. Yu, D., Silva, M. C. & Shenk, T. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12396–12401 (21 Oct. 2003).

54. Heaton, N. S., Sachs, D., Chen, C.-J., Hai, R. & Palese, P. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 20248–20253 (50 Dec. 2013).

55. Fulton, B. O. *et al.* Mutational Analysis of Measles Virus Suggests Constraints on Antigenic Variation of the Glycoproteins. *Cell Rep* **11,** 1331–1338 (June 2015).

56. Wang, L. *et al.* Generation of a Live Attenuated Influenza Vaccine that Elicits Broad Protection in Mice and Ferrets. *Cell host & microbe* **21,** 334–343 (3 Mar. 2017).

57. Iverson, E. A., Goodman, D. A., Gorchels, M. E. & Stedman, K. M. Extreme Mutation Tolerance: Nearly Half of the Archaeal Fusellovirus Sulfolobus Spindle-Shaped Virus 1 Genes Are Not Required for Virus Function, Including the Minor Capsid Protein Gene vp3. *Journal of virology* **91** (10 May 2017).

58. Fulton, B. O., Sachs, D., Schwarz, M. C., Palese, P. & Evans, M. J. Transposon mutagenesis of the Zika virus genome highlights regions essential for RNA replication and restricted for immune evasion. *Journal of Virology* (May 2017).

59. Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519,** 181–186 (Mar. 2015).

60. Adachi, A. *et al.* Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. *Journal of Virology* **59,** 284–91 (Aug. 1986).

61. Schwarz, M. C. *et al.* Rescue of the 1947 Zika Virus Prototype Strain with a Cytomegalovirus Promoter-Driven cDNA Clone. *mSphere* **1** (Sept. 28, 2016).

62. Han, E. S. *et al.* RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res* **34,** 1084–1091 (2006).

63. Goryshin, I. Y. & Reznikoff, W. S. Tn5 *in vitro* transposition. *Journal of Biological Chemistry* **273,** 7367–7374 (Mar. 1998).

64. Reznikoff, W. S., Goryshin, I. Y. & Jendrisak, J. J. Tn5 as a molecular genetics tool: In vitro transposition and the coupling of in vitro technologies with in vivo transposition. *Methods Mol Biol* **260,** 83–96 (2004).

65. Mutalik, V. K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Methods* **10,** 347–353 (Apr. 2013).

66. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* **10,** 354–360 (Apr. 2013).

67. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33,** 103–119 (1 1985).

68. Bhasin, A., Goryshin, I. Y., Steiniger-White, M., York, D. & Reznikoff, W. S. Characterization of a Tn5 pre-cleavage synaptic complex. *J Mol Biol* **302,** 49–63 (Sept. 2000).

69. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* **24,** 2033–2040 (Dec. 2014).

70. Li, D., Aaskov, J. & Lott, W. B. Identification of a cryptic prokaryotic promoter within the cDNA encoding the 5′ end of dengue virus RNA genome. *PLoS One* **6,** e18197 (3 Mar. 2011).

71. Wang, K. *et al.* Defining synonymous codon compression schemes by genome recoding. *Nature* **539,** 59–64 (7627 Nov. 2016).

72. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56 (July 2010).

73. Berto, E., Bozac, A. & Marconi, P. Development and application of replication-incompetent HSV-1-based vectors. *Gene therapy* **12 Suppl 1,** S98–102 (Oct. 2005).

74. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science (New York, N.Y.)* **288,** 1789–1796 (5472 June 2000).

75. Sharp, P. M. *et al.* The origins of acquired immune deficiency syndrome viruses: where and when? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **356,** 867–876 (1410 June 2001).

76. Zhong, P. *et al.* Cell-to-cell transmission can overcome multiple donor and target cell barriers imposed on cell-free HIV. *PloS one* **8,** e53138 (Jan. 2013).

77. Sigal, A. *et al.* Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* **477,** 95–8 (Sept. 2011).

78. Haché, G. & Harris, R. S. CEM-T4 cells do not lack an APOBEC3G cofactor. *PLoS pathogens* **5,** e1000528 (July 2009).

79. Ayinde, D., Casartelli, N. & Schwartz, O. Restricting HIV the SAMHD1 way: through nucleotide starvation. *Nature Reviews Microbiology* **10,** 675–80 (Oct. 2012).

