

**Etablierung einer Komplett-Genom-RT-PCR und
Next-Generation-Sequencing für *full-length* Genome
von HIV-1**

Masterarbeit

im Master-Studiengang Molekularbiologie und Biochemie
der Universität Potsdam

zur Erlangung des akademischen Grades eines
Master of Science (M. Sc.)

vorgelegt von

Christiane Anika Bresk

Dezember 2014

Abkürzungsverzeichnis

Acc.-Nr.	Accession Number
AIDS	<i>Acquired Immunodeficiency Syndrome</i> , Erworbenes Immundefizienzsyndrom
ART	antiretrovirale Therapie
Bp	Basenpaare
CA	<i>capsid protein</i> , Kapsidprotein
cART	Kombinationstherapie
CRF	<i>circulating recombinant form</i>
DMSO	Dimethylsulfoxid
EDTA	Ethylendiamintetraessigsäure
<i>env</i>	<i>envelope</i>
FCV	Felines Calicivirus
FKS	Fetales Kälberserum
<i>gag</i>	gruppenspezifische Antigene von HIV-1
GOBICS	Göttingen Bioinformatics Compute Server
HIV	<i>human immunodeficiency virus</i> , Humanes Immundefizienzvirus
HTLV III	humanes T-lymphotropes Virus 3
<i>int</i>	Integrase
jpHMM	<i>jumping profile Hidden Markov Model</i>
LAV	<i>Lymphadenopathy associated virus</i> , Lymphadenopathie-Virus
LI	Linkprotein
<i>LTR</i>	<i>long terminal repeat</i> , Lange terminale Repetition
MA	Matrixprotein
MENA	Middle East and North Africa, Mittlerer Osten und Nordafrika
NC	Nukleokapsid
Nef	<i>negative factor</i> , Negativer Faktor
NGS	<i>Next Generation Sequencing</i>

NNRTI	Nicht-Nukleosidische Reverse- Transkriptase- Inhibitor
NRTI	Nukleosidische Reverse-Transkriptase- Inhibitor
ORF	<i>open reading frame</i> , Offener Leserahmen
PBMC	<i>peripheral blood mononucleated cells</i> , Periphere mononukleäre Blutzellen
PIs	Protease-Inhibitoren
PHA	Phytohämagglutinin
<i>pol</i>	Polymerase-Region von HIV-1
RIP	Recombinant Identification Program
p.i.	post infectionem, nach Infektion
PR	Protease
RT	Reverse Transkriptase
SIV	<i>simian immunodeficiency virus</i> , Simianes Immundefizienzvirus
TAT	<i>transactivator of transcription</i> , Transkriptionsaktivator von HIV-1
URF	unique recombinant form
Vif	<i>virion infectivity factor</i>
Vpr	<i>viral protein R</i>
Vpu	<i>viral protein U</i>

Inhaltsverzeichnis

1. Einleitung	7
1.1 Allgemeines	7
1.2 Entdeckung von HIV	7
1.3 Morphologie und Genom von HIV-1	8
1.4 Genetische Diversität von HIV-1	9
1.4.1 Klassifizierung von HIV	11
1.4.2 Zirkulierende rekombinante Formen	11
1.4.3 Next Generation Sequencing zur Analyse der HIV Quasispezies	12
1.5 Epidemiologie von HIV-1	13
1.6 Hintergrund der Arbeit.....	13
1.6.1 Zielsetzung.....	14
2. Material und Methoden	16
2.1 Material	16
2.1.1 Technische Geräte	16
2.1.2 Kunststoffverbrauchsmaterialien.....	17
2.1.3 Chemikalien/Feinchemikalien/Reagenzien	17
2.1.4 Medien	18
2.1.5 Puffer	18
2.1.6 Kommerzielle Kits	19
2.1.7 Referenzen.....	19
2.1.8 Patientenproben.....	20
2.1.9 Blutspender.....	20
2.1.10 Oligonukleotide/PCR- und Sequenzierungsprimer	20
2.1.11 Software	22
2.2 Methoden.....	23
2.2.1 Isolierung von PBMC aus Buffy Coat.....	23
2.2.2 Virusanzucht aus Plasma.....	24
2.2.3 Quantitativer p24-Antigen-Capture-ELISA	25
2.2.4 RNA-Extraktion	27

2.2.5 full-length cDNA-Synthese	27
2.2.6 Amplifikation des Komplet-Genoms von HIV-1 - „Vier-Amplikon PCR“	28
2.2.7 Nested-pol-PCR (Inhouse).....	31
2.2.8 Ein-Amplikon full-length-PCR-Varianten	33
2.2.9 Agarose-Gelelektrophorese zur qualitativen und quantitativen DNA-Analyse ...	42
2.2.10 Präparation von DNA-Fragmenten aus Agarose-Gel.....	42
2.2.11 Aufreinigung von DNA.....	42
2.2.12 Sanger-Sequenzierung	43
2.2.13 Next Generation Sequencing.....	43
2.2.14 Phylogenetische Analysen	45
2.2.15 Analyse der rekombinanten Genomstruktur.....	46
3.Ergebnisse	48
3.1 Virusisolierung aus peripheren mononukleären Blutzellen	48
3.2 Etablierung einer ein-Amplikon HIV-1 full-length PCR.....	49
3.2 Generierung von full-length Amplikons mittels „vier-Amplikon PCR“	54
3.3 Ermittlung der Komplet-Genomsequenz für Isolate der „URF-new clade“ ...	56
3.4 Next Generation Sequencing zur Generierung der full-length Sequenz der „URF-new clade“	57
3.5 Threshold-Analyse für den Nachweis minoritärer Varianten mit NGS	61
3.6 Phylogenetische Analysen der „URF-new clade“	63
3.7 Analyse der rekombinanten Genomstruktur einer neuen zirkulierenden rekombinanten Form aus Oman	66
4.Diskussion.....	70
4.1 Virusanzucht eines omanischen Isolates auf PBMC-Spenderzellen	71
4.2 Etablierung einer full-length-RT-PCR für HIV-1 in Form einer „ein-Amplikon-PCR“	72
4.3 Amplifikation und Sequenzierung des HIV-1 Komplet-Genoms einer neuen URF-Variante aus Oman.....	73
4.4 Rekombinationsstruktur der omanischen „URF-new clade“ Viren	75
4.5 Sequenzierung von full-length Genomen mit Next Generation Sequencing ..	76

4.6 Ausblick.....	77
5.Zusammenfassung.....	79
6.Summary	81
7.Appendix	83
7.1 Übersicht der Primerbindungsstellen der full-length PCR-Variationen	83
7.2 Übersicht der Primerlokalisationen für die Sequenzierung.....	84
7.3 Stammbaum mit Genomsequenzen der „URF-new clade“ mit 170 Referenzsequenzen und Ergebnisse der BLAST-Suche.....	85
7.4 HIV-1 Referenzpanel der Los Alamos HIV Datenbank, 2010	87
7.5 Ergebnisse der BLAST-Suche	91
8.Literaturverzeichnis.....	92

1. Einleitung

1.1 Allgemeines

Seit der Entdeckung von HIV und der von diesem Virus verursachten Immunschwäche AIDS (**A**cquired **I**mmunodeficiency **S**yndrom) haben sich weltweit mehr als 60 Millionen Menschen mit HIV infiziert und mehr als 25 Millionen Menschen sind an AIDS gestorben¹⁻³. Im Jahr 2012 lebten geschätzt 35,3 Millionen Menschen mit HIV, wovon 2,3 Millionen Neuinfektionen registriert wurden, was einem Rückgang von 33 % im Vergleich zu 2001 (3,4 Millionen Neuinfektionen) entspricht. Weltweit starben im Jahr 2012 1,6 Millionen Menschen an AIDS⁴. Ein Großteil (25 Millionen, 2012) der HIV-infizierten Personen lebt in Afrika⁴.

Das Virus wird durch infektiöse Körperflüssigkeiten, wie Blut, Sperma, Vaginalsekret und Muttermilch übertragen und etabliert eine chronische, persistierende Infektion. Die Replikation von HIV kann durch eine erfolgreiche antiretrovirale Therapie (ART) gehemmt werden, wodurch die Vermehrung des Virus im Körper reduziert und somit die Krankheitsprogression verlangsamt wird. Derzeit werden Präparate fünf unterschiedlicher Wirkstoffklassen in der Kombinationstherapie (cART) eingesetzt: Nukleosidische Reverse-Transkriptase-Inhibitoren (NRTIs), Nicht-Nukleosidische Reverse-Transkriptase-Inhibitoren (NNRTIs), Protease-Inhibitoren (PIs), Fusions-Inhibitoren und Integrase-Inhibitoren⁵. Bisher konnte HIV durch cART nicht kuriert werden. Knochenmarktransplantationen waren bislang nur in einem von sechs Fällen erfolgreich⁶⁻⁹.

1.2 Entdeckung von HIV

Im Jahr 1981 wurden in den USA ungewöhnlich viele junge homosexuelle Männer mit dem Kaposi Sarkom diagnostiziert¹⁰, die schwere opportunistische Infektionen, u. a. *Pneumocystis* Pneumonia, Toxoplasmose, *Herpes simplex* Infektionen und Meningitis, entwickelten und innerhalb von 24 Monaten verstarben^{10,11}. Dieses Krankheitsbild bekam kurze Zeit später, aufgrund der einhergehenden Immunschwäche, den Namen AIDS¹²⁻¹⁴. Zwei Jahre später konnte das AIDS verursachende Retrovirus aus einem Lymphknoten eines Patienten mit Lymphadenopathie isoliert werden¹⁵ und wurde anfänglich als LAV (**L**ymphadenopathy **a**ssociated **v**irus) und HTLV III (**h**umanes **T**-lymphotropes **V**irus 3) be-

zeichnet^{16,17}. 1986 wurde das ätiologische AIDS-Agens in HIV (**h**uman **i**mmuno-**d**eficiency **v**irus) umbenannt¹⁸. Im gleichen Jahr wurde in Westafrika ein weiteres Retrovirus isoliert, das HIV-2 genannt wurde¹².

1.3 Morphologie und Genom von HIV-1

Die HIV-Partikel sind von einer Lipidmembran zellulären Ursprungs umhüllt. Im Viruspartikel befinden sich eine Vielzahl von Enzymen und Strukturproteinen. Im Inneren des konischen Kapsids ist das Genom des Retrovirus in Form von zwei einzelsträngigen RNA-Kopien lokalisiert. Jedes Genom besteht aus drei großen offenen Leserahmen, die die Strukturproteine *gag* (gruppenspezifische Antigene), *env* (Hüllglykoproteine) und die *pol*-Enzyme (Polymerase) kodieren. Diese sind gemeinsam mit Genen für zwei regulatorische Proteine (Tat und Rev) und Genen für vier akzessorische Proteine (Vpr, Vpu, Vif und Nef) im ca. 9000 Bp großen Genom¹⁹ überlappend angeordnet (Abb. 1A).

Das Genom wird von zwei nicht kodierenden *LTR*-Regionen flankiert, wobei die 5'-LTR die Promotorfunktion für die Expression der viralen Gene ausübt^{12,20,21}.

Der *gag*-Genombereich kodiert das Gag-Polyprotein, welches aus dem Matrixprotein (MA), Kapsidprotein (CA), Nukleokapsid (NC) und Linkprotein (LI) besteht²². Die *pol*-Region (Abb. 1B) beinhaltet die Geninformation für folgende Replikationsproteine: Protease (PR), Reverse Transkriptase (RT), Ribonuklease H und Integrase (INT)²⁰. Die RT schreibt die virale genomische RNA in cDNA (RNA-DNA-Hybrid) um. Der Einbau der DNA in das Wirtsgenom erfolgt durch das Enzym Integrase. Dieses entfernt ein Dinukleotid an beiden 3' Enden der DNA, sodass diese anschließend in das Wirtsgenom eingebaut werden kann (Strangtransfer). Die Protease spaltet sich autokatalytisch aus dem Gag-Vorläuferprotein ab und prozessiert dann hoch reguliert die Gag- und Pol-Vorläufer-Polyproteine in funktionale Proteine. Erst durch diesen Maturationsprozess werden die Virionen infektiös^{21,23-27}.

