Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně citovat.

Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program: Biologie

Studijní obor: Genetika, molekulární biologie a virologie

Bc. Tomáš Suchý

Charakterizace příspěvku genu gag k celkové replikační zdatnosti HIV u pacientů s různým průběhem nemoci

Contribution of gag region to overall HIV replicative fitness in patients with different disease progression

Diplomová práce

Vedoucí závěrečné práce: Mgr. Jan Weber, CSc.

Praha 2017

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval/a samostatně a že jsem uvedl/a všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 25. 4. 2017

Podpis

Zde bych rád poděkoval svému školiteli **Mgr. Janu Weberovi, CSc.** za příkladné vedení diplomové práce a nápomocnost. Dále bych mu rád poděkoval za možnost vypracování práce pod jeho vedením v Laboratoři Virologie na Ústavu Organické Chemie a Biochemie. Dále bych rád poděkoval **Ing. Lence Sácké** za to, že vedla mé první kroky v laboratoři a **Mgr. Janu Hodkovi, Ph.D**. za to, že převzal tuto štafetu a byl mi nápomocen, kdykoliv jsem potřeboval. **Mgr. Dmytro Struninovi** bych rád poděkoval za cenné rady ohledně průtokové cytometrie a statistického zpracování dat. **Haně Prouzové** bych chtěl poděkovat za obětavou péči o buněčné kultury a administrativní podporu. Dále bych rád poděkoval všem ostatním členům naší laboratoře za rady a pomoc, které mi poskytovali v celém průběhu vypracování této diplomové práce.

Abstract

Human immunodeficiency virus (HIV) is globally spread virus without available cure. Since its life-long presence, virus is carefully monitored as well as patient's immunological status. Replicative fitness of the virus is one of important aspects which can be taken into account, when monitoring HIV. Here, we are measuring HIV replicative fitness of *gag* recombinant viruses and comparing the results with replicative fitness of primary isolates. Further, we are comparing our findings of replicative fitness change over time with disease progression in the patient. We found that *gag* can be major contributor to overall fitness, although not in all cases. Additionally, we observed a correlation of replicative fitness development and slope of patient's $CD4^+$ T cells. Moreover, this relation was even more noticeable in patients with slow disease progression or in carriers of protective alleles. In summary, our results extend the understanding of replicative fitness and its role in disease progression; and pave the way to use the recombinant HIV for replicative fitness measurement in clinical practice.
Keywords: HIV, replicative fitness, recombinant virus, HIV disease progression, gag

Abstrakt

Virus lidské imunodeficience (HIV) je celosvětově rozšířen a bez možnosti vyléčení. Kvůli své celoživotní přítomnosti je virus pozorně sledován společně s imunologickým stavem pacienta. Replikační zdatnost viru je jedním z důležitých aspektů, které mohou být brány v potaz při sledování HIV. V této práci měříme replikační zdatnost HIV pomocí *gag* rekombinantních virů a srovnáváme výsledky s replikační zdatností změřenou s použitím primárních izolátů. Dále srovnáváme naše poznatky o změně replikační zdatnosti v čase s vývojem nemoci u pacienta. Zjistili jsme, že *gag* může být hlavním přispěvatelem k celkové zdatnosti, ovšem ne ve všech případech. Dále jsme pozorovali korelaci vývoje replikační zdatnosti a vývoje počtu $CD4^+$ T buněk u pacienta. Tento vztah byl patrnější u pacientů, které řadíme mezi pomalé progresory nebo u pacientů, kteří mají některou z protektivních alel. Souhrnem, naše výsledky posouvají dále porozumění replikační zdatnosti a její role v progresi nemoci a připravují cestu k použití rekombinantního HIV pro měření replikační zdatnosti v klinické praxi.
Klíčová slova: HIV, replikační zdatnost, rekombinantní virus, progrese nemoci HIV, gag

Contents

Abbreviations

AIDS	Acquired immune deficiency syndrome
AMP	Ampicillin
APOBEC3G	Apolipoprotein B mRNA editing enzyme catalytic subunit 3G
CA	Capsid protein
CCR5	C-C chemokine receptor type 5
CD	Cluster of differentiation
CRF	Circulating recombinant form
CTL	Cytotoxic T lymphocyte
CXCR4	C-X-C chemokine receptor type 4
DEAE	Diethylaminoethyl
DMEM	Dulbecco's modified Eagle's medium
DTT	Dithiothreitol
EGFP	Enhanced green fluorescent protein
Env	Envelope
5-FOA	5-Fluoroorotic acid
FSC	Forward-scattered light
Gag	Group-specific antigen
gp	Glycoprotein
HAART	Highly active antiretroviral therapy
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HEK293T	Human embryonic kidney cells 293 with SV40 Large T-antigen
IL	Interleukin
LB	Lysogeny broth
LTNP	Long term non-progressor
LTR	Long terminal repeat
MA	Matrix protein
MOI	Multiplicity of infection
MT4	Human T-cell line transformed by T-cell lymphotropic virus type 1
NC	Nucleocapsid protein
OKT3	Mouse hybridoma line producing monoclonal anti-CD3 antibody

ORF	Open reading frame
PBMC	Peripheral blood mononuclear cell
PHA	Phytohemagglutinin
Pol	Polymerase
Pr	Protease
PRO	Progressor
RPMI	Roswell Park Memorial Institute medium
RT	Reverse transcriptase
SDM	Site-directed mutagenesis
SOC	Super optimal broth with catabolite
SSC	Saline-sodium citrate
$TCID_{50}$	50% Tissue culture infectious dose
TRIM5-alpha	Tripartite motif-containing protein 5 - alpha
Trp1	Phosphoribosylanthranilate isomerase gene
TZM-bl	HeLa derived cell line expressing CD4 and CCR5
Ura3	Orotidine-5'-phosphate decarboxylase gene
URF	Unique recombinant form
VL	Viral load
YEPD	Yeast extract peptone dextrose

1. Introduction

Viruses are well-known obligatory parasites with considerable impact on human health. Occasionally viruses can be spread all over the world to cause pandemic infection. HIV is one of the viruses with global spread. Although scientists developed broad spectrum of antiretroviral drugs, clearance of the virus have not been successful until today as well as development of vaccine. Due to gradual expansion of virus, more and more research capacity was dedicated to its treatment. Today we know a lot about HIV molecular biology, gene function and pathogenesis, but large amount of knowledge yet remains to be explained. Furthermore, attention was drawn to better management of patients and personalization of treatment. For this reason it is important to characterize aspects on the intersection between virus and patient, like dynamics of $CD4^+$ T cells turnover and HIV viremia. Besides those characteristics, virus fitness is emphasized. In other words, fitness is an ability of virus to replicate in particular environment. Further knowledge of this problematics can elucidate way to efficient management of patient before and after start of antiretroviral treatment.

In our laboratory we are focusing on HIV fitness and its role in HIV pathogenesis and disease progression. Further, our interest is drawn to contribution of particular HIV genes to overall fitness of the virus. Importantly, we are using recombinant viruses for measuring fitness contribution of a particular gene. For those reasons, my thesis is oriented on contribution of *gag* to overall replicative fitness. *Gag* is one of three main genes typical for every retrovirus and its proteins are important targets of immune response; therefore it was chosen to evaluate its contribution to overall fitness. Moreover, our interest is also drawn to how HIV fitness changes in particular patient and how it influences disease progression in patients. Patients can be divided in several categories according to their immunological status. We are dividing our patients between slow progressors and typical progressors, based on their $CD4^+$ T cells decline. Research conducted for purposes of this thesis is part of a pilot study evaluating the role of HIV fitness on disease progression in the context of Czech, treatment naïve patients.

2. Aims of the thesis

Our laboratory conducts multistage research in the field of HIV replicative fitness. The work in this thesis is a part of the project about The Role of HIV Fitness on Disease Progression in the Absence of Antiretroviral Treatment. Aims of the thesis are:

1. **Preparation of *gag* recombinant virus and primary HIV isolates from untreated patients and comparison of replicative fitness values.**

   This aim includes:

   A. Insertion of patient specific *gag* in vector through method of PCR, yeast recombination, subcloning, and preparation of *gag* recombinant virus by transfection.
   B. Isolation of primary HIV isolates from patients' samples.
   C. Replicative fitness values measured by different methods using viral growth kinetics and competition experiments.

2. **Characterization of replicative fitness development in disease progression of typical progressors and slow progressors.**

   This aim includes:

   A. Insertion of patient specific *gag*, from blood samples collected in several years span, in vector and preparation of *gag* recombinant viruses.
   B. Analysis of HIV replicative fitness development in the environment of two immunologically different groups of patients.

3. Literature review

3.1. Human immunodeficiency virus

Human immunodeficiency virus (HIV) was discovered in 1983 as a pathogen causing disease, later known as Acquired immune deficiency syndrome (AIDS), which was observed two years earlier (Barre-Sinoussi *et al.* 1983; CDC 1981; CDC 1982). HIV has been categorized as a member of family *Retroviridae*, genus *Lentivirus*. All retroviruses are classified as a group VI, single-stranded RNA viruses with reverse transcription activity (ssRNA-RT) (Francki 1991). Since its discovery, HIV has become pandemic with more than 36 million currently infected people globally (UNAIDS 2016). During tracing of virus origin, another type of HIV was identified and named HIV-2, while original isolates were referred as HIV-1 (Clavel *et al.* 1986). HIV-2 never achieved pandemic status and is prevalent mainly in Western Africa (van der Loeff & Aaby 1999).

3.2. Human immunodeficiency virus-1

3.2.1. Virus origin and epidemiology

After initial confusion about suddenly appearing pandemic virus, researchers tracked down its origin. HIV-2 phylogenetic mapping revealed great similarity to viruses causing immunodeficiency in old world monkeys and apes (Guyader *et al,* 1987). New species in *Lentivirus* genus was named Simian immunodeficiency virus and are today believed to be the original source for zoonotic transfer to human (Gao *et al.* 1999). More than 40 of those viruses were identified in old world monkeys, including chimpanzee, western gorilla and sooty mangabey and those three species are considered source of HIV-1 and HIV-2 (Sharp & Hahn 2011.).

HIV-1 is categorized in four different groups. Group M (major) is associated with pandemic spread. This group is the oldest lineage with extrapolated time of transfer in 1920s or 1930s (Korber *et al.* 2000) from chimpanzee *Pan troglodytes troglodytes* in Western Africa (Gao *et al.* 1999).

Group N (non-M/O) has similar origin as group M, although zoonotic transfer was independent. In comparison with prevalent dissemination of group M, less than 20 patients with group N lineage have been found, all localized in south of Cameroon (Simon *et al.* 1998; Delaugerre *et al.* 2011).

Origin of group O (outlier) remains to be determined. This virus is greatly restricted to Cameroon, Gabon and Equatorial Guinea with ~1% prevalence (Vessiere *et al.* 2010).

Group P (putative) lineage of virus was found in two patients worldwide. Sequencing and phylogenetic analysis are displaying great similarity to SIV of western gorillas, indicating possible source of transmission (Plantier *et al.* 2009).

Thanks to pandemic spread, group M lineage had a lot of time and opportunities to gradually evolve and diverse itself. Today we recognize nine different subtypes. Those subtypes are A, B, C, D, F, G, H, J and K. Several are rare and some are the most prevalent throughout the world (Hemelaar *et al.* 2011). In many cases superinfection by two or more strains in one patient takes place, recombination links together parts of two genomes and recombinant form arises. These recombinants are named as a circulating recombinant form (CRF) if virus was confirmed in three uncoupled patients and as a unique circulating form (URF) when found only in a single patient (Robertson *et al.* 2000). Dozens of CRFs and URFs were identified worldwide and are responsible for around 20 % of infections mostly in Sub-Saharan Africa, where all known subtypes are present (Hemelaar *et al.* 2011).

In central Europe, Czech Republic specifically, subtype B is the most prevalent one, followed by subtype A and CRF01_AE (Abecasis *et al.* 2013). Subtype B is the most prevalent also in Western Europe and Northern America.

3.2.2. Molecular biology of the virus

Genomic sequence of HIV-1 contains 9180 nucleotides in form of positive single stranded RNA. This number increases in later stage of virus cycle, when RNA is reversely transcribed into the proviral genomic DNA, up to 9719 base pairs after extension of long terminal repeats (LTRs). Genome is highly organized and encodes nine genes in three open reading frames (ORFs) (fig. 1). Except encoded genes, genome contains also cis-acting RNA elements necessary for particular steps in viral life cycle (Foley *et al.* 2016). Genes can be further categorized into two groups; first group includes three main genes coding structural proteins, enzymes and envelope proteins, those are *gag, pol* and *env*. Second group consists of genes responsible for accessory and regulatory proteins; genes *vif, tat, vpr, rev, vpu* and *nef*. Among those *vpu* gene is exclusively found in HIV-1 (Cohen *et al.* 1988).

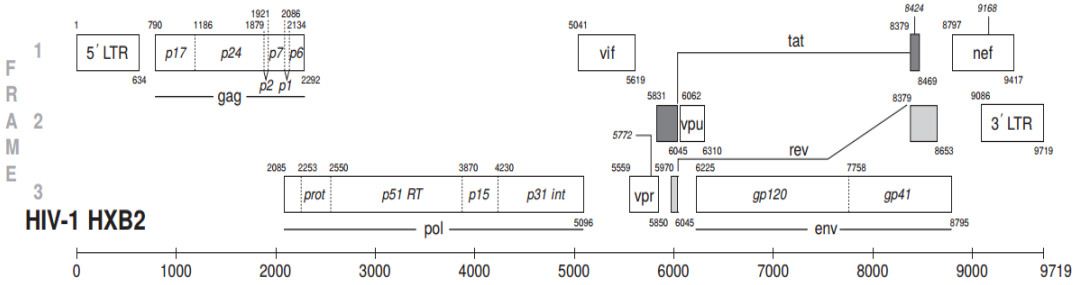


Fig. 1. HIV-1 genomic organization. All nine genes organized in the proviral DNA genome are shown. Three major genes *gag, pol* and *env* and in rectangles their encoded proteins are depicted. Genes *tat* and *rev* are depicted with pattern of spliced exons (shaded units). Numbering is based on HXB2 isolate, a reference sequence from Los Alamos National laboratory HIV database. Adapted from Foley *et al.* 2016.

In detail, *gag, pol* and *env* encode polyprotein precursors subsequently processed in particular stages of viral life cycle. Gag polyprotein is cleaved in six protein entities. Matrix protein (MA) also known as p17 is responsible for recruitment of viral and host factors. MA itself is located under envelope. Capsid protein (CA) or p24 contains C-terminal and N-terminal domains, responsible for dimerization or pentamerization and hexamerization of CA units, respectively (Pornillos *et al.* 2011). Nucleocapsid (NC) or p7 is multifunctional protein with ability to bind viral RNA and plays role in packaging and reverse transcription (Levin *et al.* 2010). Remaining proteins p1, p2 and p6 have some minor functions in capsid maturation or serve as spacer peptides (Bell & Lever 2013).

Enzymes of HIV-1 are auto-cleaved from Gag-Pol precursor polyprotein. Ratio of synthesized Gag and Gag-Pol polyproteins is critical for viral life cycle (Dinman & Wickner 1992). Pol ORF encodes protease with autocatalytic activity, reverse transcriptase and integrase (Foley *et al.* 2016). Particular functions in detail are discussed in chapter about viral life cycle.

Last important member of HIV-1 main genes is *env*. Polyprotein product of this gene, Env (or gp160), is processed by cellular enzymes. Final products are gp120 and gp41, both responsible for membrane fusion and recognition of CD4 receptor and chemokine co-receptors on target cells (Dalgleish *et al.* 1984; Clapham & McKnight 2001).

Small proteins of HIV-1 are classified into two groups. First are regulatory proteins, translated from multiple spliced mRNA (fig. 1.). These are necessary for viral replication. Tat has crucial part in early phase of viral infection. This transactivator binds stem loop of the trans-activation response element on the 5'LTR and recruits cellular kinase that phosphorylates RNA polymerase II (Nekhai *et al.* 1997). Other regulatory protein is Rev. It contains nuclear localization signal which allows its accumulation in the nucleus. There it

recognizes Rev responsible element at unspliced and partially spliced viral mRNA, binds it and facilitates transport of the unspliced transcripts out of the nucleus due to own nuclear export signal (Malim *et al.* 1989).

Second group of accessory proteins includes four members. Nef is multifunctional protein, which acts as a virulence factor. It helps to shape host environment during viral infection. Main sign of this reshaping is lower expression of CD4 on the surface of the infected cells (Aiken *et al.* 1994). It can also lower the level of cytotoxic T-lymphocyte-associated protein 4, negative regulator of T-cell activation (El-Far *et al.* 2013). Both these effects aid to narrow T-cell response for activation and enhance viral dissemination. Vif protein acts mainly as a protector of budding virions from incorporation of APOBEC3G, cellular protein with antiviral function. Vif targets APOBEC3G for ubiquitination and proteasomal degradation (Yu *et al.* 2003). Vpr is important cofactor of the genomic DNA integration. It helps to transport pre-integration complex (complex of DNA and proteins) to the nucleus. This allows viral genomic DNA to be integrated even in the genome of nondividing cells with stable nuclear membrane (Heinzinger *et al.* 1994). Vpu is responsible for degrading of CD4. Consequently, CD4 are not binding Env in endoplasmatic reticulum, allowing proper budding (Willey *et al.* 1992). Genes in genome of HIV-1 are surrounded by two long terminal repeats, 3'LTR and 5'LTR (fig. 1). They consist of U3, R and U5 regions in this order and contain secondary RNA structures responsible for integration and regulation of transcription of integrated provirus (Foley *et al.* 2016).

3.2.3. Virion structure

HIV-1 is enveloped virus with cone-shaped capsid. Although, this shape is typical for matured virion, immature version also exists (fig. 2). It was reviewed in Sundquist & Krausslich (2012) that spherical shape of immature particle is composed of Gag polyprotein. Later in maturation Gag is cleaved, CAs create cone-shaped capsid core and NCs and MAs adopt their places inside capsid or below envelope, respectively.

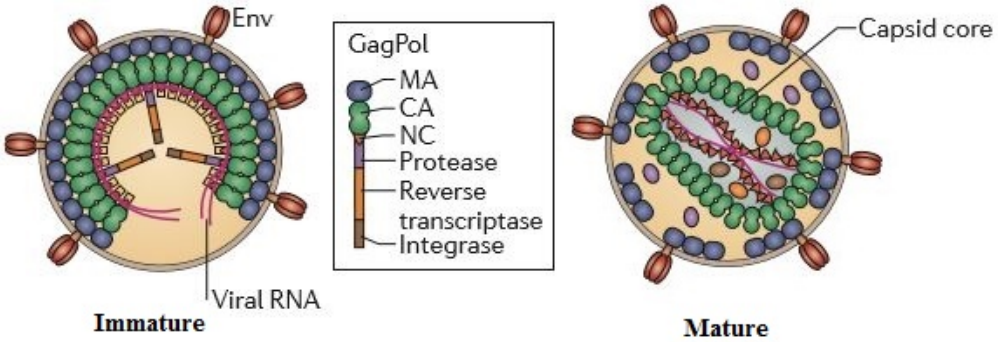


Fig. 2. HIV-1 virion structures. Structure and organization of polyproteins in the immature virion is spherical. After maturation, infective virion has polyproteins cleaved and reorganized. Difference is mainly in capsid structure with genomic RNA inside. Adapted from Freed 2015.

Capsid core consists of CA hexamers and pentamers (Ganser *et al.* 1999). Cryoelectron studies also show that minority of virions can contain several capsid cores or core with tube shape. This is due to the flexibility of the hexameric CA units during virion assembly (Briggs *et al.* 2003). Inside the capsid are two molecules of genomic RNA in almost every case (Chen *et al.* 2009). Aside from genomic RNA, virion also contains number of copies (~20) of $tRNA^{Lys}$, annealed to primer binding site inside of the 5'LTR (Huang *et al.* 1994). Ott summarized many cellular proteins within virion, but majority of their functions in viral life cycle remains to be determined (Ott 2008). Accessory and regulatory proteins of viral origin are also present in the virion (Sundquist & Krausslich 2012). Structure and function of viral envelope were recently reviewed by Merk & Subramaniam (2013). Heterodimers gp120/gp41 organize themselves into the trimeric conformation. Surface of HIV accomodates usually from 7 to 14 trimers (Sundquist & Krausslich 2012).

3.2.4. Viral life cycle

HIV-1, similar as any other virus, needs cellular host for own replication. First crucial step in viral life cycle is attachment to the host cell and entry of virion in the cell (fig. 3). HIV-1 gp120/gp41 proteins on surface of the envelope recognize CD4 receptor with high affinity (Dalgleish *et al.* 1984). The complete process from structural and functional point of view was recently reviewed (Klasse 2012). Protein gp120 interacts with CD4 and subsequently with co-receptor CCR5 or CXCR4. Correct glycosylation is necessary (Li *et al.* 1993). Binding triggers conformational change and gp41 fusion peptide is injected into host plasma membrane followed by membrane fusion (Klasse 2012).

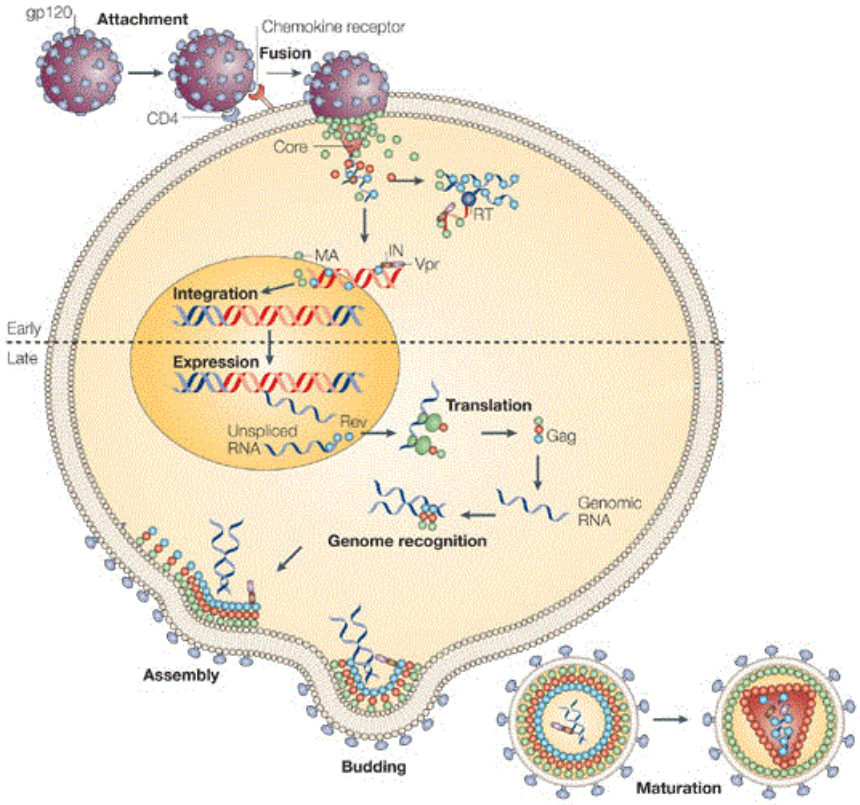


Fig. 3. HIV life cycle. Simplified picture of crucial steps in life cycle of HIV-1. After attachment and fusion, capsid is released in cytoplasm. Genomic RNA is reversely transcribed to proviral DNA and integrated in the host genome. Subsequent expression of viral proteins results in assembly of immature virion on the cell membrane. Proteins undergo proteolytic cleavage in released virion and the virion maturates. Adapted from D'Souza & Summers 2005.

After the fusion capsid is uncoated and reverse transcription complex is formed. Complex contains reverse transcriptase, integrase, protease and matrix protein with Vpr; it also includes cellular histones (Karageorgos *et al.* 1993). Its purpose is to reversely transcribe genomic RNA into the linear dsDNA. Reverse transcription is initiated by $tRNA^{Lys}$ which binds to 18 complementary nucleotides on the binding site (Arts *et al.* 1996). After successful reverse transcription the preintegration complex is formed, consisting of viral DNA, integrase, Vpr, MA and many cellular components and is transported to the nucleus via nucleopore (Rey *et al.* 1998). Process of integration was reviewed by Krishnan & Engelman (2012). HIV DNA becomes an intrinsic part of host genetic information and it can be integrated almost anywhere in the host genome. Incorporated HIV DNA is called provirus.

Transcription of proviral DNA is catalyzed by host RNA polymerase II. Tat protein is important in recruiting cellular factors enabling phosphorylation of polymerase and successful elongation of mRNA full transcript (Kao *et al.* 1987). Tat protein also helps post-transcriptionally with capping of the viral mRNA (Zhou *et al.* 2003). Further post-

transcriptional modification includes polyadenylation on 3' end (Brown *et al.* 1991). mRNA can be transported from nucleus by Rev protein and serves as genomic RNA for new virions or as a template for Gag-Pol polyprotein synthesis (Malim *et al.* 1989). Except full genomic mRNA, HIV-1 manifests multiplex splicing pattern. In work of Karn & Stoltzfus (2011) more than 40 different spliced mRNAs were reviewed. They can be divided in two classes: singly spliced mRNAs and multiply spliced mRNAs. Authors also describe main difference between early and late phase of replication. In early phase, multiply spliced mRNAs (for Rev, Tat and Nef) are exported to cytoplasm by ordinary cellular pathways. Later on, Rev protein returns to the nucleus and start exporting unspliced or singly spliced mRNAs, moving thus cycle to later phase (Karn & Stoltzfus 2011).

Translation of small proteins, Gag and Gag-Pol polyproteins occurs in cytoplasm meanwhile translation of Env takes place on ribosomes attached to ER. Synthesis of Gag or Gag-Pol from same mRNA is determined by -1 ribosomal frameshift, which allows translation of Pol ORF (Dinman *et al.* 1991) (fig. 1). Guerrero *et al.* (2015) summarized the ability of HIV-1 to utilize host translational apparatus. Initiation can start through cap dependent event or on IRES. Env and Vpu share common mRNA within different ORFs, which is common space saving mechanism of many viruses. Start of Vpu ORF is upstream of Env (fig. 1). Translation from Env ORF is enabled by ribosomal leaky scanning and missing starting codon of Vpu (Schwartz *et al.* 1990). Env, synthesized on endoplazmatic reticulum is transported to Golgi apparatus, cleaved by cellular enzymes and glycosylated (Li *et al.* 1993; McCune *et al.* 1988).

Also, assembly and maturation, final steps in viral life cycle, were recently reviewed (Freed 2015). Gag polyprotein is targeted to cytoplasmic membrane and genomic RNA, previously transferred to cytoplasm by Rev, binds to Gag polyprotein via NC domain (Malim *et al.* 1989; Freed 2015). Processed gp120/gp41 complexes are transported to the membrane by vesicular transport and via interaction of MA domain of Gag polyprotein with C-terminal end of gp41 (Freed 2015). Next, cellular complex hijacked by Gag mediates budding of the immature virion. Budded virion then proceeds to maturation, where protease first cleaves itself from Gag-Pol polyprotein, followed by stepwise cleaveage of Gag and Gag-Pol polyproteins to particular proteins (Pettit *et al.* 1994). Capsid protein afterward builds cone-shaped core and virion becomes matured and fully infective (fig. 2).

3.2.5. HIV pathogenesis

HIV-1 name was chosen to describe its ability to cause immunodeficiency in humans. Basis for this immunodeficiency is in infection of $CD4^+$ lymphocytes, known as T helper lymphocytes. HIV-1 also infects macrophages and dendritic cells, which helps the virus to disseminate into the $CD4^+$ lymphocytes (Granelli-Piperno *et al.* 1998). Acute phase of the disease is characterized with fast drop of $CD4^+$ lymphocytes due to directly or indirectly induced cell death. Infected $CD4^+$ lymphocytes can form short living syncytia (Sylwester *et al.* 1997) or they can be induced to apoptosis by $CD8^+$ T lymphocytes. Fast depletion of $CD4^+$ lymphocytes is accompanied by great increase of viral load in plasma (fig. 4). After few weeks of acute infection virus proceeds to latency, chronic state accompanied by slow decline of $CD4^+$ lymphocytes, which slowly over years progresses towards AIDS. WHO is responsible for current definition of how to diagnose AIDS based on amount of $CD4^+$ lymphocytes in patient (http://www.who.int/hiv/pub/en/). During AIDS, patients have weak immune system; therefore they easily contract secondary and opportunistic infections, often with fatal results. Although therapy is highly effective today, it is still impossible to clear the virus once latency is established. Studies confirm that in patients on highly active antiretroviral therapy (HAART), peripheral monocytes and $CD4^+$ lymphocytes (both activated and resting) are still source of the virus (Zhu *et al.* 2002).

Important part in HIV pathogenesis is viral tropism. We distinguish HIV based on co-receptor they use for entry: R5 (utilizing CCR5 co-receptor) and X4 (utilizing CXCR4 co-receptor). Mixed and dual tropic viruses (R5/X4) with ability to bind both co-receptors also exist (Xiang *et al.* 2013). Recognition of co-receptor is mostly based in V3 hypervariable loop of gp120 (Sharon *et al.* 2003). Study by Ochsenbauer and her colleagues (2012) analyzes founder effect of predominantly R5-utilizing T-tropic virus (infects $CD4^+$ lymphocytes through CCR5 co-receptor) during beginning of infection. Virus quickly adapts to R5-utilizing M-tropism, ability to infect macrophage/monocyte cells, extending range of target cells (Ochsenbauer *et al.* 2012). Most of the virus population starts switching towards X4 tropism in later stage of infection. Desired effect is ability to infect naïve $CD4^+$ lymphocytes with high level of CXCR4 co-receptor (Shankarappa *et al.* 1999). Switch is accompanied with drastic loss in $CD4^+$ lymphocytes due to the syncytia formation and followed by progress of disease (Berger *et al.* 1999).

3.2.6. Disease progression

Based on the disease progression, we can divide patients in several groups. The most common group are progressors (PROs) with slow decline of $CD4^+$ T cells count and slow increase in viral load over several years (fig. 4). Rare group are long-term nonprogressors (LTNPs) that have asymptomatic HIV over 10 years after infection with stable or very slowly decreasing $CD4^+$ T cells count in the absence of antiretroviral therapy. LTNPs can be further divided according to level of control of viremia into viremic controllers, noncontrollers and elite controllers. LTNPs are able to maintain $CD4^+$ T cell level above 500 per microliter of blood for at least eight years of infection without therapy and low level of viral RNA in plasma (Madec *et al.* 2009). Elite controllers in addition to preserving stable $CD4^+$ T cell count maintain undetectable levels of viral RNA in plasma (below 50 copies per milliliter) (Grabar *et al.* 2009). Last group, the rapid progressors, develops AIDS very quickly, usually within 3 years (Casado *et al.* 2010).