80. Usami, Y., Wu, Y. & Göttlinger, H. G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526,** 218–223 (7572 Oct. 8, 2015).

81. Rosa, A. *et al.* HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526,** 212–217 (7572 Oct. 8, 2015).

82. Blanc, D., Patience, C., Schulz, T. F., Weiss, R. & Spire, B. Transcomplementation of VIF- HIV-1 mutants in CEM cells suggests that VIF affects late steps of the viral life cycle. *Virology* **193,** 186–192 (1 Mar. 1993).

83. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272,** 263–7 (Apr. 1996).

84. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nature biotechnology* **15,** 871–5 (Sept. 1997).

85. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J Virol* **72,** 8463–8471 (Nov. 1998).

86. Ocwieja, K. E. *et al.* Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Research* **40,** 10345–10355 (Nov. 2012).

87. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460,** 711–6 (Aug. 2009).

88. Clever, J., Sassetti, C. & Parslow, T. G. RNA secondary structure and binding sites for *gag* gene products in the 5′packaging signal of human immunodeficiency virus type 1. *Journal of Virology* **69,** 2101–2109 (4 Apr. 1995).

89. Abbink, T. E. M. & Berkhout, B. RNA structure modulates splicing efficiency at the Human Immunodeficiency Virus Type 1 major splice donor. *Journal of Virology* **82,** 3090–8 (Mar. 2008).

90. Houzet, L. *et al.* HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic acids research* **35,** 2695–704 (Jan. 2007).

91. McBride, M. S. & Panganiban, A. T. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **70,** 2963–2973 (May 1996).

92. McBride, M. S., Schwartz, M. D. & Panganiban, A. T. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J Virol* **71,** 4544–4554 (June 1997).

93. McBride, M. S. & Panganiban, A. T. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J Virol* **71,** 2050–2058 (Mar. 1997).

94. Keating, C. P. *et al.* The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. *Nucleic Acids Res* **37,** 945–956 (Feb. 2009).

95. Chamanian, M. *et al.* A *cis*-acting element in retroviral genomic RNA links Gag-Pol ribosomal frameshifting to selective viral RNA encapsidation. *Cell Host & Microbe* **13,** 181–92 (Feb. 2013).

96. Charneau, P. & Clavel, F. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J Virol* **65,** 2415–2421 (May 1991).

97. Zennou, V. *et al.* HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101,** 173–185 (Apr. 2000).

98. Nikolaitchik, O. A. & Hu, W.-S. Deciphering the role of the Gag-Pol ribosomal frameshift signal in HIV-1 RNA genome packaging. *J Virol* **88,** 4040–4046 (Apr. 2014).

99. Ho, Y.-C. *et al.* Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. *Cell* **155,** 540–551 (Oct. 2013).

100. Bernier, R. & Tremblay, M. Homologous interference resulting from the presence of defective particles of human immunodeficiency virus type 1. *Journal of virology* **69,** 291–300 (1 Jan. 1995).

101. An, D. S. *et al.* An inducible human immunodeficiency virus type 1 (HIV-1) vector which effectively suppresses HIV-1 replication. *Journal of virology* **73,** 7671–7 (Sept. 1999).

102. Dropulić, B., Hěrmánková, M. & Pitha, P. M. A conditionally replicating HIV-1 vector interferes with wild-type HIV-1 replication and spread. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 11103–8 (Oct. 1996).

103. Bukovsky, A., Song, J.-P. & Naldini, L. Interaction of human immunodeficiency virus-derived vectors with wild-type virus in transduced cells. *Journal of virology* **73,** 7087–92 (Aug. 1999).

104. Klimatcheva, E. *et al.* Defective lentiviral vectors are efficiently trafficked by HIV-1 and inhibit its replication. *Molecular Therapy* **3,** 928–39 (June 2001).

105. Morris, K. V. & Looney, D. J. Characterization of Human Immunodeficiency Virus (HIV)-2 vector mobilization by HIV-1. *Hum Gene Ther* **16,** 1463–1472 (Dec. 2005).