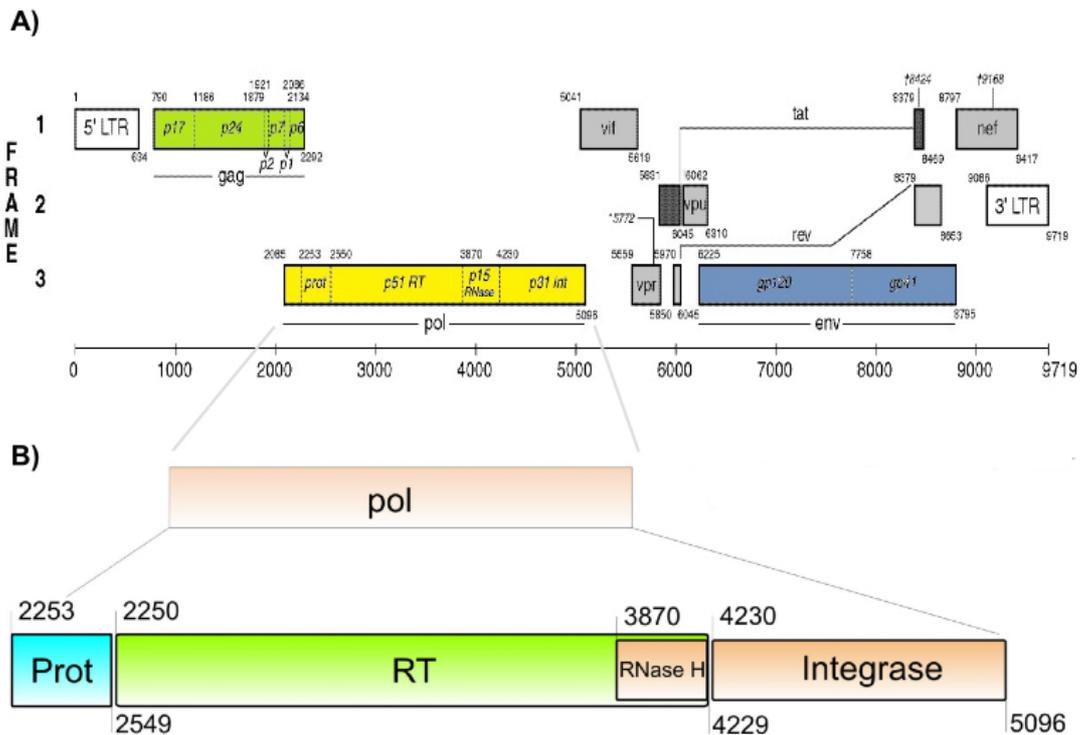


Abb. 1 Schematische Struktur des Genoms von HIV-1

- A) Genom des Referenzvirus HxB2 (Acc.-Nr.: K03455, Quelle (<http://www.hiv.lanl.gov/>) bearbeitet)
 Offene Leserahmen (*frames*) sind als Rechtecke dargestellt. Das jeweilige Startcodon wird durch die Position, welche das A des ATG markiert, an der oberen linken Ecke angezeigt und durch die Nummer (letzte Position des Stop-Codons) an den unteren rechten Ecken beendet
- B) Vergrößerter Ausschnitt der *pol*-Region mit den einzelnen offenen Leserahmen Protease (blau), Reverse Transkriptase mit RNase H (grün) und Integrase (orange)

Im *env*-Genombereich sind die Hüllglykoproteine kodiert, bestehend aus gp120 und gp41. Diese werden ebenfalls aus einem Vorläuferprotein, gp160, prozessiert²⁸. Die regulatorischen Proteine Tat (*transactivator of transcription*) und Rev (*regulator of expression of virion*) sind essentiell für die Replikation des Virus²⁸. Die akzessorischen Proteine Vif (*virion infectivity factor*), Vpr (*viral protein R*), Vpu (*viral protein U*) und Nef (*negative factor*) haben keine essentielle Bedeutung für die Replikation von HIV-1, sondern erhöhen die Infektiosität der Viruspartikel (Nef) und unterstützen die Freilassung und Übertragung der Viren aus den infizierten Zellen (Vpu, Vpr, Vif)²⁹.

1.4 Genetische Diversität von HIV-1

Während einer HIV-Infektion befindet sich nach der Übertragung eine relativ homogene Viruspopulation im Organismus, da nur wenige Varianten übertragen werden. Im Verlauf der Infektion divergieren die Virusvarianten, sodass es auf

Nukleotidebene beispielsweise des *env*-Genomabschnittes zu einer Variation von 10 % kommen kann³⁰. Diese Gesamtpopulation (Gesamtheit aller Varianten) von HIV in einem infizierten Individuum wird als Quasispezies bezeichnet³⁰.

Die Diversität von Virusvarianten entwickelt sich durch den natürlichen Selektionsdruck, welcher zur Adaption des Virus an die immunologische Antwort des infizierten Individuums führt. Auch der Einsatz von antiretroviralen Medikamenten kann zur Veränderung der Divergenz der viralen Quasispezies führen. Die einzelnen Genombereiche sind unterschiedlich konserviert. Hoch divergierende Bereiche sind die *env*-Region mit einer Sequenzvariabilität auf Aminosäureebene von durchschnittlich 17 % und der *gag*-Genomabschnitt mit durchschnittlich 8 % Aminosäuredivergenz. Zwischen HIV-1 Subtypen, wie bspw. Subtyp A und B, kann es zu größeren Unterschieden kommen. Hier beträgt der Unterschied im *env*-Bereich 25 % und im *gag*-Bereich im Durchschnitt 17 %³⁰. Neben variablen Bereichen gibt es ebenfalls hoch konservierte Regionen, wie die *LTR*-Region (long terminal repeat), das aktive Zentrum der Reversen Transkriptase und Integrase sowie Teile des p24-Proteins³¹.

Die hohe genetische Diversität wird hauptsächlich durch die fehleranfällige Reverse Transkriptase verursacht. Dieser fehlt die Korrekturfunktion, wodurch es zum Einbau fehlgepaarter Nukleotide (5-10 pro Genom^{32,33}) und dadurch zu Mutationen im HIV-Genom kommt, wodurch neue Varianten entstehen^{31,33-35}. Die Fehlerrate der Reversen Transkriptase beträgt 1 in 10.000 Nukleinsäuren³¹.

Aufgrund der hohen Generationsrate von einer Milliarde Viruspartikel pro Tag pro Patient³⁶ können Mutationen, je nach ihrem Einfluss auf die Überlebensfähigkeit, sehr schnell akkumulieren.

Eine weitere Eigenschaft der Reversen Transkriptase, die zur Diversität durch Rekombination beiträgt, ist die schwache Bindungsaffinität zum RNA-*Template*^{37,38}. Durch das Vorhandensein von zwei RNA-Kopien im Virus kann es zum intermolekularen Wechsel des RNA-*Templates* kommen³⁹. Dieser Vorgang des Wechsels wird auch als „*template switching*“ bezeichnet³⁵. Zusätzlich kann es zu Insertionen/Deletionen durch diese intermolekularen Sprünge der RT kommen.

1.4.1 Klassifizierung von HIV

Das humane Immundefizienz-Virus hat seinen Ursprung in West- und Zentralafrika^{2,40}. Dort ist das SIV (*simian immunodeficiency virus*) durch Zoonosen von Affen auf den Menschen übertragen worden⁴¹. Durch die Übertragung des Virus von der Raubkatze auf den Menschen entstand HIV-2 (HIV Typ 2)⁴¹. Durch unabhängige Transmissionsereignisse von großen Menschenaffen auf den Menschen entstand HIV-1 (HIV Typ 1), welches anhand von phylogenetischen Analysen, in 4 Gruppen unterteilt wird^{2,3}.

Die größte und für die globale Epidemie verantwortliche Gruppe ist die Gruppe M (*major*) von HIV-1. Die Gruppe-M Viren sind nach Schätzungen anhand phylogenetischer Analysen Mitte des 19. Jahrhunderts⁴⁰ vom Schimpansen auf den Menschen übertragen worden⁴¹. Weitere Übertragungen von Mensch zu Mensch brachten unterschiedliche Subtypen und zirkulierende rekombinante Formen (CRF) der Gruppe-M hervor. Diese unterscheiden sich, in Abhängigkeit der Genregion und des Subtypen auf Aminosäureebene, genetisch aber nicht serologisch voneinander. So können Divergenzunterschiede zwischen den Subtypen von durchschnittlich 17 % (*env*), 15 % (*tat*) und 8 % (*gag*) auf der Aminosäureebene beobachtet werden^{2,30}. Derzeit sind 9 Subtypen bekannt, die mit A-D, F-H, J und K bezeichnet werden².

Die Gruppe-O (*outlier*) ist mit einigen zehntausend bekannten Infektionen in West- und Zentralafrika die zweitgrößte Gruppe von HIV-1 und entstand durch eine Transmission von SIV vom Gorilla auf den Menschen (ca. 1741)^{3,40-42}. In Kamerun sind einige wenige Menschen mit dem HIV-1 Gruppe-N Virus (*non-M, non-O*) infiziert³, das ca. 1963^{40,43} ebenfalls vom Schimpansen auf den Menschen übertragen wurde. Die kleinste und zuletzt klassifizierte Gruppe-P (*pending the identification of further human cases*) wird von bisher nur zwei bekannten Isolaten in Kamerun gebildet^{3,44,45}. Es wird vermutet, dass das P-Virus genau wie das O-Virus vom Gorilla auf den Menschen übertragen wurde, durch die wenigen bekannten Sequenzinformationen kann jedoch keine verlässliche Aussage darüber getroffen werden.

1.4.2 Zirkulierende rekombinante Formen

In Populationen, in denen mehrere HIV-Varianten zirkulieren, kommt es zu Rekombinationen zwischen diesen Viren^{2,46,47}. Rekombinante Varianten entstehen

während einer Ko- bzw. Superinfektion in einer Zelle eines Individuums. Von einer Koinfektion spricht man, wenn sich ein Patient während eines Transmissionsereignisses mit zwei verschiedenen Virus-Varianten infiziert. Kommt es nach einer Serokonversion (Bildung spezifischer Antikörper) zu einer zweiten Infektion mit HIV spricht man von einer Superinfektion. Der Großteil dieser Rekombinanten ist jedoch singulär und wird als singuläre rekombinante Form (URF, **u**nique **r**ecombinant **f**orm) bezeichnet^{2,46}. Kommt es jedoch in mehreren, epidemisch nicht direkt verknüpften Individuen, zur Identifizierung einer URF, kann sie als zirkulierende rekombinante Form (CRF, **c**irculating **r**ecombinant **f**orm) klassifiziert werden. Derzeit sind 72 CRF in der Los Alamos HIV Datenbank eingetragen (Los Alamos HIV Datenbankstand: 18.11.2014). Um als zirkulierende rekombinante Form klassifiziert zu werden, müssen bestimmte Kriterien erfüllt sein. Die *full-length* Sequenzen von mindestens zwei Isolaten und eine partielle Sequenz müssen vorliegen sowie eine signifikante genetische Distanz zu anderen Subtypen aufzeigen^{2,46}.