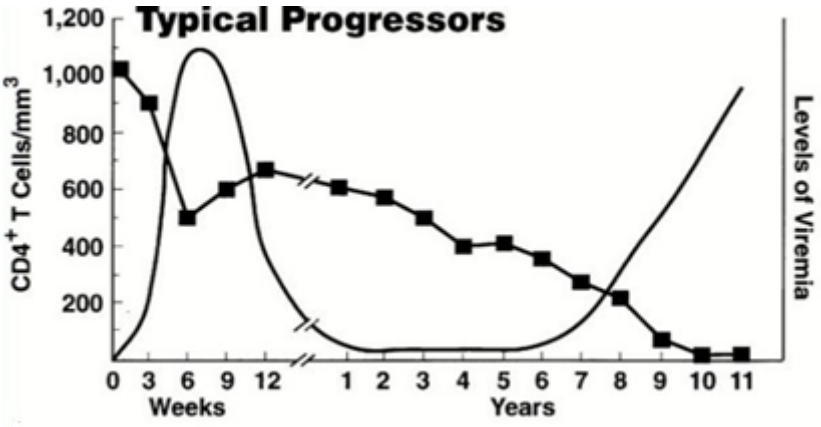


Fig. 4. Curves of disease progression in typical progressor. Time is plotted on axis x, on axis y1 is plotted $CD4^+$ T cells count and on y2 is plotted level of viral RNA in blood. Curve with squares represents $CD4^+$ T cells. Solid line represents viremia. Primary infection constitutes peak in viral load. High viremia is accompanied with rapid level drop of $CD4^+$ T cells. Latent phase is established and viral load drops again. In subsequent years, level of $CD4^+$ T cells is slowly declining eventually accompanied with increase in viral load. Adapted from Pantaleo & Fauci 1996.

Disease progression varies significantly between patients and can be influenced by many host and viral factors. One of the most protective effects is deletion in CCR5 co-receptor. 32bp long deletion introduces stop codon, resulting in nonfunctional CCR5 co-receptor (Dean *et al.* 1996). Protective mechanism is not yet completely clear, but this deletion provides natural resistence for entry of CCR5 tropic viruses (Agrawal *et al.* 2004). Another host protective factor in AIDS development are HLA class I alleles. Particularly

HLA-B*27, HLA-B*57 and HLA-B*13 are associated with strong cytotoxic T-lymphocytes response (Migueles *et al.* 2000; Brettle *et al.* 1996, Shahid *et al.* 2015) leading to slow disease progression. Individuals expressing protective HLA I alleles exhibited strong induction of cytotoxic T-lymphocytes response early in the infection targeted against a specific epitope within HIV-1 capsid protein (Streeck *et al.* 2007). On the other hand, APOBEC3G, well-characterized antiviral restriction factor known to inhibit also HIV-1, is not elevated in elite controllers and cannot explain control of disease (Gandhi *et al.* 2008). Another example of host restriction factor is TRIM5-alpha. This protein carries several functions contributing to HIV-1 restriction, including receptor for viral capsid and signal transducer (Pertel *et al.* 2011).

Many studies have been focusing on viral factors in disease progression. Early research proved link between virus genotype and disease progression (Koot *et al.* 1992). Koot's study opened new opportunities to examine particular mutations in genes and their effect on disease progression.

Recent study showed effect of viral single nucleotide polymorphism in LTR of provirus. LTR orchestrates viral gene expression together with host transcription factors and other viral regulators. Study stresses out presence of several viral polymorphisms, one at position 108 in particular and their effect on disease severity. Effect is due to elevated binding of host transcription factors in binding site on LTR caused by mutations at position 108 (Nonnemacher *et al.* 2016). Genotype variation of Env was evaluated for effect in disease progression, particularly in brain microglia. Brain microglias have lower level of CD4 than PBMCs. Mutant N283 is associated with lower CD4 dependence due to new hydrogen bonding with this receptor. This means that higher affinity of Env N283 mutant to CD4 receptor mediates efficient infection of brain microglia. It leads to high rate of HIV-associated dementia among patients with HIV Env variant N283 (Dunfee *et al.* 2006). Role of Gag in disease progression is associated with host HLA class I alleles. Specifically, broad spectrum of CTL (cytotoxic T-lymphocytes) escape mutations was identified and anlalyzed (Boutwell *et al.* 2013). Further, escape mutations are reverted or compensated by secondary mutations in order to re-elevate fitness of the virus (Sunshine *et al.* 2015). To summarize, host factors and viral factors are both valid participators in disease progression.

3.2.7. Quasispecies

HIV population does not consist of homogenous species but rather forms diverse swarm of species termed quasispecies. Concept of quasispecies was introduced to the field of RNA viruses in the work by Steinhauer & Holland (1987). Two major aspects contribute to

HIV-1 quasispecies formation. First, RT has high error rate due to low fidelity and absence of proofreading. Mutation rate of RT is around 10^{-4} per nucleotide per replication cycle (Hu & Hughes 2012). Second, HIV replication is very dynamic, allowing HIV-1 to reach vast amount of possible mutants in few cycles.

Quasispecies theory has implication for HIV transmission and disease progression. During initial infection virus goes through bottleneck, resulting in sequence variability drop. This effect is also accompanied by drop of viral fitness. However, virus quickly establishes dynamic range of similar sequences, which are derived from introducing sequences and form new quasispecies (fig. 5) (Lauring & Andino 2010). We have to consider virus always as a swarm of diversified sequences. They are constantly mutating, evolving and quickly adapting to environment changes and consequently influence progress of infection. For example, virulence is enhanced if broad quasispecies are present in contrast with presence of few even highly specialized variants (Vignuzzi *et al.* 2006).

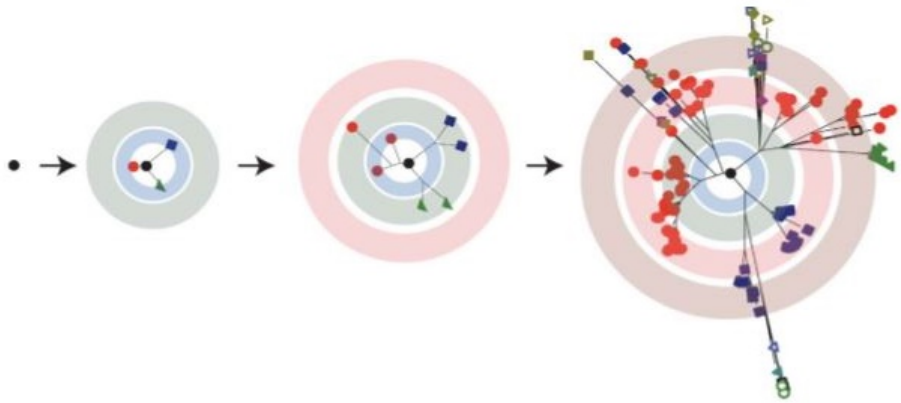


Fig. 5. Development of quasispecies. Simplified diagram of quasispecies development over time. Black spot is original introducing sequence marking center of sequential space. Further from center, sequence is more mutated from the original sequence. Although original sequence does not have to be preserved in host and can be outcompeted, it is kept here for better orientation. Concentric circles portrait individual rounds of replication. In every round new centers are established surrounded by quasispecies in subsequent round of replication (colored spots). Adapted from (Lauring & Andino 2010).

3.2.8. Viral fitness

Viral fitness is complex term, describing ability of virus to survive, develop and reproduce itself in particular environment (Domingo & Holland 1997). Viral replicative fitness is subset of viral fitness, focusing on virus ability to replicate in certain conditions. Another concept, used by many authors, is replicative capacity. Concept focuses on replicative ability of virus however, only in single round of infection.

3.2.8.1. Methodology of viral fitness

Different methods are used for viral fitness determination. It includes *in vivo* and *ex vivo* methods (Quiñones-Mateu & Arts 2001). Growth kinetics and competition experiments are mainly in particular interest, also accompanied today with deep sequencing to detect rare variants in virus population (Zanini *et al.* 2015). Method of growth kinetics portrays amount of virus in each time point. Virus amount can be quantitated by analyzing reverse transcriptase activity or by detecting of p24 protein levels, although this method can examine productivity but not survival of the virus and results can differ (Weber *et al.* 2003). Since growth kinetics is method based on mono-infection, it expresses fitness diference with lesser distinction than competition kinetics (fig. 6). Therefore dual-infection based competition kinetics can reveal small differences in fitness (Dykes & Demeter 2007).

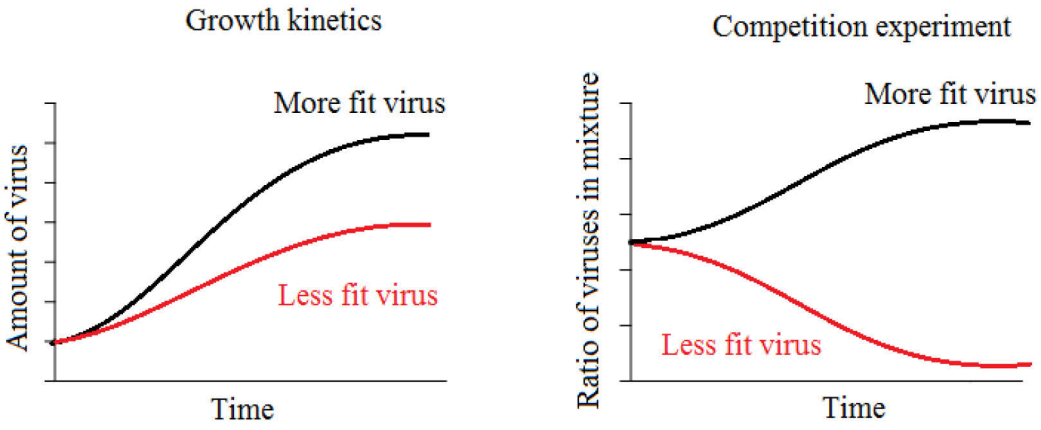


Fig. 6. Growth kinetics and competition experiment diagrams. Growth kinetic assays determine HIV fitness in parallel infection whereas growth competition assays in one dual infection. As it is depicted here, difference between two viruses is more profound in competition experiment.

Deep sequencing method is useful method for detecting genetic variation in quasispecies and fitness of drug resistant HIV. Virus has to react on many host restriction factors and, of course, to potential drug therapy. With deep sequencing it is possible to observe changes in minority of virus population and evaluate effects of restriction factors. It is used also in longitudinal studies observing development of sequence in patients' context (Zanini *et al.* 2015). Deep sequencing is now a high-throughput method and has been successfully used for evaluation of fitness of drug resistant HIV-1 mutants. This method can compare representation of different sequences in presence or absence of particular antiretroviral drugs and draw conclusion of viral fitness advantage of particular mutated sequence over another (Brumme *et al.* 2013).

Another important issue for HIV-1 fitness determination is to consider whether to use HIV-1 primary isolates from patients or recombinant viruses. Both have advantages and disadvantages. Firstly, primary isolates need to be isolated from patients, which is laborious, expensive and time consuming. Also, it was discovered that during isolation shifts in quasispecies may occur (Kusumi *et al.* 1992). On the other hand, HIV fitness determined with recombinant viruses disregards fitness effects of regions outside of the cloned gene(s) and has to be carefully evaluated in context of overall fitness (Weber *et al.* 2006; Dykes & Demeter 2007).

To obtain recombinant virus, several approaches were utilized. Homologous recombination of particular gene and vector was done in mammalian cells (Covens *et al.* 2009). Direct approach of subcloning HIV gene amplicon into the vector carrying rest of HIV sequence was also adapted (Claiborne *et al.* 2015). Further, homologous recombination was done in yeast cells (Dudley *et al.* 2009). Additionally, recombinant viruses were marked by reporter gene. First studies were altering ORFs of HIV genes to introduce reporter gene. First method, introducing two alternative reporter genes in HIV genome without HIV ORFs alteration was done in 2006. Resulting recombinant virus is detectable thanks to fluorescent proteins, either EGFP or DsRed2 (Weber *et al.* 2006). Both Brumme with colleagues (2011) and Covens with colleagues (2009) used green fluorescent protein to detect presence of recombinant virus as well. Absence of probe is not limiting to use viral growth kinetics if virus output is measured by radioactive RT or p24 ELISA assay. Using recombinant virus with reporter gene was proven to be valid approach for HIV-1 replicative fitness evaluation (Weber *et al.* 2006).

3.2.8.2. Transmission and fitness

In transmission, infection is built on few genetic variants of HIV. In addition, it was shown that virus has to be fit in order to be transmitted and its need for fitness is growing with obstacles virus has to overcome during transmission (Carlson *et al.* 2014). Transmitted virus needs to adapt to new environment such as pH, density of target cells or shape of host immune system (Tebit *et al.* 2007). Severe bottlenecks on the onset of every infection are important, because they select HIV species that will establish infection into the new host. In general, transmission bottlenecks influence fitness of HIV-1 and subsequently also pathogenesis (Tebit *et al.* 2007).

Later in infection, virus starts to select variants to adapt to the host environment. Earlier works discovered, that older HIV-1 isolates from 1986-89 outcompeted newer isolates

from 2002-2003 (Arien *et al.* 2005). Although first conclusions were focused on attenuation because of new host adaptation, other works provided model of HIV-1 quasispecies behavior in fitness landscape (Rolland *et al.* 2007). In Rolland's work HIV-1 is defined as gaining robustness instead of losing fitness. This follows theoretical model 'survival of the flattest' (Wilke *et al.* 2001). Fitness landscape is theoretical space, where two dimensions express sequence variability and third dimension express fitness. Isolates on high narrow peak have excellent fitness however; they are very sensitive to any mutation which can push them down from the peak. Flatter and much broader peak occupies lot of space in two sequential dimensions although its fitness is not so high (fig. 7). This flat population is versatile and can easily and quickly search near sequential space for advantage spots (local peaks) and move through space. It is also constant in unstable mutating environment. In overall, fitness of flatter population is on average higher than fitness of population on high narrow peaks and better represents behavior of virus after transmission (Rolland *et al.* 2007; Lauring & Andino 2010).

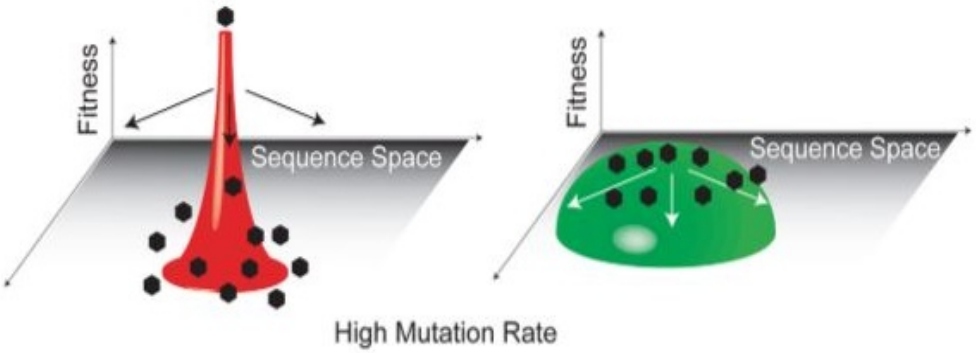


Fig. 7. Survival of the 'flattest'. Consequences of high mutational rate in narrow population are often detrimental. On the other hand, flat population is not so impaired and has time to move in sequence space and search for other local fitness peaks. This demonstrates advantage of keeping flat population as a highly mutating virus. Adapted from Lauring & Andino 2010.

3.2.8.3. Drug resistance

Since introduction of first antiretroviral drug, it was possible to observe how HIV-1 quickly adapts and becomes resistant (Iyidogan & Anderson 2014). HIV-1 is adapted to host and when new strong selective pressure is introduced, e.g. antiretroviral drug, it has to change rapidly. Usually, HIV-1 resistant mutations are selected in the drug target region (reverse transcriptase, protease, integrase or envelope proteins), and these mutations are generally accompanied with decrease of fitness (Yang *et al.* 2015). Virus then has to compensate for fitness loss. With introduction of HAART, combination of several antiretroviral drugs, it has been for virus very difficult to select resistant mutations for each drug and simultaneously

keep sufficient fitness (Iyidogan & Anderson 2014), although not impossible. HIV usually compensates for fitness loss by selection of additional compensatory mutations in the drug target region. Interestingly, in case of HIV-1 protease, resistant mutations in protease are compensated by mutations selected in *gag* and *pol* regions near the processing sites. Mutated polyproteins are then effectively cleaved by mutated protease so virus maintains its resistance and concurrently compensates for fitness loss due to poor catalytic effectivity of mutated protease (Kozisek *et al.* 2012).

If fitness price from resistant mutation is not high, compensatory or reverse mutations develop slowly or not at all after transmission to new patient (Yang *et al.* 2015). This has a big influence on drug resistance persistence in population. It was shown that transmitted drug resistance is maintained mainly in therapy naïve patients (Drescher *et al.* 2014). In general, fitness is important factor for survival of HIV-1 during therapy.

3.2.8.4. Fitness and disease progression

Amount of $CD4^+$ T cells is widely used as gold standard to evaluate HIV disease progression (Fahey *et al.* 1990). Other important marker of disease progression is number of HIV RNA copies per ml of blood (viral load; VL). Other studies suggested link between viral fitness and disease progression, stating that VL can not predict disease progression unless reaching extreme values (both low and high) (Prince *et al.* 2012). Authors followed up and used Gag recombinant viruses to elucidate role of viral replicative fitness as a predictor of disease progression. Furthermore, they showed replicative fitness, VL and presence of protective HLA alleles as factors in prediction of disease progression (Claiborne *et al.* 2015). Moreover, higher viral fitness was linked not only to $CD4^+$ T cells decline but also to higher $CD8^+$ T cells activation with lower cytotoxic effect and higher proviral burden in $CD4^+$ T cells (Claiborne *et al.* 2015).

Considering host immunological factors, viral replicative fitness and disease progression, several studies are important to mention. CTL response escape mutations, accumulated in *gag* have noticeable effect on virus fitness. Virus is escaping CTLs by mutating epitopes of Gag protein and evades presentation of epitope on HLA. HLA driven selection has considerable effect on mutagenesis of *gag*, mainly in the first year of infection (Brumme *et al.* 2008). Boutwell and his colleagues showed that majority of escape mutations cause also reduction in replicative fitness. Pattern of mutations is linked with genetical status of the patient, namely with specific HLA class I alleles (Boutwell *et al.* 2013, Shahid *et al.* 2015) and accompanied by compensatory mutations. Thus, fitness cost is apparent in acute

phase; later in the disease it can dissipate. In chronic infection, HIV-1 has time to acquire compensatory mutations and balance fitness cost (Gijsbers *et al.* 2013, Liu *et al.* 2014). Furthermore, development of compensatory mutations happens more frequently than reversion of escape mutations (Sunshine *et al*. 2015). Effort to compensate original lost in fitness is driving HIV-1 evolution during later stage of infection. Virus fitness cost in patient with protective alleles has long-lasting effect up to chronic stage (Brockman *et al.* 2010).

Study of elite controllers provides extensive perspective for effect of viral fitness on disease progression. Persistent low viremia in elite controllers is related to fitness of virus and immune response of the host. Strong pressure from host CTL is one of examples, how elite controllers maintain low fit virus (Troyer *et al.* 2009). However, poor viral fitness can not be explained solely by presence of protective HLA alleles associated with vigorous CTL response. This is exemplified by missing protective alleles in half of elite controllers. Thus, other factors have to contribute (fig. 8). *Gag-protease* derived from elite controllers and recombined in HIV-1 vector (pNL4-3) showed lower replicative fitness than Gag-protease derived from chronic progressors (Lobritz *et al.* 2011). On the contrary, virus isolates from elite controllers displayed full replication potential (Blankson *et al.* 2007).

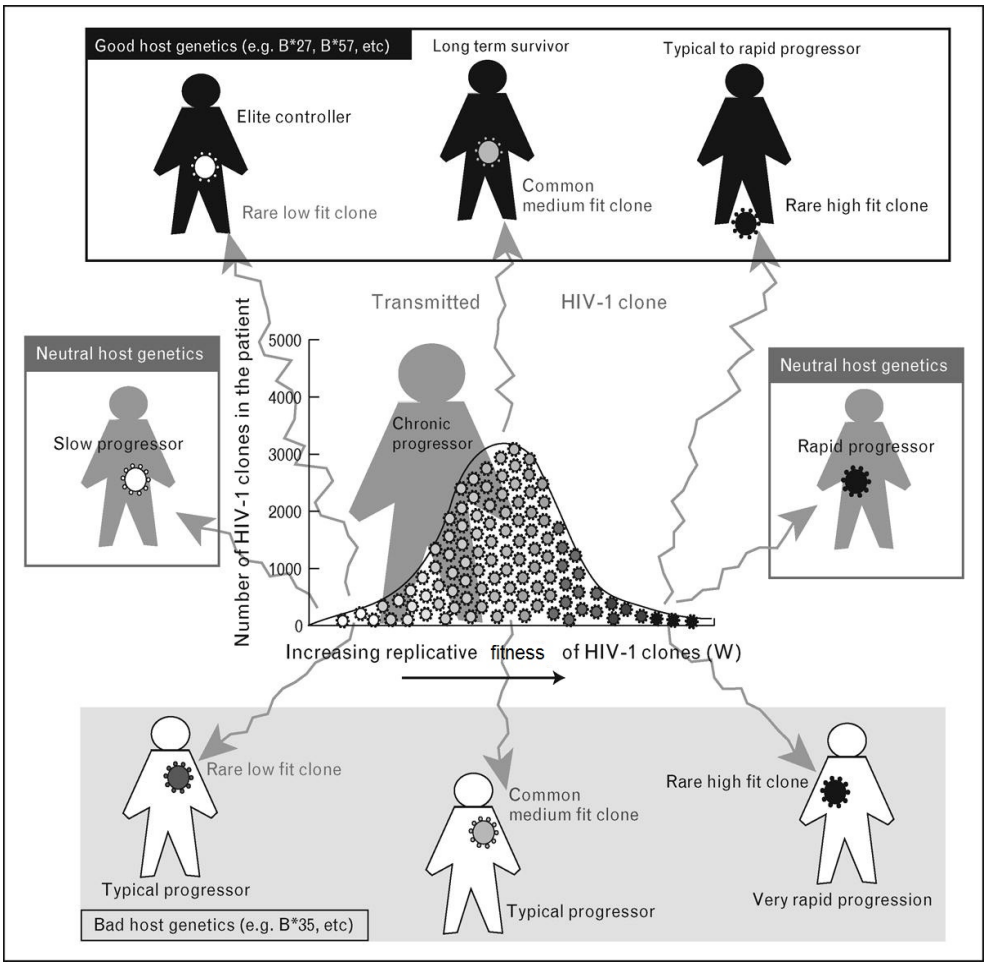


Fig. 8. Mutual relationship between virus fitness and host genetics. Distribution of replicative fitness is following Gaussian distribution, where only few clones are introduced to new patient during transmission. Elite controller is depicted as a result of both, low fit clone transmitted and good host genetics (illustrated by protective alleles). Most common example is classical chronic progressor accounting for ~90% of all infections. Adapted from Lobritz *et al.* 2011.

To summarize, viral replicative fitness is important characteristics for disease progression. Importance of *gag* region for virus replicative fitness was shown in several studies, highlighting the impact of CTL escape mutation. Following previous research, emphasis of this thesis is given to study of *gag* contribution to overall replicative fitness. Replicative fitness is measured through recombinant viruses and primary isolates. Furthermore, replicative fitness development, evaluated by *gag* recombinant viruses, and disease progression are analyzed.

4. Material and methods

4.1. Plasmids

pNL4-3-EGFP: plasmid containing complete sequence of HIV laboratory strain NL4-3 with green fluorescence probe between *env* and *nef*. Contains ampicillin resistance gene.

p83.2: NL4-3 strain derived plasmid containing region from 5'LTR up to part of *vpr* gene. Contains ampicillin resistance gene.

pREC$_{nfl}$-TRP-Δgag/URA3: plasmid containing near full length of HIV where *gag* is replaced by *Ura3* gene. Contains *Trp1* gene and ampicillin resistance gene.

4.2. DNA primers

Sequences are written in direction from 5' to 3'.

Primers for reverse transcription of viral RNA in cDNA and amplification of *gag* with surrounding sequence in external PCR.

GAG P17 FWD 8M	GCGAAAGTAAAGCCAGAGGAGATCTCTCGACGCAVGRCTCGGCTTGCT
POL PRO REC CON BWD 7	ACTAATGCTTTTATTTTTTCTTCTGTCAATGGCCAYTGTTTRAYYYTTGG

Primers for amplification of *gag* with surrounding sequence in nested PCR.

GAG P17 FWD 7M	GTAAAGCCAGAGGAGATCTCTCGACGCAGGACTCGGCTTGCTGARGYGCG
POL PRO REC CON BWD 8	TTTCTTCTGTCAATGGCCATTGTTTAACTTTTGGNCCATCCATHCCTGGY

Primers for introduction of MluI restriction site by site directed mutagenesis. Substituted nucleotides are underlined.

MluI-SDM-F	GTACTGGATGTGGGCGACGCGTATTTTTCAGTTCCC
MluI-SDM-R	GGGAACTGAAAAATACGCGTCGCCCACATCCAGTAC

Primers for introduction of PspXI restriction site by site directed mutagenesis. Substituted nucleotides are underlined.

PspXI-SDM-F	GCCCGAACAGGGACTCGAGGGCGAAAGTAAAGCCAGAGG
PspXI-SDM-R	CCTCTGGCTTTACTTTCGCCCTCGAGTCCCTGTTCGGGC

Primers for deletion of native MluI restriction site in pREC$_{nfl}$-TRP-Δgag/URA3-PM plasmid by site directed mutagenesis. Substituted nucleotides are underlined.

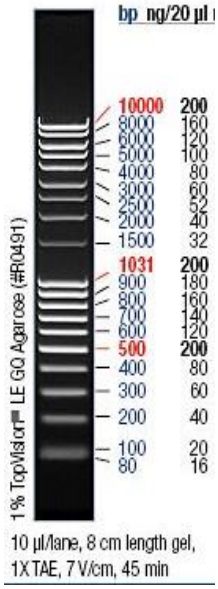
MluI-delSDM-F	CAGTGAGCGCGCGTATTCGTGTTGACATTGATTATTG
MluI-delSDM-R	CAATAATCAATGTCAACACGAATACGCGCGCTCACTG

Primers to insert linker in pNL4-3-EGFP-PM-linker plasmid.

PspXI-MluI-F	TCCAGTVCTCGAGBGGCGCGCCACGCGTATAGCA
PspXI-MluI-R	TGCTATACGCGTGGCGCGCCVCTCGAGBACTGGA

4.3. Molecular markers

DNA molecular marker: MassRuler™ DNA Ladder Mix, ready-to-use (ThermoFisher).



4.4. Bacterial and yeast strains, viruses and cell lines

Bacterial strain TOP10 Electrocomp™ *E. coli* cells: *F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ-* (ThermoFisher)

Yeast strain SL1463: *Saccharomyces cerevisiae MATα leu2 ura3-52 trp1 his3-Δ200* (Case Western Reserve University)

HIV-1: NL4-3 strain (obtained from NIH AIDS Reagent Program)

HIV-1: B2 primary isolate 92US076 (obtained from NIH AIDS Reagent Program)

MT4: Human T-cell line transformed by Human T-cell lymphotropic virus type 1 (obtained from NIH AIDS Reagent Program).

TZM-bl: HeLa derived cell line expressing CD4 and CCR5. Contains luciferase and ß-galactosidase genes under control of the HIV-1 promoter (obtained from NIH AIDS Reagent Program).

HEK293T: Human embryonic kidney cell line transformed with large T antigen from Simian virus 40 (Stanford University).

OKT3: Mouse x mouse hybridoma, producing IgG2a against the human T cells CD3 antigen (European Collection of Cell Cultures).

4.5. Culture media

4.5.1. Culture media for bacteria

LB medium: 2% (w/v) LB broth (Sigma-Aldrich)

SOC medium (Invitrogen): 2% (w/v) Tryptone, 0.5% (w/v) Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

LB agar plates: 2% (w/v) LB broth 1.2% (w/v) Agar, Ampicillin (all Sigma-Aldrich)

4.5.2. Culture media for yeast

YEPD medium: 1% (w/v) Yeast Extract, 2% (w/v) Peptone, 2% (w/v) Dextrose (all Sigma-Aldrich)

YEPD plates: 1% (w/v) Yeast Extract, 2% (w/v) Peptone, 2% (w/v) Dextrose, Bacto agar (all Sigma-Aldrich)

4.5.3. Culture media for mammalian cells

DMEM medium with 10% serum: DMEM (Dulbecco's Modified Eagle's Medium), 10% (v/v) Fetal Bovine Serum HI (both Biotech), 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (Sigma-Aldrich)

DMEM without serum: DMEM (Dulbecco's Modified Eagle's Medium, Biotech), 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (Sigma-Aldrich)

Freezing medium: 89% (v/v) Fetal Bovine Serum HI (Biotech), (v/v) 10% Dimethyl sulfoxide, 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (both Sigma-Aldrich)

RPMI medium with 10% serum: RPMI 1640 with L-Glutamine, 10% (v/v) Fetal Bovine Serum HI, 1% (v/v) 1M HEPES (all Biotech), 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (Sigma-Aldrich)

RPMI medium - phenol RED free: RPMI 1640 Medium, no glutamine, no phenol red, 1% (v/v) 200mM L-glutamine (both Sigma-Aldrich), 10% (v/v) Fetal Bovine Serum HI, 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml), 1% (v/v) 1M HEPES (all Biotech)

PBMCs medium: RPMI 1640 with L-Glutamine, 10% (v/v) Fetal Bovine Serum HI, 1% (v/v) 1M HEPES (all Biotech), 1% (v/v) PHA (100U/ml), 0.001% (v/v) IL-2 (100 ng/µl) (both Life Technologies), 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (Sigma-Aldrich)

PBMCs without PHA: RPMI 1640 with L-Glutamine, 10% Fetal Bovine Serum HI, 1% (v/v) 1M HEPES (all Biotech), 1% (v/v) Penicillin (10,000 U/ml) Streptomycin (10 mg/ml) (Sigma-Aldrich), 0.001% (v/v) IL-2 (100 ng/µl) (Life Technologies),

OKT3 medium (Sigma Aldrich): DMEM, 2mM glutamine, 20% Fetal Bovine Serum

4.6. Antibiotics

Ampicillin (Sigma-Aldrich) (working concentration 100 µg/ml).