106. Franz, K. M. *Design and Characterization of a Therpeutic Interfering Particle to Halt the Spread of Infectious Disease* MA thesis (University of California, San Diego, 2011).

107. Weinberger, L. S. & Notton, T. J. *U.S. pat. req.* US20160015769A1 (2016). Pending.

108. Al-Mawsawi, L. Q. *et al.* High-throughput profiling of point mutations across the HIV-1 genome. *Retrovirology* **11,** 124 (Dec. 2014).

109. Miller-Jensen, K., Skupsky, R., Shah, P. S., Arkin, A. P. & Schaffer, D. V. Genetic selection for context-dependent stochastic phenotypes: Sp1 and TATA mutations increase phenotypic noise in HIV-1 gene expression. *PLoS Computational Biology* **9,** e1003135 (7 2013).

110. Gibbs, J. S., Regier, D. A. & Desrosiers, R. C. Construction and *in vitro* properties of HIV-1 mutants with deletions in "nonessential" genes. *AIDS Res Hum Retroviruses* **10,** 343–350 (Apr. 1994).

111. Gibbs, J. S., Regier, D. A. & Desrosiers, R. C. Construction and *in vitro* properties of SIVmac mutants with deletions in "nonessential" genes. *AIDS Res Hum Retroviruses* **10,** 607–616 (May 1994).

112. Miyoshi, I. *et al.* Type C virus-producing cell lines derived from adult T cell leukemia. *Gann Monogr* **28** (1 1982).

113. Vermeire, J. *et al.* Quantification of Reverse Transcriptase Activity by Real-Time PCR as a Fast and Accurate Method for Titration of HIV, Lenti- and Retroviral Vectors. *PLoS ONE* **7** (ed Tachedjian, G.) e50859 (Dec. 2012).

114. Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518,** 55–60 (Feb. 2015).

115. Li, D. *et al.* Defective interfering viral particles in acute dengue infections. *PloS One* **6,** e19447 (4 Apr. 2011).

116. Jayaraman, B. *et al.* RNA-directed remodeling of the HIV-1 protein Rev orchestrates assembly of the Rev-Rev response element complex. *eLife* **3,** e04120 (Dec. 2014).

117. Kutluay, S. B. & Bieniasz, P. D. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog* **6,** e1001200 (2010).

118. Kutluay, S. B. *et al.* Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell* **159,** 1096–1109 (Nov. 2014).

119. Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* **101,** 4204–4209 (Mar. 2004).

120. Pollack, R. A. *et al.* Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell host & microbe* **21,** 494–506.e4 (4 Apr. 2017).

121. Hosmane, N. N. *et al.* Proliferation of latently infected CD4(+) T cells carrying replication-competent HIV-1: Potential role in latent reservoir dynamics. *The Journal of experimental medicine* **214,** 959–972 (4 Apr. 2017).

122. Bruner, K. M. *et al.* Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nature medicine* **22,** 1043–1049 (9 Sept. 2016).

123. Perelson, A. S. & Nelson, P. W. Mathematical Analysis of HIV-1 Dynamics in Vivo. *SIAM Review* **41,** 3–44 (Jan. 1999).

124. Alcaraz-Estrada, S. L., Yocupicio-Monroy, M. & del Angel, R. M. Insights into dengue virus genome replication. *Future Virology* **5,** 575–592 (Sept. 2010).

125. Grubaugh, N. D. *et al.* Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* **546,** 401–405 (7658 June 2017).

126. Metsky, H. C. *et al.* Zika virus evolution and spread in the Americas. *Nature* **546,** 411–415 (7658 June 2017).

127. Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K. & Holmes, E. C. Long-term transmission of defective RNA viruses in humans and Aedes mosquitoes. *Science* **311,** 236–8 (Jan. 2006).

128. Li, D. & Aaskov, J. Sub-genomic RNA of defective interfering (D.I.) dengue viral particles is replicated in the same manner as full length genomes. *Virology* **468-470,** 248–255 (Nov. 2014).

129. Alcaraz-Estrada, S. L., Manzano, M. I. M., Del Angel, R. M., Levis, R. & Padmanabhan, R. Construction of a dengue virus type 4 reporter replicon and analysis of temperature-sensitive mutations in non-structural proteins 3 and 5. *The Journal of General Virology* **91,** 2713–2718 (Pt 11 Nov. 2010).