1.4.3 Next Generation Sequencing zur Analyse der HIV Quasispezies

In den letzten 10 Jahren wurden neue Methoden zur DNA-Sequenzierung entwickelt, welche als *Next Generation Sequencing* bezeichnet werden und mit denen es möglich ist, minoritäre Varianten der Quasispezies zu detektieren. Im Vergleich zur herkömmlichen Sanger-Sequenzierung ermöglichen die Next Generation Sequenzierungstechnologien (NGS) durch mehr Sequenzinformation eine schnellere, sensitivere und dabei auch kostengünstigere Analyse des HIV-1 Genoms. Aktuell sind die NGS-Systeme der Firmen Illumina, 454 Life Sciences, Life Technologies und Pacific Biosciences marktführend, deren unterschiedliche Sequenzierungstechnologien für unterschiedliche Anwendungen geeignet sind⁴⁸. Für die Sequenzierung des HIV-1 Hüllproteins zeigte eine Studie, dass die Illumina-Technologie weniger fehlerhafte Sequenzen (3 %) erzeugt als die Sequenzierungstechnologien von 454 Life Sciences (16 %), Life Technologies (44 %) und Pacific Biosciences (64 %). Nach einer umfassenden Sequenzdatenbereinigung konnten jedoch vergleichbare Häufigkeiten von Virusvarianten mit allen NGS-Systemen detektiert werden⁴⁹.

1.5 Epidemiologie von HIV-1

Global ist die Verteilung der einzelnen HIV-1 Subtypen sehr heterogen. Mit ca. 50 % ist Subtyp C weltweit dominant und kommt in Sub-Sahara Afrika, Ost-Afrika, Indien und Süd-Brasilien vor. Subtyp A ist mit 12 % in Zentral-Afrika, Iran, Zentralasien und Osteuropa anzutreffen, seltener jedoch in Europa. Hier sowie in Nord-Amerika, Japan und Australien ist Subtyp B (11 %) dominant. Die verbleibenden Subtypen umfassen 10 % aller HIV-Infektionen und sind überwiegend in Afrika verbreitet. Einige rekombinante Formen sind epidemisch relevant, wie CRF01_AE (Südost-Asien) und CRF02_AG (West-Afrika)^{2,33}.

1.6 Hintergrund der Arbeit

Seit 2011 besteht eine Kooperation zwischen der Arbeitsgruppe (Fachgebiet 18 „HIV und andere Retroviren“) und dem „Central Public Health Laboratory“ (CPHL) des omanischen Gesundheitsministeriums „Ministry of Health“ in Muscat. Der Oman gehört neben 10 weiteren Ländern laut UNAIDS zu der MENA-Region (**M**iddle **E**ast and **N**orth **A**frica)⁴ in denen derzeit 260.000 Personen mit HIV leben, was einer HIV-Prävalenz von 0,1 % entspricht⁴. Für das Jahr 2012 wurden die HIV-Neuinfektionen auf 32.000 und die Anzahl an AIDS-Toten auf 17.000 Personen geschätzt. Für Oman bewegen sich die Schätzungen für Personen die mit HIV leben zwischen 2800-5700⁴. Hier gibt es keine offiziellen Daten über die jährliche Neuinfektionsrate, AIDS-Tote und die Verteilung der Subtypen^{4,33}. Basierend auf noch nicht veröffentlichten Daten einer Pilotstudie des Robert Koch-Instituts ist der Subtyp C im Oman am häufigsten. Dem schließt sich der Subtyp A1 sowie die rekombinante Form CRF01_AE an.

Ziel der Kooperation war die genotypische Resistenzbestimmung HIV-1-infizierter Patienten mit Therapieversagen. Anhand der ermittelten *pol*-Sequenzen wurde auch der Subtyp der HIV-1 Infektion bestimmt. Dabei konnten ca. 50 % der Isolate keinem bekannten Subtyp oder einer zirkulierenden rekombinanten Form zugewiesen werden. Diese Isolate wurden zunächst als singuläre rekombinante Form (URF) klassifiziert. In weiteren Analysen stellte sich heraus, dass sich 13 dieser Isolate signifikant von allen bekannten Subtypferenzviren und rekombinanten Formen der HIV-Datenbank unterscheiden⁵⁰⁻⁵². Diese 13 Isolate waren in einer gemeinsamen monophyletischen Gruppe („clade“) des Stammbaumes (*Neighbor-Joining* Methode)

lokalisiert und wurden zunächst als „URF-*new clade*“ bezeichnet. Nach phylogenetischen Analysen handelt es sich nicht um epidemisch verknüpfte Infektionen. Daher bestand die Vermutung, dass es sich um eine eigenständige, bisher noch unbekannte CRF handelt. Um eine neue CRF klassifizieren zu können, müssen zwei Komplet-Genomsequenzen und eine dritte partielle Sequenz eines potentiell neuen Subtypen oder CRF vorliegen und hinsichtlich ihrer Genomstruktur übereinstimmen². Phylogenetische Analysen mit dem aktuellen Referenzpanel der Los Alamos HIV Datenbank müssen eine signifikante Abspaltung zu bekannten Subtypen aufzeigen und eine mögliche Infektkette muss ausgeschlossen werden.

In vorherigen Arbeiten wurde die Komplet-Genom-Amplifikation aus viraler RNA anhand einer *full-length* cDNA-Synthese mit anschließender Amplifikation von vier überlappenden Fragmenten etabliert („vier-Amplikon PCR“)⁵¹.

Zur Etablierung wurde der Laborstamm HTLVIIIB und der rekombinante molekulare HIV-Klon pNL4.3wt (Acc.-Nr. M19921) verwendet. Die „vier-Amplikon PCR“ wurde bereits erfolgreich für die Untersuchung einer EDTA-Plasmaprobe (13-5995, Primärmaterial) als auch mit einem angezüchteten Virusisolat eines Patienten (13-0346), die beide dieser putativen neuen CRF *clade* angehören, eingesetzt^{51,52}. Die Populationssequenz des Komplet-Genoms des Patienten 13-0346 konnte bereits durch Sanger-Sequenzierung ermittelt werden⁵².

1.6.1 Zielsetzung

Zu Beginn dieser Masterarbeit lag von zwei Patienten (13-0346 und 13-5995) eine vollständige Sequenz sowie eine partielle Sequenz eines dritten Patienten (14-0875) vor. Die vorliegende Populationssequenz 13-5995 enthielt jedoch viele Mischpositionen.

Ziel dieser Arbeit war es, ausreichende Sequenzinformationen zu etablieren. Anhand der Komplet-Genomsequenz sollte eine Strukturanalyse zur Aufklärung der rekombinanten Genomstruktur vergleichend durchgeführt werden, um eine neue zirkulierende rekombinante Form anhand der „URF-*new clade*“ endgültig zu definieren. Zusätzlich sollte eine „ein-Amplikon *full-length* HIV-1 PCR“ etabliert werden, da man bei der Populationssequenz, die aus vier unabhängig erzeugten PCR-Amplifikaten erstellt wurde, nicht beurteilen kann, ob die finale Sequenz aus unterschiedlichen Varianten der viralen Quasispezies zusammengesetzt wurde. Durch Einsatz dieser PCR soll für den Patienten 13-5995 eine eindeutige

Populationssequenz generiert werden und die Vervollständigung der Komplett-Genomsequenz für den Patienten 14-0875 erfolgen. Um ausreichend Material für die Etablierung, Optimierung der ein-Amplikon *full-length* PCR und weitere Untersuchungen zur Verfügung zu haben, soll eine Virusanzucht aus Plasma versucht werden.

Um eine rasche und kostengünstigere Sequenzanalyse zu ermöglichen, sollte über *Next Generation Sequencing* (NGS) ebenfalls die Ermittlung der Komplett-Genom-Konsensussequenz erfolgen. Diese Methode wird in der Arbeitsgruppe zurzeit etabliert. Die Sanger-Populationssequenzen sollen mit der Konsensussequenz aus NGS verglichen werden.

2. Material und Methoden

2.1 Material

2.1.1 Technische Geräte

Brutschrank	CO ₂ -Inkubator CB 150, Binder, Tuttlingen
Gelelektrophoresekammer	Horizon 58, Invitrogen, Karlsruhe
Kapillarelektrophorese	Agilent 2100 Bioanalyzer, AgilentTechnologies, Waldbronn
Lunimometer	Tecan INFINITE F200, Tecan, Männedorf, Schweiz
Mikroplatten-Waschgerät	Columbus ProTM, Tecan, Männedorf, Schweiz
Neubauer Zählkammer	Roth, Karlsruhe
Präzisionswaage	L610D, Satorius, Elk Grove Village, USA
Sicherheitswerkbank	Serien Nr.12100722, Bleyemehl Reinraumtechnik HeraSafe, Heraeus, Hanau
Thermozykler	Mastercycler proS, Eppendorf AG, Hamburg T3000 Thermicycler, Biometra GmbH, Göttingen
Ultrazentrifuge	Optima XL-90 Ultrazentrifuge, Beckman Coulter GmbH, Krefeld
Rotor Ultrazentrifuge	SW 60Ti, Beckman Coulter GmbH, Krefeld
Einsätze Ultrazentrifuge	SW 60, Beckman Coulter GmbH, Krefeld
Videodokumentationssystem	
E.A.S.Y RH-3	Herolab GmbH, Wiesloch
Zentrifugen	Heraeus Fresco 21, Thermo Scientific, USA Heraeus Multifuge X3 FR, Thermo Scientific, USA

Microcentrifuge 5417R, Eppendorf AG,
Hamburg

2.1.2 Kunststoffverbrauchsmaterialien

24-Lochplatte	TPP Techno Plastic Products AG, Trasadingen, Schweiz
96-Lochplatte	Corning Coster, Corning Life Science, Tewksbury MA, USA
Einzelblattfolie	3M™ High-Strength Acrylic Adhesive 300LSE, St. Paul, USA
Freezing Container „Mr. Freeze“	SIGMA-ALDRICH, St. Louis, USA
Zellkulturflaschen 25 cm ²	TPP Techno Plastic Products AG, Trasadingen, Schweiz
Zellkulturflaschen 75 cm ² , 125 cm ²	Nunc, Roskilde, Dänemark

2.1.3 Chemikalien/Feinchemikalien/Reagenzien

Bovine serum albumin	SIGMA-ALDRICH, St. Louis, USA
Chloarmin-T reinst	VRW International, Pennsylvania, USA
DMSO	SIGMA-ALDRICH, St. Louis, USA
EDTA 0,5 pH 8,0	MERCK, Darmstadt
Empigen BB 30 %	MERCK, Darmstadt
Essigsäure 99-100 %	MERCK, Darmstadt
Ethanol 96 %	Carl Roth GmbH + Co. KG, Karlsruhe
Ethidiumbromid	Amresco, Solon, OH, USA
Fetales Kälberserum (FKS)	Biochrom AG, Berlin
Ficoll Paque Plus	GE Healthcare Europe GmbH, Freiburg
Freeze-Mix	90 % FKS, 10 % DMSO
Gene Ruler™ 1 kb DNA Ladder	Fisher Scientific GmbH, Scherte
Interleukin-2	SIGMA-ALDRICH, St. Louis, USA
Konjugat-Lösung	Alkaline Phosphatase Conjugate of Anti-HIV-1-p24 Mouse Monoclonal, Aalto Bio Reagents, Dublin, Irland