Penicillin (10,000 U/ml), Streptomycin (10 mg/ml) mixture for mammalian cells cultivation (Sigma-Aldrich).

4.7. Sterilization

Used solutions and media were sterilized by liquid cycle at 121°C in autoclave for 15 minutes. Laboratory glass was sterilized at same conditions. Solutions unsterilized by autoclave were sterilized through filter with 0.22µm pores. Sterilization of inoculating loops and cell spreaders was done by denatured ethanol and flame. All liquids and liquid

contaminated tools and containers in BSL-3 laboratory were disinfected in SAVO diluted with water at ratio 1:2 and autoclaved.

4.8. Work with DNA and RNA

4.8.1. Isolation of HIV RNA from plasma

Isolation of HIV RNA from plasma was done by commercial QIAamp Viral RNA Mini Kit (Qiagen) according to provided protocol. RNA was eluted once by 60 µl of AVE buffer.

4.8.2. Isolation of RNA from virus particles

Isolation of RNA from virus particles was done QIAamp Viral RNA Mini Kit (Qiagen) according to provided protocol. RNA was isolated from 70 µl of virus supernatant and eluted once in 60 µl of AVE buffer.

4.8.3. Polymerase chain reaction

For amplification of viral genetic information three step PCR in Mastercycler® pro (Eppendorf) was used.

Reverse transcription annealing was performed using primer POL PRO REC CON BWD 7. Reaction was done in 12.5 µl reaction volume including 0.5µM primer, 1mM dNTPs and 5 µl of RNA. Annealing was performed at 65°C for 5 minutes and 4°C for 5 minutes. cDNA synthesis was done with annealed mixture, 1x FS buffer, 100 U SuperScript™ III Reverse transcriptase (both Invitrogen), 10mM DTT and 40 U RNAse inhibitor (New England BioLabs). Synthesis was performed at 50°C for 60 minutes and 70°C for 15 minutes.

External PCR reaction was performed using primers POL PRO REC CON BWD 7 and GAG P17 FWD 8M. Amplification was done in 50 µl reaction volume including 1mM $MgCl_2$, 200µM dNTPs, 0.4µM both primers, 1x Pfu buffer, 3.75 U of PfuTurbo polymerase (both Agilent) and 5 µl of cDNA template. PCR reaction performed at 95°C for 2 minutes for denaturation followed by 35 cycles of 92°C for 30 seconds denaturation, 55°C for 30 seconds annealing and 72°C for 5 minutes extension. PCR was finished by last extension at 72°C for 10 minutes.

Nested PCR reaction was performed using primers POL PRO REC CON BWD 8 and GAG P17 FWD 7M. Amplification was done in 50 µl reaction volume including 0.5mM $MgCl_2$,

200μM dNTPs, 0.4μM both primers, 0.31 U of PfuTurbo polymerase (Agilent), 1x Taq buffer and 1.875 U Taq DNA polymerase (both New England BioLabs) and 5 μl of DNA from external PCR as template. PCR reaction performed at 95°C for 2 minutes for denaturation followed by 35 cycles of 92°C for 30 seconds denaturation, 55°C for 30 seconds annealing and 72°C for 3 minutes extension. PCR was finished by last extension at 72°C for 10 minutes.

4.8.4. Agarose electrophoresis

Relevant amount of agarose was dissolved in 1x TAE buffer (40mM Tris, 20mM acetic acid, 1mM EDTA). During my experiments 0.8% to 3% agarose gels were prepared. Mixture of agarose and TAE buffer was boiled in microwave. DNA samples were mixed with 6X DNA Loading Dye (ThermoFisher Scientific). Gels run at 6.25 V/cm. Finished gel was stained with ethidium bromide solution (0.5 μg per ml of distilled H_2O) for 20 minutes and thereafter photographed.

4.8.5. PCR purification

Purification of PCR products was done by commercial QIAquick PCR Purification Kit (Qiagen). Procedure was following provided protocol. DNA was eluted in 30 μl of ddH_2O.

4.8.6. Site-directed mutagenesis

Site-directed mutagenesis was done by commercial kit QuikChange II XL (Agilent Technologies). Reaction was done analogously to PCR reaction using one set of mutating primers. Total volume of reaction was 50 μl, including 1x reaction buffer, 1μl of provided dNTP mix, 0.2 μM both primers, 3 μl of QuikSolution and 2.5 U of PfuTurbo polymerase (Agilent). Template was 50 ng of plasmid dsDNA. Mutagenesis reaction performed at 95°C for 1 minutes for denaturation followed by 18 cycles of 95°C for 50 seconds denaturation, 60°C for 50 seconds annealing and 68°C for 15 minutes extension. PCR was finished by last extension at 68°C for 7 minutes. Each sample was treated with DpnI at 37°C for 1 hour to degradate parental, nonmutated DNA. DpnI treated mixture was transformed in *E.coli* by electroporation.

4.8.7. DNA isolation from gel

Isolation of DNA from agarose gels was done by two different methods.

4.8.7.1. Isolation of DNA from gel by excision

DNA was resolved on agarose gel and correct size bands were excised under UV light. Sample was processed by QIAEX II Gel Extraction Kit (Qiagen). Procedure followed provided protocol and DNA was dissolved in 20 µl of ddH_2O.

4.8.7.2. Isolation of DNA by E-gel technique

Isolation of majority of bands was done by E-Gel™ Electrophoresis System (ThermoFisher). Pre-casted 0.8% E-Gel CloneWel gels stained with SYBR® with 8 wells in two rows were used (ThermoFisher). Samples of DNA were loaded in upper row of wells with E-Gel 96 High Range DNA Marker in designated well. Samples run for required amount of time, until correct size band reached second row of wells, filled with ddH_2O. When band disappeared in water filled well pure DNA was collected in a tube. Presence of correct band was verified by standard agarose gel electrophoresis.

4.8.8. DNA isolation from bacteria

Isolation of DNA from bacterial cells was done by QIAprep Spin Miniprep, Midiprep or Maxiprep Kits (Qiagen). Method was done according to provided protocol and DNA was eluted in 50, 100 or 500 µl of ddH_2O, respectively. Presence of DNA was verified by standard electrophoresis.

4.8.9. DNA isolation from yeast

Yeast colonies grown on CSM-TRP 5-FOA plates were resuspended in 3 ml of YEPD media. Then, 1.5 ml of mixture was centrifuged for 5 seconds to collect yeast at the bottom of the tube. Pellet was resuspended in 200 µl of breaking buffer (20 % (v/v) 10% Triton X-100, 10 % (v/v) 10% SDS, 2 % (v/v) 5M NaCl, 1 % (v/v) 1M Tris-Cl (pH=8), 0.2 % (v/v) 0.5M EDTA). Suspension was mixed with 0.3 g of glass beads (425 - 600 µm in diameter) and 200 µl of UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v:v; all Sigma-Aldrich). Mixture was vortexed for two minutes and centrifuged at 18000 g for 5 minutes at room temperature. 50 µl from aqueous phase was collected and precipitated with 700 µl of absolute ethanol. Mixture was centrifuged for 10 minutes at previous conditions and supernatant discarded. Pellet was washed by 70% ethanol and dissolved in 20 µl of ddH_2O. DNA was transformed in One Shot TOP10 Electrocomp *E. coli* (ThermoFisher) by electroporation.

4.8.10. DNA sequencing

DNA was sequenced by GATC Biotech company, using Sanger sequencing methodology. Mixture for each sequencing contained: DNA sample (at least 100 ng), 10μM appropriate primer (5 μl), ddH_2O up to 10 μl. Results were analyzed in SeqMan Pro program™ from DNASTAR® and full sequence contigs were constructed.

4.8.11. Measuring concentration of DNA

Concentration of DNA was measured using NanoDrop™ ND-1000 (ThermoFisher). Operating program was NanoDrop 1000 Operating Software, version 3.8.1.

4.8.12. DNA restriction by enzymes

All used restriction enzymes were from New England BioLabs. All mixtures consist of: DNA, restriction buffer (New England BioLabs) 5 μl, 10 U of restriction enzyme per μg of DNA and ddH_2O up to 50 μl. Digest proceeds in temperature suitable for specific restriction enzyme. Digested DNA was evaluated by agarose gel electrophoresis.

4.8.13. Partial digestion

Partial digestion method is analogous to standard digestion by restriction enzymes. However, restriction enzyme was diluted to achieve partial digestion of DNA. Time of digestion was precisely designed. Partial digestion was done with enzyme in serial dilution and enzyme was inactivated to prevent further digestion. Purpose of this method was to ensure that enzyme will not digest plasmid completely at both restriction sites.

4.8.14. Dephosphorylation of 5' end of DNA

To remove 5'end phosphate vector was treated by phosphatase. Suitable amount of DNA was mixed with 5 μl of Antarctic Phosphatase Reaction Buffer (New England BioLabs), 10U/μg of DNA of Antarctic Phosphatase (New England BioLabs) and ddH_2O up to 50 μl. Mixture was incubated at 37°C for 1 hour and deactivated at 80°C for 5 minutes. Dephosphorylated vector was further used in ligation.

4.8.15. Ligation

Suitable vector was treated by restriction enzymes and dephosphorylated. 100 ng of vector was mixed with amount of insert corresponding to desired molar ratio.

$$\frac{kb\ of\ insert}{kb\ of\ vector} \times ng\ of\ vector = ng\ of\ insert\ for\ 1{:}1\ ratio$$

Vector and insert were mixed with 2 µl of 10x T4 DNA Ligase Reaction Buffer (New England BioLabs) and 200U of T4 DNA Ligase (New England BioLabs). Mixture was filled with ddH_2O up to 20 µl and incubated at 16°C overnight. One tenth of ligation product was transformed in One Shot TOP10 Electrocomp *E. coli* (ThermoFisher) by electroporation.

4.9. Work with bacteria

4.9.1. Cultivation of bacterial cultures

One Shot TOP10 Electrocomp *E. coli* were plated on agar plates with LB media and ampicillin as selective marker and incubated at 34°C. Same cells were also incubated in LB media with ampicillin at 34°C with shaking at 225 rpm. Incubation proceeded overnight.

4.9.2. Electroporation

DNA vectors designated to be amplified in bacterial cells were electroporated in One Shot TOP10 Electrocomp *E. coli* cells. Electroporation cuvette of width of 0.1 cm (Bio-Rad) was placed on ice and filled with 62 µl of ddH_2O, 1.5 µl of DNA and 20 µl of thawed electrocompetent cells. Mixture was incubated on ice for 5 minutes. Cooled cells/DNA mixture was subjected to a pulse at 1.8kV, 25 µF and 200 Ω in Multiporator electroporator (Eppendorf). 480 µl of SOC media (ThermoFisher) were added immediately after pulse. Whole mixture was transferred in snap-cap tube and incubated at 34°C and 225 rpm for 1 hour. After incubation, 250 µl of mixture were plated on prewarmed agar plate with LB media and ampicillin.

4.9.3. Harvesting bacterial colonies from agar plates

Colonies grown on LB agar plates with ampicillin were counted. 5 ml of LB media with ampicillin was applied on agar plate and colonies were washed down. Four ml were collected in snap cap tube for further cultivation. In case, where only one colony from plate was needed, such a colony was transferred by sterile toothpick in 5 ml of LB media with ampicillin.

4.10. Work with yeast

4.10.1. Preparation of CSM-TRP 5-FOA plates

Composition of mixture was: 0.074 % (w/v) Complete supplement mixture w/o tryptophan, 0.167 % (w/v) Yeast nitrogen base w/o ammonium sulfate, 2 % (w/v) Dextrose, 0.5 % (w/v)

Ammonium sulfate, 0.003 % (w/v) Uracil, 2 % (w/v) Bacto agar; (all Sigma-Aldrich). Mixture was sterilized in autoclave. After sterilization, solution was stirred at 55°C for 30 minutes. Later, 1 % (v/v) 100x 5-FOA was added and mixed thoroughly. FOA is light sensitive. Liquid was dispensed in 10cm petri dishes, 25 ml each. Plates were stored in dark place to solidify.

4.10.2. Yeast cells transformation

Single colony of yeast cells was inoculated in 5 ml of YEPD media and incubated overnight at 30°C and 250 rpm. Following day, optical density at 600 nm of overnight culture (diluted ten times with fresh YEPD) was measured. Plain YEPD was used as a blank. Required amount of YEPD, 5 ml per sample, was inoculated by overnight culture to a final OD_{600nm} of 0.3. Culture was incubated at 30°C and 250 rpm until OD_{600nm} reached 0.7 to 0.8. Grown culture was centrifuged at 2100 g at 4°C for 10 minutes. Pellet was resuspended in 5 ml (per 50 ml of yeast culture) of cold LiAc/TE buffer (10 % (v/v) 10x LiAc, 10 % (v/v) 10x TE buffer (both Sigma-Aldrich)). Suspension was centrifuged at the same conditions for 5 minutes and pellet was resuspended with 500 µl of cold LiAc/TE buffer (per 50 ml of yeast culture). Competent yeast cells (50 µl per transfection) were added as last to a mixture containing following ragents: 5 µl of boiled salmon sperm DNA (10 mg/ml) (Sigma-Aldrich), 2 to 8 µg of PCR purified DNA (patient specific *gag* amplicon), 4 µg of SacII-linearized vector $pREC_{nfl}$-TRP-Δgag/URA3-PM-Mdel, 300 µl of liquidized PEG 3350 (Sigma-Aldrich). Prepared samples were vortexed and shaken in horizontal position at 30°C, 250 rpm for 30 minutes. After incubation, 35 µl of DMSO (Sigma-Aldrich) were added to samples and mixed by vortexing. Mixture was heat shocked at 42°C for 20 minutes, briefly centrifuged and pellet resuspended in 250 µl of warm YEPD. Samples were incubated at 30°C, 250 rpm for 2.5 hour. After final incubation, mixture of each sample was plated on warmed CSM-TRP 5-FOA plates and incubated for 3 days at 30°C. Number of colonies was counted.

4.11. Work with mammalian cells

4.11.1. TZM and HEK293T cells passage

TZM and HEK293T cells were incubated in DMEM with 10% serum at 37°C and 5% CO_2. Cells were passaged every 3-4 days at ration 1:10. Medium was aspirated from confluent cells and they were washed with 1x PBS/EDTA (Biotech). Cells were treated with 2 ml of 1x Trypsin solution (Sigma-Aldrich) and let to incubate until cells are released in medium. Afterward, 8 ml of DMEM with serum were added to cells for incubation.

4.11.2. MT4 cells passage

MT4 cells were incubated in RPMI medium with 10% serum at 37°C and 5% CO_2. Cells were passaged with fresh media at ratio 1:4 every 2-3 days.

4.11.3. Counting cells

200 µl of cell suspension were mixed with 600 µl of 1x PBS (Biotech) and 200 µl of 0.4% Trypan blue (Sigma-Aldrich). 10 µl of the mixture were added in Neubauer improved cell counting chamber and cells were counted. At least two fields, representing 0.1 mm^3 of volume were used for calculation and sum was divided by number of counted fields to reach average number of cells per field. The final number of cells was multiplied by dilution factor 5 and volume factor 10000 to get number of cells per ml.

4.11.4. Transfection

$2x10^6$ of HEK293T cells in 5 ml of DMEM with 10% serum were seeded on 6cm petri dish. Next day, cells were examined under microscope for 90-95% confluency and medium was replaced with fresh 5 ml of DMEM with 10% serum. Cells were incubated at 37°C and 5% CO_2 for 30 minutes. During this interval, 8 µg of plasmid DNA was mixed with 500 µl of DMEM without serum. Concurrently, 20 µl of Lipofectamine 2000 (Life Technologies) reagent were gently mixed in 480 µl of DMEM without serum. DNA mixture was added to Lipofectamine 2000 mixture, mixed gently in a falcon tube and incubated at room temperature for 15 minutes. After 30 minutes incubation 1 ml of the DNA-Lipofectamine 2000 mixture was slowly added dropwise to HEK293T cells and the dish was swirled gently to distribute the mixture evenly. Transfected cells were incubated at 37°C and 5% CO_2 for 48 hours and virus was harvested.

4.12. Work with virus

4.12.1. Harvesting virus

Virus culture was aspirated with pipette from the HEK293T cells. Virus supernatant was filtered through 0.45µm filter unit (P-Lab) and aliquoted.

4.12.2. Virus titer determination using TZM-bl cells

TZM-bl cells were counted and $3x10^4$ of TZM-bl cells were seeded per well in 96 well plate with flat bottom in 100 µl of DMEM with 10% serum. Plates were incubated at 37°C and 5% CO_2 overnight. Following day cells were checked for 90-95% confluency under microscope

and 50 µl of DMEM with 10% serum and DEAE-dextran (Sigma-Aldrich) with final concentration of 40 µg/ml was added. Then, 50 µl of four-fold serial diluted viruses (12-step dilution) in DMEM medium were added to the cells in triplicate. Three negative controls containing 50 µl of DMEM without serum and three positive controls containing known amount of NL4-3 were added. Cells were incubated at 37°C and 5% CO_2 for 48 hours. To determine which virus dilutions still resulted in infection, commercial Steady-Glo® Luciferase Assay System (Promega) was used according to the manufacturer's protocol. Firefly luciferase activity was measured in VICTOR X Multilabel Plate Reader (PerkinElmer) and 50% tissue culture infectious dose ($TCID_{50}$) were analyzed using the Reed and Muench method (Reed and Muench 1938). Virus titers were expressed as infectious units per milliliter (IU/ml).

4.12.3. Isolation of PBMCs from HIV-1 negative blood

Blood sediment, collected at Institute of Hematology and Blood Transfusion and screened there for absence of HIV-1, were mixed with PBS at 1:1 ratio. 15 ml of ficoll-paque (Scintila) was added in 50ml falcon tube, then, slowly overlaid by 30 ml of buffy coats/PBS mixture. Falcon tube was held at angle of approximately 45° during addition. Samples were centrifuged at room temperature, 600 g for 25 minutes. Acceleration and brake were reduced (level 3) (Beckman Coulter, Allegra X15R centrifuge) to avoid disruption of the layers. After centrifugation, tube contained yellowish layer of PBMCs, separated between phase of blood plasma and ficoll. PBMCs layer was aspirated in new 50ml-tube and washed with PBS. Tube was centrifuged at room temperature, 600 g for 7 minutes. Pelleted PBMCs were resuspended in 25 ml of PBS and centrifuged again at the same conditions. Supernatant was discarded again and PBMCs were resuspended in 30 ml of warm PBMCs medium and counted.

4.12.4. Isolation of PBMCs from HIV-1 positive blood

Procedure was done at same conditions as in the case of HIV negative blood. Except no PBS dilution step is necessary at the beginning of the procedure because starting material is whole blood. In addition, 6 ml of plasma (layer above PBMCs layer) were collected in cryotubes and stored at -80°C. Collected plasma was later used for HIV RNA isolation. PBMCs collected in falcon tube were processed by centrifugation as in previous protocol (4.12.3.).

4.12.5. Preparation of CD3 treated cultivation flask for PBMC stimulation

CD3 antibody-producing OKT3 cells were cultivated at 37°C and 5% CO_2 until in OKT3 medium until confluent. Culture supernatant was harvested, filtered and aliquoted. To prepare CD3 coated flask an empty 75 cm^2 cultivation flask bottom was covered with 8 ml of harvested supernatant and incubated at 37°C and 5% CO_2 overnight. Next day, supernatant was discarded and CD3 treated flask was used for stimulation of $CD8^+$ T cells depleted PBMCs.

4.12.6. $CD8^+$ T cells depletion and stimulation of HIV-1 negative PBMCs

Isolated HIV-1 negative PBMCs need to be depleted of $CD8^+$ T cells to improve isolation and propagation of primary isolates. Required amount of PBMCs were centrifuged at 300 g at room temperature for 10 minutes. Afterward, pellets were resuspended in PBS/EDTA/BSA buffer (pH7.2, 2mM EDTA, 0.5% BSA) at final cell concentration of $1x10^8$ cells per 500 µl of buffer, stained with human CD8 MicroBeads (Miltenyi Biotec), transferred on LD columns and exposed to magnetic field in MidiMACS separator (Miltenyi Biotec). Collected $CD8^+$-depleted cells were resuspended in 30 ml of PBS and counted. Amount of cells was divided to three equal batches and centrifuged at 300 g at room temperature for 10 minutes. First $CD8^+$-depleted PBMCs pellet was resuspended in PBMCs medium, second $CD8^+$-depleted PBMCs pellet was resuspended in PBMCs medium with 5-times lower PHA concentration and third $CD8^+$-depleted PBMCs pellet was resuspended in PBMCs medium without PHA and transferred to CD3 coated flask. Final concentration of cells was $4x10^6$ per ml. Cells were incubated at 37°C and 5% CO_2 for 3 days. After incubation, all cells were combined, centrifuged at 300 g at room temperature for 10 minutes and resuspended in 30 ml of PBMCs medium without PHA. Cells were counted and used for isolation and culturing of primary isolates.

4.12.7. Isolation of HIV-1 from HIV^+ PBMCs

HIV-1 primary isolates were prepared by co-cultivation with stimulated $CD8^+$-depleted HIV-negative PBMCs. Briefly, $10x10^6$ stimulated $CD8^+$-depleted HIV-negative PBMCs were mixed with $5x10^6$ PBMCs isolated from HIV-1 positive patient. Cell mixture was incubated at 37°C and 5% CO_2 for up to 4 weeks. Every 3 to 4 days fresh PBMCs medium without PHA was added and every 7th day 5 x 10^6 HIV-1 negative, stimulated $CD8^+$ -depleted PBMCs were added. Virus culture was monitored twice a week for cytopathic effect under microscope

and for virus activity in TZM-bl assay. Negative cultures were discarded and strong HIV positive cultures were pooled, filtered through 0.45μm filter and aliquoted.

4.12.8. Evaluation of HIV activity in TZM-bl assay using X-gal staining

Steps were the same as previously described in 4.12.2. chapter except X-galactosidase staining was employed instead of determination of firefly luciferase activity. After 48h long incubation, supernatant was discarded and cells were washed with 100 μl of PBS. Cells were fixed with 1% glutaraldehyde (Sigma-Aldrich) for 20 minutes and afterward, fixative was discarded. Each well was stained with 50 μl of 2% X-gal staining solution (2% (w/v) X-Gal; Dimethylformamide (both Sigma-Aldrich) and plates were incubated at 37°C for 2 hours. After incubation, staining solution was discarded and cells were washed by 100 μl of PBS. Reaction was stopped by addition of 50 μl of 70% glycerol (Sigma-Aldrich) and stored at 4°C for 10 minutes. Blue spots in each well were counted under microscope and compared to number of blue spots produced by NL4-3 control.

4.12.9. Viral growth kinetics of recombinant virus

Amount of virus necessary for infection of $5x10^4$ MT4 cells with MOI = 0.005 in triplicate was calculated from $TCID_{50}$ and diluted with RPMI medium with 10% serum up to 1 ml for each sample. MT4 cells were centrifuged in a snap-cap tube at 600g for 5 minutes at room temperature, supernatant was discarded and cells were resuspended with RPMI medium containing virus inoculum. Cell/virus mixture in snap-cap tube was incubated at 37°C, 5% CO_2 for 2 hours. During incubation mixture was gently shaken few times by tapping on the tube. After incubation, infected cells were centrifuged at above conditions and washed twice by 3 ml of 1x PBS. After last centrifugation, cells were resuspended in RPMI medium with 10% serum and pipetted in triplicate in 96-well plate. Medium only was used as a mock control and three wells were used as a positive control, where MT4 were infected by NL4-3-EGFP virus. Infected cells were incubated at 37°C, 5% CO_2 for 10 days. At days 0, 3, 4, 5, 6, 7 and 10, 50 μl from each well were saved in new 96-well round bottom plate and stored at -80°C. After each collection, volume was replenished with 50 μl of fresh RPMI medium with 10% serum. At day 5, 50 μl of the fresh RPMI medium with $5x10^4$ MT4 cells were added. Stored samples were used in the reverse transcriptase assay to determine HIV activity.

4.12.10. Viral growth kinetics of primary isolates

Analogously to 4.12.9., $5x10^5$ of fresh PBMCs were infected at MOI = 0.005. Cells were treated by PBMC medium without PHA and with 0.5% IL-2. Medium only was used as mock control. Viruses B2 (92US076), A2 (92UG029) and E1 (CMU06) served as positive controls of infection. Growth kinetics was done for 17 days with collections at days 0, 3, 4, 5, 6, 7, 10, 14, 17. Removed media was replaced by fresh media. At days 7 and 14 wells were replenished with $2.5x10^5$ stimulated PBMCs.

4.12.11. Growth competition experiments

MT4 cells were resuspended, mixed with trypan blue and automated Countess® cell counter (ThermoFisher Scientific) was used to enumerate number of cells per ml and to verify if viability is higher than 93 %. $2x10^6$ MT4 cells in 1 ml of RPMI medium were added in a snap-cap tube and viruses, both sample and competing control were added in volume corresponding to the MOI = 0.0001 per virus (total MOI in snap-cap tube was equal to 0.0002). As a wild-type control to recombinant virus containing EGFP probe, NL4-3-Crimson virus was used. Infection was incubated in thermostat at 37°C and 5% CO_2 for 2 hours, snap-cap tubes were shaken occasionally. After incubation, snap-cap tubes were centrifuged at 600g at room temperature for 5 minutes and cells were rinsed by 3 ml of 1x PBS twice. After rinsing, cells were resuspended in RPMI medium - phenol RED free to final concentration $5x10^5$ cells per ml. $5x10^4$ of cells were added in triplicate to wells on the 96-well plate with flat bottom. Plate was further incubated in thermostat another 72 hours. At day 3, cells were resuspended and 75 µl were transferred into the wells on 96-well plate with round bottom and joined with 75 µl of PBS. Next, Guava® easyCyte HT flow cytometer with guavaSoft 2.2.3 software and InCyte™ program was used for analysis. Number of recorded events was set to 10000 and plot of Side-scattered light and Forward-scattered light (FSC) was used to subduct debris. At a next plot, FSC-Height and FSC-Area were used to select only singlet cells. Additionally, rest 25 µl in the 96-well plate with flat bottom were replenished with RPMI medium – phenol RED free with fresh MT4 cells. Final volume in replenished wells was 100 µl with $5x10^4$. Plate was incubated another 24 hours. After incubation, at day 4 samples were processed analogously to day 3.

4.12.12. RT assay

Virus samples were quickly thawed. Three wells on one plate were selected for positive control and filled with 50, 5 and 0.5 µl of NL4-3 virus, respectively. Another three cells were

selected for negative control and filled with 50 µl of PBS. 15 µl of unlabeled RT buffer (5 % (v/v) 1M Tris-HCl pH 7.8; 3.75 % (v/v) 2M KCl; 0.2 % (v/v) 1M DTT; 2.5 % (v/v) 200mM $MgCl_2$; 0.5 % (v/v) NP-40 (all Sigma-Aldrich)) was added per each well and incubated at 37°C, 5% CO_2 for 30 minutes to lyse the virions. Then, 10 µl of labeled RT buffer containing 10µmCi/ml α-32P labeled dTTP were added to each well and incubated at 37°C, 5% CO_2 for 2 hours. After incubation, 10 µl from each well was spotted on separate field of DEAE stamped blotting paper. Blotting papers were dried at room temperature and then washed with 1x SSC buffer (15mM Sodium citrate tribasic dihydrate, 150mM NaCl (both Sigma-Aldrich)) five times for 5 minutes on a plate shaker. After this step, papers were washed with 85% ethanol two times times for 5 minutes on a plate shaker. Blotting papers were air dried and put in phosphor screen storage cassette (Life Technologies) for overnight incubation. Next day phosphor screen was scanned using Typhoon phosphorimager 8600 (GE Healthcare). Density of particular spots, which corresponds to the HIV activity, was measured by array analysis in ImageQuant TL program and data were used to construct growth kinetics curves.

4.13. Statistical analysis

Graphs were constructed in GraphPad Prism program version 6.02 (GraphPad Software, La Jolla, CA). Values are depicted with standard deviations or as normalized averages. Pearson's and Spearman's correlations were used to evaluate linear and monotonic relationship between variables, respectively. Values with $p < 0.05$ were considered statistically significant.

5. Results

5.1. Construction of plasmids used for the gag recombinant virus production

Successful production of gag recombinant HIV required homologous recombination step in yeast followed by a cloning step in bacteria. For that we had to prepare two specific plasmids pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel and pNL4-3-EGFP-PM. During their preparation we used following plasmids: (1) yeast vector pREC$_{nfl}$-TRP-Δgag/URA3 derived from pREC$_{nfl}$-TRP-Δp2-int/URA3 plasmid that was proved to be a convenient tool for generation of p2-int-recombinant viruses (Weber *et al.* 2011); (2) pNL4-3-EGFP plasmid, originally prepared for generation of fluorescently labelled viruses for HIV-1 fitness research (Weber *et al.* 2006); and (3) p83-2 plasmid containing the 5'-half of pNL4-3 (i. e. 5'LTR, complete *gag, pol, vif* coding regions and part of the *vpr* gene) from NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH from Dr. Desrosiers. All plasmids were further modified in the course of this work.

5.1.1. Construction of plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel

Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel was prepared in a serie of successive steps. It originated from pRECnfl-TRP-Δgag/URA3 plasmid that contains near full length of HIV-1 from pNL4-3, except 5`LTR and *gag*, which was replaced by *Saccharomyces cerevisiae* uracil biosynthesis gene *ura3* (fig. 9). Plasmid was used as a vector for DNA recombination in yeast cells, therefore second LTR would be contradictory. 3`LTR and 5`LTR sequences are the same and their linkage during recombination leads toward undesirable excision of whole HIV-1 sequence from plasmid. Gene *gag* was replaced by *ura3*, which was used as a negative selective marker for successful recombination in yeast.