130. Sirohi, D. *et al.* The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **352,** 467–470 (6284 Apr. 2016).

131. Dai, L. *et al.* Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody. *Cell Host & Microbe* **19,** 696–704 (5 May 2016).

132. Song, H., Qi, J., Haywood, J., Shi, Y. & Gao, G. F. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. *Nature Structural & Molecular Biology* **23,** 456–458 (5 May 2016).

133. Wang, D. *et al.* A Mutation Identified in Neonatal Microcephaly Destabilizes Zika Virus NS1 Assembly in Vitro. *Scientific reports* **7,** 42580 (Feb. 2017).

134. Zhao, B. *et al.* Structure and function of the Zika virus full-length NS5 protein. *Nature communications* **8,** 14762 (Mar. 2017).

135. Baer, A. & Kehn-Hall, K. Viral concentration determination through plaque assays: using traditional and novel overlay systems. *Journal of Visualized Experiments : JoVE,* e52065 (93 Nov. 2014).

136. Bujalowski, P. J., Bujalowski, W. & Choi, K. H. Interactions between the Dengue Virus Polymerase NS5 and Stem-Loop A. *Journal of Virology* **91** (11 June 2017).

137. Khromykh, A. A., Varnavski, A. N. & Westaway, E. G. Encapsidation of the flavivirus kunjin replicon RNA by using a complementation system providing Kunjin virus structural proteins in trans. *Journal of Virology* **72,** 5967–5977 (7 July 1998).

138. Villordo, S. M. & Gamarnik, A. V. Differential RNA sequence requirement for dengue virus replication in mosquito and mammalian cells. *Journal of virology* **87,** 9365–9372 (16 Aug. 2013).

139. Tripathi, S. *et al.* A novel Zika virus mouse model reveals strain specific differences in virus pathogenesis and host inflammatory immune responses. *PLoS pathogens* **13,** e1006258 (3 Mar. 2017).

Appendix A

Listing of plasmids

This dissertation work generated more than 100,000 defined and unique plasmids, of which $\approx$ 100 are available as clonal bacterial stocks. The remainder are available as polyclonal bacterial libraries. I've annotated each plasmid or plasmid library with a name, a lab strain designation for the Weinberger lab (with prefix BTN), a page reference in this dissertation, a literature reference, antibiotic-resistance phenotype, and a brief description.

All plasmids or plasmid libraries are available as *E. coli* DH10B stocks in 7%(v/v) DMSO in LB. The DH10B genotype is: Δ(*ara–leu*) *7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14*$^-$ ϕ80d*lacZ*Δ*M15 recA1 relA1 endA1 nupG rpsL* ($\mathrm{Str^R}$) *rph spoT1* Δ(*mrr-hsdRMS-mcrBC*).

Antibiotic resistance abbreviations are the following: amp indicates resistance to ampicillin or carbenicillin, kan indicates resistance to kanamycin, and cam indicates resistance to chloramphenicol.

Table A.1 is a partial listing of plasmids used in this work. Table A.2 lists all the molecular clones (wildtype and mutant) of HIV-1 generated in this work. Table A.3 lists the polyclonal libraries of insertion/deletion mutants of HIV-1 and ZIKV.