Lambda DNA/HindIII Marker, 2	Thermo Scientific, Schwerte
L-Glutamin	Gibco®, Life Technologies GmbH, Darmstadt
Low Mass Ladder	Fermentas - Life sciences, St. Leon, Rot
Mineralöl	SIGMA-ALDRICH, St. Louis, USA
NaCl	SIGMA-ALDRICH, St. Louis, USA
NaHCO ₃	SIGMA-ALDRICH, St. Louis, USA
Penicillin/Streptomycin 100x	GE Healthcare Europe GmbH, Freiburg
Phytohämagglutinin (PHA)	SIGMA-ALDRICH, St. Louis, USA
Schaf- α -HIV-1-p24 Antikörper	Affinity Purified Sheep α -HIV-1-p24gag, Aalto Bio Reagents, Dublin, Irland
Schafserum	SIGMA-ALDRICH, St. Louis, USA
TE-Puffer	Promega, San Luis Obispo, USA
Tris-HCl	SIGMA-ALDRICH, St. Louis, USA
Tris Ultra Qualität	Carl Roth GmbH + Co. KG, Karlsruhe
TROPIX CSPD Sapphire Substrat	ABI, Weiterstadt
Tween®20	Fluka, Schweiz
Ultra Pure™ Agarose	Invitrogen, Karlsruhe

2.1.4 Medien

RPMI-1640	Gibco BRL, Eggenstein
RPMI- Vollmedium (RPMI ⁺):	
440ml	RPMI 1640
5 ml	Glutamin (final 2 mM)
50 ml	FKS (100 %)
5 ml	Pen/Strep (100u/ml)

2.1.5 Puffer

10x Assay Buffer TROPIX	ABI, Weiterstadt
6x Ladepuffer	Fisher Scientific GmbH, Schwerte
1x TE-Puffer	Promega, Madison, USA

Phosphatgepufferte Salzlösung PBS „ohne“

NaCl 137 mM

KCl 2,7 mM

Na₂HPO₄ 4,3 mM

KH₂PO₄ 1,47 mM

Red Cell Lysis Buffer

8,290 g NH₄Cl

1,000 g KHCO₃

0,372 g EDTA (Titriplex)

Lösen in acq. Bdest, auf 1 L auffüllen, in 4 x 250 ml Portionen autoklavieren,

Lagerung bei 4°

2.1.6 Kommerzielle Kits

BigDye Terminator v 3.1 Cycle

Sequencing Kit

Applied Biosystems, Weiterstadt

Expand Long Template PCR

Roche, Mannheim

GeneJet Gel Extraction Kit

Thermo Scientific, Schwerte

High Sensitivity DNA Analysis Chip

Agilent 2100 Bioanalyzer,

Agilent Technologies, Waldbronn

Kapa Library Quantification Kit for Illumina

Kapa Biosystems, Cape Town, South Africa

High Fidelity PCR System

Roche, Mannheim

MSB Spin PCRapace

Invitek, Berlin

Long PCR Enzyme Mix

Thermo Scientific, Schwerte

QIAamp Viral RNA Mini Kit

Qiagen, Hilden

Superskript® III First-Strand Synthesis

System

Invitrogen, Karlsruhe

2.1.7 Referenzen

Positivkontrolle

HIV-Laborstamm HTLVIII B

(1x10⁴ Kopien/mL) aus Virusstock in H-9

Zellen vermehrt; Tag der Abnahme: Tag 21

Negativkontrolle

HIV-negatives Plasma, Pool aus 26
Einzelspenden, Paul-Ehrlich-Institut

FCV

Felines Calicivirus, Fachgebiet 15, Robert
Koch-Institut

2.1.8 Patientenproben

Für diese Arbeit wurden Isolate bzw. Blutproben von HIV-1-infizierten Patienten aus Oman verwendet. Bei einer Probe handelt es sich um ein Virusisolat (13-0346), das aus PBMCs einer EDTA-Blutprobe als Co-Kultur auf Spender-PBMCs in vorangegangenen Arbeiten der AG angezüchtet worden war⁵¹. Zwei weitere Proben (13-5995, 14-0875) lagen als EDTA-Plasma vor (Primärmaterial, nicht kultiviert) und wurden in dieser Arbeit untersucht. Es liegt ein positives Ethikvotum aus dem Oman für die Untersuchung der Proben aus 2011 vor und somit auch für die Verlaufsprobe 13-0346.

2.1.9 Blutspender

In dieser Arbeit wurde versucht, Virus aus den Plasmaproben auf mononukleären Blutzellen (PBMCs) zweier verschiedener Blutspender anzuzüchten.

Tab.1 Übersicht der erhaltenen Buffy Coats von Blutspenden vom Deutschen Roten Kreuz (DRK)

DRK-Spender-Nr.	Scout	HIV-1 ELISA	Zellen/10 mL	Virusisolat
852 14 637 391	14-2008	negativ	3,175 x 10 ⁸	/
852 14 657 305	14-2009	negativ	4,850 x 10 ⁸	14-0875
853 14 637 399	14-2010	negativ	5,175 x 10 ⁸	13-6001

2.1.10 Oligonukleotide/PCR- und Sequenzierungsprimer

Primer, die für die Amplifikation und die Sequenzierung der „URF-*new clade*“ verwendet wurden, sind in folgender Tabelle aufgelistet.

Tab. 2 Oligonukleotide für die Amplifikation und die Sequenzierung des Komplett-Genoms von HIV-1 verwendet wurden

Name	URF full 1-4		Quelle	Sequenz 5'-3'	Lokalisation auf HxB2 (Acc.Nr. K03455)
	PCR	Sequenzierung			
Pan-HIV-1_1F	1	1	<i>Gall (2012)</i>	AGCCTGGGAGCTCTCTG	479-495
1.U5Cc	1		<i>Rousseau (2006)</i>	CCTTGAGTGCTCAAGTAGTGTGTGCCCGTCTGT	537-570
2.U5Cd	1		<i>Rousseau (2006)</i>	AGTAGTGTGTGCCCGTCTGTTGTGTGACTC	551-580
URF_full_5'-LTRs	full-length		FG18	CAGTGGCGCCCGAACAGGGACC	633-654
URF_full_5'-LTRs_neu	full-length	1	diese Arbeit	TCTCTAGCAGTGGCGCCCGAACAG	625-648
full_lenght_727 F	full-length	1	diese Arbeit	CGAGGGGCGGCGACTGGTGAG	727-747
full_lenght_782 F	full-length	1	diese Arbeit	GAGAGAGATGGGTGCGAGAG	782-801
G20s		1	FG18	GTATGGGCAAGCAGGGAGCTAGAA	891-914
URF_G20as		1	FG18	TTCTAGCTCCCTGCTTGCCCATAC	891-914
Gag_1259as		1	FG18	TTTACCCATGCATTTAAAGTTCTAGGTGA	1230-1258
Gag_1322s		1	FG18	ATACCCATGTTTTTCAGCATTATCAGAAGG	1293-1321
G45as		1	FG18	TTGGACCAACAAGTTTCTGTC	1739-1760
Pol_2001s	2	1,2	FG18	CTAGGAAAAAGGGCTGTT	2011-2028
Pol_2036s		1,2	FG18	GTGGAAAGGAAGGACACCAAATGAAAG	2035-2061
Pol_2322s		1,2	FG18	TTAGATACAGGAGCAGATGA	2321-2340
Pol_2402as		1,2	FG18	AATCCCCCTATCATTTTTGG	2381-2401
Pol_2696s		1,2	FG18	ATTGGGCCTGAAAAATCCATA	2696-2715
Pol_3034s		1,2	FG18	GTAGCATGACAAAAATCTTAGAG	3033-3055
Pol_3119as		1,2	FG18	TCCTACATACAAATCATCCAT	3098-3118
Pol_3454as		1,2	FG18	TCTGCTTCTTCTGTTAGTGGTA	3426-3447
Pol_3532as		1,2	FG18	TTCTGCTATTAAGTCTTTTGATGGGTCA	3505-3532
Pan-HIV-1_3F	3	2,3	<i>Gall (2012)</i>	TTAAAAGAAAAGGGGGATTGGG	4782-4804
ENVoutF1		3	FG18	AGAGGATAGATGGAACAAGCCCCAG	5549-5573
Pan-HIV-1_4F	4	3,4	<i>Gall (2012)</i>	CCTATGGCAGGAAGAAGCG	5966-5984
env_5901s		3	FG18	ATTGTGGGTCACAGTCTATTATGGGGTACCT	6322-6352
E40s_SF11		3	FG18	CATGTGGAAAAATGACATGGTAGAACA	6505-6531
URF_6881as		3,4	FG18	GGCACAATAATGTATGGGAATTGG	6857-6880
Env_6537s		3,4	FG18	AATGTCAGCACAGTACAATGTACAC	6944-6968
Env_2as		3,4	FG18	AGAAAAATCCCCTCCACAATTAA	7350-7373
Env_2s		3,4	FG18	TTAATTGTGGAGGGGAATTTTTCT	7350-7373
Env_5as		3,4	FG18	TCCACTGATGGGAGGGGCATACATTGC	7520-7546
Env_7254asc		3,4	FG18	TCATATCTCCTCCTCCAGGTCTGAA	7625-7649
Env_7407asc		3,4	FG18	CATAGTGCTTCTGCTGCTCCCAAGAACC	7785-7813
Pan-HIV-1_3R	3	3,4	<i>Gall (2012)</i>	TGGCCTGTACCGTCAGCG	7830-7847
Gp46_F2		4	FG18	ACAATTATTGTCTGGTATAGTGCAGCAGCA	7849-7878
E180s_SF7		4	FG18	GTCTGGTATAGTGCAGCAGCA	7858-7878
Gp48_R2		4	FG18	TCCTACTATCATTATGAATAATTTTATATA	8264-8293
Gp41_R1		4	FG18	AACGATAATGGTGAATATCCCTGCCTAA	8346-8373

Name	URF full 1-4		Quelle	Sequenz 5'-3'	Lokalisation auf HxB2 (Acc.Nr. K03455)
	PCR Sequenzierung				
URF_full_3'-LTRas	full-length	4	FG18	GTCATTGGTCTTAAAGGTACCTGAGGTGTGA	9004-9034
URF-nef_9104as		4	FG18	GAGTGAATTAGCCCTTCCAGTCCC	9080-9103
full_lenght_9,372 R	full-length		diese Arbeit	GATGAAATGCTAGGCGGCTG	9353-9372
full_lenght_9,396 R	full-length		diese Arbeit	CCGGATGCAGCTCTCGGGCCA	9376-9396
2.3'3'PICb	full-length		<i>Rousseau (2006)</i>	TGAAGCACTCAAGGCAAGCTTTATTGAGGCTTA	9603-9635
LTR-full-length-as	4	4	FG18	AGCACTCAAGGCAAGCTTTATTGAGGC	9606-9632
1.3'3'PICb	full-length		<i>Rousseau (2006)</i>	ACTACTTGAAGCACTCAAGGCAAGCTTTATTG	9610-9641

2.1.11 Software

BioEdit Version 7.2.5

Tom Hall, Sequence Alignment Editor⁵³

BLAST-Search

<http://www.hiv.lanl.gov>

contentsequence/BASIC_BLAST

basic_blast.html

DNA Lasergene SeqMan Pro

DNASTAR, Madison, Winconsin, USA

DNA Lasergene PrimerSelect

DNASTAR, Madison, Winconsin, USA

FastQC Version 0.11.2

Simon Andrews, Pierre Lindenbaum, Brian

Howard, Phil Ewls 2011-2014

FLASH

Tanja Magnoc, Assambly-Programm⁵⁴

Geneious

Biomatters Ltd. Auckland, Neuseeland

EASY Win32-System

Herolab GmbH, Wiesloch

Magellan 7.1

Tecan, Männedorf, Schweiz

Microsoft Excel

Microsoft Deutschland GmbH,

Deutschland

PHYLIP package 3.69

Joe Felsenstein

Segminator II

John Archer, Sequenzassambly-

Programm⁵⁵

Trimmomatic

Anthony M. Bolger, Trimmer für

Illumina-Sequenzen⁵⁶

2.2 Methoden

2.2.1 Isolierung von PBMC aus Buffy Coat

Die Isolierung von mononukleären Zellen des peripheren Blutes (PBMC) basiert auf der Methode der Dichtegradientenzentrifugation. Buffy Coat wurde vom Blutspendedienst DRK Heckshorn bezogen. Ein Buffy Coat geht auf eine 500 mL Blutspende zurück, aus der Plasma und Erythrozyten nach Zentrifugation weitestgehend entnommen wurden und periphere mononukleäre Blutzellen angereichert sind.