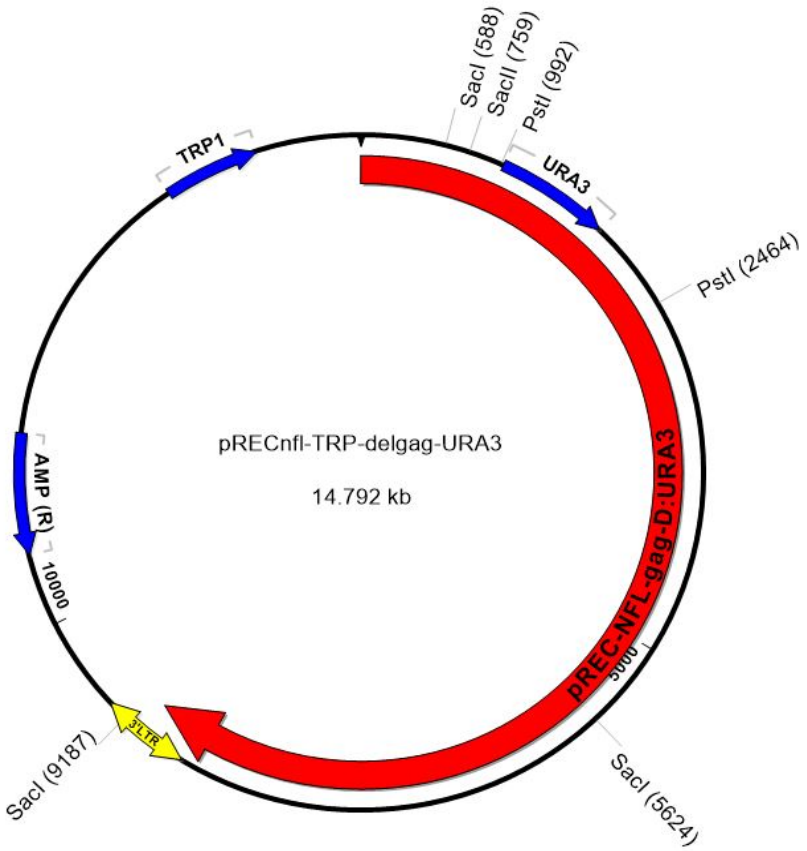


Figure 9. Plasmid pREC$_{nfl}$-TRP-Δgag/URA3. Original plasmid used for recombination of DNA in yeast cells. Plasmid contains near full length of HIV-1 except 5`LTR and *gag*. This gene was replaced by selective marker, gene *ura3*. Apart from the HIV-1 sequence, plasmid contains also gene for ampicillin resistance and gene *Trp1* for tryptophan biosynthesis, providing positive selective markers for growth in bacteria and yeast, respectively. Restriction sites of PstI, SacI and SacII enzymes are depicted. Picture was prepared in program SeqBuilder (DNASTAR v.10.0.1).

Plasmid further contains beta-lactamase gene *amp* providing ampicillin resistance and gene *Trp1* for tryptophan biosynthesis, providing positive selection of yeast with incorporated plasmid during homologous recombination.

DNA integrity of the original plasmid pREC$_{nfl}$-TRP-Δgag/URA3 was verified by restriction with three separate enzymes: PstI, SacI and SacII. These restriction enzymes digest this plasmid with different frequency resulting in specific restriction pattern. Each enzyme was used in separate restriction reaction and DNA restriction pattern was verified by 1% agarose electrophoresis (fig. 10). All restriction reactions resulted in correct number and size of bands.

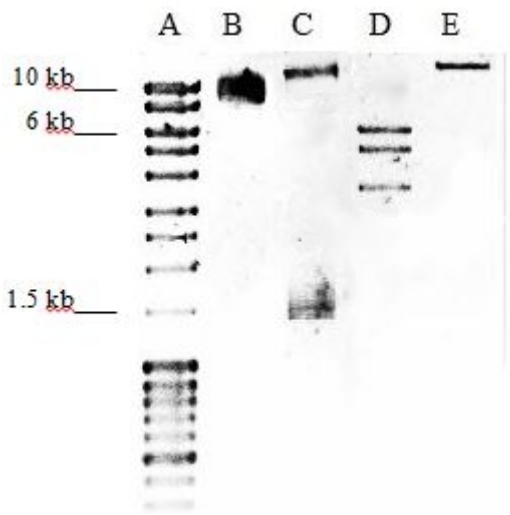


Figure 10. Agarose gel or restricted plasmid pREC$_{nfl}$-TRP-Δgag/URA3. A) Marker. B) Unrestricted plasmid. C) Plasmid restricted by PstI resulting in 13.3 kb and 1.5 kb long fragments. D) Plasmid restricted by SacI resulting in 6.2, 5 and 3.5 kb long fragments. E) Plasmid restricted by SacII resulting in single linear fragment of size 14.8 kb.

During the gag-recombinant virus preparation it is necessary to remove recombined gag regions from yeast vector and insert them in pNL4-3-EGFP vector (complete vector with both LTR regions) that is able to produce replication-competent virus upon transfection in HEK293T cells. As we did not find any suitable naturally occurring recognition sites in these plasmids we decided to introduce restriction sites by site directed mutagenesis. The most important selection criterion for these restriction sites was that their recognition sites are not (or very rarely) naturally present in HIV-1 gag regions. For that we analyzed over 10000 HIV-1 gag sequences maintained in HIV sequence database at Los Alamos laboratory website (https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html). Restriction enzymes PspXI and MluI were selected because their recognition sites occurred in less than 1% of available HIV-1 gag sequences. Both sites were unique in pNL4-3-EGFP plasmid but MluI recognition site was present in pREC$_{nfl}$-TRP-Δgag/URA3, but outside of any important regions. We still decided to use MluI site and silence the native MluI site by site directed mutagenesis.

5.1.1.1. Introduction of MluI restriction site

First step in pREC$_{nfl}$-TRP-Δgag/URA3 plasmid modification was to establish new MluI restriction site at position 2508 to generate new plasmid designated as pREC$_{nfl}$-TRP-Δgag/URA3-M. This position was chosen because only two nucleotide substitutions were necessary to achieve MluI recognition site without change of any amino acids. MluI site was introduced in 5' end region of reverse transcriptase gene. This site became 3' end of fragment

later during ligation step. Plasmid pREC$_{nfl}$-TRP-Δgag/URA3 contains native site for MluI and this site was removed in later steps. Introduction of MluI site was done by site directed mutagenesis with primers MluI-SDM-F and MluI-SDM-R and resulting DpnI-treated DNA was transformed in TOP10 Electrocompetent *E.coli* cells by electroporation. Ten single colonies were scraped off and transferred in LB media with ampicillin. Isolated DNA of single colonies was verified by restriction with enzyme MluI only and positive clone was verified by restriction with MluI and SalI-HF and visualized by gel electrophoresis on 1% agarose gel (fig. 11).

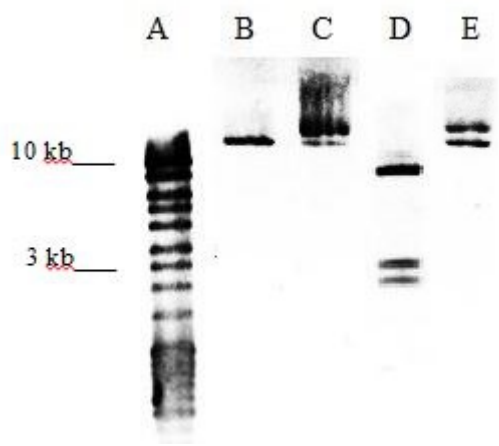


Figure 11. Verification of plasmid pREC$_{nfl}$-TRP-Δgag/URA3-M. A) Marker. B) Control plasmid pREC$_{nfl}$-TRP-Δgag/URA3 restricted by MluI. C) Control unrestricted plasmid pREC$_{nfl}$-TRP-Δgag/URA3. D) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-M (introduced MluI site) restricted by MluI and SalI-HF enzymes resulting in 9.3, 2.9 and 2.5 kb fragments. E) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-M unrestricted. Two bands display different species of plasmid DNA.

Product of site directed mutagenesis restriction by MluI and SalI-HF resulted in three fragments of size 9.3 kb, 2.9 kb and 2.5 kb as expected consequence of successful mutagenesis. As a control, non-mutated plasmid pREC$_{nfl}$-TRP-Δgag/URA3 was only linearized by MluI enzyme utilizing native MluI site in the plasmid. Introduced site was verified by sequencing and aligning with original sequence of unmutated plasmid (fig. 12). Sequencing confirmed presence of the new, MluI recognition site with motif ACGCGT.

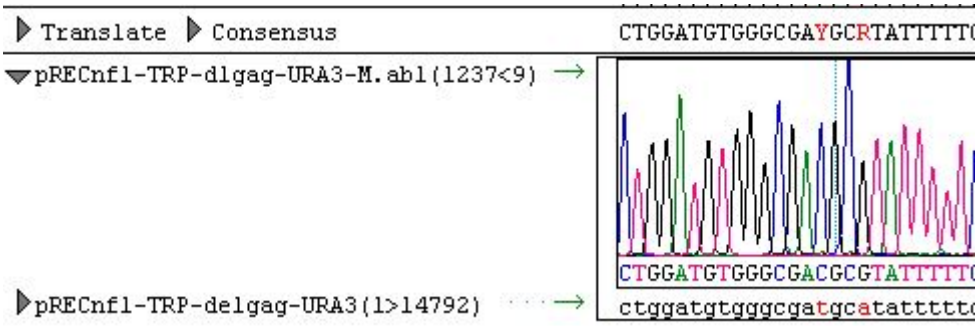


Figure 12. Verification of introduced MluI site by sequencing. Two substitutions as a result of site directed mutagenesis. Top line of sequence of mutated plasmid contains ACGCGT motif, recognized by MluI enzyme. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

5.1.1.2. Introduction of PspXI restriction site

Along with MluI site, plasmid pREC$_{nfl}$-TRP-Δgag/URA3 has to contain second, specific restriction site. We selected restriction site for PspXI that was introduced in position 619 in vector, thus upstream of *ura3* gene to generate plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM. Position of described restriction site was chosen carefully to be in the non-coding region between 5'LTR and *gag* gene. As in previous procedure, also here was new restriction site introduced by site directed mutagenesis; three nucleotides were substituted, using primers PspXI-SDM-F and PspXI-SDM-R. Mutated plasmid was transformed in TOP10 Electrocompetent E.coli cells by electroporation and isolated DNA was verified by restriction with enzymes MluI and PspXI and DNA fragments were separated by gel electrophoresis on 1% agarose gel (fig. 13).

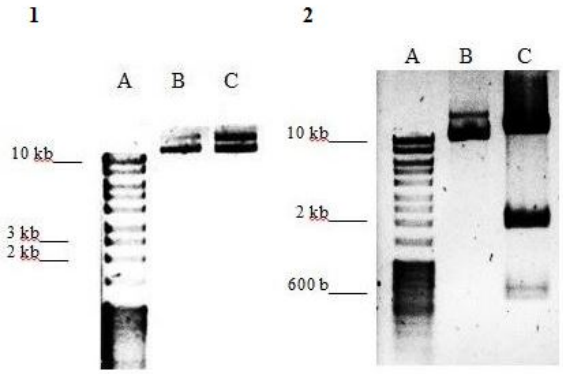


Figure 13. Verification of plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM. 1) A) Marker. B) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3 unrestricted. C) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3 restricted by enzyme PspXI. Original plasmid contains no native site for PspXI, therefore was not restricted. **2)** A) Marker, B) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM (newly introduced site for PspXI) unrestricted. C) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM, restricted by MluI and PspXI, resulting in three visible fragments, smallest fragment (size 600 b) is present due to native MluI site in the plasmid.

Original plasmid pREC$_{nfl}$-TRP-Δgag/URA3 was restricted with enzyme PspXI as a control of presence of native PspXI sites. No sites were detected by restriction or sequence analysis (fig. 13, 1; B and C). After successful mutagenesis, final plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM had newly introduced site for PspXI enzyme (fig. 13, 2; C). Three individual bands are visible. Furthermore, existence of PspXI site with motif CCTCGAGT was confirmed by sequencing (fig. 14), where all three substituted nucleotides are displayed.

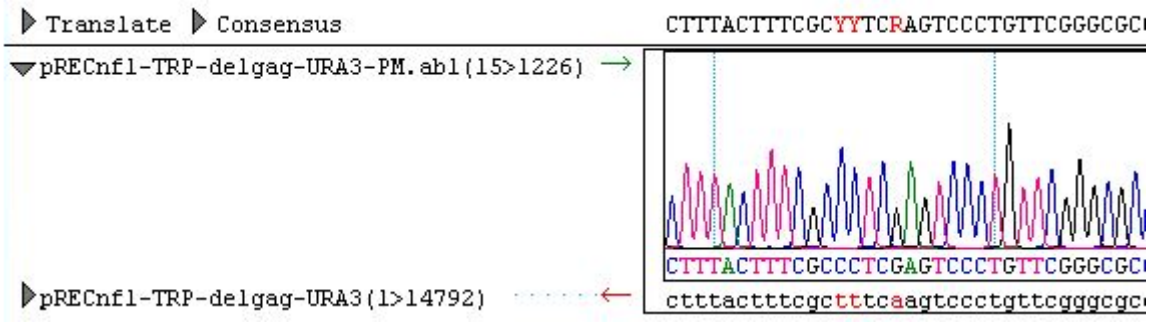


Figure 14. Verification of introduced PspXI site by sequencing. Original unmutated plasmid pREC$_{nfl}$-TRP-Δgag/URA3 was aligned with sequence of new plasmid with introduced site for enzyme PspXI. Sequence newly contained motif CCTCGAGT. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

5.1.1.3. Substitution of native MluI restriction site

As a last step, we silenced the native MluI site in plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM. For the further cloning, it was fundamental to secure clear excision of recombined patient *gag* gene therefore, additional MluI site at the beginning of promotor in the plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM was removed. Site directed mutagenesis with primers MluI-delSDM-F and MluI-delSDM-R were used for substitution of two nucleotides in plasmid sequence to generate pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel. Mutated plasmid was transformed in TOP10 Electrocompetent *E.coli* cells by electroporation and amplified DNA was isolated and verified on 1% agarose gel (fig. 15).

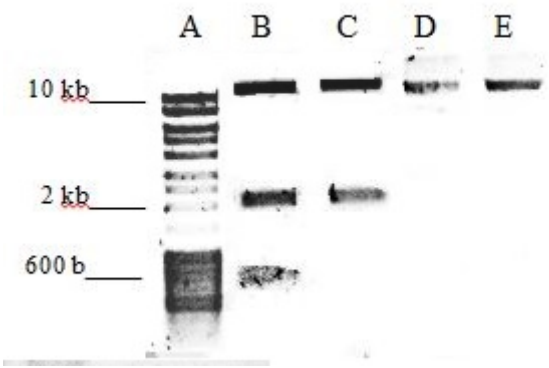


Figure 15. Verification of plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel. A) Marker. B) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM restricted by enzymes PspXI and MluI. This plasmid still contains native MluI site hence, restriction resulted in presence of 621 b long fragment. C) Plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel restricted by enzymes PspXI and MluI. New, mutated plasmid has native MluI site substituted and 621 b long fragment is absent. D) Unrestricted plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM. E) Unrestricted plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel.

Original plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM after restriction by enzymes PspXI and MluI excised 621 bases and 2.2 kb long fragments. New plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel contains just one MluI site and after restriction by same enzymes under same conditions exhibits only 2.2 kb long fragment, which contains our target gag region replaced by *ura3* gene. In addition, Sanger sequencing was used to confirm substitution of native MluI site with TCGTGT sequence unrecognizable by MluI enzyme (fig 16.).

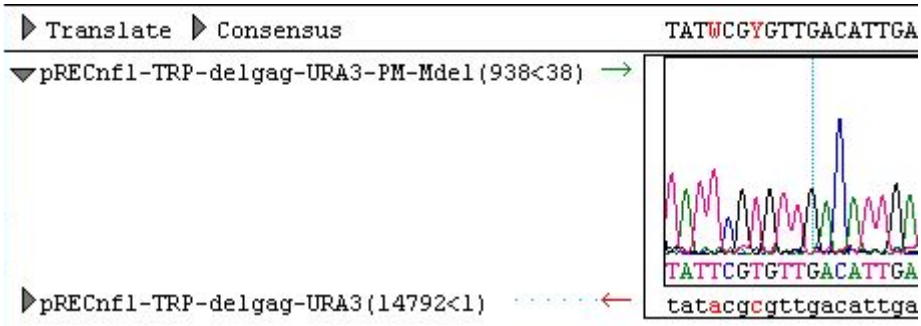


Figure 16. Verification of substituted native MluI site by sequencing. Original unmutated plasmid pREC$_{nfl}$-TRP-Δgag/URA3 was aligned with sequence of new plasmid with substituted MluI native site. Plasmid has MluI recognition motif ACGCGT replaced by neutral sequence TCGTGT. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

After last step of mutagenesis, plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel was ready to be used as a vector in yeast transformation. Two introduced sites, recognized by enzymes MluI and PspXI, surrounds *ura3* gene, which was replacing sequence of *gag* in this plasmid (fig. 17). Purpose of this mutagenesis was to facilitate excision of patient specific *gag* region after it was recombined in plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel during yeast transformation. Later, the gag region is excised with PspXI and MluI and introduced in vector pNL4-3-PM-linker.

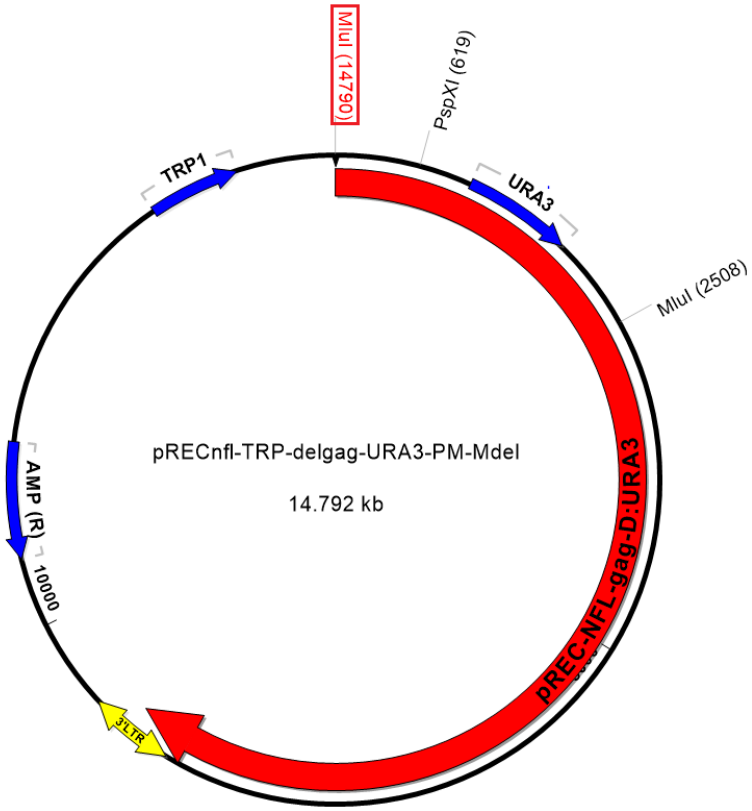


Figure 17. Map of plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel. Final product of site directed mutagenesis of plasmid pREC$_{nfl}$-TRP-Δgag/URA3. Newly introduced sites recognized by enzymes PspXI and MluI are depicted with position number. Native site recognized by MluI, formerly present in plasmid pREC$_{nfl}$-TRP-Δgag/URA3, was substituted with neutral sequence in plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel. Specified site is highlighted with red frame. Picture was prepared in program SeqBuilder (DNASTAR v.10.0.1).

5.1.2. Construction of plasmid pNL4-3-EGFP-PM

Another important plasmid for my work was plasmid pNL4-3-EGFP-PM with introduced PspXI and MluI sites and deleted gag region. This plasmid was used as a vector for transfection into the HEK293T cells to produce HIV-1 recombinant virus. Original plasmid, pNL4-3-EGFP has been prepared by members of our laboratory during earlier research of HIV-1 fitness from plasmid pNL4-3. Plasmid contains full sequence of HIV-1 with probe EGFP (enhanced green fluorescent protein) sequence inserted between *env* and *nef* ORFs. Since plasmid pNL4-3-EGFP-PM is over 15 kb long, above recommended length for efficient mutagenesis, we decided to use shorter 8kb long plasmid p83.2 for site-directed mutagenesis and subsequently ligate fragment with introduced PspXI and MluI sites in pNL4-3-EGFP scaffold. Plasmid p83.2 is 8151 bp long and except gene providing ampicillin resistance, contains also 5' half of HIV-1$_{NL4-3}$ sequence. Included sequence is from 5' LTR up to *vpr*, although this gene is contained only partially (fig. 18).

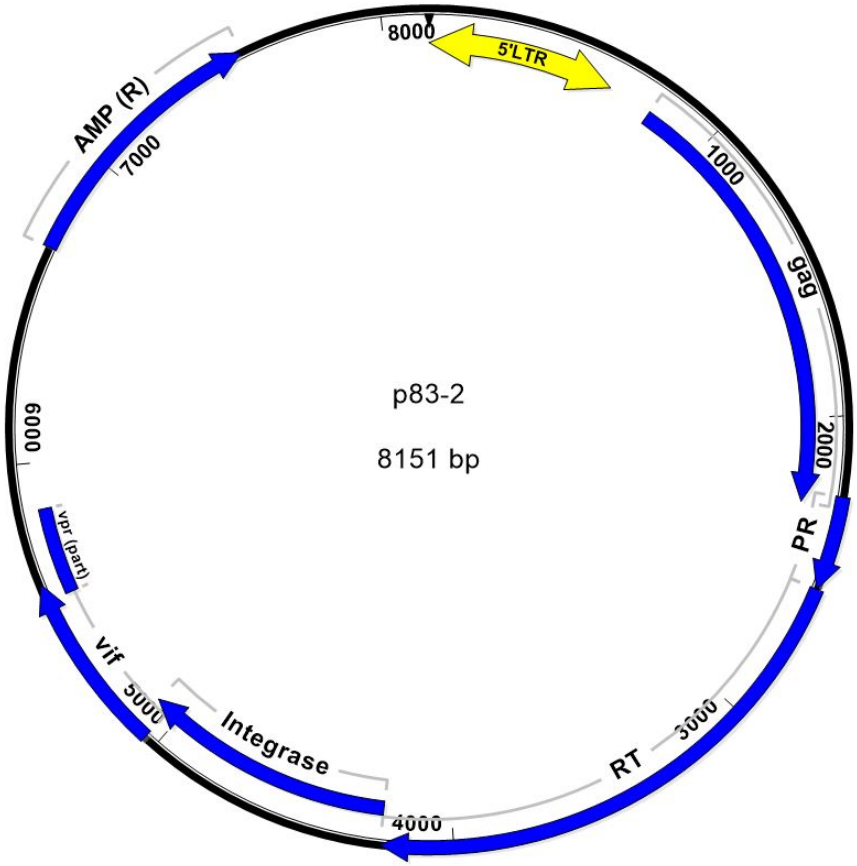


Figure 18. Plasmid 83.2 Plasmid p83.2 is known also as HIV-1_{NL4-3} 5′ Clone. Plasmid carries 5' LTR, *gag*, *pol*, *vif* and part of *vpr* gene. Plasmid also contains gene for ampicillin resistance. Plasmid was used for mutagenesis Picture was prepared in program SeqBuilder (DNASTAR v.10.0.1).

Absence of MluI and PspXI restriction sites was verified by restriction of specific enzymes and confirmed by 1% agarose gel electrophoresis (fig. 19).

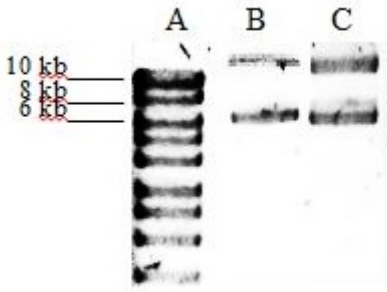


Figure 19. Control restriction of plasmid p83.2. A) Marker. B) p83.2 unrestricted. C) p83.2 treated by enzymes MluI and PspXI. Double bands display supercoiled and relaxed circular forms of plasmid DNA.

5.1.2.1. Introduction of MluI and PspXI restriction sites in plasmid p83.2.

Introduction of mutated sites in plasmid p83.2 was analogous to procedure of establishing restriction sites in plasmid pREC$_{nfl}$-TRP-Δgag/URA3. Restriction sites were introduced into the corresponding positions; 654 (PspXI) and 2888 (MluI). PspXI site is located in non-coding region between 5'LTR and *gag* gene; and MluI site is located in 5' end of gene for reverse transcriptase (fig. 20).

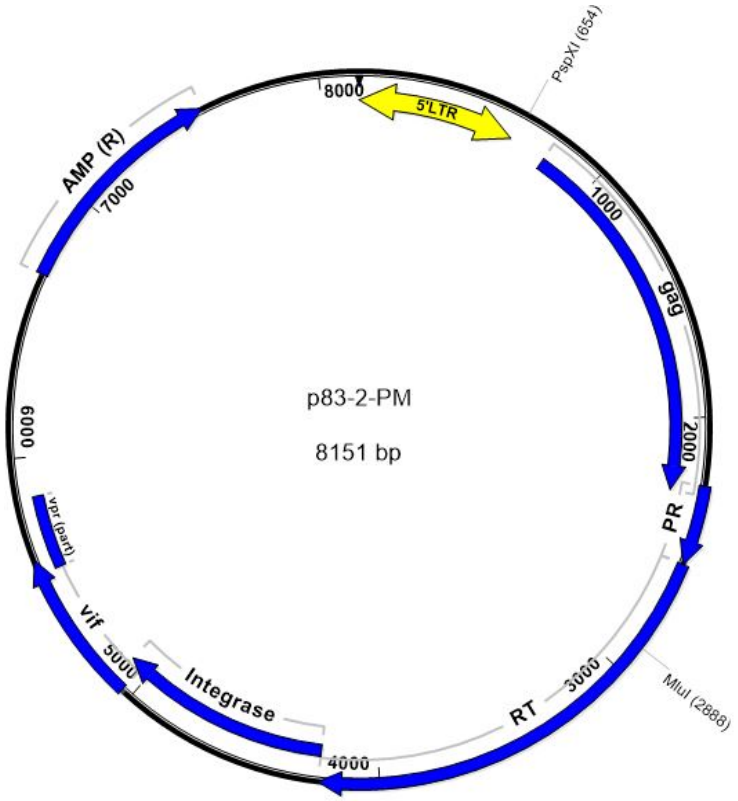


Figure 20. Plasmid p83.2-PM. Introduced restriction sites are depicted with corresponding position. Restriction sites PspXI and MluI are surrounding *gag* gene. Picture was prepared in program SeqBuilder (DNASTAR v.10.0.1).

MluI restriction site was introduced by site directed mutagenesis with primers MluI-SDM-F and MluI-SDM-R, following the procedure as for plasmid pREC$_{nfl}$-TRP-Δgag/URA3 mutagenesis. After transformation of DpnI-treated DNA in electrocompetent *E. coli* cells through electroporation, two colonies were scraped off the plate and restricted by enzyme MluI (fig. 21).

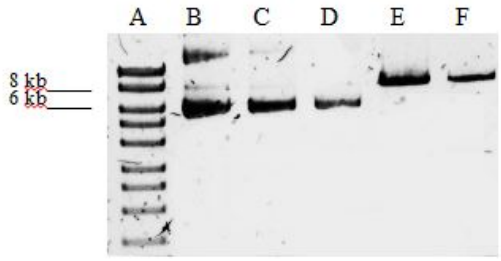


Figure 21. Verification restriction of plasmid p83.2-M. A) Marker. B) Plasmid p83.2 unrestricted. C & D) DNA from colonies 1 and 2, unrestricted. E & F) DNA from colonies 1 and 2, restricted by enzyme MluI. Both restrictions resulted in linearized DNA of size around 8 kbp, newly prepared plasmid p83.2-M.

DNA from both colonies resulted in one correct-sized band after MluI restriction. One of the positive colonies was used as source plasmid p83.2-M for introduction of PspXI restriction site (fig. 21; E). Also, presence of MluI restriction site in plasmid p83.2-M was verified by sequencing (fig. 22). Motif recognized by enzyme MluI, ACGCGT was present in plasmid p83.2-M.

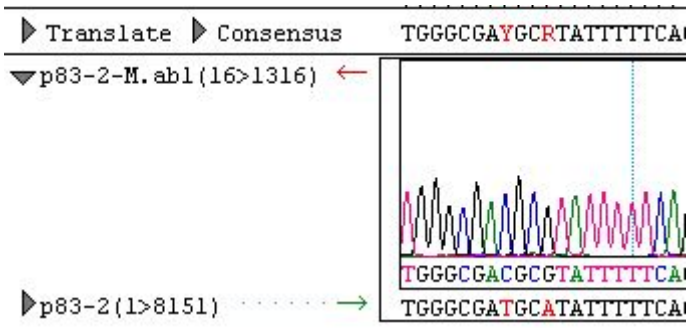


Figure 22. Sequence verification of p83.2-M. Plasmid p83.2-M was sequenced and aligned with original plasmid p83.2. Restriction site for MluI with motif ACGCGT is newly introduced. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

Next, restriction site for PspXI was introduced in plasmid p83.2-M by synonymous procedure as in case of the pREC$_{nfl}$-TRP-Δgag/URA3-M. In brief, primers PspXI-SDM-F and PspXI-SDM-R were used and DNA was transformed in *E. coli*. Positive clone was verified by restriction (fig. 23; 1). PspXI restricted plasmid p83.2-PM was linearized and restriction by both MluI and PspXI resulted in excised fragment of size 2.2 kbp. Presence of motif CCTCGAGT was verified by sequencing (fig. 23; 2).