name	strain	page	source	drug	description
pNL4-3	BTN9	18	[60]	amp	HIV-1 mol. clone, subtype B, `AF324493.2`
pTN5MC	BTN520	26	this work	amp/cam	TN5 transposon cassette, cam^{R}
pTN5MK	BTN522	26	this work	amp/kan	TN5 transposon cassette, kan^{R}
pMR766(+)	BTN248	58	[61]	amp	ZIKV molecular clone, strain MR766, pol(+)
pMR766(−)	BTN252	58	[61]	amp	ZIKV molecular clone, strain MR766, pol(−)
pCMVΔR8.91	BTN11	158	[84]	amp	HIV-1 pkg. plasmid ($\Delta\{\Psi$,*vif*,*vpr*,*vpu*,*env*,*nef*$\}$)
pMD.G	BTN10	158	[83]	amp	for pseudotyping with VSV envelope (VSV-G)
pNLENG1-IRES	BTN2	157	[119]	amp	NL4-3 EGFP reporter virus (*nef*::*EGFP*-IRES-*nef*)
pUC19Δ2-NL43_A00	BTN535	154	this work	amp	HIV-1 subsequence, A, WT
pUC19Δ2-NL43_B00	BTN536	154	this work	amp	HIV-1 subsequence, B, WT
pUC19Δ2-NL43_B01	BTN537	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B02	BTN538	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B03	BTN539	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B04	BTN540	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B05	BTN541	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B06	BTN542	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_B07	BTN543	154	this work	amp	HIV-1 subsequence, B, Δ
pUC19Δ2-NL43_C00	BTN544	154	this work	amp	HIV-1 subsequence, C, WT
pUC19Δ2-NL43_C01	BTN545	154	this work	amp	HIV-1 subsequence, C, Δ
pUC19Δ1-NL43_D00	BTN546	154	this work	amp	HIV-1 subsequence, D, WT
pUC19Δ1-NL43_D01	BTN547	154	this work	amp	HIV-1 subsequence, D, Δ
pUC19Δ1-NL43_D02	BTN548	154	this work	amp	HIV-1 subsequence, D, Δ
pUC19Δ1-NL43_D03	BTN549	154	this work	amp	HIV-1 subsequence, D, Δ
pUC19Δ1-NL43_D04	BTN550	154	this work	amp	HIV-1 subsequence, D, Δ
pUC19Δ2-NL43_E00	BTN551	154	this work	amp	HIV-1 subsequence, E, WT
pUC19Δ2-NL43_F00	BTN552	154	this work	amp	HIV-1 subsequence, F, WT
pUC19Δ2-NL43_F01	BTN553	154	this work	amp	HIV-1 subsequence, F, Δ
pUC19Δ2-NL43_G00	BTN554	154	this work	amp	HIV-1 subsequence, G, WT

Table A.1: Partial listing of plasmids used in this study

Table A.2: generated HIV-1 molecular clones

name	strain	page	source	drug	description
pUC19-NL43_A00B01C00D00E00F00G00	BTN400	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C00D01E00F00G00	BTN401	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C00D04E00F00G00	BTN402	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C01D01E00F00G00	BTN403	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C00D00E00F00G00	BTN404	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C00D01E00F00G00	BTN405	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C00D04E00F00G00	BTN406	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C01D00E00F00G00	BTN407	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C01D01E00F00G00	BTN408	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B02C01D04E00F00G00	BTN409	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C00D01E00F00G00	BTN410	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C00D04E00F00G00	BTN411	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C00D00E00F00G00	BTN412	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C00D01E00F00G00	BTN413	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C00D04E00F00G00	BTN414	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C01D00E00F00G00	BTN415	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C01D01E00F00G00	BTN416	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B04C01D04E00F00G00	BTN417	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C00D00E00F00G00	BTN418	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C00D01E00F00G00	BTN419	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C00D04E00F00G00	BTN420	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C01D00E00F00G00	BTN421	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C01D01E00F00G00	BTN422	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B05C01D04E00F00G00	BTN423	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C00D01E00F00G00	BTN424	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C00D04E00F00G00	BTN425	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C01D01E00F00G00	BTN426	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C01D04E00F00G00	BTN427	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B07C00D00E00F00G00	BTN428	155	this work	amp	HIV-1 deletion mutant

Table A.2: generated HIV-1 molecular clones (continued from previous page)

name	strain	page	source	drug	description
pUC19-NL43_A00B07C00D00100F00G00	BTN429	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B07C00D04E00F00G00	BTN430	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B07C01D00E00F00G00	BTN431	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B07C01D01E00F00G00	BTN432	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B07C01D04E00F00G00	BTN433	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C01D00E00F00G00	BTN434	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C01D00E00F00G00	BTN435	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C01D04E00F00G00	BTN436	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C01D04E00F00G00	BTN437	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C00D00E00F00G00	BTN438	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C00D00E00F00G00	BTN439	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D00E00F00G00	BTN440	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D00E00F00G00	BTN441	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D01E00F00G00	BTN442	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D01E00F00G00	BTN443	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D04E00F00G00	BTN444	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B03C01D04E00F00G00	BTN445	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C01D00E00F00G00	BTN446	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B06C01D00E00F00G00	BTN447	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B01C00D00E00F00G00	BTN454	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D01E00F00G00	BTN477	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D02E00F00G00	BTN478	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D03E00F00G00	BTN479	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D04E00F00G00	BTN480	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D01E00F01G00	BTN481	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D02E00F01G00	BTN482	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D03E00F01G00	BTN483	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D04E00F01G00	BTN484	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D04E00F01G00	BTN485	155	this work	amp	HIV-1 deletion mutant