Plasma für einen HIV-ELISA wurde, wenn möglich, aus den Schläuchen des Blutbeutels entnommen. War dies nicht möglich, wurde 1 mL Buffy Coat separat zentrifugiert (2800 rpm, Microcentrifuge 5417R, 10 min, RT) und Plasma abgenommen. Der Hauptteil des Buffy Coat wurde aus dem Beutel in einen Messzylinder überführt und mit einem Volumenanteil (ca. 80 mL) PBS „ohne“ (ohne Magnesium und Kalium) durchmischt. In vier bis sechs 50 mL Falcon-Röhrchen wurden je 15 mL Ficoll Paque Plus (RT) vorgelegt und vorsichtig mit 30 mL Buffy Coat-PBS-Gemisch überschichtet.

Durch Zentrifugation (45 min, 1600 rpm, Heraeus Multifuge X3 FR, ohne Bremse, RT) reichern sich die PBMC oberhalb des Ficoll-Kissens an.

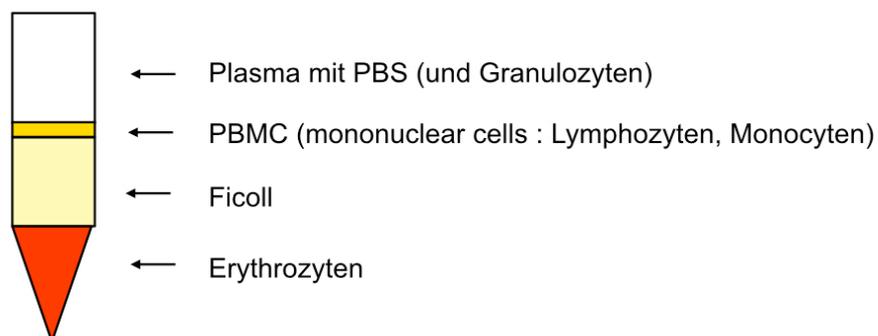


Abb. 2 Ausbildung des Ficoll-Gradienten nach der Zentrifugation

Phase 1- Plasma, Phase 2-PBMC, Phase 3-Ficoll-Kissen, Phase 4-Erythrozyten

Anschließend wurde der weißliche PBMC-Ring (Abb. 2) vorsichtig mit einer sterilen Pasteurpipette entnommen und in ein neues Falcon-Röhrchen überführt. Es folgte

ein Waschschrift, zur Entfernung des Ficoll, bei dem auf 50 mL mit PBS „ohne“ mit 3 %-igen FKS aufgefüllt und nach Zentrifugation (10 min, 1600 rpm, Heraeus Multifuge X3 FR, mit Bremse, RT) der Überstand vorsichtig abpipettiert wurde. Das Pellet wurde erneut in wenig PBS „ohne“ mit 3 %-igem FKS resuspendiert, anschließend auf 50 mL aufgefüllt und abzentrifugiert (10 min, Heraeus Multifuge X3 FR, 1200 rpm, RT). Ist nach Abnahme des Überstandes eine rötliche Färbung des Pellets durch Erythrozyten zu erkennen, wurde eine Erythrozytenlyse mit RCLB (**R**ed **C**ell **L**ysis **B**uffer) durchgeführt. Dazu wird das Pellet mit 5-10 mL RCLB maximal für 10 Minuten bei 4 °C inkubiert, anschließend mit PBS „ohne“ mit 3 %-igem FKS auf 50 mL aufgefüllt, abzentrifugiert (10 min, Heraeus Multifuge X3 FR, 1200 rpm) und ein weiteres mal mit PBS „ohne“ mit 3 %-igem FKS gewaschen. Nach Abnahme des Überstandes wurde das Pellet in 10 mL RPMI mit 10 % FKS, 2 mM Glutamin und Antibiotika (Penicillin/Streptomycin, 100 u/mL) resuspendiert. Anschließend wurden die Zellen in einer Neubauer Zählkammer gezählt und auf 4×10^6 Zellen/mL je Zellkulturflasche eingestellt. Die Berechnung der Zellzahl nach Neubauer ergibt sich aus folgender Formel:

$$\text{Berechnung der Zellzahl} = \frac{\text{Mittelwert der Gesamtzellzahl}}{\text{Verdünnung} \left(\frac{1}{100} \right) \cdot \text{Kammerhöhe} \left(\frac{1}{10} \right)} \cdot 10000 = \frac{\text{Zellzahl}}{\text{mL}}$$

Für eine DNA-Isolierung wurden 2×10^6 Zellen abzentrifugiert und das Zellpellet trocken bei -20 °C eingefroren. Durch Zugabe von Phytohämagglutinin (PHA, 5 µg/mL) in die Zellkulturflaschen wurden die Zellen für 48 h bei 37 °C mitogen stimuliert. Nicht benötigte Zellen werden zu 1 mL Aliquots á $2-3 \times 10^7$ Zellen in Freeze-Mix (siehe 2.1.3) aufgenommen und im Mr. Frosty kontrolliert bei -80 °C eingefroren und gelagert.

2.2.2 Virusanzucht aus Plasma

Die Anzuchtversuche aus Plasma wurden wegen des begrenzten Primärmaterials in 24-Lochplatten begonnen. Wenn Virus sich vermehrt, werden die Zellen unter Zufütterung von weiteren Spender-PBMCs in 75 ml Zellkulturflaschen transferiert. Von jedem Spender wurden in 3 Kavitäten einer 24-Lochplatte 4×10^6 Zellen ausgesät, wobei zwei Kavitäten für den Anzuchtversuch dienten und die dritte als

Negativkontrolle verwendet wurde. Während der Virusvorbereitung wurden die 24-Lochplatten mit den Zellen im 37 °C Brutschrank inkubiert.

Für die Virusvorbereitung wurden die Plasmaproben 450 µL (14-0875) und 480 µL (13-6001) auf Eis aufgetaut und anschließend zur Entfernung von Kryopräzipitaten zentrifugiert (10 min, Microcentrifuge 5417R, 3000 g, 4 °C). Der Überstand wurde in Zentrifugenröhrchen überführt und mit RPMI mit 2 mM Glutamin und 10 % FKS aufgefüllt, sodass das Röhrchen bis 0,5 cm unter dem Rand befüllt war. Die Viren wurden durch Ultrazentrifugation pelletiert (60 min, Optima XL-90 Ultracentrifuge, 25.000 g, 4 °C, ohne Bremse), um eine Anreicherung der Viruspartikel zu erzielen. Nachdem der Überstand bis auf 100 µL abgenommen wurde, erfolgte die Zugabe von 400 µL RPMI⁺. Nach 60 Minuten Inkubation bei 4 °C mit gelegentlichem Mischen wurden weitere 500 µL RPMI⁺ hinzupipettiert und das gesamte Volumen vorsichtig gemischt. Zwei Kavitäten der 24-Lochplatte, die der Virusanzucht dienen, wurden mit 500 µL Virussuspension versetzt und anschließend bei 37 °C für 12-16 h inkubiert.

Alle drei bis vier Tage erfolge ein Mediumwechsel. Am Tag 6 nach der Infektion (p.i.) wurde 1 mL aus einer beimpften Kavität mit Zellen der 24-Lochplatte in eine neue Kavität mit 1×10^6 Zellen überführt. Am Tag 11 p.i. wurden der Inhalt der Platte vereinigt, alles bis auf 1 mL in eine 25 cm² Zellkulturflasche überführt und frisches Medium hinzugegeben. 4 Tage später wurden frische Zellen zu dieser Flasche hinzugefüttert (4×10^6). Die Kavitäten der 24-Lochplatte sowie die potentiell infizierte Zellkulturflasche wurde am Tag 22 p.i. geteilt und mit 2×10^6 Zellen (Platte) bzw. 6×10^6 Zellen (Flasche) gefüttert. Am Tag 27 p.i. wurden die Zellen der Platte sowie die der infizierten Flasche vereinigt und mit neuen frisch aufgetauten 1×10^7 Zellen gefüttert. An den Tagen 0, 4, 6, 8, 11, 15, 19, 21, 22, 26 und 29 p.i. wurde Zellkulturüberstand entnommen und im p24-Antigentest auf Virusreplikation überprüft.

2.2.3 Quantitativer p24-Antigen-Capture-ELISA

Der p24-Antigen-Test wird verwendet um, das Wachstum von HIV in Zellkultur zu überprüfen und zu quantifizieren.

Beim quantitativen p24-Antigen-Capture-ELISA handelt es sich um ein inhouse-Protokoll mit kommerziellen Bestandteilen der Firma Aalto. Das Testprinzip beruht auf einen Sandwich-ELISA, bei dem das p24-Antigen des HIV an immobilisierte Schaf- α -HIV-1-p24-Antikörper bindet und über einen, mit alkalischer Phosphatase

markierten, Antikörper detektiert wird. Dies geschieht durch Zugabe des Substrates und Messung der Lumineszenz, die bei Substratumsatz entsteht.

96-Lochplatten wurden mit dem polyklonalen Schaf- α -HIV-1-p24-Antikörper beschichtet. Dieser wurde dazu nach Herstellerangaben gelöst und anschließend in 400 mL 0,1 M NaHCO₃ (pH=8,5) verdünnt. Nach 2-3 Minuten Rühren wurden 100 μ L der Antikörper-lösung in jede Kavität pipettiert, die Platten mit Einzelblattfolie abgeklebt, in Alufolien gegen Verdunstung eingewickelt und über Nacht bei Raumtemperatur inkubiert. Am Folgetag wurden die Platten zweimal mit 200 μ L 1xTBS (25 mM Tris-HCl, 144 mM NaCl, pH 7,5) pro Kavität gewaschen und anschließend mit 100 μ L 2 %-iges BSA in 1xTBS geblockt. Nach Abkleben der Platten mit Einzelblattfolie wurden diese bis zum Gebrauch bei -20 °C gelagert.

Die Abnahme der Überstände aus der Virusanzucht erfolgte an den Tagen 0, 4, 6, 8, 11, 15, 19, 21, 22, 26 und 29. Diese wurden inaktiviert (50 μ L 5 %-iges Empigen/ 200 μ L Überstand) und bei -20 °C gelagert.

Für die Testdurchführung wurde der Standard rekombinantes HIV-p24gag in 2,5 mL 1xTBS mit 1% FKS gelöst (40 μ g/mL), auf 4 μ g/mL und 80 ng/mL verdünnt und aliquotiert. Aus einem 80 ng/mL Aliquot wurden je Testdurchführung zwei unabhängig voneinander hergestellte Standardreihen 0,04-40 ng/mL in RPMI mit 10 % FKS und 2 mM Glutamin pipettiert. Nach Auftauen und zweimaligem Waschen der Platte mit je 200 μ L 1xTBS wurden je 125 μ L des Standards (0,078-10 ng/mL), der Proben sowie Medium (RPMI mit 10 % FKS und 2 mM Glutamin) als Negativkontrolle in die Platte pipettiert und für 3 Stunden abgeklebt und bei Raumtemperatur inkubiert. Anschließend wurde die Platte zweimal mit 200 μ L 1xTBS gewaschen.