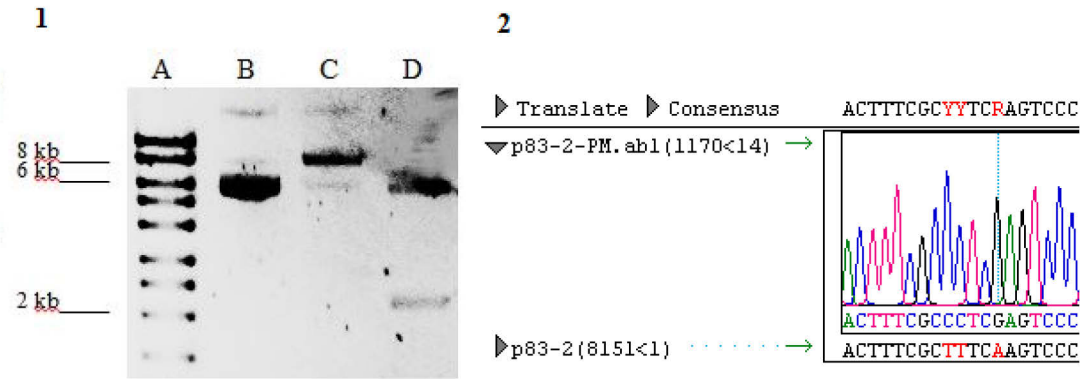


Figure 23. Verification of plasmid p83.2-PM. 1) Restriction verification of p83.2-PM A) Marker. B) plasmid p83.2-PM, unrestricted. C) plasmid p83.2-PM linearized by enzyme PspXI. D) plasmid p83.2-PM restricted by enzymes MluI and PspXI. Lower band displays excised fragment of size 2.2 kbp. **2)** Sequence verification of CCTCGAGT motif in plasmid p83.2-PM. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

5.1.2.2. Ligation of EcoRI-p83.2-PM-BspEI fragment in pNL4-3-EGFP vector

Since plasmid p83.2-PM contains only part of the HIV clone NL4-3 sequence, it is not suitable for transfection to produce recombinant viruses. Therefore, fragment of the p83.2-PM containing introduced restriction sites was ligated in pNL4-3-EGFP vector. As we did not find two restriction enzymes that would restrict both plasmids just once, we had to employ partial digest method in case of pNL4-3-EGFP vector preparation. We selected enzyme EcoRI that digests both plasmid just once and BspEI that digests once p83.2 and twice pNL4-3-EGFP (fig. 24). Initially, plasmid p83.2-PM was fully digested by enzymes BspEI and EcoRI to prepare fragment containing introduced PspXI and MluI sites.

Next, plasmid pNL4-3-EGFP was fully digested by enzymes EcoRI and SphI. We used additional digest by SphI to get better separation between individual bands on agarose gel after partial digest. Both enzymes utilize single restriction sites inside pNL4-3-EGFP sequence. Finally, we used BspEI partial digest to preserve inner BspEI site (at position 10125, fig. 24) to generate vector fragment BspEI-EcoRI.

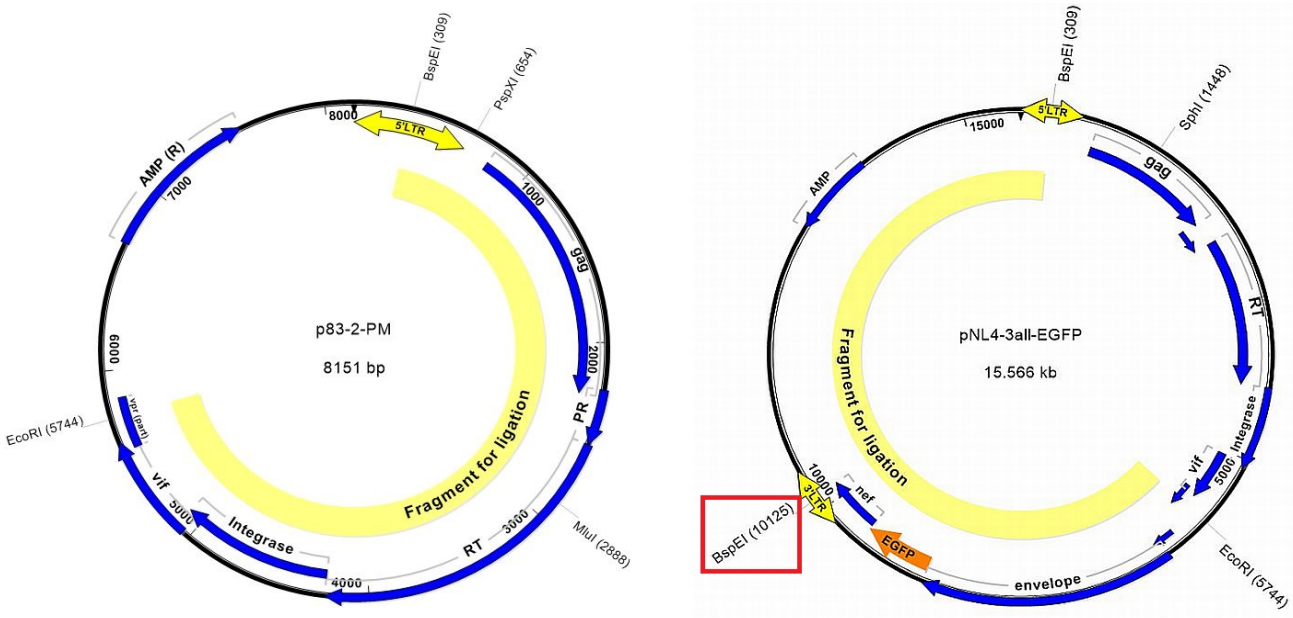


Figure 24. Partial digestion of plasmids p83.2-PM and pNL4-3-EGFP. Plasmid p83.2-PM was digested by enzymes BspEI and EcoRI, utilizing single restriction sites. Excised fragment is displayed by light yellow box; this fragment has BspEI and EcoRI overhangs on 3' and 5' end. Excised fragment contains only sequence of clone NL4-3 with introduced restriction sites for PspXI and MluI. Plasmid pNL4-3-EGFP was digested by enzymes SphI and EcoRI a partially digested by enzyme BspEI to preserve second BspEI restriction site; highlighted in red box. Excised vector fragment for ligation is displayed by light yellow box and contains BspEI and EcoRI overhangs on 3' and 5' end. Picture was prepared in program SeqBuilder (DNASTAR v.10.0.1).

In order to achieve successful partial digestion of pNL4-3-EGFP, the optimal amount of BspEI enzyme was determined. BspEI was serially diluted (concentrations between 6 U and 0.005 U per µg of DNA) and used in 15-min partial digestion followed by 20 minutes of inactivation at 80°C. Concentrations 3.35 U and 1.1 U per µg of DNA were chosen for final partial digestion (fig. 25; 1). Incomplete digestion resulted in presence of several different bands, including our fragment of interest with size 10,131 bp with BspEI and EcoRI overhangs. Fragment of interest was purified on 0.8% E-Gel and isolated by E-Gel Electrophoresis system. Next, the presence of purified fragment was verified by 1% gel electrophoresis (fig. 25; 1). Digested plasmid p83.2-PM was also isolated by E-Gel Electrophoresis system; fragment of interest has size 5,435 bp (fig. 25; 2).

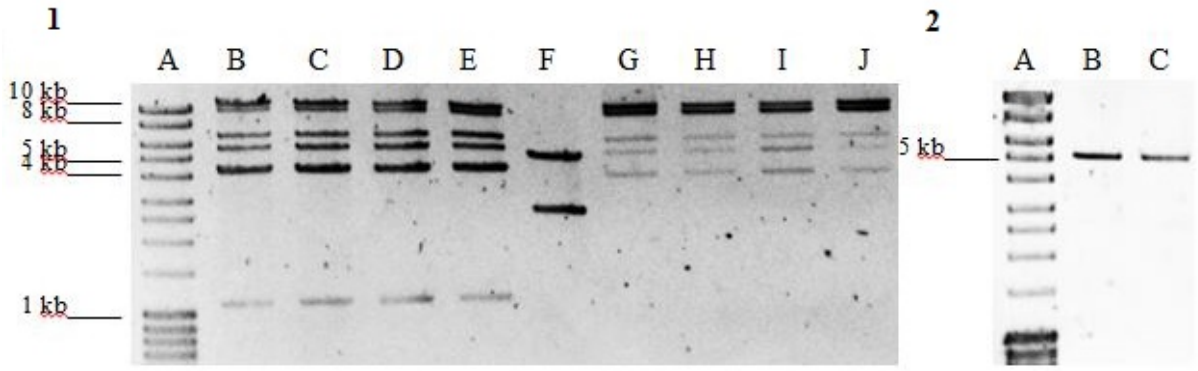


Figure 25. Agarose gel of partial digestion and DNA purification. 1) A) Marker. B-C) plasmid pNL4-3-EGFP fully digested by enzymes SphI and EcoRI and subsequently partially digested by 3.35 U per µg of DNA of enzyme BspEI for 15 minutes. D-E) plasmid pNL4-3-EGFP fully digested by enzymes SphI and EcoRI and subsequently partially digested by 1.1 U per µg of DNA of enzyme BspEI for 15 minutes. F) plasmid p83.2-PM digested by BspEI and EcoRI. G-J) Isolation of 10,131 bp long fragment from samples displayed at lines B-E, respectively. **2)** A) Marker. B-C) Isolation of 5,435 bp long fragment from digested p83.2-PM.

E-gel purification of digested p83.2-PM plasmid resulted in one pure 5,435 bp long fragment. E-gel purification of 10,131 bp long fragment from pNL4-3-EGFP resulted in mixture of correct and still contained minority of other fragments. Mainly fragment of size 11,270 bp was not removed, because of limited separation power of E-Gel Electrophoresis system. Although fragment of size 11,270 bp was present in equal amount to fragment of interest, its presence did not hamper ligation as larger fragment had SphI and EcoRI overhangs.

To generate pNL4-3-EGFP-PM fragments BspEI-p83.2-PM-EcoRI and BspEI-pNL4-3-EGFP-EcoRI were ligated in molar ratio 4:1 and one tenth of ligation mixture was transformed in *E. coli* by electroporation.

Four colonies were scraped-off the plate and isolated DNA was digested by enzymes MluI and PspXI for one hour and verified by 1% agarose gel electrophoresis (fig. 26).

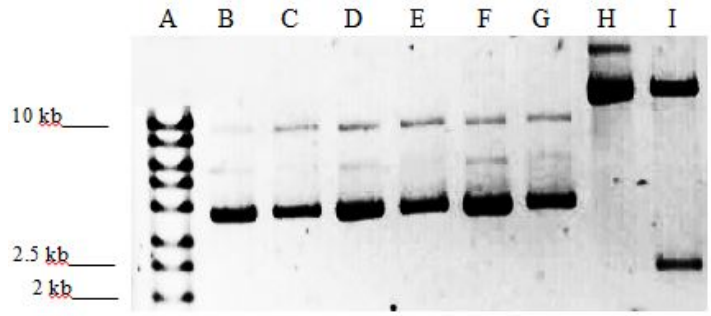


Figure 26. Agarose gel of plasmid pNL4-3-EGFP-PM. A) Marker. B, D & F) DNA isolated from three different colonies, unrestricted. C, E & G) DNA isolated from same three colonies, restricted by enzymes MluI and PspXI. No digestion occurred, also bands have incorrect size. H) DNA isolated from fourth colony, unrestricted. I) DNA from the fourth colony restricted by MluI and PspXI resulted in excised band of 2.2 kbp size. Bands have correct size expected from plasmid pNL4-3-EGFP-PM.

Gel electrophoresis displayed correct restriction pattern expected from pNL4-3-EGFP-PM plasmid in one isolated clone. Next, we performed sequence analysis of whole BspEI-p83.2-PM-EcoRI region to verify that we did not introduce any other substitution during site-directed mutagenesis step. Plasmid pNL4-3-EGFP-PM was further modified by introducing small linker sequence instead of NL4-3 gag region.

5.1.3. Construction of plasmid pNL4-3-EGFP-PM-linker

In order to generate patient HIV gag-recombinant viruses without the possibility of contamination with NL4-3 gag sequence the original *gag* region in plasmid pNL4-3-EGFP-PM was replaced with short linker.

Plasmid pNL4-3-EGFP-PM was digested by enzymes PspXI and MluI overnight and large fragment was isolated by excision from gel, as any contamination by smaller fragment would be counterproductive (fig. 27). Linker was prepared from primers PspXI-MluI-F and PspXI-MluI-R. Primers were mixed together, heated up for 95°C for 3 minutes and let to cool. Prepared dsDNA was digested by enzymes MluI and PspXI and product PspXI-MluI linker was verified on 3% agarose gel electrophoresis (fig. 27).

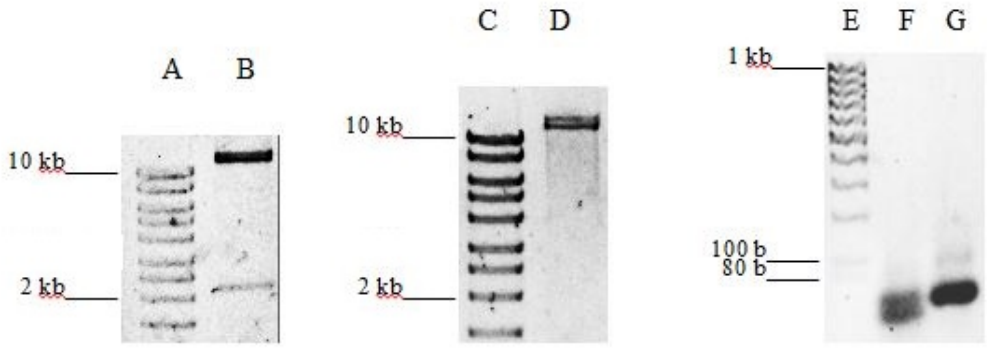


Figure 27. Verification of isolated vector pNL4-3-EGFP-PM and linker PspXI-MluI. A, C & E) Marker. B) plasmid pNL4-3-EGFP-PM digested by MluI and PspXI. D) Large fragment isolated from digested plasmid pNL4-3-EGFP-PM. F) dsDNA of linker PspXI-MluI digested by MluI and PspXI. G) dsDNA of linker PspXI-MluI.

Isolated large fragment of pNL4-3-EGFP-PM was dephosphorylated and mixed with digested linker PspXI-MluI in molar ratio 1:3. One tenth of ligation mixture was transformed in *E. coli* and several colonies were scraped off the plate.

DNA isolated from colonies was digested by enzymes AatII and SphI. Restriction site for SphI is present in replaced *gag* gene; therefore plasmid with linker cannot utilize it. Digested DNA was verified by 1% agarose gel electrophoresis (fig. 28).

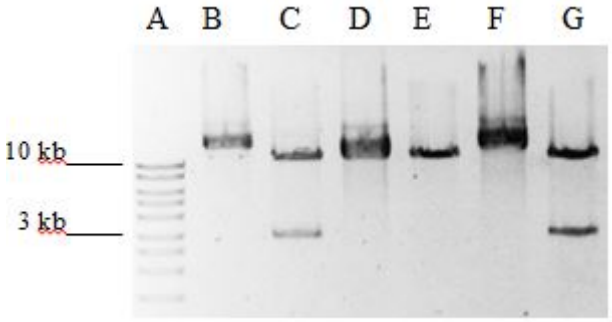


Figure 28. Digestion verification of ligation. A) Marker. B, D & F) DNA isolated from three different colonies, unrestricted. C, E & G) DNA isolated from those three colonies restricted by enzymes AatII and SphI. Presence of 3 kbp long fragment displays presence of both restriction sites.

Presence of 3 kbp long fragment on gel indicated occurrence of gag and absence of the linker sequence. One of the samples from single colony demonstrated only linearization after digestion (fig. 28; E). In addition, this clone had correct size as expected from pNL4-3-EGFP-PM-linker plasmid. Finally, presence of the linker, replacing *gag* gene and surrounding sequence, was verified by Sanger sequencing (fig. 29). Both restriction sites; PspXI and MluI are now separated only by short linker.

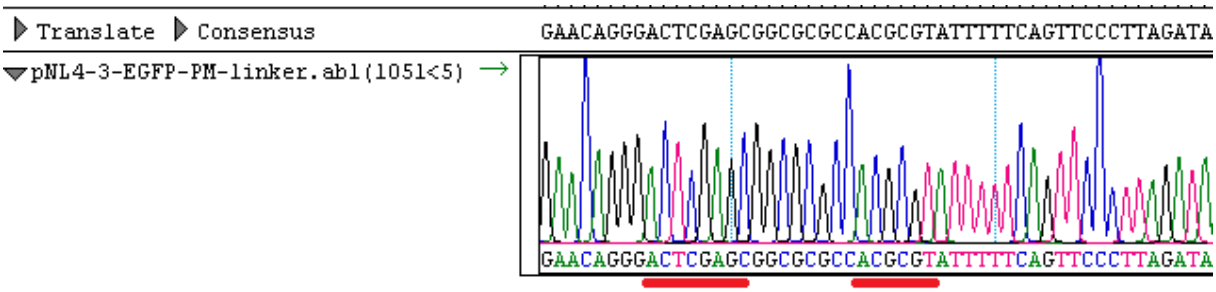


Figure 29. Sequencing of introduced linker PspXI-MluI. Both, PspXI and MluI restriction sites (highlighted red) are present in short sequence of linker replacing *gag* gene and surrounding sequence. Picture was prepared in program SeqMan Pro (DNASTAR v.10.0.1).

As a final result, plasmid pNL4-3-EGFP-PM-linker has replaced *gag* gene with surrounding sequence. This plasmid is used for incorporation of patient derived *gag* and production of recombinant virus upon transfection (fig. 30). In summary, together with plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel, pNL4-3-EGFP-PM-linker are both prepared for generation of recombinant viruses.

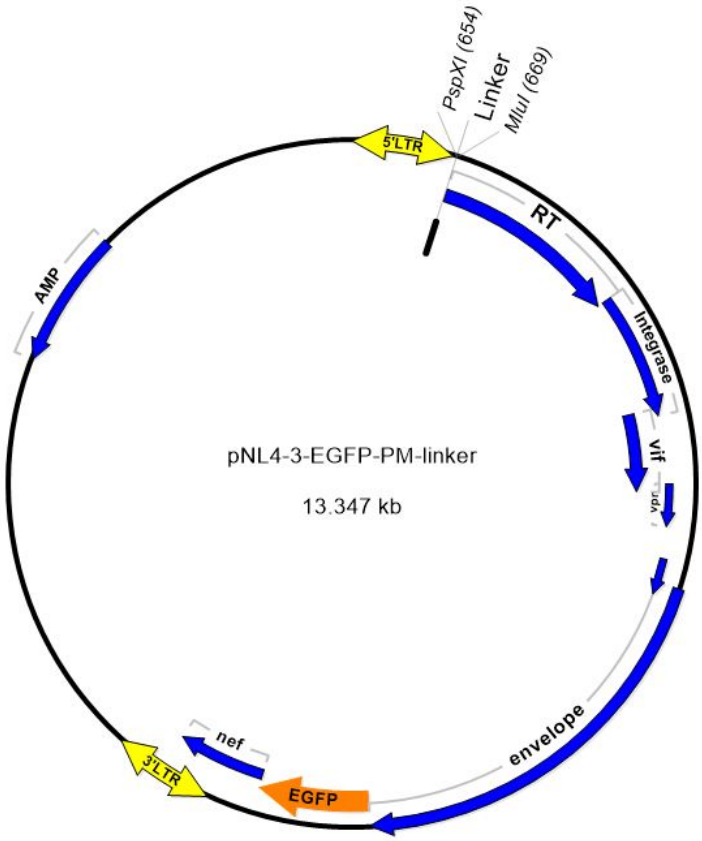


Fig. 30. Plasmid pNL4-3-EGFP-PM-linker. Plasmid carries sequence of HIV with *gag* gene replaced by short linker (15 bp). Linker ir surrounded by restriction sites PspXI and MluI.

5.2. Generation of recombinant viruses

To begin with replicative fitness evaluation, set of recombinant viruses had to be generated. Preparation of recombinant virus from patient-derived plasma samples and constructed plasmids was crucial part of my research. In order to achieve production of replication-competent recombinant virus a series of consecutive steps was performed (fig. 31). Briefly, RNA was isolated from patient-derived plasma and *gag* region (gag-protease) was amplified by PCR. Next, through steps of homologous recombination in yeast, transformation into E.coli, digestion and ligation of patient's *gag* region into linker vector a complete replication-competent plasmid was produced. Finally, 48 h after transfection we harvested recombinant virus.

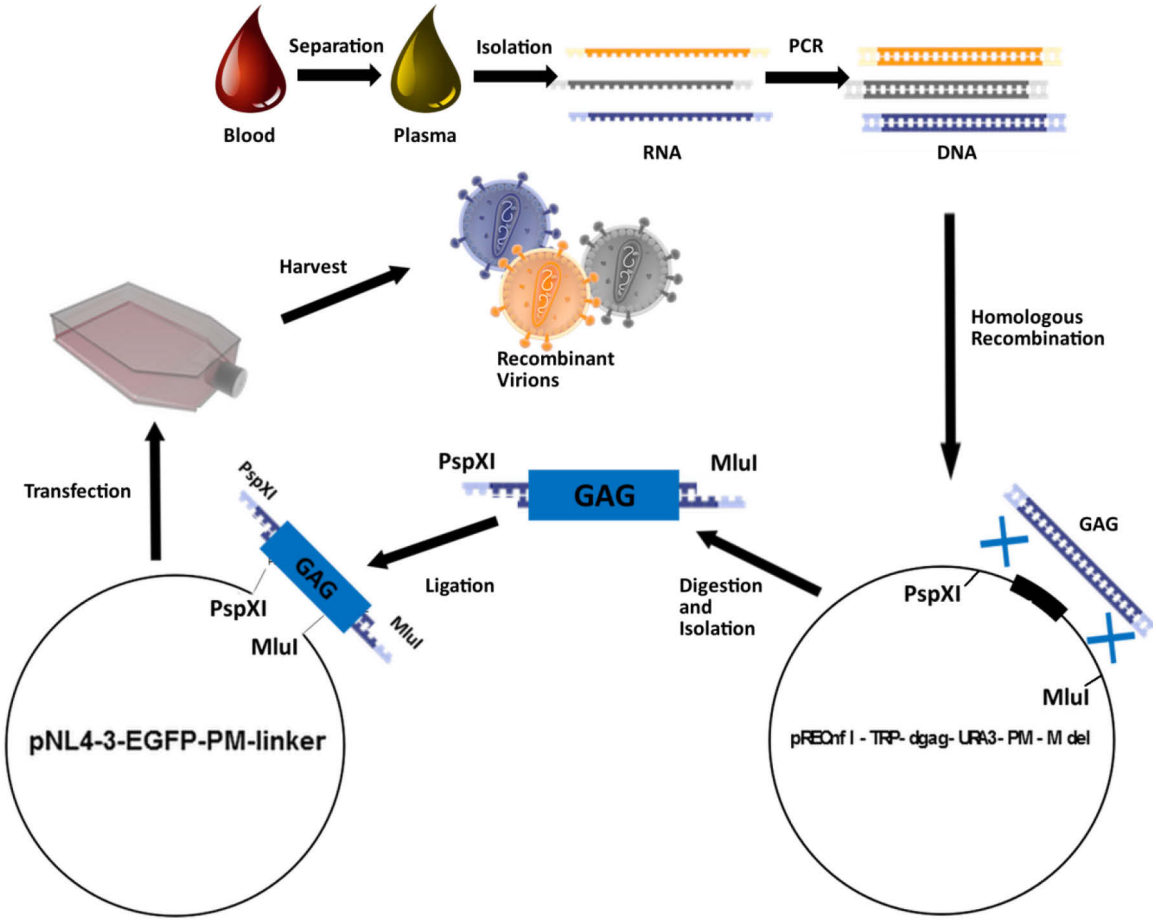


Figure 31. Generation of recombinant virus with patient-derived *gag*. First, blood of HIV positive patient is separated into cell pellet and plasma. RNA is isolated from plasma and *gag* region is reverse transcribed to cDNA and amplified by PCR. Afterward, amplicon is recombined into plasmid $pREC_{nfl}$-TRP-Δgag/URA3-PM-Mdel by homologous recombination in yeast. After recombination, incorporated *gag* is excised with surrounding sequence by PspXI and MluI. Next, excised fragment is ligated into PspXI and MluI digested plasmid pNL4-3-EGFP-PM-linker, resulting in pNL4-3-based plasmid with patient-derived *gag* and EGFP probe. Finally, ligated plasmid is transfected into HEK293T cells and recombinant virus is produced.

5.2.1. PCR amplification of patient *gag* sequence

To begin with generation of recombinant virus, *gag* specific region is amplified by PCR. Patient-derived RNA was reverse transcribed into cDNA using primer POL PRO REC CON BWD 7. In addition, cDNA was amplified by two-step PCR using POL PRO REC CON BWD 7 and GAG P17 FWD 8M pair of primers in first, external PCR step; and POL PRO REC CON BWD 8 and GAG P17 FWD 7M pair of primers in second, nested PCR step. Second step of PCR is necessary to enhance PCR product yield.

As a matter of fact, all primers are degenerate to overcome sequence variability of HIV and relatively long to facilitate yeast homologous recombination. It increased our success rate of PCR to 81 % (Table 1). In general, purpose of PCR was to amplify ~1.9 kb long fragment (fig. 32) containing patient's *gag* and surrounding sequence, utilized during homologous recombination.

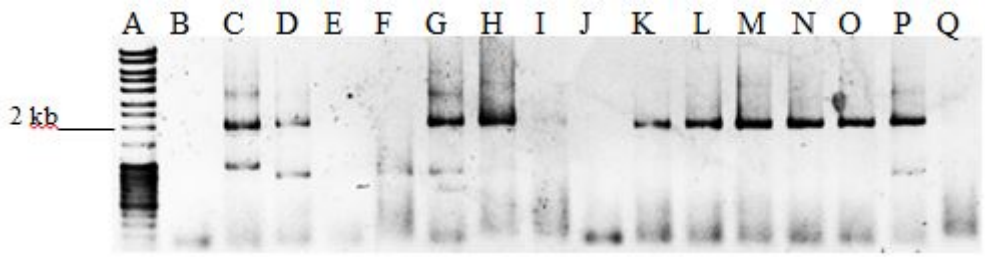


Figure 32. PCR amplification. A) Marker. B-Q) Patient-derived RNA was amplified by PCR into dsDNA. Columns B-Q displays particular patient-derived samples. Indeed, some of the samples are absent of ~1.9 kb long fragment. In addition, several samples display presence of non-specific band.

Although most of PCRs were successful, samples associated with viral load fewer than 10,000 copies per ml of blood were hardly amplified. Moreover, some of the samples also exhibited non-specific bands. Eventually, presence of such a non-specificity hampered subsequent steps of recombinant virus production.

Number of samples in PCR	Number of amplified samples
105	85

Table 1. Success rate of PCR. PCR amplification was successful in 81 % of the samples.

5.2.2. Homologous recombination in yeast cells

First, PCR-amplified *gag* amplicon is purified to obtain pure DNA for homologous recombination. Step of homologous recombination in yeast is crucial for maintaining

sufficient quasispecies of virus. In fact, number of yeast colonies after homologous recombination was deciding factor if the procedure was continued.

To begin with homologous recombination, plasmid pREC$_{nfl}$-TRP-Δgag/URA3-PM-Mdel was linearized by enzyme SacII and used as a vector. Next, purified *gag* amplicon was introduced into the vector by homologous recombination. When homologous recombination occurs negative selection marker *ura3* gene is replaced by patient-derived sequence.

Yeast colonies were counted and only plates with more than 100 colonies were used for DNA preparation. DNA extracted from yeast cells contains large amounts of genomic DNA and only minority of plasmid DNA, therefore DNA isolated from all yeast colonies was transformed in *E. coli*. Again, only plates with at least 100 bacterial colonies were used for preparation of plasmid DNA. Finally, size and quality of DNA isolated from all colonies was verified by 0.8% agarose gel electrophoresis (fig. 33).

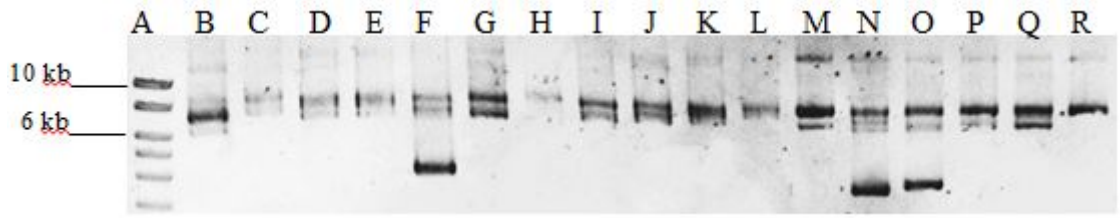


Figure 33. Homologous recombination in yeast. DNA recombined in yeast cells and amplified in bacterial cells was purified and electrophoresed. A) Marker. B-Q) Samples of DNA purified from particular bacterial colonies. R) Positive size control (pREC$_{nfl}$-TRP-HIV).

As a result of successful recombination, the size of the plasmid increases by 345 bp. Agarose gel electrophoresis provides first control of recombination since the size of recombined plasmid pREC$_{nfl}$-TRP-'patient-derived-gag'-PM-Mdel should match the size of positive control pREC$_{nfl}$-TRP-HIV. After size verification, plasmid was digested by enzymes PspXI and MluI overnight and result was verified by 1% agarose gel electrophoresis (fig. 34).

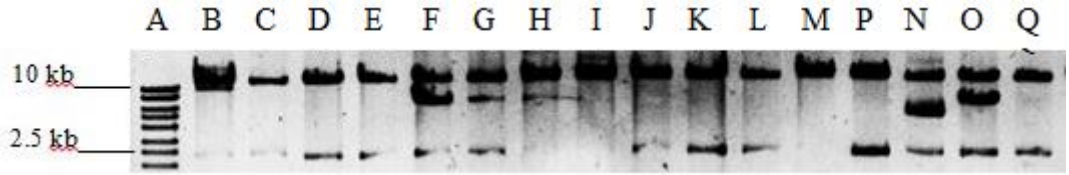


Figure 34. Digestion of plasmids by enzymes PspXI and MluI. DNA purified from samples was digested and subsequently, presence of 2.2 kb long fragment was verified by agarose gel electrophoresis. A) Marker. B-Q) Samples of plasmid DNA digested by enzymes PspXI and MluI.

Samples with weak correct band (fig. 33) yield weak or no excised fragment while digestion of strong correct samples usually results in the presence of excised band (fig. 34). Nevertheless, several samples with correct size band in fig. 33 did not yield correct excised fragment in fig. 34. Presence of larger bands than excised fragment is not detrimental, since fragment of interest was purified by 0.8% E-Gel Electrophoresis system (fig. 35).