Table A.2: generated HIV-1 molecular clones (continued from previous page)

name	strain	page	source	drug	description
pUC19-NL43_A00B00C00D04E00F01G00	BTN486	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D00E00F00G00	BTN487	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D00E00F00G00	BTN488	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D01E00F00G00	BTN489	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D02E00F00G00	BTN490	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D02E00F00G00	BTN491	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D03E00F00G00	BTN492	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D03E00F00G00	BTN493	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D04E00F00G00	BTN494	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D04E00F00G00	BTN495	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D00E00F01G00	BTN496	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D01E00F01G00	BTN497	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D02E00F01G00	BTN498	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D03E00F01G00	BTN499	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D03E00F01G00	BTN500	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D04E00F01G00	BTN501	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C01D04E00F01G00	BTN502	155	this work	amp	HIV-1 deletion mutant
pUC19-NL43_A00B00C00D00E00F00G00	BTN503	155	this work	amp	wildtype HIV-1 (NL4-3 provirus)
pUC19-NL43_A00B00C00D00E00F01G00	BTN504	155	this work	amp	HIV-1 deletion mutant (Δnef)

Table A.2: HIV-1 molecular clones generated in this study

Table A.3: polyclonal random insertion and random deletion libraries

name	strain	page	drug	description
pNL4-3^{+}	BTN238	45	amp/kan	pNL4-3 (HIV-1) insertion library (TN5MK)
pNL4-3Δ_1	BTN245	55	amp	pNL4-3 (HIV-1) deletion library
pMR766(+)$^{+}$	BTN256	59	amp/kan	pMR766(+) (ZIKV) insertion library (TN5MK)
pMR766(−)$^{+}$	BTN257	59	amp/kan	pMR766(−) (ZIKV) insertion library (TN5MK)
pMR766(+)Δ_1	BTN258	63	amp	pMR766(+) (ZIKV) deletion library, “short” chew
pMR766(+)Δ_2	BTN259	63	amp	pMR766(+) (ZIKV) deletion library, “long” chew
pMR766(−)Δ_1	BTN260	63	amp	pMR766(−) (ZIKV) deletion library, “short” chew
pMR766(−)Δ_2	BTN261	63	amp	pMR766(−) (ZIKV) deletion library, “long” chew

Table A.3: Polyclonal random insertion/deletion plasmid libraries

Appendix B

Listing of oligonucleotides

DNA oligonucleotides used in this dissertation are listed in Table B.1. Each row of the table gives a Weinberger oligo number (prefixed with a "y"), and oligo name, a page reference in the dissertation, and the primary DNA sequence listed in $5' \rightarrow 3'$ orientation.

Abbreviations are the following: `/5Phos/` indicates that the $5'$ end is phosphorylated, `N` indicates a machine-mixed base, where `N` can be `A`, `C`,`G`, or `T` with roughly equal probability.