Der monoklonale Maus- α -HIV-1-p24-Antikörper, gekoppelt an alkalischer Phosphatase, wurde in Konjugat-Verdünnungslösung (1xTBS, 20 % Schafserum, 0,05 % Tween) 1:8 vorverdünnt und 1:8000 final verdünnt in den ELISA eingesetzt. Je Kavität wurden 100 μ L der Konjugat-Lösung hinzupipettiert. Nach 60 minütiger Inkubation wurde die Platte viermal mit 200 μ L 0,1 % Tween in PBS und anschließend zweimal mit 200 μ L 1xTROPIX Wash Puffer gewaschen und ausgeklopft. Nun wurden 50 μ L des Substrates CSPD mit Sapphire Substrat, 1:4 in 1xTROPIX Wash Puffer verdünnt, hinzugegeben und für mindestens 50 Minuten inkubiert.

Die Lumineszenz wurde durch die Abspaltung von Phosphat durch die alkalische Phosphatase hervorgerufen und mit dem Tecan Reader und der dazugehörigen Software Magellan gemessen. Die Berechnung der p24-Konzentration erfolgte, nach

Normierung der Werte über die Negativkontrolle, anhand der Regressionsgeraden aus dem Konzentrationsstandard mit Hilfe der TREND-Funktion in Microsoft Excel.

2.2.4 RNA-Extraktion

Zur Extraktion viraler RNA wurde das QIAamp Viral RNA Mini Kit der Firma Qiagen verwendet. Als Positivkontrolle diente der Laborstamm HTLVIII B (1×10^4 Kopien/mL) und HIV-negatives Plasma als Negativkontrolle.

Zur Entfernung von Kryopräzipitat wurden 500 μ L eines Plasma-Aliquots bei $3.000 \times g$ 10 Minuten lang bei $4 \text{ }^\circ\text{C}$ zentrifugiert. Nach Abnahme von 450 μ L Überstand und Über-führung in ein neues Reaktionsgefäß wurden die Viren durch Zentrifugation bei $20.800 \times g$ für weitere 90 Minuten (bei $4 \text{ }^\circ\text{C}$) pelletiert. Anschließend wurden 315 μ L Überstand entnommen, verworfen und 5 μ L FCV (interner Standard) hinzu pipettiert, um bei Bedarf eine Viruslastbestimmung durchzuführen. Zu dem verbleibenden Viruspellet mit 140 μ L Überstand wurde 560 μ L AVL-Lysepuffer mit t-RNA hinzupipettiert. Nach einer Inkubationszeit von 10 Minuten erfolgte die Zugabe von 560 μ L 99,8% Ethanol ($4 \text{ }^\circ\text{C}$). Die Proben wurden, nach kurzer Abzentrifugation, anschließend auf die Mikrospinsäule pipettiert. Die folgenden Schritte wurden nach dem Herstellerprotokoll durchgeführt. Die RNA wurde in 60 μ L AVE-Puffer eluiert und anschließend sofort gekühlt ($0-4 \text{ }^\circ\text{C}$). Die Aliquots wurden zu je 10 μ L in vorgekühlte Reaktionsgefäße (0,5 mL) portioniert und mit 1 Tropfen Mineralöl überschichtet. Die RNA wird bei $-70 \text{ }^\circ\text{C}$ für max. sechs Monate gelagert. Während der RNA-Extraktion wurde auf eine DNase-Behandlung verzichtet, da diese bei der üblichen genotypischen Resistenztestung nicht vorgesehen ist.

2.2.5 full-length cDNA-Synthese

Für die Synthese von cDNA wurden 5 μ L der extrahierten RNA mit einem Denaturierungs-Mix aus Primer und dNTP versehen und für 5 Minuten bei $65 \text{ }^\circ\text{C}$ inkubiert.

Mix für die Denaturierung:

Reagenz	Konz. Stocklösung	Konz. final	x1
Primer LTR-full-length-as	2 µM	0,2 µM	1 µL
dNTP	2,5 mM	1 mM	4 µL
			Σ 10 µL

Anschließend wurde der gesamte Mix aus der Denaturierung (10 µL) mit einem RT-Mastermix versehen und für 2 h bei 50 °C gefolgt von 5 Minuten bei 85 °C inkubiert.

RT-Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	1x
10x RT Puffer	10 x	1 x	2 µL
MgCl ₂	25 mM	4,76 mM	4 µL
DTT	0,1 M	0,0095 M	2 µL
RNAse OUT	40 U/µL	1,9 U/µL	1 µL
Superscript III	200 U/µL	400 U	2 µL
			11 µL RT-Mix + 10 µL RNA/Primer Mix
			Σ 21 µL

Zur Entfernung des RNA-*Templates* erfolgte im Anschluss die Zugabe von 1 µL RNase H (Superskript® III First-Strand Synthesis System, Invitrogen) je Probe. Der Ansatz wurde 20 min bei 37 °C inkubiert und dann für 15 Minuten bei 70 °C inaktiviert. Die cDNA ist zwischen -30 bis -10 °C zu lagern.

2.2.6 Amplifikation des Komplet-Genoms von HIV-1 - „Vier-Amplikon PCR“

Die Amplifikation des Komplet-Genoms von HIV-1 Isolaten der „URF-*new clade*“ erfolgte in Anlehnung an die Literatur anhand von vier überlappenden Fragmenten⁵⁷. Die „vier-Amplikon PCR“ besteht aus subtypgenerischen PCRs, die in einer vorangegangenen Arbeit unabhängig voneinander etabliert wurden⁵¹. Die verwendeten Primerpaare für die vier Fragmente sind in Tabelle 2 (2.1.10) aufgelistet. In der Abbildung 3 sind die überlappenden Fragmente und die entsprechenden Primerpaare relativ zum Genom von HxB2 grafisch dargestellt.

Für die „vier-Amplikon PCR“ wurden 9 µL der zuvor synthetisierten cDNA mit folgendem Mix versehen. Das verwendete Enzym war das Long PCR Enzyme Mix der Firma Thermo Scientific/Fermentas. Anschließend durchlief der Ansatz das unten aufgeführte PCR Programm.

Mastermix

Reagenz	Konz. Stocklösung	Konz. final	x1
Nuclease free water		ad 41 µL	26,2 µL
10x Long PCR Puffer ohne MgCl ₂	10 x	1 x	5,0 µL
MgCl ₂	25 mM	1,2 mM	2,4 µL
dNTP	2,5 mM	0,2 mM	4,0 µL
Primer sense	25 µM	0,5 µM	1,0 µL
Primer antisense	25 µM	0,5 µM	1,0 µL
DMSO	100 %	2 %	1,0 µL
Long PCR Enzyme Mix	5 U/mL	2 U	0,4 µL
			41 µL Mastermix + 9 µL cDNA
			Σ 50 µL

In Abhängigkeit vom zu amplifizierenden Fragment wurden folgende Primer eingesetzt:

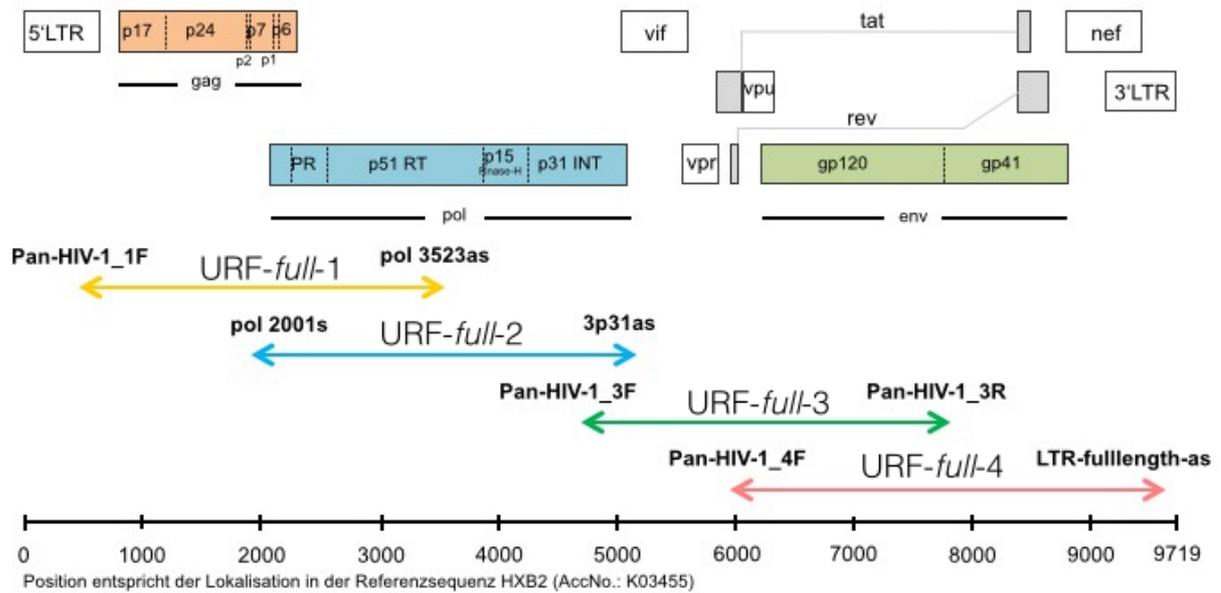


Abb. 3 Grafische Darstellung der vier überlappenden Fragmente der „vier-Amplikon PCR“ anhand des Referenzgenoms HxB2 (Acc.-Nr.: K03455) (Quelle: ⁵² bearbeitet)

Die einzelnen Fragmente sind mit URF-*full* 1-4 benannt und in unterschiedlichen Farben markiert. Die flankierenden Primerpaare entsprechen den Primer für die jeweilige PCR. Die jeweilige Länge der Fragmente beträgt: URF-*full*-1: 3053 Bp, URF-*full*-2: 3062 Bp, URF-*full*-3: 3066 Bp und URF-*full*-4: 3666 Bp

PCR-Programm

Temperatur (°C)	Zeit	Zyklen
96	5 min	1
96	20 sec	
58	30 sec	10
68	4 min	
96	20 sec	
58	30 sec	30
68	4 min (+10 sec/ Zyklus)	
68	10 min	1
4	∞	

2.2.7 Nested-pol-PCR (Inhouse)

Bei einer Veränderung des Endvolumens der *full-length* cDNA-Synthese wurde zur Kontrolle die inhouse-nested-PCR für das *PR/RT*-Fragment von HIV-1 eingesetzt.

Für die inhouse-nested-PCR wurde das Expand High Fidelity PCR System eingesetzt. Dafür wurden 10 µL der cDNA mit 40 µL des entsprechenden PCR-Mastermixes für die äußere PCR versetzt.

Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	x1
Wasser		ad 40 µL	25,7 µL
PCR-Puffer (3)	10 x	1 x	5,0 µL
MgCl ₂	50 mM	3,9 mM	3,3 µL
dNTP	2,5 mM	0,2 mM	4,0 µL
Primer sense (pol2001s)	25 µM	0,25 µM	0,5 µL
Primer antisense (pol3532as)	25 µM	0,25 µM	0,5 µL
DMSO	100 %	1 %	0,5 µL
Polymerase (1)	3,5 U/µL	1,75 U	0,5 µL
			40 µL Mastermix + 10 µL cDNA
			Σ 50 µL

Der Ansatz wurde in folgendem Zyklerprogramm amplifiziert:

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
95	2 min	1
95	20 sec	
64	30 sec	10
72	2 min	
95	20 sec	
64	30 sec	20
72	2 min (+20 sec/ Zyklus)	
72	10 min	1
4	∞	

2,5 µL des erhaltenen Produktes wurden mit dem Mastermix für die innere PCR versetzt.

Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	x1
Wasser		ad 47,5 µL	33,2 µL
PCR-Puffer (3)	10 x	1 x	5,0 µL
MgCl ₂	50 mM	2,5 mM	2,3 µL
dNTP	2,5 mM	0,2 mM	4,0 µL
Primer sense (pol2001s)	25 µM	0,5 µM	1,0 µL
Primer antisense (3454)	25 µM	0,5 µM	1,0 µL
DMSO	100 %	1 %	0,5 µL
Polymerase (1)	3,5 U/µL	1,75 U	0,5 µL
			47,5 µL Mastermix + 2,5 µL DNA
			∑ 50 µL

Anschließend durchlief der Ansatz folgendes Zyklusprogramm.