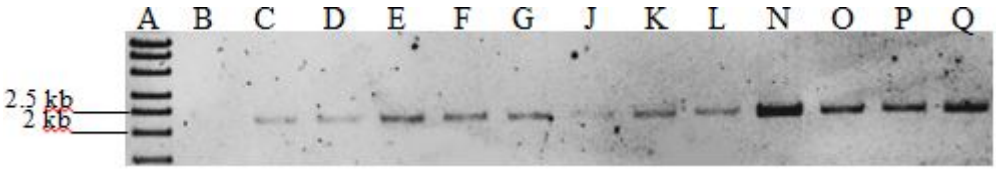


Figure 35. Excised fragments from pREC$_{nfl}$-TRP-'patient-derived-gag'-PM-Mdel. DNA of excised fragment was purified by 0.8% E-Gel electrophoresis. A) Marker. B-N) Isolated fragments from particular samples. 10 µl of fragment was loaded on gel.

Success rate of homologous recombination is 81 % (Table 2). In summary, after successful homologous recombination and maintenance of quasispecies, patient-derived *gag* fragments with surrounding NL4-3-derived sequence were excised and purified. Purified fragments were used in ligation to obtain replication-competent plasmid.

Number of samples initiating homologous recombination	Number of samples with successful excision of fragment
85	**69**

Table 2. Success rate of homologous recombination. Homologous recombination in yeast was successful in 81 % of the samples.

5.2.3. Ligation

In the next step of recombinant virus production, we had to transfer excised fragment into the plasmid with full HIV sequence however, with *gag* replaced by short linker. Excised *gag* fragments were ligated in pNL4-3-EGFP-PM-linker. Initially, vector pNL4-3-EGFP-PM-linker was digested by enzymes MluI and PspXI overnight and verified by 0.8% agarose gel electrophoresis (fig. 36).

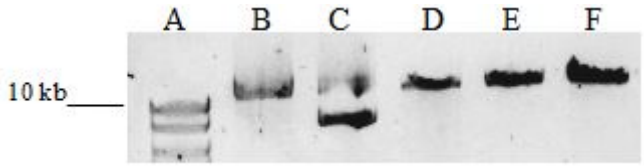


Figure 36. Linearization of vector pNL4-3-EGFP-PM-linker. A) Marker. B, C) Vector pNL4-3-EGFP-PM-linker isolated by QIAprep Spin Midiprep (B) and Maxiprep (C) Kits. D, E) Vector pNL4-3-EGFP-PM-linker restricted by enzymes MluI and PspXI. F) Positive size control of linearization.

After restriction verification, vector was dephosphorylated to minimize self-ligation. Next, ligation was performed with 30 ng of linearized vector and purified fragment from pREC$_{nfl}$-TRP-'patient-derived-gag'-PM-MdeI in molar ratio 1:3. Subsequently, 1.5 µl of mixture was transformed in bacteria by electroporation. As previously described, number of colonies is important factor for maintaining quasispecies variation. Samples with less than 100 colonies on the plate were not processed. Finally, DNA was purified from all colonies and verified by 0.8% agarose gel electrophoresis (fig. 37).

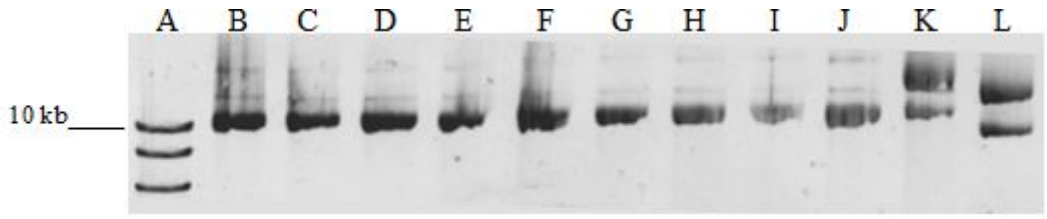


Figure 37. DNA preparation of pNL4-3-EGFP-PM-linker. A) Marker. B-J) DNA isolated from colonies of particular samples. K) Size control of successful ligation. L) Size control of unsuccessful ligation.

80% success rate of ligation was limited by number of colonies and correct size of isolated plasmid (Table 3). Whereas plasmid pNL4-3-EGFP-PM provides positive size control, plasmid pNL4-3-EGFP-PM-linker serves as negative size control to ligation (fig. 37; K and L). Plasmid concentration was determined for each sample and 8 µg of DNA were used in transfection of HEK293T cells.

Number of samples initiating subcloning	Number of samples with successful subcloning
69	55

Table 3. Success rate of subcloning. Depicted is number of samples going into the subcloning and number of samples yielding at least 100 colonies after ligation step.

5.2.4. Titer measurement of recombinant virus and selection of samples

After subcloning step, plasmid pNL4-3-EGFP-PM carrying patient-derived *gag* was transfected. Next, virus was used for titer evaluation by endpoint dilution assay. Fifty percent tissue culture infective dose ($TCID_{50}$) was calculated. Success rate of transfection was 100 %, thus we prepared 55 recombinant viruses. Although we measured $TCID_{50}$ for all prepared viruses, only 24 samples were further analyzed (Table 4).

Sample	TCID50 (10^)/ml	Sample	TCID50 (10^)/ml	Sample	TCID50 (10^)/ml	Sample	TCID50 (10^)/ml
1*1	4	7*1	4.5	13*1	4.6	7*2	4
2*1	3.9	8*1	3.9	14*1	4.3	8*2	4.5
3*1	4	9*1	3.9	2*2	3.9	9*2	4.2
4*1	3.9	10*1	4.3	3*2	4.5	10*2	4.5
5*1	4.3	11*1	4.6	4*2	4.2	11*2	4
6*1	4.1	12*1	4.8	5*2	4.2	14*2	3.6

Table 4. Overview of recombinant viruses with titer. $TCID_{50}$ was calculated for each sample by endpoint dilution method. First number states particular patient-derived sample, number separated by asterisk marks earlier (*1) or later sampling (*2).

Concerning samples selection, they were divided according to two criteria in individual batches. First, to compare fitness contribution of *gag* to overall fitness, we selected ten viruses with corresponding primary isolate in one batch (Table 5). Second, we selected matching pairs of recombinant viruses derived from the same patient from two separate time points and organized them in another batch (Table 5). Purpose was to compare replicative fitness development using early- and late-sampling derived *gag* recombinant virus. Distribution of samples goes along with particular aims of my thesis. Nomenclature of samples is marked by numbers to distinguish particular patients. Specifically, samples are marked by *1 add-on to indicate earlier patient-derived sample and *2 to indicate later patient-derived sample. Between earlier and later sampling is span of at least two years.

First batch of samples		Second batch of samples	
1*1	1 – primary isolate	2*1	2*2
2*1	2 – primary isolate	3*1	3*2
5*1	5 – primary isolate	4*1	4*2
6*1	6 – primary isolate	5*1	5*2
8*1	8 – primary isolate	7*1	7*2
9*1	9 – primary isolate	8*1	8*2
10*1	10 – primary isolate	9*1	9*2
11*1	11 – primary isolate	10*1	10*2
12*1	12 – primary isolate	11*1	11*2
13*1	13 – primary isolate	14*1	14*2

Table 5. Samples distribution in batches. First batch of samples was used to compare recombinant virus and primary isolate replicative fitnesses. Second batch was used to determine replicative fitness development over time.

5.2.5. Sequence integrity verification

Before we can proceed with replicative fitness characterization it is necessary to verify sequence integrity of prepared recombinant viruses. For this reason, RNA was purified from 24 recombinant viruses. Similarly, we isolated RNA directly from plasma of patients. Next, *gag* gene region was sequenced and sequences were aligned by ClustalW method. Sequence of NL4-3 virus was included as a control to exclude any contamination issues. Phylogenetic analysis was performed two times according Table 5.

First, *gag* recombinant virus sequences were aligned with corresponding primary isolate sequences. Primary isolate sequence represents here HIV genetic information from RNA isolated from patient's plasma. Phylogenetic analysis demonstrates close sequence proximity between patient-derived *gag* recombinant virus and same patient-derived primary isolate (fig 38.).

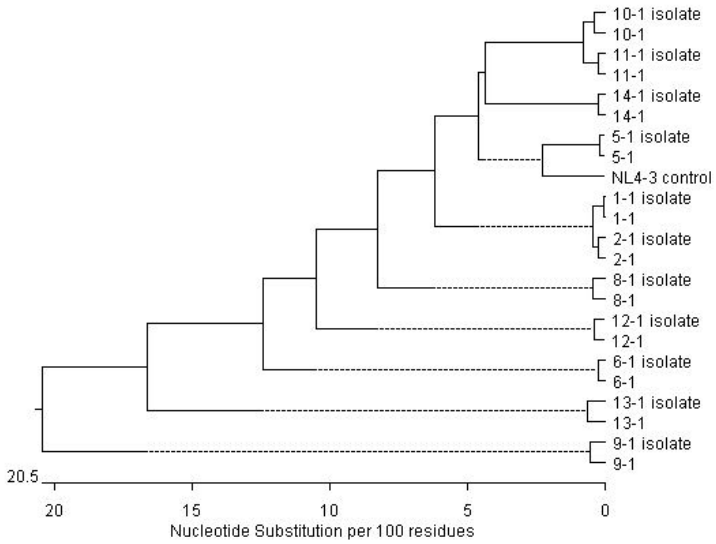


Figure 38. Phylogenetic tree of primary isolates and corresponding recombinant viruses. Alignment was done by ClustalW method. Numbers indicates particular patients; number followed by 'isolate' indicates sequence of primary isolate. NL4-3 is a control sequence.

Likewise, *gag* recombinant virus sequence from earlier patient-derived recombinant virus and from later patient-derived recombinant virus were aligned. Phylogenetic tree displays nucleotide substitution rate in *gag* region between early and late sampling (fig. 39).

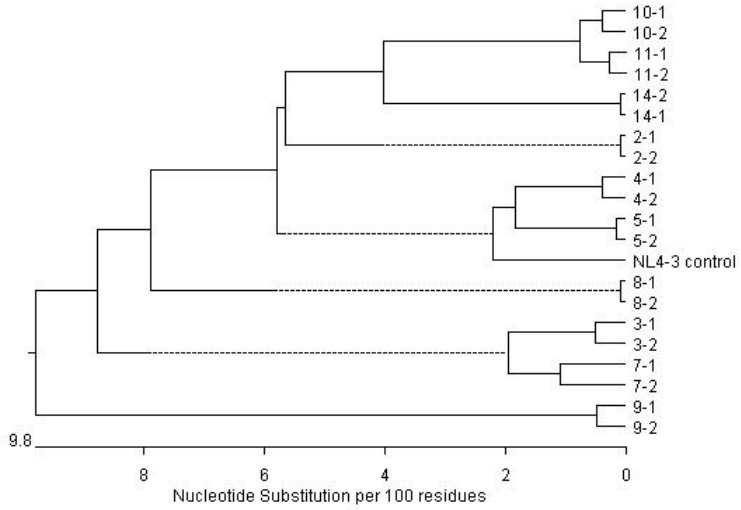


Figure 39. Phylogenetic tree of early- and late-sampling derived recombinant viruses. Alignment was done by ClustalW method. Numbers indicates particular patients. Marks '-1' and '-2' point on earlier and later sampling-derived recombinant virus. NL4-3 is a control sequence.

After alignment, phylogenetic trees were constructed using MegAlign program (DNASTAR v.10.0.1). Samples from the same patients are clustering together in the tree. Expectedly, sequence distance between first and second sampling-derived recombinant virus from the same patient (fig 39.) is larger than between first sampling-derived recombinant

virus and primary isolate sequence from the same patient (fig. 38). NL4-3 sequence is not identical with any of sequenced samples; hence contamination of samples is excluded. Sequence verification is final step in the preparation of recombinant viruses. In summary, recombinant viruses were prepared with 52% success rate in total. From pool of prepared 55 viruses, 24 candidates were chosen for further replicative fitness measurement.

5.3. Patients' characterization

For purpose of further data analysis we collected demographic, clinical and virological parameters for all fourteen patients. $CD4^+$ T cells counts were determined at Hospital Na Bulovce and viral loads were measured in National Institute of Public Health. $CD4^+$ T cells counts and viral loads were determined during routine medical examination usually two or three times per year. Characteristics are shown in Table 6.

Patients' characterization							
Patient	Sex	Age in the time of last sampling	Follow-up (months)	Average $CD4^+$ T cell per µl	Slope of $CD4^+$ T cells	Average viral load average (log of RNA copies per ml)	Protective allele
1	Male	57	51	505	**3.1**	5.24	None
2	Male	26	73	681	**2.4**	4.67	HLA-B*13
3	Male	36	74	447	**-0.6**	4.13	HLA-B*27
4	Male	44	60	690	-6.6	4.35	HLA-B*57
5	Male	32	31	398	-9.6	5.38	None
6	Male	39	20	608	-9.2	4.4	None
7	Male	32	166	544	-1.3	4.58	None
8	Male	39	58	986	-3.7	4.15	HLA-B*13
9	Male	44	72	813	-11.6	4.5	None
10	Male	38	55	674	**1**	3.98	None
11	Male	37	57	705	-2.5	5.25	None
12	Male	30	24	650	-8	5.09	None
13	Male	40	35	660	**7.8**	4.53	None
14	Male	38	32	335	-4.2	4.67	None

Table 6. Patients' characterization. In table are present summarized data about patients. $CD4^+$ T cells counts were provided by Hospital Na Bulovce and viral loads are from National Institute of Public Health. Slopes of $CD4^+$ T cells were calculated by linear regression of all counts of $CD4^+$ T cell per µl during the follow-up. Bold numbers of $CD4^+$ T cell slope marks slow progressors. HLA typization was done by Inno-Train GmbH.

Slopes of $CD4^+$ T cells were calculated as a $CD4^+$ T cell decline determined as cells/µl per month using all $CD4^+$ T cell counts during the clinical follow-up. This slope displays general trend of patient's immunological status development during infection. Further, we used $CD4^+$ T cell slopes to divide patients in arbitrary categories of slow progressors (slope > -1) and typical progressors (slope < -1). All collected samples and analyzed data were from untreated patients infected by subtype B. Age of patients is displayed at time of later sampling.

5.4. Replicative fitness measurement

5.4.1. Viral growth kinetics evaluation

Initially, replicative fitness of all recombinant viruses was determined by viral growth kinetics. Experiment was evaluated by RT assay performed on cell-free supernatants from individual collection days. All viral growth kinetics wereperformed in triplicate and aliquots were harvested from each well. After RT assay assessment, curves of viral growth (fig. 40) were drawn and slopes were calculated through linear regression from ln transformed day 0 values to day 3 and day 4 values respectively. In my experiments, peak of infection was usually reached during day 4.

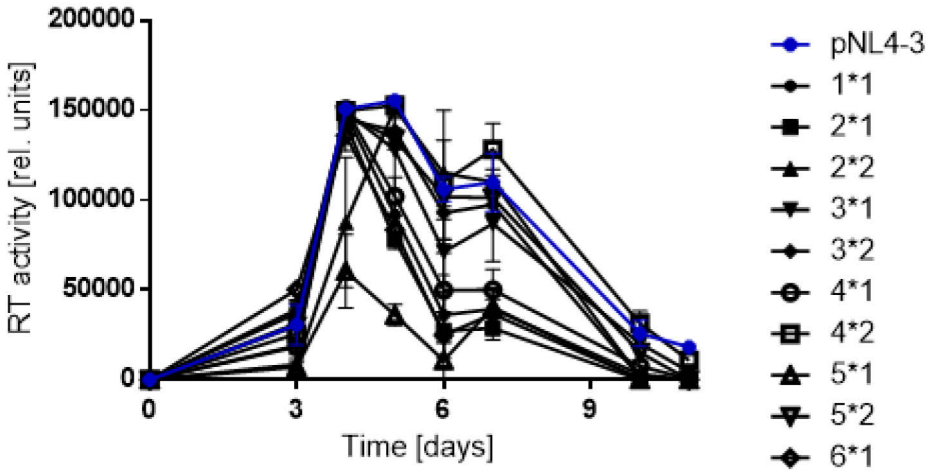


Figure 40. Viral growth curves. Virus growing in cells was collected in particular days. Reverse transcriptase activity of collected samples was measured by radioactive assay and growth kinetics was constructed. Displayed are samples of recombinant viruses used for comparison of replicative fitness with primary isolate.

Sample 11*2 is exception in calculation of slopes average. Its slope from day 0 to day 3 is absent in calculation due to impossibility to drawn ln transformation of zero. Slopes

distribution was plotted in graph using mean with minimum to maximum dispersion of values (fig. 41).

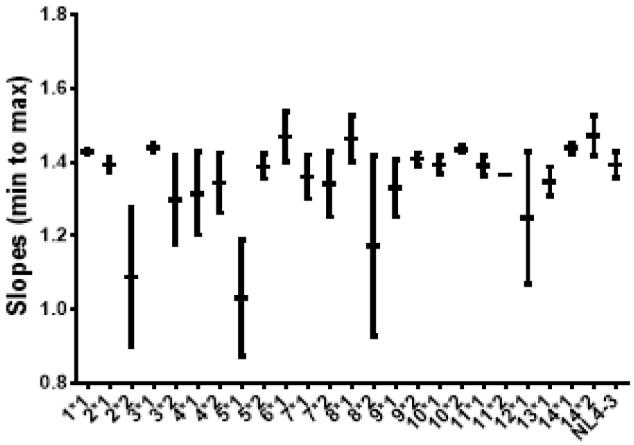


Figure 41. Mean of slopes of recombinant viruses evaluated from viral growth kinetics. Value for each sample is drawn from slopes calculated from day 0 to days 3 and 4. Value is displayed as mean of slopes of included days; whiskers indicate minimal and maximal dispersion of values. NL4-3 is a control.

Furthermore, for purpose of subsequent analysis, values from figure 41 were normalized to highest value (referred as 100%) in the set, excluding NL4-3-EGFP control. Span of samples is 79.9 % to 100 %, thus the weakest virus is 1.25 times weaker than the strongest.

Replicative fitness evaluation of primary isolates was done analogously (fig. 42). However, due to usage of different type of cells for viral growth kinetics of primary isolates, different control virus was chosen. Namely, strain B2 (92US076) is preferred for control infection in PBMCs, since NL4-3 is performing weakly in this type of cells. Normalized values of sample were between 51 % and 100 %. Therefore, replicative fitness measured on primary isolates shows greater range of values than replicative fitness of recombinant viruses, since the weakest virus is 2 times weaker than the strongest one.

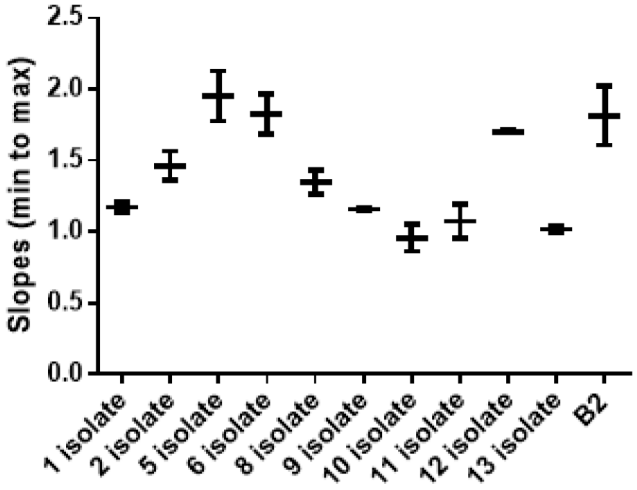


Figure 42. Mean of slopes of primary isolates evaluated by viral growth kinetics. Value for each sample is drawn from slopes calculated from day 0 to days 3 and 4. Value is displayed as mean of slopes of included days; whiskers display min – max distribution of values. B2 is a control virus.

From comparison between figure 41 and 42 is apparent, that values are more dispersed in the case of recombinant virus growth kinetics, especially for recombinant viruses 2*2, 5*1, 8*2 and 12*1. This effect is due to inequalities in triplicate measurement (days 3 and 4). However, those values were used for further analysis, since any outliers were not detected by ROUT method in GraphPrism program (Q = 1 %). Next, values drawn from figure 42 were also normalized to the highest value in the set (excluding B2 control). Normalized percentage values were used in correlations.

5.4.2. Viral competition evaluation

Additionally, viral competition experiment was used to evaluate replicative fitness of recombinant viruses. Viral competition was done at MOI = 0.0001 (according to $TCID_{50}$ value measured in TZM-bl cells) for each virus participating in dual infection. Dual infection was done in triplicate for each recombinant/control virus pair. Competition in each well was between sample of recombinant virus with EGFP probe and control NL4-3 virus with red E2-Crimson probe. Dual infections were analyzed by flow cytometry. Number of counted events was set up to 10,000 within manually defined gating. Manual gating filtered only living single cells to be counted as an event. For further data analysis, automated gating approach was adopted (Finak *et al.* 2014). Next, data obtained after automated gating processing were used for calculation of selection coefficients (*s*).

$$s = \frac{\ln[\frac{H(T)}{H(0)}]}{ln16+\ln[\frac{W(T)}{W(0)}]+ \delta T}$$

Where, W(T)/W(0) is fold change of NL4-3 virus with E2-Crimson probe from day 4 to day 3. H(T)/H(0) is fold change in EGFP/E2-Crimson carrying viruses ration from day 4 to day 3. Ln16 in denominator stains for 16x dilution between day 3 and 4 and δT is dying coefficient of cells, stated arbitrarily as 0.5, multiplied by time of experiment. Equation for *s* calculation was adapted from work of Marée and his colleagues (2000). Relation between day 3 p. i. and day 4 p. i. was chosen to represent viral competition development, based on our previous experiments. We diluted dual infections by fresh media to repress number of double-infected cells in day 4 measurement. Finally, relative fitness was derived according to formula 1 + *s*. Relative fitness calculation was done for every well separately. Three values, corresponding to a triplicate of sample, were plotted in figure 43 as mean with minimum to maximum value distribution. Relative fitness was normalized to highest value (excluding control). Values were spanning between 59 % and 100 %. Thus, the weakest virus was competing with control 1.7 times worser than the strongest virus.

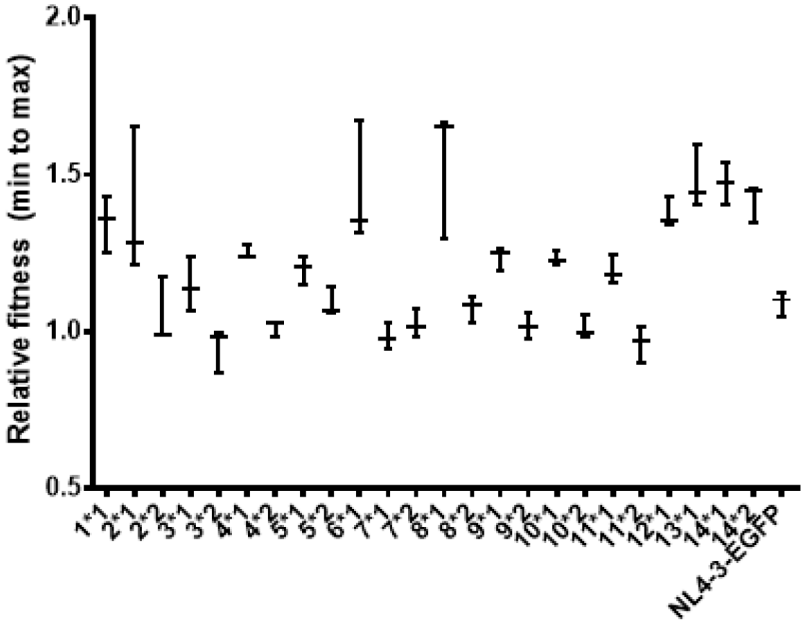


Figure 43. Relative fitness of recombinant viruses measured by viral competition. Relative fitness was derived from selection coefficient (*s*); *s* was calculated according to formula used previously (Marée *et al*. 2000). Relative fitness was calculated in triplicate and plotted as mean with min to max value distribution. NL4-3-EGFP is control dual infection, where NL4-3 virus with green probe competed with NL4-3 with red probe.

Furthermore, presence of positively infected cells was verified by confocal microscopy. Measurement was done in day 5 of undiluted dual infection. Noticeably, without

dilution, almost all cells are exhibiting both colors (fig. 44). We repressed numbers of double infected cells by dilution, since presence of double infected cells is adversely influencing calculation of selective coefficient.

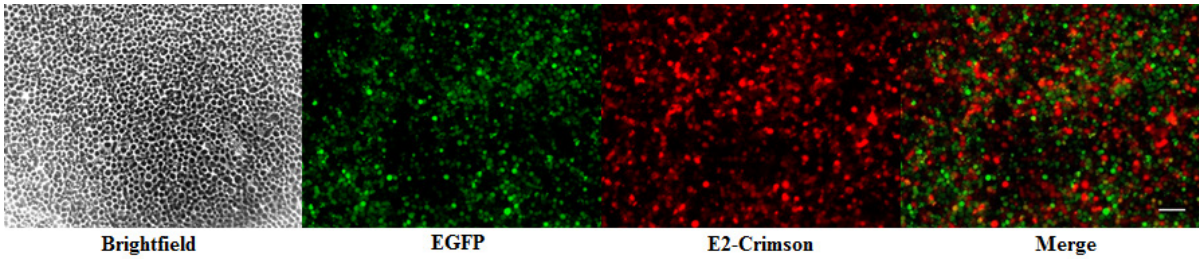


Figure 44. Microscopy of dual infection. MT4 cells were observed under microscope at day 5 p. i. in undiluted sample of dual infection between NL4-3-EGFP and NL4-3-E2-Crimson. At day 5, signal of double infected cells is visible. Scale bar of 10 µm.

5.4.3. Comparison of replicative fitness measurement methods

Next, we were interested in how replicative fitness results from viral growth kinetics and viral competition compare to each other. For that, we plotted values obtained through viral growth kinetics against values of relative fitness obtained from competition experiment (fig. 45).

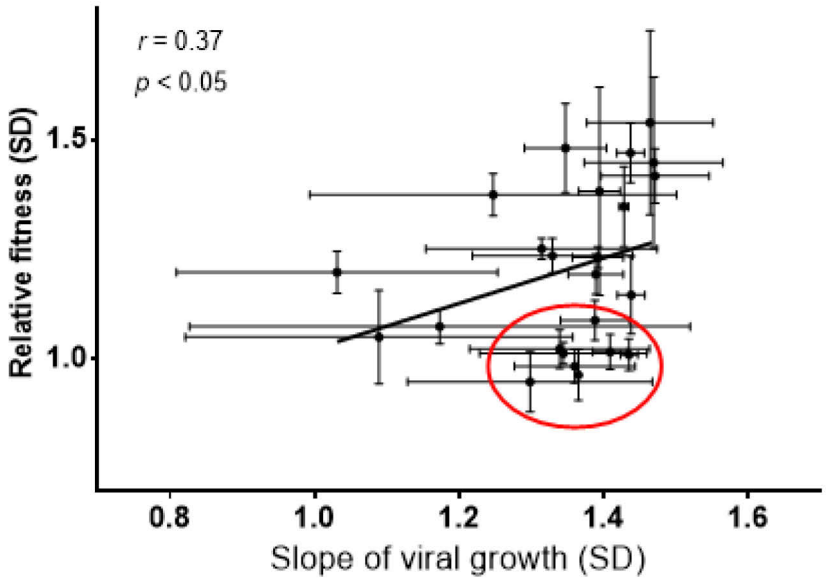


Figure 45. Comparison of replicative fitness measurement outcomes. Relative fitness was measured in triplicate and plotted with standard deviation against slope of viral growth with standard deviation. Values of all 24 recombinant viruses were plotted. Trend describing curve was drawn through regression and is describing monotonic trend with $p < 0.05$ (Spearman's correlation). Red circle marks group of recombinant viruses with high slope in viral growth kinetics but poor relative fitness in competition experiments.

Values of replicative fitness from whole set of 24 recombinant viruses were included and plotted in the graph with standard deviations. We observed statistically significant association between slopes from growth kinetics and relative fitness from competition experiments. Interestingly, group of recombinant viruses with relatively high slope of viral growth but somewhat low value of relative fitness can be observed in the graph (fig. 45; red circle). This group represents recombinant viruses that replicate fairly well in mono-infections but they are outcompeted with our control NL4-3 virus in competition assays.

5.4.4. Comparison of replicative fitness between *gag* recombinant virus and primary isolate

According to first aim of this thesis, replicative fitness was compared between primary isolates and *gag* recombinant virus. Replicative fitness of recombinant virus was obtained through viral growth kinetics experiment and through viral competition experiment. On the other hand, replicative fitness of primary isolates was measured only by viral growth kinetics. Normalized values were plotted in two separate graphs (fig. 46). For comparison 10 pairs of *gag* recombinant and primary isolate were used.

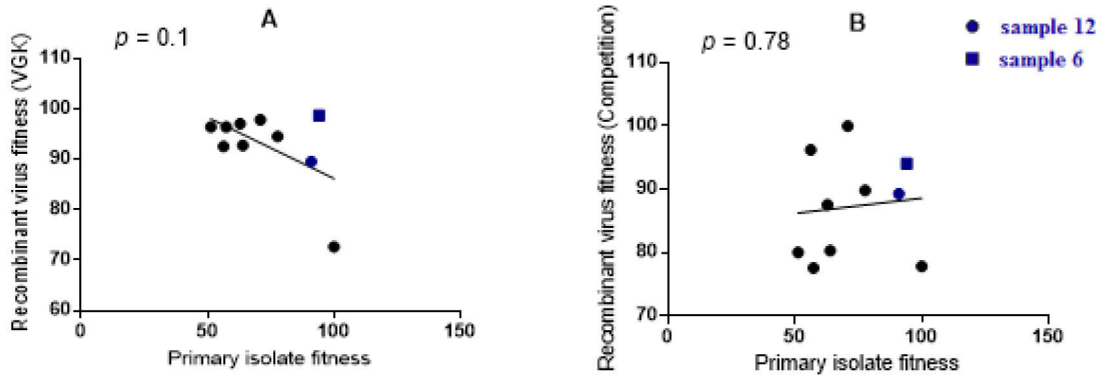


Figure 46. Replicative fitness comparison between *gag* recombinants and primary isolates. A) Replicative fitness values obtained through viral growth kinetics are plotted against replicative fitness values of primary isolates. **B)** Replicative fitness values obtained through viral competition are plotted against replicative fitness values of primary isolates. All values were normalized to highest value in particular set. Trend line is drawn through linear regression; correlation was calculated by Pearson's correlation, without statistical significance. Blue square and circle mark samples of patients 6 and 12.