Table B.1: DNA oligonucleotides used in this dissertation

num	name	page	sequence
y530	oTN5-F	26	/5Phos/CTGTCTCTTATACACATCTGCGGCCGC
y531	oTN5-R	26	/5Phos/CTGTCTCTTATACACATCTTTAATTAATTCGCTACC
y468	oBC20v1-F	36	/5Phos/CCGTCCATGAAGGGTTCGAT
y469	oBC20v1-R	36	/5Phos/ACGAATCTGCCGTTGCCATA
y456	oBC20v1-T	36	CCGTCCATGAAGGGTTCGATNNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT
y538	oBC20v2-F	42	/5Phos/CCGTCCATGAAGCGTTCGAT
y539	oBC20v2-R	42	/5Phos/ACGAATCTGCCGTTGCCATA
y540	oBC20v2-T	42	CCGTCCATGAAGCGTTCGATNNNNNNNNNNNNNNNNNNNNNTATGGCAACGGCAGATTCGT
y541	opUC19_BamHI-F	151	TCTAGAGTCGACCTGCAGGCATGC
y542	opUC19_BamHI-R	151	CCGGGTACCGAGCTCGAATTCACT
y126	oNL43pol-F	97	GAGACAGGGCAAGAAACAGC
y127	oNL43pol-R	97	AACAGGCGGCCTTAACTGTA
y229	oMS2-F	97	TCCTGCTCAACTTCCTGTCGAG
y230	oMS2-R	97	CAGGTCAAACCTCCTAGGAATG
y534	oMR766-C-F	131	TTGGTCATGATACTGCTGATTGC
y535	oMR766-C-R	131	CCTCCATGTTCCAAGACAAC
y470	opUC19Δ1-F	150	AGTGTAAAGCCTGGGGTGCCT
y471	opUC19Δ1-R	159	TGACTGGGAAAACCCTGGCGT
y472	opUC19Δ2-F	159	ATTGCGTTGCGCTCACTGCC
y473	opUC19Δ2-R	159	GCCCTTCCCAACAGTTGCG
y474	oNL43_A0_inner-F	151	TTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGAGGCCTCCACCTGGGTCTTG
y475	oNL43_A0_outer-F	151	GCGGCCGCAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGGCGCGCCTTGTAAAACGACGGCCAGTGAATTC
y476	oNL43_A0-R	151	GCGGCCGCAGGCCCTGCATGCACTGGATGCAATCTATCCCATTCTGCA
y477	oNL43_B0-F	151	GCGGCCGCTGCAGAATGGGATAGATTGCATCCAGTGCATGCAGGGCCT
y478	oNL43_B0-R	151	GCGGCCGCAGCTGTCCTTTTCTGGCAGCACTATAGGCTGTACTGTCCA
y479	oNL43_C0-F	151	GCGGCCGCTGGACAGTACAGCCTATAGTGCTGCCAGAAAAGGACAGCT
y480	oNL43_C0_inner-R	151	CATTAGGCACCCCAGGCTTTACACTGGCGCGCCATCTTGTATTACTACTGCCCCTTCACCTTTCCAGAGGAGC
y481	oNL43_C0_outer-R	151	GCGGCCGCGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACT
y482	oNL43_D0-F	151	AGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGGCGCGCCGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGAT
y483	oNL43_D0-R	151	GTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTGGCGCGCCCATTTACCAATACTACTTCTTGTGGGTTGGGGTCTGTGGG
y484	oNL43_E0_inner-F	151	GTAACGCCAGGGTTTTCCCAGTCACGGCGCGCCCCCACAGACCCCAACCCACAAGAAGTAGTATTGGTAAATG
y485	oNL43_E0_outer-F	151	GCGGCCGCAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCAC
y486	oNL43_E0-R	151	GCGGCCGCGCACAGGCTCCGCAGATCGTCCCAGATAAGTGCTAAGGAT
y487	oNL43_F0-F	151	GCGGCCGCATCCTTAGCACTTATCTGGGACGATCTGCGGAGCCTGTGC
y488	oNL43_F0-R	151	GCGGCCGCGCAGCTCTCGGGCCACGTGATGAAATGCTAGGCGGCTGTC

Table B.1: DNA oligonucleotides utilized (continued from previous page)

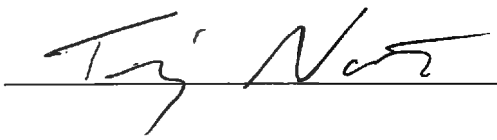
num	name	page	sequence
y489	oNL43_G0-F	151	GCGGCCGCGACAGCCGCCTAGCATTTCATCACGTGGCCCGAGAGCTGC
y490	oNL43_G0_inner-R	151	TGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGCGCGATCTTTGGCTCACTGC
y491	oNL43_G0_outer-R	151	GCGGCCGCGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTGGCGCGCCTGATTACGCCAAGCTTGCATGCCTG

Table B.1: Oligonucleotides used in this study

Publishing agreement

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

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

Author  Date 07 Sep 2017