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
95	2 min	1
95	20 sec	
60	30 sec	10
72	2 min	
95	20 sec	
60	30 sec	12
72	2 min (+20 sec/ Zyklus)	
72	10 min	1
4	∞	

2.2.8 Ein-Amplikon *full-length*-PCR-Varianten

Für die Etablierung einer ein-Amplikon *full-length*-PCR wurden mehrere PCR-Varianten mit dem Long PCR Enzyme Mix der Firma Thermo Scientific/Fermentas durchgeführt.

Variante 1:

9 µL cDNA einer HTLVIIIB-Verdünnungsreihe (10^3 - 10^6) wurden mit folgendem PCR-Mastermix versehen und durchliefen anschließend folgendes PCR-Programm:

Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	x1
Nuclease free water		ad 41 µL	26,6 µL
10x Long PCR Puffer mit 15 mM MgCl ₂	10 x	1 x	5,0 µL
dNTP	2,5 mM	0,2 mM	4,0 µL
Primer sense (PAN-HIV-1_1F)	25 µM	1,0 µM	2,0 µL
Primer antisense (LTR-full-length-as)	25 µM	1,0 µM	2,0 µL
DMSO	100 %	2 %	1,0 µL
Long PCR Enzyme Mix	5 U/mL	2 U	0,4 µL
			41 µL Mastermix + 9 µL cDNA
			Σ 50 µL

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
96	5 min	1
96	20 sec	
58	30 sec	10
68	10 min	
96	20 sec	
58	30 sec	30
68	10 min (+10 sec/ Zyklus)	
68	10 min	1
4	∞	

Bei allen folgenden PCR-Varianten handelte es sich um *semi-nested* und *nested*-PCRs.

Variante 2:

In dieser Variante wurden unterschiedliche Mengen an cDNA (aus HTLVIIIB 10^6 cop/mL) eingesetzt: 1 μ L 1:10 verdünnt, 1 μ L und 5 μ L. Der Mastermix sowie das PCR-Programm stimmen mit Variante 1 überein.

Für die innere PCR wurde der gleiche Mastermix, jedoch mit anderen Primern (URF_full_5'-LTR, URF_full_3'-LTR), verwendet. Es wurden 2 μ L des Produktes der äußeren PCR eingesetzt und das gleiche PCR-Profil, wie in der äußeren PCR, verwendet.

Variante 3:

In dieser Variante wurden unterschiedliche Mengen an cDNA (HTLVIIIB 10^6 cop/mL) eingesetzt: 1 μ L 1:10 verdünnt, 1 μ L und 5 μ L. Die Menge an eingesetzten Enzym wurde auf 1,5 U/Ansatz reduziert, der restliche Mastermix entspricht der Variante 1.

Bei dieser Variante wurde ein doppelter Ansatz pipettiert und zwei unterschiedliche PCR-Programme durchgeführt.

PCR-Programm 1
2006)

PCR-Programm 2 (Rousseau *et al.*

Temperatur (°C)	Zeit	Zyklen	Temperatur (°C)	Zeit	Zyklen
96	5 min	1	94	2 min	1
96	20 sec		94	10 sec	
58	30 sec	10	68	38 sec	10
68	10 min				
96	20 sec		94	10 sec	
58	30 sec	30	68	8,5 min (+20 sec/Zyklus)	20
68	8 min (+10 sec/Zyklus)				
68	10 min	1	68	20 min	1
4	∞		4	∞	

Der Mastermix der inneren PCR variierte bezüglich der eingesetzten Primer (URF_full_5'-LTR, URF_full_3'-LTR), die PCR-Programme waren dieselben, wie in der äußeren PCR. Für die innere PCR wurden je 2 µL der äußeren PCR eingesetzt.

Variante 4:

Diese PCR-Variante wurde mit einem anderen Puffer, zusätzlichem MgCl₂ und mit 9 µL cDNA einer pNL4.3wt-Verdünnungsreihe (10¹-10³, 10⁶) durchgeführt. Weiterhin wurde der Elongations-Schritt in der zweiten Hälfte der PCR verkürzt.

Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	x1
Nuclease free water		ad 41 µL	22,6 µL
10x Long PCR Puffer (ohne MgCl ₂)	10 x	1 x	5,0 µL
MgCl ₂	25 mM	2 mM	4,0 µL
dNTP	2,5 mM	0,2 mM	4,0 µL
Primer sense (Pan-HIV-1_1F)	25 µM	1,0 µM	2,0 µL
Primer antisense (LTR-full-length-as)	25 µM	1,0 µM	2,0 µL
DMSO	100 %	2 %	1,0 µL
Long PCR Enzyme Mix	5 U/mL	2 U	0,4 µL
			41 µL Mastermix + 9 µL Plasmid
			Σ 50 µL

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
96	5 min	1
96	20 sec	
58	30 sec	10
68	10 min	
96	20 sec	
58	30 sec	30
68	8 min (+8 sec/Zyklus)	
68	10 min	1
4	∞	

Für die innere PCR wurden 2 µL der äußeren PCR eingesetzt, die Primer URF_full_5'-LTR und URF-nef 9140as verwendet und das PCR-Profil der äußeren PCR benutzt.

Variante 5:

Der Mastermix dieser PCR-Variante entspricht dem der Variante 1, beinhaltet jedoch die Primer full_length_727F und full_length_9396R. Im verwendeten PCR-Programm wurde im Annealing-Schritt ein Temperaturgradient eingestellt. Es wurden 9 µL (HTLVIIIB 10⁶) für jeden aufgeführte Annealing-Temperatur eingesetzt.

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
96	5 min	1
96	20 sec	
57,9 / 59,0 / 60,4 / 62,7	30 sec	10
68	7 min	
96	20 sec	
57,9 / 59,0 / 60,4 / 62,7	30 sec	30
68	7 min (+5 sec/Zyklus)	
68	10 min	1
4	∞	

Die innere PCR besitzt, bis auf die eingesetzten Primer (hier full_length_782F und full_length_9372R), den gleichen Mastermix. Ebenfalls wurde im PCR-Programm ein Temperaturgradient eingeführt (T=47,8 / 50,5 / 54,5 / 57,9). Das restliche Programm unterscheidet sich nicht von dem der äußeren PCR. Jedes Produkt der unterschiedlichen Annealing-Temperaturen der äußeren PCR (2 µL) durchlief bei jeder Annealing-Temperatur der inneren PCR das PCR-Programm.

Variante 6:

In die Variante Nr. 6 wurden ebenfalls 9 µL (HTLVIIIB 10⁶) der cDNA eingesetzt. Des Weiteren besteht diese Variante aus dem gleichen Mastermix wie Variante 5. Der einzige Unterschied ist im Temperaturgradient des Annealing-Schrittes zu finden. In der äußeren PCR handelt es sich um die Temperaturen: 63,7 / 67,2 / 70,1 / 73,1 °C und in der inneren PCR um 55,1 / 58,6 / 62,1 / 65,4 / 68,0 °C. Weiterhin wurde jede Annealing-Temperatur der äußeren PCR (2 µL) gegen jede Annealing-Temperatur der inneren PCR getestet.

Variante 7:

Es wurden 9 µL cDNA (HTLVIIIB 10⁴ cop/mL und pNL.4.3wt 10⁶ cop/mL) in die äußeren PCR eingesetzt und der Mastermix dieser PCR entsprach dem gleichen wie in der Variante 1. Die Primer URF_full_5'-LTR_neu und LTR-full-length-as wurden hier für den Mix der äußeren PCR verwendet. Das PCR-Programm unterscheidet sich nur in den verwendeten Annealing-Temperaturen (55,4 / 58,5 / 61,6 / 64,0 °C) von Variante 5.

Für die innere PCR wurden 2 µL der äußeren PCR und folgende Primer eingesetzt, URF_full_5'-LTR_neu und URF_full_3'-LTR. Der restliche Mastermix entspricht dem der äußeren PCR. Das PCR-Programm der inneren PCR unterscheidet sich nicht von dem der äußeren PCR. Auch hier erfolge eine Testung jeder Annealing-Temperatur der äußeren PCR gegen jede der inneren PCR.

Variante 8:

Der Mastermix dieser Variante entspricht dem der Variante 1 mit folgenden Primern: URF_full_5'-LTR_neu und URF-nef 9104as. Das eingesetzte PCR-Programm ist das gleiche wie in Variante 5, jedoch mit veränderten Annealing-Temperaturen (60,1 / 64,0 / 67,9 / 69,9 °C).

Für die innere PCR wurden die Primer URF_full_5'-LTR_neu und URF_full_3'-LTR eingesetzt, bei sonst gleichem Mastermix, wie Variante 1. Es wurden 2 µL der äußeren PCR bei folgenden Annealing-Temperaturen verwendet, 58,1 / 61,3 / 64,5 / 67,6 / 70,0 °C.

Die folgenden drei PCR-Varianten wurden mit einem anderen Enzym durchgeführt, es handelt sich hierbei um das Expand Long Template PCR System der Firma Roche.

Variante 9

Es wurden 9 µL (HTLVIIIB 10⁶ cop/mL) mit folgendem Mastermix versehen und bei dem aufgeführten PCR-Programm eingesetzt.

Mastermix:

Reagenz	Konz. Stocklösung	Konz. final	x1
Nuclease free water		ad 41 µL	26,05 µL
10x PCR Puffer mit 17,5 mM MgCl ₂ (1)	10 x	1 x	5,00 µL
dNTP	2,5 mM	0,35 mM	7,00 µL
Primer sense (Pan-HIV-1_1F)	25 µM	0,3 µM	0,60 µL
Primer antisense (LTR-full-length-as)	25 µM	0,3 µM	0,60 µL
DMSO	100 %	2 %	1,00 µL
Long PCR Enzyme Mix	5 U/mL	3,75 U	0,75 µL
			41 µL Mastermix +9 µL cDNA
			Σ 50 µL

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
94	5 min	1
95	20 sec	
58	30 sec	10
68	7 min	
94	20 sec	
58	30 sec	25
68	8 min (+20 sec/ Zyklus)	
68	7 min	1
4	∞	

Der Mix der inneren PCR unterschied sich nur in den verwendeten Primern, als sense Primer wurde URF_full_5'-LTR_neu und als antisense wurden URF_full_3'-LTR oder URF-nef 9104as, in einem parallelen Ansatz, bei gleichem Zyklusprogramm eingesetzt.

Variante 10:

Diese Variante war der Veröffentlichung von Rousseau *et al.* (2006)⁵⁸ nachempfunden. Der Mastermix und das PCR-Programm entsprechen der Veröffentlichung, außer dass keine „wax beads“ eingesetzt wurden. In der äußeren PCR wurden 1 / 5 / 9 µL cDNA (HTLVIIIB 10⁶ cop/mL) und in der inneren 1 µL der äußeren PCR eingesetzt.

Der gleiche Ansatz wurde zusätzlich mit den eigenen Primern Pan_HIV-1_1F und URF-nef 9104as für die äußere und mit Primern URF_full_5'-LTR_neu und URF_full_3'-LTR und, in einem parallelen Ansatz, URF-nef 9104as für die innere PCR durchgeführt. Hier wurden 9 µL cDNA in die äußeren und in die innere PCR eingesetzt.

Variante 11:

Um die PCR von Rousseau *et al.* komplett nachzugenerieren, wurde das PCR-Programm so verändert, dass keine „wax beads“ benötigt wurden. Es wurden 2 Mixe hergestellt, wobei Mix 1 sich aus Wasser, 10x PCR Buffer mit 17,5 mM MgCl₂ (1), dNTPs, Primer sowie 1 µL / 3 µL / 9 µL cDNA (HTLVIIIB) zusammensetzte. Zu diesem Mix wurde bei einem zusätzlichen Temperaturschritt (siehe PCR-Programm) ein zweiter Mix aus Wasser, Enzym und Puffer hinzupipettiert.