Furthermore, linear trends were drawn in graphs through linear regression. Trend of replicative fitness of recombinants and primary isolates, measured by growth kinetics, is without statistical significance (fig. 46; A). Similarly, trend of replicative fitness measured by

competition experiment is also statistically insignificant (fig. 46; B). However, we identified two samples, namely 6 and 12. Those samples display similar replicative fitness values, measured on recombinant virus same as on primary isolate. Importantly, selected samples exhibit similar values not only if they are measured by viral growth kinetics, but also measured by competition experiment. However, due to low number of samples (n=2), correlation is not supported with statistical test.

5.4.5. Characterization of replicative fitness development in disease progression

In addition to comparison of replicative fitness between primary isolates and recombinant viruses, we were interested in determining if fitness of gag recombinant viruses changes over time and if it correlates with disease progression in patients without antiretroviral treatment. In order to quantify change of replicative fitness over time, 10 pairs of recombinant viruses were prepared (Table 5). Each pair was derived from same patient with a span of at least two years between samplings. We hypothesized that two years' span will be enough to observe change in replicative fitness of *gag* recombinant viruses.

Initially, slopes obtained from viral growth kinetics for early- and late-sampling derived recombinant virus (fig. 41) were converted to natural logarithm. Slopes were drawn by linear regression between two values, one for each time-derived sampling. Calculated slopes are describing fitness development in patients. Disease progression in patient was evaluated, as $CD4^+$ T cells counts over time, displayed as slope of $CD4^+$ T cells in Table 6. Finally, slope of $CD4^+$ T cells was plotted against slope describing change of replicative fitness over time (fig. 47). Calculated values are in Table 7.

Patient	Fitness change	Fold change	Patient	Fitness change	Fold change
2	**-0.305**	**-1.28x**	8	-0.291	-1.25x
3	**-0.140**	**-1.11x**	9	+0.08	+1.06x
4	+0.03	+1.02x	**10**	**+0.043**	**+1.03x**
5	+0.357	+1.35x	11	-0.024	-1.02x
7	-0.021	-1.02x	14	+0.034	+1.02x

Table 7. Replicative fitness change. Replicative fitness change of *gag* recombinant viruses over time. Fitness change is expressed as difference between late and early virus slope from viral growth kinetics. Fold change is expressed as ratio between higher and lower replicative fitness values. Plus sign denotes increase in replicative fitness in late sample, negative sign denotes decrease in replicative fitness in late sample. Bold are marked slow progressors. Blue letters mark carriers of protective allele.

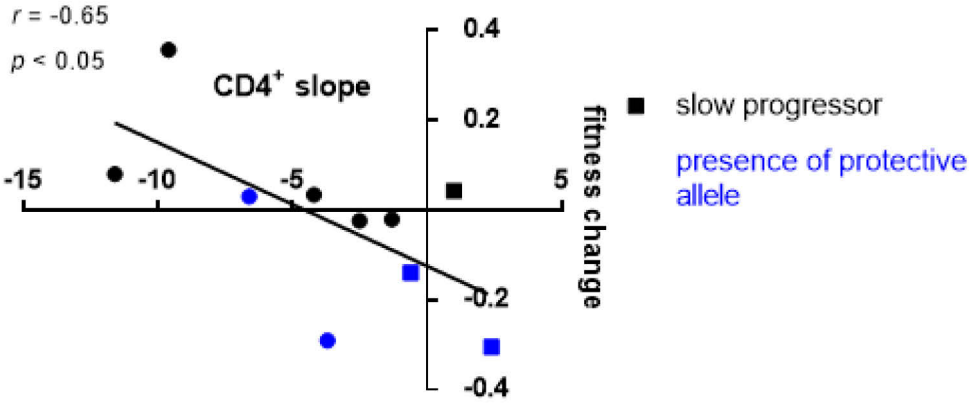


Figure 47. Fitness change and disease progression. Fitness change is calculated from replicative fitness of two recombinant viruses from consecutive blood sample collections from same patient separated by at least two years. Replicative fitness change is calculated from viral growth kinetics; linear regression between values from early- and late-sampling derived patient was used for calculation. Trend describing lines are drawn through linear regression with displayed *p* value and *r* value calculated by Pearson's correlation. Squares display patients who are in category of slow progressors. Presence of blue color marks patients carrying protective allele.

$CD4^+$ T cells counts are widely used as a marker for disease progression; therefore we also adopted this marker. Fitness change calculated from viral growth kinetics, shows to be descriptive factor correlating with disease progression with statistical significance ($p < 0.05$). Fold change of replicative fitness was in the range of -1.28 to +1.35. Furthermore, 2 of 3 slow progressors displayed noticeable decline of replicative fitness over time (-1.28- and -1.11-fold change) together with distinctively stable $CD4^+$ T cells slope. Additionally, virus derived from late sampling from patient 5 was the only one that displayed marked increase of replicative fitness (+1.35-fold change). Carrier of protective allele HLA-B*13, patient 8 displays -1.25-fold decrease of replicative fitness. Other viruses practically did not display any changes in replicative fitness (fold change from -1.02 to +1.06).

5.4.6. Fitness cost characterization

For further investigation of genetic background behind fitness development, I analyzed genetic polymorphism and divergence of each viral isolate. Time span between early- and late-sampling derived recombinant viruses is 2.2 year. In contrast, among 10 compared pairs were three pairs included with larger time span. Namely, samples derived from patients 3, 4 and 7. Specifically, pairs from patient 3 and 4 are separated by four year-long span and pair from patient 7 is spanned by thirteen years. We included those samples

with intention to observe fitness change in larger time scale. Finally, sequence divergence was calculated (fig. 48) from phylogenetic tree by program MegAlign (DNASTAR v.10.0.1).

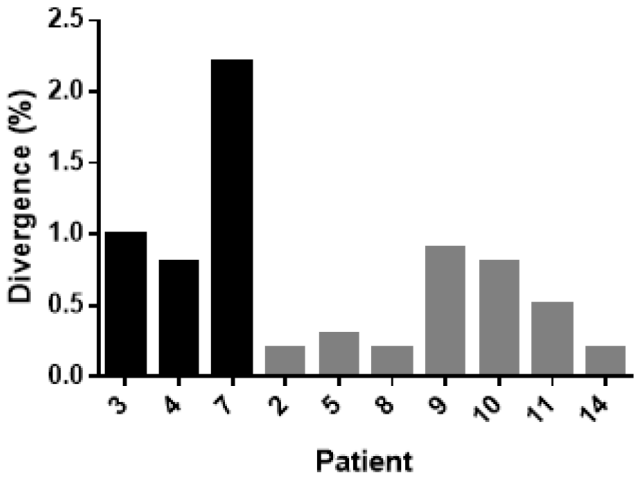


Figure 48. Sequence divergence of recombinant viruses. Pairwise percent identity and divergence were calculated for all 10 pairs of early- and late-sampling derived recombinant viruses. Noticeably, divergence between pairs with longer span, derived from patients 3, 4 and 7 (black bars) is in average higher than between pairs derived from other patients (gray bars).

Distinctly, samples with large span between pair members display higher divergence of sequence. Divergence between 3*1 and 3*2 is 1 % and between 4*1 and 4*2 is 0.8 %, while divergence between 7*1 and 7*2 is 2.2 %. Patient 7 derived pair with thirteen year-long span is expressing the highest level of divergence. In average, samples with longer time span have average divergence 1.3 % meanwhile samples with shorter span have in average 0.4 % divergence. For further interest of fitness development in patient, I examined functional changes in sequence. Specifically, sequences were translated to proteins Gag corresponding to Gag ORFs and searched for known fitness changing mutations linked with CTL escape.

Mutations in Gag protein sequence were analyzed through Epitope Variant and Escape Mutation Database (www.hiv.lanl.gov). Furthermore, work of Boutwell and his colleagues was taken in account (Boutwell *et al.* 2013). As a matter of fact, in early infection, virus is acquiring escape mutations to avoid immune system. Those mutations often reduce replicative fitness of the virus and are later in infection reverted or compensated. All analyzed samples were detected to be carrying escape mutations. Interestingly, some of the escape mutations were present only in sequence of one member of patient-derived pair of recombinant viruses. Escape mutations are displayed in Table 8.

Patient	Early CTL escape mutations	Late CTL escape mutations
2	A146P	A146P
3	**K28R**, Y79F, **R264K**, E312D	Y79F, R91G, **E211D**, **R264K**, E312D, G357S, E412D
4	**K28R**, I147L, T242N, **G248A**, E312D, T427N, E482D	**K28R**, I147L, T242N, **G248A**, E312D, T427N
5	G357S, E482D	K26R, S310T, G357S, E482D
7	A146P, A223I	Y79F, A146P
8	**K28Q**, R91G, A146P, I147L**, E211D,** E312D**,** G357S	**K28Q**, R91G, A146P, I147L, **E211D**, E312D, G357S
9	**K28R**, Y79F, I147L, **T186M**, **T303V**	L75I, Y79F, I147L, **T186M**, **T303V**
10	**K28R**, R91G, **R264K**, **T303V**, E312D	**K28R**, R91G, **R264K**, **T303V**, E312D
11	**R264K**, L268M, **T303V**, E312D, G357S, T427N	**R264K**, L268M, **T303V**, E312D, G357S, T427N
14	**K28R**	**K28R**, I147L

Table 8. Escape mutations in recombinant viruses. Escape mutations with reducing effect on replicative fitness (Boutwel *et al.* 2013). Bold mark mutations have significantly greater reduction of replicative fitness than replicative fitness of virus with RT mutation M184V (serves as benchmark in HIV fitness field).

Some of the samples contain several escape mutations, namely 3*2, and both viruses representing patients 4, 8, 9, 10 and 11. Mutations with statistically significant effect on replicative fitness are marked by bold letters. Effect of mutations on replicative fitness was measured by Boutwell and his colleagues, as a control was used mutation M184V. This mutation is linked with resistance to nucleoside analog reverse transcriptase inhibitors and to reduction of fitness.

In table are also described changes in mutation, as they developed over time. Most of the samples did not change their mutation pattern, although crippling mutations can be reverted or compensated. Interestingly, sample 3*2 developed four new mutations and lost only one in comparison to 3*1. Accumulation of those mutations can account for lower fitness of 3*2 in comparison to 3*1, as it is depicted in Table 7. Interestingly, sample 4 is the only one which reduced absolute number of mutations in late sample (4*2) in comparison to early sample (4*1). More interestingly, both 4*1 and 4*2 viruses are maintaining mutations T242N and G248A, both linked with HLA-B*57. Additionally, virus derived from patient 4 is only one carrying those two escape mutations.

Moreover, we were looking for presence of known compensatory mutations accumulated in Gag as a response to reduced replicative fitness by escape mutations. We

detected mutation N252H in virus 4*1. This compensatory mutation was previously described to have positive effect on replicative fitness restoration if virus develops T242N escape mutation (Gijsbers *et al.* 2013). More surprisingly, compensatory mutation N252H is not present in virus 4*2. Further, we identified putative compensatory mutation V190I in patient's 9 derived samples (Liu *et al.* 2014); this mutation is present in majority of viruses with escape mutation T186M. Also, Ile present in position 147 is linked with escape mutation A146P and such amino acid we identified at viruses from patients 2 and 7. Patient's 8 virus, carrier of A146P has Leu present in position 147. This mutation is described in minority of A146P carriers (Liu *et al.* 2014). In summary, we observed changes of escape mutation pattern in HIV *gag* region between early and late time point. Interestingly, we did not observe accumulation of previously described compensatory mutations nor reversion of escape mutations during our monitoring period.

6. Discussion

Up to this day, thirty years since its discovery, HIV is still spread world-wide with 36 million people infected. Although incredible amount of research was conducted on HIV, we are still unable to eliminate the virus. Many drugs were discovered that can effectively inhibit HIV replication. However, virus exhibited quick development of resistance. Thus, interest was drawn to virus population behavior and development in particular environment and term virus fitness was coined (Domingo & Holland 1997). Consequently, interest was drawn to link HIV fitness, host immunologic, and genetic factors with their roles in disease progression.

Two basic approaches are used for replicative fitness evaluation; first approach measures replicative fitness of whole viruses isolated from patient's cells meanwhile second approach utilizes recombinant virus (Dykes & Demeter 2007). Although nowadays majority of studies is assessing fitness using recombinant viruses, interest of those studies is usually focused on drug resistant mutants (Iyidogan & Anderson 2014). Therefore, identification of whole *gag* contribution to overall fitness of nonresistant HIV may help to determine if recombinant virus can be used to measure HIV replicative fitness in untreated patients.
Additionally, information about intra-patient replicative fitness change over time can further improve personalized treatment of patients. Thereby, analysis of whole *gag* development in single patient and its contribution to replicative fitness change will help to evaluate replicative fitness as another marker of disease progression.

In theory, replicative fitness of virus is well represented by primary isolate, since virus contains all genes and can take their contribution to replicative fitness into account. However, primary isolates preparation through co-cultivation is laborious and can lead to alteration of viral quasispecies (Kusumi *et al.* 1992). Particularly, the maintenance of quasispecies is a vital aspect of replicative fitness determination in complex viral population. Therefore, recombinant viruses are frequently utilized for measuring of the replicative fitness in the last decade. General approach for production of recombinant virus has evolved over the years. For analysis of HIV replicative fitness, we adapted strategy of introduced reporter gene between ORFs of *env* and *nef* (Weber *et al.* 2006). Presence of reporter gene allowed us to measure amount of infected cells through flow cytometry or under the microscope. Importantly, assay developed by Weber and colleagues measures replicative fitness of recombinant virus without altering expression of HIV-1 genes. Further, our approach adapted homologous recombination step in yeast cells described by Dudley (Dudley *et al.* 2009) and successfully

used by Weber and colleagues (Weber *et al.* 2013) for replicative fitness measurement. My thesis is a part of project focused on role of HIV fitness in disease progression. The project evaluates contribution of all three main genes (*gag*, *pol*, and *env*) to replicative fitness of the virus. Specifically, *gag*-recombinant viruses were utilized in my work. Besides *gag* contribution to replicative fitness, we were interested in the correlation of disease progression and fitness development. For this reason, patients were arbitrarily divided in two different groups according their immunological status, represented by $CD4^+$ T cells count development (Weber *et al.* 2017). Moreover, great emphasis was given to analyze samples from only treatment naïve patients.

Although recombinant viruses are utilized in many studies, majority consider only single mutation introduced in NL4-3 scaffold (for example Boutwell *et al.* 2013). As we are measuring contribution of whole *gag* gene to overall fitness, we were amplifying mentioned region by PCR from RNA isolated from patient's plasma. Importantly, we were utilizing two pairs of primers in three-step PCR to secure amplification of sufficient amount of DNA of *gag* region. Moreover, all four primers are degenerate at 5' end to secure amplification of broad extent of quasispecies (Tsibris *et al.* 2009). Thanks to the effect of degenerative primers, we succeeded in amplification of more than 80% (n=105) of all samples. Additionally, samples with unsuccessful PCR were often with viral load under 10,000 copies per ml of blood. Presence of non-specific amplicons in final mixture is, in our opinion, due to degenerate nature of primers.

Regarding homologous recombination in yeast cells, we established success rate of procedure to be 81 %. Although subcloning using intrinsic restriction sites in amplified HIV sequence was used for recombinant virus generation by ligation previously (Weber *et al.* 2006, Claiborne *et al.* 2015), we adopted approach of recombination as intermediate step to maintain broad quasispecies. Furthermore, we adopted method of homologous recombination in yeast rather than approach through bacteria or mammalian cells, since those alternatives are less efficient (Dudley *et al.* 2009, Weber *et al.* 2011). During procedure, we encountered inability of few samples to excise *gag* fragment after MluI and PspXI digestion. We conclude that homologous recombination performed weakly. Since we are amplifying highly variable sequences of HIV, we assume that recombination could be negatively influenced by non-specific PCR products.

To produce transfection competent plasmid, we established subcloning of *gag* fragment to pNL4-3-EGFP scaffold. Final subcloning for purpose to produce replication competent plasmid yielded 80% success rate. We conclude that without complete removal of

phosphate on 5' end of the vector by dephosphorylation, we could not prevent religation of linearized vector. However, all isolated plasmids were replication competent as showed by our 100 % success rate of transfection. Importantly, isolated viruses exhibited $TCID_{50} > 1x10^4$ IU/ml in most of the cases. Strength of the virus is important for correct set-up of replicative fitness experiments, as a very low $TCID_{50}$ can influence virus performance even after recalculation to MOI. Certainly, we clearly demonstrate complexity of *gag* recombinant virus production with overall 52% success rate. Nonetheless, we were successful to generate pool of 55 *gag* recombinant viruses. Next, we selected 24 candidates for further analysis and sequenced their whole *gag* region. Sequencing of *gag* was done also for 10 samples from HIV RNA isolated directly from the patient's plasma. It is believed that isolated HIV RNA from plasma best represents the sequence variability of circulating primary isolates. To deal with sequence variability inside the *gag* gene, we adopted primer design recommended in HIV Sequence Compendium from Los Alamos National Laboratory (Sanders-Buell *et al.* 1995).

To determine contribution of *gag* gene to overall replicative fitness, we established one batch of samples consisting of early-derived *gag* recombinant virus from patient and corresponding late primary isolate from the same patient. Similar approach, although oriented on *env* and *pol* genes was done previously (Rangel *et al.* 2003, Weber *et al.* 2003). Specifically, sequence analysis of primary isolates and analogous recombinant viruses showed us more than 98% identity between primary isolate and recombinant virus through all ten pairs. However, we concluded that most of the changes were in ambiguous positions. Thus, we can assume that *gag* sequence of circulating viruses is almost identical with sequence of recombinant viruses.

Since Gag polyprotein is a strong immunogenic target of CTL response (Troyer *et al.* 2009), we wondered if we would see particular influence of *gag* gene to overall fitness of the virus. We have not found significant relationship between replicative fitness of primary isolates and replicative fitness of analogous *gag* recombinant virus. However, since we are taking into account only one of three main genes; we were not expecting to see significant contribution of *gag* on overall fitness in all samples. Indeed, we detected 2 samples with corresponding replicative fitness measured on primary isolate and on recombinant virus. We hypothesize that overall fitness of samples 6 and 12 is determined by the fitness input from *gag* gene. Other study oriented on *env* gene (Rangel *et al.* 2003), showed significant relationship between primary isolate fitness and *env* recombinant virus fitness in all samples (n=7). However, they included viruses with resistance mutations in the *env* region that exerted strong impact on overall fitness. Furthermore, study performed on *pol* recombinant viruses

concluded that replicative capacity of *pol* recombinant viruses can be used to estimate overall fitness. However, study limits this conclusion to patients treated by protease or reverse transcriptase inhibitors (Weber *et al*. 2003). In other words, if *pol* accumulates resistance mutations, it becomes major contributor to replicative fitness (or replicative capacity measured by Weber and colleagues). Likewise, my conclusion based on fitness comparison of samples 6 and 12 is, that *gag* recombinant virus can be considered to estimate overall replicative fitness only in case, when the *gag* gene has a major impact on fitness. The *gag* gene contribution does not have to be necessarily linked with drug resistance to strongly influence overall fitness. Instead CTL escape mutations in the *gag* region can be such major fitness contributors (Shahid *et al.* 2015, Sunshine *et al.* 2015, Brockman *et al.* 2010) and therefore, we can observe major *gag* contribution to overall fitness even in treatment naïve patients.

HIV infection is developing over time and adapting itself to the patient (Tebit *et al.* 2007). For this reason we were interested to see how HIV replicative fitness, based on *gag* recombinant virus, changes together with clinical characteristics over time. According to clinical characteristics of patients, $CD4^+$ T cell counts development is widely used predictor of disease progression (Fahey *et al.* 1990). Our patients were divided into two groups based on their slopes of $CD4^+$ T cells. Arbitrary groups were: slow progressors with slope > -1 and typical progressors with slopes < -1 cell/month/µl (Weber *et al.* 2017). Whereas progression is typically established over many years, our on average 3 year follow-up does not have to necessarily represent true status of the patient. Specifically, in second batch of our samples, established for purpose to measure change of replicative fitness over time, we designated patients 2, 3 and 10 as slow progressors. Moreover, patients 2 and 3 are carriers of protective alleles HLA-B*13 and HLA-B*27, respectively. Presence of protective alleles can explain stable slope of $CD4^+$ T cells as well as declining replicative fitness over time (Shahid *et al.* 2015, Brettle *et al.* 1996). Calculation of replicative fitness change between early- and late-sampling derived recombinant viruses was done from viral growth kinetics. Fitness change is represented as slope between replicative fitness values measured on earlier and later sample, respectively. Correlation of fitness change and slope of $CD4^+$ T cells of all ten patients displayed significant negative correlation ($p < 0.05$). More importantly, slow progressors, patients 2 and 3 and typical progressor patient 8, carrier of protective allele HLA-B*13, displayed the largest drop in HIV replicative fitness. We conclude that presence of protective allele seems to have negative effect on replicative fitness over time. In other words, we observe virus attenuation in carriers of protective alleles. Conclusion corroborates previous

studies about protective alleles (Boutwell *et al.* 2013, Gijsbers *et al.* 2013, Shahid *et al.* 2015). Moreover, we observed statistically significant association between patient clinical status and replicative fitness change over time, thus suggesting that *gag* recombinant viruses can be used for fitness change characterization.

In order to characterize association of replicative fitness development with genetic polymorphisms and sequence diversity, we sequenced whole *gag* region in all consecutive HIV samples. We could observe sequence divergence between early and late sample to be in the range of 0.2 % to 2.2 %. Additionally, we translated the sequences in Gag polyprotein *in silico*. Primarily, we were looking for *gag* mutations associated with fitness decline in literature (Boutwel *et al.* 2013, Liu *et al.* 2014). As we observed broad spectrum of escape mutations among our samples, we decided to analyze their impact on replicative fitness. Interestingly, comparing early and late viruses derived from patient 3, we noticed four new CTL escape mutations introduced in later sample. Since patient 3 is carrier of protective HLA-B*27, we assume that early sampling was obtained relatively briefly after acute infection and new CTL escape mutations were further selected. We also observed drop in replicative fitness of the later sampling-derived virus and we believe, that further accumulation of escape mutations lead to reduction of replicative fitness. Highly dynamic CTL escape was previously observed (Sunshine *et. al.* 2015). Noticeably, several samples preserved CTL linked mutations over our time scale. Namely samples 4, 8, 10 and 11 were carrying at least five CTL escape mutations in early as well as in late virus. Although patients 4 and 8 are carriers of protective alleles, patients 10 and 11 are not. Therefore, we assume that persistence of mutations does not have to be necessarily linked with protective allele. Furthermore, except virus in patient 8, replicative fitness of those viruses displayed practically no change. Moreover, Sunshine (2015) noted that reverse mutations occurs rarely and virus rather develops compensatory mutations. Unfortunately compensatory mutations are not characterized as well as escape mutations. Previously, analysis was done on escape mutations specifically linked with CTL response of HLA-B*57 carrying patient (Brockman *et al.* 2007, Gijsbers *et al.* 2013). Namely, mutations T242N and G248A are both occurring in patient 4 thus we were looking for known compensatory mutations in *gag*. Although we have not detected any compensatory mutation described by Brockman (2007), we have found compensatory mutation N252H (Gijsbers *et al.* 2013). Surprisingly, compensatory mutation occured only at early-sampling derived virus and is absent in late-sampling derived virus. Furthermore, we did not observe notable change in replicative fitness. We conclude that N252H loss had no critical effect on improvement of replicative fitness in this particular virus

genetic background. From other compensatory mutations described in literature, we have found V190I mutation, linked with escape mutation T186M. Escape mutation T186M is present in viruses derived from patient 9. Additionally, Ile presented in position 147 is compensating escape mutation A146P (Liu *et al.* 2014). All viruses carrying A146P had compensating Ile in position 147 except viruses from patient 8. Since viruses from patient 8 are carrying another escape mutation I147L in this position, we conclude this is due to presence of protective allele HLA-B*13 in patient 8. More importantly, I147L is linked in escape with HLA-B*13 (Boutwell *et al.* 2013). Viruses derived from patient 5 display emergence of K26R. Mutation was shown to have mild negative effect on replicative fitness (Boutwell *et al.* 2013). However, we observed the largest increase of replicative fitness over time in this particular virus. We assume that unknown compensatory mutation(s) is/are responsible for this effect. In summary, virus in patients with protective alleles evolves by accumulating of CTL escape mutations however; this rule is not general since virus from patient 2 is present with only one known CTL escape mutation. We also showed that emergence of multiple escape mutations can have negative effect on replicative fitness as i tis displayed in viruses from patient 3, on the other hand, viruses from patient 5 with emerging mutation K26R showed the opposite effect. Viruses without change in pattern of escape mutations display stable replicative fitness over time in most of the cases. However, further research is required to clarify emergence of compensatory mutations and their effect on HIV fitness of *gag* recombinant viruses. Additionally, even though effect of compensatory mutations in *gag* on fitness was presented in literature, we do not observe this trend in our pool of samples probably because of short time span between sampling. However, only few compensatory mutations are well described in literature and are linked with specific escape mutations (Brockman *et al.* 2010, Gijsbers *et al.* 2013, Sunshine *et al.* 2015). Thus we could identify only limited number of them. Furthermore, we assume that detailed sequence analysis through deep sequencing, planned in the future for this project, could enlighten pattern of evolved mutations in whole quasispecies and indicate new putative compensatory mutations.

As demonstrated here *gag* region can be major contributor to overall HIV fitness. Further research about role of HIV replicative fitness in disease progression would benefit from similar analysis about *pol* and *env* contribution to overall fitness. In detail, unveiling if *pol* and *env* can influence overall fitness more than *gag* would outline future path for recombinant virus application in replicative fitness measurement. Furthermore, deep sequencing analysis can provide important details about behavior of quasispecies over time.

Although longitudinal study of HIV sequence development was done (Zanini *et al.* 2015) it was not correlated with replicative fitness changes.

In conclusion, by comparison of replicative fitness using viral growth kinetics and competition experiments on *gag* recombinant viruses and primary isolates, we proved *gag* can be major contributor to overall fitness, but not exclusively. Furthermore, we shoved that development of fitness, measured on *gag* recombinant viruses is correlating with patient's slope of $CD4^+$ T cells. More importantly, we recognized that patients carrying protective alleles and/or slow progressors displayed attenuation of replicative fitness over time. Moreover, we detected many CTL escape mutations, but rather small number of reversion or appearance of CTL escape mutations during our monitoring period. We believe that our findings are paving the way to application of recombinant viruses to measure replicative fitness for clinical purposes.

7. Conclusions

In this chapter is provided summary of results corresponding to the aims of my thesis.

1. **Preparation of *gag* recombinant virus and primary HIV isolates from untreated patients and comparison of replicative fitness values.**

Production of *gag* recombinant viruses was established. Pool of recombinant viruses was generated (n=55) as well as isolation of primary isolates (n=10), all derived from treatment naïve patients. Replicative fitness of ten *gag* recombinant viruses was measured likewise replicative fitness of ten corresponding primary isolates. Comparison of results showed that *gag* recombinant virus can estimate overall replicative fitness in cases when *gag* is major contributor to overall fitness.

2. **Characterization of replicative fitness development in disease progression of typical progressors and slow progressors.**

From pool of produced recombinant viruses (n=55), ten pairs were selected. Pairs were consisting of early-sampling derived HIV and late-sampling derived HIV with at least two years span between samplings. We observed positive and negative change in replicative fitness as well as no change at all during our monitoring period. Furthermore, we observed that gradual increasement of replicative fitness leads to decrease of $CD4^+$ T cells in the patient. We showed that not only slow progressors but also carriers of protective alleles can attenuate virus replicative fitness. Additionally, we detected escape mutations previously linked with replicative fitness attenuation and observed a decline of replicative fitness after accumulation of new escape mutations.

8. References

Abecasis, A. B., Wensing, A. M. J., Paraskevis, D., Vercauteren, J., Theys, K., de Vijver, D., Albert, J., Asjo, B., Balotta, C., Beshkov, D., *et al*. 2013. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology,* 10(1), 7

Agrawal, L., Lu, X. H., Qingwen, J., VanHorn-Ali, Z., Nicolescu, I. V., McDermott, D. H., Murphy, P. M. & Alkhatib, G. 2004. Role for CCR5 Delta 32 protein in resistance to R5, R5X4, and X4 human immunodeficiency virus type 1 in primary CD4(+) cells. *Journal of Virology,* 78(5)**,** 2277-2287.

Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E. & Trono, D. 1994. Nef induces CD4 endocytosis - requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell,* 76(5)**,** 853-864.

Arien, K. K., Troyer, R. A., Gali, Y., Colebunders, R. L., Arts, E. J. & Vanham, G. 2005. Replicative fitness of historical and recent HIV-1 isolates suggests HIV-1 attenuation over time. *Aids,* 19(15)**,** 1555-1564.

Arts, E. J., Ghosh, M., Jacques, P. S., Ehresmann, B. & LeGrice, S. F. J. 1996. Restoration of tRNA(3)(Lys)-primed (-)-strand DNA synthesis to an HIV-1 reverse transcriptase mutant with extended tRNAs - Implications for retroviral replication. *Journal of Biological Chemistry,* 271(15)**,** 9054-9061.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axlerblin, C., Vezinetbrun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune-deficiency syndrome (AIDS). *Science,* 220(4599)**,** 868-871.

Bell, N. M. & Lever, A. M. L. 2013. HIV Gag polyprotein: processing and early viral particle assembly. *Trends in Microbiology,* 21(3)**,** 136-144.

Berger, E. A., Murphy, P. M. & Farber, J. M. 1999. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. *Annual Review of Immunology,* 17(1), 657-700.

Blankson, J. N., Bailey, J. R., Thayil, S., Yang, H. C., Lassen, K., Lai, J., Gandhi, S. K., Siliciano, J. D., Williams, T. M. & Siliciano, R. F. 2007. Isolation and characterization of replication-competent human immunodeficiency virus type 1 from a subset of elite suppressors. *Journal of Virology,* 81(5)**,** 2508-2518.

Boutwell, C. L., Carlson, J. M., Lin, T. H., Seese, A., Power, K. A., Peng, J., Tang, Y. H., Brumme, Z. L., Heckerman, D., Schneidewind, A. & Allen, T. M. 2013. Frequent and Variable Cytotoxic-T-Lymphocyte Escape-Associated Fitness Costs in the Human Immunodeficiency Virus Type 1 Subtype B Gag Proteins. *Journal of Virology,* 87(7)**,** 3952-3965.

Brettle, R. P., McNeil, A. J., Burns, S., Gore, S. M., Bird, A. G., Yap, P. L., MacCallum, L.,

Leen, C. S. L. & Richardson, A. M. 1996. Progression of HIV: Follow-up of Edinburgh injecting drug users with narrow seroconversion intervals in 1983-1985. *Aids,* 10(4)**,** 419-430.

Briggs, J. A. G., Wilk, T., Welker, R., Krausslich, H. G. & Fuller, S. D. 2003. Structural organization of authentic, mature HIV-1 virions and cores. *Embo Journal,* 22(7)**,** 1707-1715.

Brockman, M. A., Brumme, Z. L., Brumme, C. J., Miura, T., Sela, J., Rosato, P. C., Kadie, C. M., Carlson, J. M., Markle, T. J., Streeck, H., *et al.* 2010. Early Selection in Gag by Protective HLA Alleles Contributes to Reduced HIV-1 Replication Capacity That May Be Largely Compensated for in Chronic Infection. *Journal of Virology,* 84(22)**,** 11937-11949.

Brockman, M. A., Schneidewind, A., Lahaie, M., Schmidt, A., Miura, T., DeSouza, I., Ryvkin, F., Derdeyn, C. A., Allen, S., Hunter, E., *et al*. 2007. Escape and compensation from early HLA-B57-Mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. *Journal of Virology,* 81(22)**,** 12608-12618.

Brown, P. H., Tiley, L. S. & Cullen, B. R. 1991. Efficient polyadenylation within the human-immunodeficiency-virus type-1 long terminal repeat requires flanking U3-specific sequences. *Journal of Virology,* 65(6)**,** 3340-3343.

Brumme, C. J., Huber, K. D., Dong, W., Poon, A. F. Y., Harrigan, P. R. & Sluis-Cremer, N. 2013. Replication Fitness of Multiple Nonnucleoside Reverse Transcriptase-Resistant HIV-1 Variants in the Presence of Etravirine Measured by 454 Deep Sequencing. *Journal of Virology,* 87(15)**,** 8805-8807.

Brumme, Z. L., Brumme, C. J., Carlson, J., Streeck, H., John, M., Eichbaum, Q., Block, B. L., Baker, B., Kadie, C., Markowitz, M., *et al*. 2008. Marked epitope- and allele- specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *Journal of Virology,* 82(18)**,** 9216-9227.

Brumme, Z. L., Li, C., Miura, T. Y., Sela, J., Rosato, P. C., Brumme, C. J., Markle, T. J., Martin, E., Block, B. L., Trocha, A., *et al*. 2011. Reduced Replication Capacity of NL4-3 Recombinant Viruses Encoding Reverse Transcriptase-Integrase Sequences From HIV-1 Elite Controllers. *Jaids-Journal of Acquired Immune Deficiency Syndromes,* 56(2)**,** 100-108.

Carlson, J. M., Schaefer, M., Monaco, D. C., Batorsky, R., Claiborne, D. T., Prince, J., Deymier, M. J., Ende, Z. S., Klatt, N. R., DeZiel, *et al*. 2014. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science*, *345*(6193), 178

CDC. 1981. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men—New York City and California. *Morbidity and mortality weekly report*, 30(25), 305-8

CDC. 1982. Current Trends Update on Acquired Immune Deficiency Syndrome (AIDS) - United States. *Morbidity and mortality weekly report*, 31(37), 507.

Chen, J. B., Nikolaitchik, O., Singh, J., Wright, A., Bencsics, C. E., Coffin, J. M., Ni, N., Lockett, S., Pathak, V. K. & Hu, W. S. 2009. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proceedings of the National Academy of Sciences of the United States of America,* 106(32)**,** 13535-13540.

Claiborne, D. T., Prince, J. L., Scully, E., Macharia, G., Micci, L., Lawson, B., Kopycinski, J., Deymier, M. J., Vanderford, T. H., Nganou-Makamdop, K., *et al*. 2015. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4(+) T cells, and disease progression. *Proceedings of the National Academy of Sciences of the United States of America,* 112(12)**,** 1480-1489.

Clapham, P. R. & McKnight, A. 2001. HIV-1 receptors and cell tropism. *British Medical Bulletin,* 58(1), 43-59.

Clavel, F., Guetard, D., Brunvezinet, F., Chamaret, S., Rey, M. A., Santosferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L. & Montagnier, L. 1986. Isolation of a new human retrovirus from west-african patients with aids. *Science,* 233(4761)**,** 343-346.

Cohen, E. A., Terwilliger, E. F., Sodroski, J. G. & Haseltine, W. A. 1988. Identification of a protein encoded by the vpu gene of HIV-1. *Nature,* 334(6182)**,** 532-534.

Covens, K., Dekeersmaeker, N., Schrooten, Y., Weber, J., Schols, D., Quinones-Mateu, M. E., Vandamme, A. M. & Van Laethem, K. 2009. Novel Recombinant Virus Assay for Measuring Susceptibility of Human Immunodeficiency Virus Type 1 Group M Subtypes To Clinically Approved Drugs. *Journal of Clinical Microbiology,* 47(7)**,** 2232-2242.

D'Souza, V. & Summers, M. F. 2005. How retroviruses select their genomes. *Nature Reviews Microbiology,* 3(8)**,** 643-655.

Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. 1984. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature,* 312(5996)**,** 763-767.

Dean, M., Carrington, M., Winkler, C., Huttley, G. A., Smith, M. W., Allikmets, R., Goedert, J. J., Buchbinder, S. P., Vittinghoff, E., Gomperts, E, *et al*. 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science,* 273(5283)**,** 1856-1862.

Delaugerre, C., De Oliveira, F., Lascoux-Combe, C., Plantier, J. C. & Simon, F. 2011. HIV-1 group N: travelling beyond Cameroon. *Lancet,* 378(9806)**,** 1894-1894.

Dinman, J. D., Icho, T. & Wickner, R. B. 1991. A -1 ribosomal frameshift in a double-stranded-RNA virus of yeast forms a gag pol fusion protein. *Proceedings of the National Academy of Sciences of the United States of America,* 88(1)**,** 174-178.

Dinman, J. D. & Wickner, R. B. 1992. Ribosomal frameshifting efficiency and gag gag-pol

ratio are critical for yeast M(1) double-stranded-RNA virus propagation. *Journal of Virology,* 66(6)**,** 3669-3676.

Domingo, E. & Holland, J. J. 1997. RNA virus mutations and fitness for survival. *Annual Review of Microbiology,* 51(1), 151-178.

Drescher, S. M., von Wyl, V., Yang, W. L., Boni, J., Yerly, S., Shah, C., Aubert, V., Klimkait, T., Taffe, P., Furrer, H., *et al.* 2014. Treatment-Naive Individuals Are the Major Source of Transmitted HIV-1 Drug Resistance in Men Who Have Sex With Men in the Swiss HIV Cohort Study. *Clinical Infectious Diseases,* 58(2)**,** 285-294.

Dudley, D. M., Gao, Y., Nelson, K. N., Henry, K. R., Nankya, I., Gibson, R. M. & Arts, E. J. 2009. A novel yeast-based recombination method to clone and propagate diverse HIV-1 isolates. *Biotechniques,* 46(6)**,** 458-467.

Dunfee, R. L., Thomas, E. R., Gorry, P. R., Wang, J. B., Taylor, J., Kunstman, K., Wolinsky, S. M. & Gabuzda, D. 2006. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. *Proceedings of the National Academy of Sciences of the United States of America,* 103(41)**,** 15160-15165.

Dykes, C. & Demeter, L. M. 2007. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clinical Microbiology Reviews,* 20(4)**,** 550-578.

El-Far, M., Isabelle, C., Chomont, N., Bourbonniere, M., Fonseca, S., Ancuta, P., Peretz, Y., Chouikh, Y., Halwani, R., Schwartz, O., *et al*. 2013. Down-Regulation of CTLA-4 by HIV-1 Nef Protein. *Plos One,* 8(1)**,** e54295.

Fahey, J. L., Taylor, J. M. G., Detels, R., Hofmann, B., Melmed, R., Nishanian, P. & Giorgi, J. V. 1990. The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type-1. *New England Journal of Medicine,* 322(3)**,** 166-172.

Finak, G., Frelinger, J., Jiang, W. X., Newell, E. W., Ramey, J., Davis, M. M., Kalams, S. A., De Rosa, S. C. & Gottardo, R. 2014. OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis. *Plos Computational Biology,* 10(8)**,** e1003806.

Foley B., Leitner T., Apetrei C., Hahn B., Mizrachi I., Mullins J., Rambaut A., Wolinsky S. and Korber B., (eds.). 2016. HIV Sequence Compendium 2016. *Theoretical Biology and Biophysics Group*, Los Alamos, NM.

Francki, R. I. B., (ed). 1991. Virus taxonomy: classification and nomenclature of viruses: Fifth Report of the International Committee on Taxonomy of Viruses. *Virology Division of the International Union of Microbiological Societies*. *Springer Science & Business Media*, New York, NY.

Freed, E. O. 2015. HIV-1 assembly, release and maturation. *Nature Reviews Microbiology,* 13(8)**,** 484-496.

Gandhi, S. K., Siliciano, J. D., Bailey, J. R., Siliciano, R. F. & Blankson, J. N. 2008. Role of

APOBEC3G/F-mediated hypermutation in the control of human immunodeficiency virus type 1 in elite suppressors. *Journal of Virology,* 82(6)**,** 3125-3130.

Ganser, B. K., Li, S., Klishko, V. Y., Finch, J. T. & Sundquist, W. I. 1999. Assembly and analysis of conical models for the HIV-1 core. *Science,* 283(5398)**,** 80-83.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y. L., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M. & Hahn, B. H. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature,* 397(6718)**,** 436-441.

Gijsbers, E. F., Feenstra, K. A., van Nuenen, A. C., Navis, M., Heringa, J., Schuitemaker, H. & Kootstra, N. A. 2013. HIV-1 Replication Fitness of HLA-B*57/58:01 CTL Escape Variants Is Restored by the Accumulation of Compensatory Mutations in Gag. *Plos One,* 8(12)**,** e81235.

Grabar, S., Selinger-Leneman, H., Abgrall, S., Pialoux, G., Weiss, L. & Costagliola, D. 2009. Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV. *Aids,* 23(9)**,** 1163-1169.

Granelli-Piperno, A., Delgado, E., Finkel, V., Paxton, W. & Steinman, R. M. 1998. Immature dendritic cells selectively replicate macrophagetropic (M-tropic) human immunodeficiency virus type 1, while mature cells efficiently transmit both M- and T-Tropic virus to T cells. *Journal of Virology,* 72(4)**,** 2733-2737.

Guerrero, S., Batisse, J., Libre, C., Bernacchi, S., Marquet, R. & Paillart, J. C. 2015. HIV-1 Replication and the Cellular Eukaryotic Translation Apparatus. *Viruses-Basel,* 7(1)**,** 199-218.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. 1987. Genome organization and transactivation of the human-immunodeficiency-virus type-2. *Nature,* 326(6114)**,** 662-669.

Heinzinger, N. K., Bukrinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M. A., Gendelman, H. E., Ratner, L., Stevenson, M. & Emerman, M. 1994. The vpr protein of human-immunodeficiency-virus type-1 influences nuclear-localization of viral nucleic-acids in nondividing host-cells. *Proceedings of the National Academy of Sciences of the United States of America,* 91(15)**,** 7311-7315.

Hemelaar, J., Gouws, E., Ghys, P. D. & Osmanov, S. 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *Aids,* 25(5)**,** 679-689.

Hu, W. S. & Hughes, S. H. 2012. HIV-1 Reverse Transcription. *Cold Spring Harbor Perspectives in Medicine,* 2(10)**,** a006882.

Huang, Y., Mak, J., Cao, Q., Li, Z., Wainberg, M. A. & Kleiman, L. 1994. Incorporation of excess wild-type and mutant tRNA(3)(lys) into human-immunodeficiency-virus type-1. *Journal of Virology,* 68(12)**,** 7676-7683.

Iyidogan, P. & Anderson, K. S. 2014. Current Perspectives on HIV-1 Antiretroviral Drug

Resistance. *Viruses-Basel,* 6(10)**,** 4095-4139.

Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene-product. *Nature,* 330(6147)**,** 489-493.

Karageorgos, L., Li, P. & Burrell, C. 1993. Characterization of HIV replication complexes early after cell-to-cell infection. *Aids Research and Human Retroviruses,* 9(9)**,** 817-823.

Klasse, P. J. 2012. The molecular basis of HIV entry. *Cellular Microbiology,* 14(8)**,** 1183-1192.

Koot, M., Vos, A. H. V., Keet, R. P. M., Degoede, R. E. Y., Dercksen, M. W., Terpstra, F. G., Coutinho, R. A., Miedema, F. & Tersmette, M. 1992. HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. *Aids,* 6(1)**,** 49-54.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science,* 288(5472)**,** 1789-1796.

Kozisek, M., Henke, S., Saskova, K. G., Jacobs, G. B., Schuch, A., Buchholz, B., Muller, V., Krausslich, H. G., Rezacova, P., Konvalinka, J., *et al*. 2012. Mutations in HIV-1 gag and pol Compensate for the Loss of Viral Fitness Caused by a Highly Mutated Protease. *Antimicrobial Agents and Chemotherapy,* 56(8)**,** 4320-4330.

Krishnan, L. & Engelman, A. 2012. Retroviral Integrase Proteins and HIV-1 DNA Integration. *Journal of Biological Chemistry,* 287(49)**,** 40858-40866.

Kusumi, K., Conway, B., Cunningham, S., Berson, A., Evans, C., Iversen, A. K. N., Colvin, D., Gallo, M. V., Coutre, S., Shpaer, E. G., *et al*. 1992. Human-immunodeficiency-virus type-1 envelope gene structure and diversity invivo and after cocultivation invitro. *Journal of Virology,* 66(2)**,** 875-885.

Lauring, A. S. & Andino, R. 2010. Quasispecies Theory and the Behavior of RNA Viruses. *Plos Pathogens,* 6(7)**,** e1001005.

Levin, J. G., Mitra, M., Mascarenhas, A. & Musier-Forsyth, K. 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *Rna Biology,* 7(6)**,** 754-774.

Li, Y., Luo, L. Z., Rasool, N. & Kang, C. Y. 1993. Glycosylation is necessary for the correct folding of human immunodeficiency virus-gp120 in CD4 binding. *Journal of Virology,* 67(1)**,** 584-588.

Liu, D. L., Zuo, T., Hora, B., Song, H. S., Kong, W., Yu, X. H., Goonetilleke, N., Bhattacharya, T., Perelson, A. S., Haynes, B. F., *et al*. 2014. Preexisting compensatory amino acids compromise fitness costs of a HIV-1 T cell escape mutation. *Retrovirology,* 11(1), 101.

Lobritz, M. A., Lassen, K. G. & Arts, E. J. 2011. HIV-1 replicative fitness in elite controllers. *Current Opinion in Hiv and Aids,* 6(3)**,** 214-220.

Madec, Y., Boqfassa, F., Avettand-Fenoel, V., Hendou, S., Melard, A., Boucherit, S., Surzyn, J., Meyer, L., Rouzioux, C. & Grp, A. S. H. S. 2009. Early Control of HIV-1 Infection in Long-Term Nonprogressors Followed Since Diagnosis in the ANRS SEROCO/HEMOCO Cohort. *Jaids-Journal of Acquired Immune Deficiency Syndromes,* 50(1)**,** 19-26.

Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral messenger-RNA. *Nature,* 338(6212)**,** 254-257.

Maree, A. F. M., Keulen, W., Boucher, C. A. B. & De Boer, R. J. 2000. Estimating relative fitness in viral competition experiments. *Journal of Virology,* 74(23)**,** 11067-11072.

McCune, J. M., Rabin, L. B., Feinberg, M. B., Lieberman, M., Kosek, J. C., Reyes, G. R. & Weissman, I. L. 1988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell,* 53(1)**,** 55-67.

Merk, A. & Subramaniam, S. 2013. HIV-1 envelope glycoprotein structure. *Current Opinion in Structural Biology,* 23(2)**,** 268-276.

Migueles, S. A., Sabbaghian, M. S., Shupert, W. L., Bettinotti, M. P., Marincola, F. M., Martino, L., Hallahan, C. W., Selig, S. M., Schwartz, D., Sullivan, J., *et al*. 2000. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the National Academy of Sciences of the United States of America,* 97(6)**,** 2709-2714.

Nekhai, S., Shukla, R. R. & Kumar, A. 1997. A human primary T-lymphocyte-derived human immunodeficiency virus type 1 Tat-associated kinase phosphorylates the C-terminal domain of RNA polymerase II and induces CAK activity. *Journal of Virology,* 71(10)**,** 7436-7441.

Nonnemacher, M. R., Pirrone, V., Feng, R., Moldover, B., Passic, S., Aiamkitsumrit, B., Dampier, W., Wojno, A., Kilareski, E., Blakey, B., *et al*. 2016. HIV-1 Promoter Single Nucleotide Polymorphisms Are Associated with Clinical Disease Severity. *Plos One,* 11(4)**,** e0150835.

Ochsenbauer, C., Edmonds, T. G., Ding, H. T., Keele, B. F., Decker, J., Salazar, M. G., Salazar-Gonzalez, J. F., Shattock, R., Haynes, B. F., Shaw, G. M., *et al*. 2012. Generation of Transmitted/Founder HIV-1 Infectious Molecular Clones and Characterization of Their Replication Capacity in CD4 T Lymphocytes and Monocyte-Derived Macrophages. *Journal of Virology,* 86(5)**,** 2715-2728.

Ott, D. E. 2008. Cellular proteins detected in HIV-1. *Reviews in Medical Virology,* 18(3)**,** 159-175.

Pantaleo, G. & Fauci, A. S. 1996. Immunopathogenesis of HIV infection. *Annual Review of Microbiology,* 50(1), 825-854.

Pertel, T., Hausmann, S., Morger, D., Zuger, S., Guerra, J., Lascano, J., Reinhard, C., Santoni, F. A., Uchil, P. D., Chatel, L., *et al*. 2011. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature,* 472(7343)**,** 361-365.

Pettit, S. C., Moody, M. D., Wehbie, R. S., Kaplan, A. H., Nantermet, P. V., Klein, C. A. & Swanstrom, R. 1994. The p2 domain of human-immunodeficiency-virus type-1 gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *Journal of Virology,* 68(12)**,** 8017-8027.

Plantier, J. C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemee, V., Damond, F., Robertson, D. L. & Simon, F. 2009. A new human immunodeficiency virus derived from gorillas. *Nature Medicine,* 15(8)**,** 871-872.

Pornillos, O., Ganser-Pornillos, B. K. & Yeager, M. 2011. Atomic-level modelling of the HIV capsid. *Nature,* 469(7330)**,** 424-427.

Prince, J. L., Claiborne, D. T., Carlson, J. M., Schaefer, M., Yu, T. W., Lahki, S., Prentice, H. A., Yue, L., Vishwanathan, S. A., Kilembe, W., *et al*. 2012. Role of Transmitted Gag CTL Polymorphisms in Defining Replicative Capacity and Early HIV-1 Pathogenesis. *Plos Pathogens,* 8(11)**,** e1003041.

Quiñones-Mateu, Miguel E., and Eric J. Arts. 2001. HIV-1 fitness: implications for drug resistance, disease progression, and global epidemic evolution. *In*: Kuiken C., Foley B., Hahn B., Marx P., McCutchan F., Mellors J., Wolinsky S., and Korber B. (eds.): HIV sequence compendium 2001. *Theoretical Biology and Biophysics Group*, Los Alamos, NM, 134-170.

Rangel, H. R., Weber, J., Chakraborty, B., Gutierrez, A., Marotta, M. L., Mirza, M., Kiser, P., Martinez, M. A., Este, J. A. & Quinones-Mateu, M. E. 2003. Role of the human immunodeficiency virus type 1 envelope gene in viral fitness. *Journal of Virology,* 77(16)**,** 9069-9073.

Reed, L. J. & Muench, H. 1938. A simple method of estimating fifty per cent endpoints. *Am J Epidemiol,* 27(3)**,** 493-497.

Rey, F., BouHamdan, M., Navarro, J. M., Agostini, I., Willetts, K., Bouyac, M., Tamalet, C., Spire, B., Vigne, R. & Sire, J. 1998. A role for human immunodeficiency virus type 1 Vpr during infection of peripheral blood mononuclear cells. *Journal of General Virology,* 79(5), 1083-1087.

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., *et al*. 2000. HIV-1 nomenclature proposal. *Science,* 288(5463)**,** 55-57.

Rolland, M., Brander, C., Nickle, D. C., Herbeck, J. T., Gottlieb, G. S., Campbell, M. S., Maust, B. S. & Mullins, J. I. 2007. HIV-1 over time: fitness loss or robustness gain? *Nature Reviews Microbiology,* 5(9)**,** 2006-2007.

Sanders-Buell, E., Salminen, M.O., McCutchan, E. 1995. Sequencing primers for HIV-1. *In:*

Myers G., Korber B., Hahn B., Jeang K.-T., Mellors J. W., McCutchan F. E., Henderson L. E. and Pavlakis G. N. (eds.): Human Retroviruses and AIDS 1995. *Theoretical Biology and Biophysics Group*, Los Alamos, NM, III15-21.

Schwartz, S., Felber, B. K., Fenyo, E. M. & Pavlakis, G. N. 1990. Env and vpu proteins of human-immunodeficiency-virus type-1 are produced from multiple bicistronic messenger-RNAs. *Journal of Virology,* 64(11)**,** 5448-5456.

Shahid, A., Olvera, A., Anmole, G., Kuang, X. T., Cotton, L. A., Plana, M., Brander, C., Brockman, M. A. & Brumme, Z. L. 2015. Consequences of HLA-B*13-Associated Escape Mutations on HIV-1 Replication and Nef Function. *Journal of Virology,* 89(22)**,** 11557-11571.

Shankarappa, R., Margolick, J. B., Gange, S. J., Rodrigo, A. G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C. R., Learn, G. H., He, X, *et al.* 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology,* 73(12)**,** 10489-10502.

Sharon, M., Kessler, N., Levy, R., Zolla-Pazner, S., Gorlach, M. & Anglister, J. 2003. Alternative conformations of HIV-1V3 loops mimic beta hairpins in chemokines, suggesting a mechanism for coreceptor selectivity. *Structure,* 11(2)**,** 225-236.

Sharp, P. M. & Hahn, B. H. 2011. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine,* 1(1)**,** a006841.

Simon, F., Mauclere, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M. C., Saragosti, S., Georges-Courbot, M. C., Barre-Sinoussi, F. & Brun-Vezinet, F. 1998. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nature Medicine,* 4(9)**,** 1032-1037.

Steinhauer, D. A. & Holland, J. J. 1987. Rapid evolution of RNA viruses. *Annual Review of Microbiology,* 41(1)**,** 409-433.

Streeck, H., Lichterfeld, M., Alter, G., Meier, A., Teigen, N., Yassine-Diab, B., Sidhu, H. K., Little, S., Kelleher, A., Routy, J. P., *et al*. 2007. Recognition of a defined region within p24 Gag by CD8(+) T cells during primary human immunodeficiency virus type I infection in individuals expressing protective HLA class I alleles. *Journal of Virology,* 81(14)**,** 7725-7731.

Sundquist, W. I. & Krausslich, H. G. 2012. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harbor Perspectives in Medicine,* 2(7)**,** a006924.

Sunshine, J. E., Larsen, B. B., Maust, B., Casey, E., Deng, W. J., Chen, L., Westfall, D. H., Kim, M., Zhao, H., Ghorai, S., *et al*. 2015. Fitness-Balanced Escape Determines Resolution of Dynamic Founder Virus Escape Processes in HIV-1 Infection. *Journal of Virology,* 89(20)**,** 10303-10318.

Sylwester, A., Murphy, S., Shutt, D. & Soll, D. R. 1997. HIV-induced T cell syncytia are self-perpetuating and the primary cause of T cell death in culture. *Journal of Immunology,* 158(8)**,** 3996-4007.

Tebit, D. M., Nankya, I., Arts, E. J. & Gao, Y. 2007. HIV diversity, recombination and disease progression: How does fitness "Fit" into the puzzle? *Aids Reviews,* 9(2)**,** 75-87.

Troyer, R. M., McNevin, J., Liu, Y., Zhang, S. C., Krizan, R. W., Abraha, A., Tebit, D. M., Zhao, H., Avila, S., Lobritz, M. A., *et al*. 2009. Variable Fitness Impact of HIV-1 Escape Mutations to Cytotoxic T Lymphocyte (CTL) Response. *Plos Pathogens,* 5(4)**,** e1000365.

Tsibris, A. M. N., Korber, B., Arnaout, R., Russ, C., Lo, C. C., Leitner, T., Gaschen, B., Theiler, J., Paredes, R., Su, Z. H., *et al*. 2009. Quantitative Deep Sequencing Reveals Dynamic HIV-1 Escape and Large Population Shifts during CCR5 Antagonist Therapy In Vivo. *Plos One,* 4(5)**,** e5683.

UNAIDS. 2016. Global AIDS Update. *UNAIDS*. Geneva, Switzerland.

van der Loeff, M. F. S. & Aaby, P. 1999. Towards a better understanding of the epidemiology of HIV-2. *Aids,* 13(Suppl. A)**,** 69-84.

Vessiere, A., Rousset, D., Kfutwah, A., Leoz, M., Depatureaux, A., Simon, F. & Plantier, J. C. 2010. Diagnosis and Monitoring of HIV-1 Group O-Infected Patients in Cameroun. *Jaids-Journal of Acquired Immune Deficiency Syndromes,* 53(1)**,** 107-110.

Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature,* 439(7074)**,** 344-348.

Weber, J., Gibson, R. M., Sácká, L., Strunin, D., Hodek, J., Weberová, J., Pávová M., Alouani D., J., Asaad R., Rodriguez B., *et al.* 2017. Impaired human immunodeficiency virus type 1 replicative fitness in atypical viremic non-progressor individuals. *AIDS research and therapy*, *14*(1)**,** 15.

Weber, J., Rangel, H. R., Chakraborty, B., Tadele, M., Martinez, M. A., Martinez-Picado, J., Marotta, M. L., Mirza, M., Ruiz, L., Clotet, B., *et al*. 2003. A novel TaqMan real-time PCR assay to estimate ex vivo human immunodeficiency virus type 1 fitness in the era of multi-target (pol and env) antiretroviral therapy. *Journal of General Virology,* 84(8)**,** 2217-2228.

Weber, J., Rose, J. D., Vazquez, A. C., Winner, D., Margot, N., McColl, D. J., Miller, M. D. & Quinones-Mateu, M. E. 2013. Resistance Mutations outside the Integrase Coding Region Have an Effect on Human Immunodeficiency Virus Replicative Fitness but Do Not Affect Its Susceptibility to Integrase Strand Transfer Inhibitors. *Plos One,* 8(6)**,** e65631.

Weber, J., Vazquez, A. C., Winner, D., Rose, J. D., Wylie, D., Rhea, A. M., Henry, K., Pappas, J., Wright, A., Mohamed, N., *et al*. 2011. Novel Method for Simultaneous Quantification of Phenotypic Resistance to Maturation, Protease, Reverse Transcriptase, and Integrase HIV Inhibitors Based on 3' Gag(p2/p7/p1/p6)/PR/RT/INT - Recombinant Viruses: a Useful Tool in the Multitarget Era of Antiretroviral Therapy. *Antimicrobial Agents and Chemotherapy,* 55(8)**,** 3729-3742.

Weber, J., Weberova, J., Carobene, M., Mirza, M., Martinez-Picado, J., Kazanjian, P. & Quinones-Mateu, M. E. 2006. Use of a novel assay based on intact recombinant viruses expressing green (EGFP) or red (DsRed2) fluorescent proteins to examine the contribution of pol and env genes to overall HIV-1 replicative fitness. *Journal of Virological Methods,* 136(1-2)**,** 102-117.

Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature,* 412(6844)**,** 331-333.

Willey, R. L., Maldarelli, F., Martin, M. A. & Strebel, K. 1992. Human-immunodeficiency-virus type-1 vpu protein induces rapid degradation of CD4. *Journal of Virology,* 66(12)**,** 7193-7200.

Xiang, S. H., Pacheco, B., Bowder, D., Yuan, W. & Sodroski, J. 2013. Characterization of a dual-tropic Human immunodeficiency virus (HIV-1) strain derived from the prototypical X4 isolate HXBc2. *Virology,* 438(1)**,** 5-13.

Yang, W. L., Kouyos, R. D., Boni, J., Yerly, S., Klimkait, T., Aubert, V., Scherrer, A. U., Shilaih, M., Hinkley, T., Petropoulos, C., *et al.* 2015. Persistence of Transmitted HIV-1 Drug Resistance Mutations Associated with Fitness Costs and Viral Genetic Backgrounds. *Plos Pathogens,* 11(3)**,** e1004722.

Yu, X. H., Yu, Y. K., Liu, B. D., Luo, K., Kong, W., Mao, P. Y. & Yu, X. F. 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science,* 302(5647)**,** 1056-1060.

Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J. & Neher, R. A. 2015. Population genomics of intrapatient HIV-1 evolution. *Elife,* 4**,** e11282.

Zhou, M. S., Deng, L. W., Kashanchi, F., Brady, J. N., Shatkin, A. J. & Kumar, A. 2003. The Tat/TAR-dependent phosphorylation of RNA polymerase IIC-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proceedings of the National Academy of Sciences of the United States of America,* 100(22)**,** 12666-12671.

Zhu, T. F., Muthui, D., Holte, S., Nickle, D., Feng, F., Brodie, S., Hwangbo, Y., Mullins, J. I. & Corey, L. 2002. Evidence for human immunodeficiency virus type 1 replication in vivo in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *Journal of Virology,* 76(2)**,** 707-716.