PCR-Programm (Abweichung zu Rousseau *et al.* sind grau)

Temperatur (°C)	Zeit	Zyklen
94	2 min	1
68	1 min	1
94	10 sec	10
68	38 sec	
94	10 sec	
68	8,5 min (+20 sec/ Zyklus)	20
68	20 min	1
4	∞	

2.2.9 Agarose-Gelelektrophorese zur qualitativen und quantitativen DNA-Analyse

Die Produkte der PCR wurden mittels qualitativer Gelelektrophorese auf einem Ethidiumbromid (0,5 µg/mL) gefärbtem Agarosegel dargestellt.

Das portionierte Agarosegel (25 mL, 0,6% für *full-length* Produkte, 1,0% für vier-Amplikon Produkte) wurde in einer Mikrowelle aufgelöst, in eine horizontale Elektrophoresekammer gegossen und vor der Polymerisierung mit einem bzw. zwei Kämmen versehen. Die Größe des PCR-Produktes wurde mit Hilfe des mitgeführten Größenstandards (Lambda DNA/HindIII Maker und Gene Ruler™ 1kB DNA-Ladder, 100 ng/µL) ermittelt.

Vom PCR-Ansatz wurden 3 µL mit 6x Ladepuffer versetzt und 5 µL auf das Gel aufgetragen. Für das qualitative Gel wurden 3 µL des Größenstandards aufgetragen. Nach der Produktaufreinigung (Methode 2.2.11) wurde für das quantitative Gel 2 µL Probe mit 6x Ladepuffer versetzt und anschließend ebenfalls 5 µL auf das Gel aufgetragen. Als Konzentrations- und Größenstandard wurde der Low Mass Ladder eingesetzt. Nach Auftrennung der PCR-Fragmente für 90 Minuten bei 70 Volt wurde die DNA anhand einer Eichkurve des Konzentrationsstandards quantifiziert. Die Auswertung der qualitativen als auch der quantitativen Gele erfolgte mit dem System E.A.S.Y RH-3 und der dazugehörigen Software EASY Win32-System.

2.2.10 Präparation von DNA-Fragmenten aus Agarose-Gel

Diese Methode wurde angewandt, um definierte DNA-Fragmente aus einem Gemisch unterschiedlich großer Fragmente zu isolieren und aufzureinigen. Nach gelelektrophoretischer Auftrennung der DNA-Fragmente (47 µL DNA+ 9 µL 6x Ladepuffer) wurde das Zielfragment unter UV-Bestrahlung ausgeschnitten, gewogen und nach dem Protokoll „GeneJET™ Gel Extraction Kit“ aufgereinigt. Das eluierte PCR-Produkt (50 µL) wurde bei 4 °C gelagert.

2.2.11 Aufreinigung von DNA

Die Aufreinigung der PCR-Fragmente erfolgte nach dem qualitativen Gel mit dem MSB Spin PCRapace Kit der Firma Invitex. Das restliche Volumen (ca. 47 µL) der PCR-Reaktion wurde nach dem Herstellerprotokoll aufgereinigt. Die DNA wurde

anschließend mit 50 µL Elutionspuffer in einem neuen Gefäß eluiert. Die aufgereinigte DNA wurde mittels Gelelektrophorese quantifiziert und bei 4 °C gelagert.

2.2.12 Sanger-Sequenzierung

Für die Sequenzierung wurde ein 10 µL Ansatz aus ca. 40 ng Probe, 2 µL BigDye 3.1 und 10 µM Primer eingesetzt.

Die Ansätze durchliefen anschließend folgendes Zyklusprogramm und wurden danach ins Sequenzierungslabor gegeben.

PCR-Programm:

Temperatur (°C)	Zeit	Zyklen
96	2 min	1
96	10 sec	
55	5 sec	25
60	4 min	
4	∞	

Die Sequenzierungsprodukte wurden im hauseigenen Sequenzierungslabor gereinigt und in der Kapillarelektrophorese dargestellt. Die Elektropherogramme wurden dem Anwender als ABI-Files vom Sequenzierungslabor zur Auswertung übergeben.

Die Auswertung der Elektropherogramme erfolgte mit Hilfe des Programms Lasergene SeqMan Pro Version 10.0.1. Es wurde eine Contiq-Sequenz im fasta-Format erzeugt.

2.2.13 Next Generation Sequencing

Die Amplikons werden mittels Transposomen fragmentiert und markiert, wodurch sogenannte DNA-Bibliotheken generiert werden. Diese DNA-Fragmente binden über Adapter an eine Fließzelle und werden dort klonal zu Clustern amplifiziert (*solid phase bridge amplification*). Die Sequenzierung erfolgt über fluoreszenzmarkierte Nukleotide mit reversiblen Terminatoren (*sequencing by synthesis*). Mit dem Illumina

MiSeq können 300 Bp in beide Orientierungen (*paired-end* Modus) sequenziert werden und es werden pro Lauf bis zu 15 Gigabasen und 25 Millionen Sequenzen erzeugt.

Alle Schritte wurden gemäß den jeweiligen Herstellerprotokollen durchgeführt. Die PCR-Produkte wurden in 1x TE (Promega) auf 5 ng/μL vorverdünnt und danach auf die finale Konzentration von 0,2 ng/μL eingestellt. Von jeder Probe wurden 5 μL in die Probenpräparation mit dem Nextera® XT DNA Sample Preparation Kit (Illumina) eingesetzt. Zuerst erfolgte die *tagmentation*, eine Transposomen-basierte parallele Fragmentierung und Markierung der Amplikons. Durch eine nachfolgende 12-Zyklen PCR wurden weitere Markierungen an die DNA-Fragmente angefügt. Nach diesen zwei Schritten lagen DNA-Fragmente vor, die beidseitig von Sequenzierprimern, Indices und Adaptoren flankiert waren und als DNA-Bibliotheken bezeichnet werden. Die DNA-Bibliotheken wurden mit einem „Solid Phase Reversible Immobilization“-Verfahren (AMPure XP Beads, Agilent) aufgereinigt, um Salze, dNTPs und Primer aus der vorangegangenen PCR zu entfernen. Durch die Zugabe der 0,5-fachen Menge an AMPure XP Beads erfolgte gleichzeitig eine Größenselektion, um Primerdimere und zu kurze DNA-Fragmente zu entfernen. In einem Normalisierungsschritt wurden die DNA-Bibliotheken auf vergleichbare Konzentrationen gebracht und die DNA-Fragmente mit 0,1 M NaOH einzelsträngig gemacht. Äquivalente Volumina der normalisierten DNA-Bibliotheken wurden zu einem Probenpool vereinigt. In einer qPCR wurden drei Verdünnungsstufen (1:25, 1:50, 1:100) des Probenpools im Vergleich zu einem Standard mit bekannter Größe und Konzentration gemessen (Kapa Library Quantification Kit for Illumina, PeqLab). Gleichzeitig wurden Aliquots jeder gereinigten DNA-Bibliothek vereinigt und die Fragmentlängenverteilung im Probenpool mit einem High Sensitivity DNA Analysis Chip (Agilent) auf dem Bioanalyzer 2100 (Agilent) untersucht. Aus der mittleren Fragmentlänge und dem CT-Wert wurde die Konzentration des Probenpools anhand der Standardkurve errechnet. Der Probenpool wurde verdünnt und mit weiteren Proben auf dem Illumina MiSeq sequenziert. Die erzeugten Daten wurden im fastq-Format ausgegeben und unter einem Linux-Betriebssystem weiter verarbeitet. Zuerst wurde die Qualität der Sequenzen mit dem Programm FastQC (Version 0.11.2, Simon Andrews) geprüft. Im Anschluss wurden die Sequenzen mit den Programmen Trimmomatic⁵⁶ und Flash⁵⁴ vorprozessiert. Mit Trimmomatic erfolgte ein Adapter-Clipping, da bei DNA-Fragmenten unter 300 Basen in den Illumina Adapter hineinsequenziert wird, sowie eine Entfernung der ersten 15 Basen der Sequenz, da diese noch nicht

vertrauenswürdig sind. Mit Flash wurden paarige Sequenzen, die zwischen 15 und 300 Basen überlappten, zu längeren Sequenzen fusioniert. Unter erneuter Verwendung von Trimmomatic⁵⁶ wurden die Sequenzen anhand der Basenqualität ($\geq Q30$) getrimmt und zu kurze Sequenzen (< 50 Basen) eliminiert. Die vorprozessierten Sequenzen wurden dann mit dem Programm Segminator II⁵⁵ gegen eine Referenzsequenz kartiert. Segminator II ermittelt die Häufigkeiten der kartierten Basen an jeder Position bezogen auf die Referenzsequenz und gibt diese in einem Bericht aus. Ein speziell in der Arbeitsgruppe entwickeltes Python-Skript (Britta Altmann) ermöglichte das Auslesen der Populationssequenzen mit einem variierbaren Grenzwert für die Berücksichtigung von Basenmischungen (Ambiguitäten) im fasta-Format. Des Weiteren gibt Segminator II einen Bericht zu den detektierten Insertionen und Deletionen im Vergleich zur Referenzsequenz aus.

2.2.14 Phylogenetische Analysen

Mit Hilfe der Phylogenie (Stammesgeschichte) lassen sich Verwandtschaftsbeziehungen zwischen Populationen oder Individuen, bei denen man einen gemeinsamen Vorfahren vermutet, feststellen. Geschieht dies mit Hilfe von molekularen Merkmalen (z. B. DNA- oder Protein-Sequenzen), wird die Analyse als molekulare Phylogenie bezeichnet⁵⁹.

Eine häufig eingesetzte Methode ist das *Neighbor-Joining* Verfahren. Dieser Algorithmus ist ein Distanz-basiertes Verfahren, bei dem ein Alignment, nach zu wählendem Evolutionsmodell (Nukleotidsubstitutionsmodell z. B. Kimura-2-Parameter), in eine Matrix aller paarweisen genetischen Distanzen zwischen den Sequenzen umgewandelt wird⁵⁹. Anhand der berechneten Distanzmatrix wird dann im *Neighbor-Joining* Algorithmus die Verzweigungsstruktur des Stammbaumes berechnet.

Dafür erfolgt zunächst die Alignierung der eingeladenen Sequenzen (fasta-Format) mit Hilfe der Software ClustalW (BioEdit). Alle im Alignment enthaltenen Sequenzen werden anschließend auf die gleiche Länge editiert. PHYLIP erstellt über das Tool Dnadist (Kimura 2-Parameter) eine Distanzmatrix, in der alle Positionen im Alignment paarweise verglichen werden. Basierend auf den Distanzen in dieser Matrix wird mit der *Neighbor-Joining* Methode die Verwandtschaft von Sequenzen in einer phylogenetischen Baumstruktur errechnet. Die Knoten repräsentieren den letzten gemeinsamen Vorfahren von zwei sich von diesem Vorfahren abzweigenden

Sequenzen. Die Länge der Äste ist ein Maß für die, im Laufe der Evolution stattgefunden Mutationen (% Nukleotidaustausche pro Position im Alignment). Die Darstellung des Baumes als Phylogramm liefert, im Gegensatz zum Kladogramm, ein Größenmaß für die Anzahl von Mutationen über die Astlänge, d. h. je länger ein Ast, desto mehr Mutationsereignisse haben stattgefunden. Bei einem Stammbaum sind die horizontalen Astlängen proportional zu dem Unterschied zwischen zwei Isolaten. Der Skalenwert von 0,1 steht für den Austausch von 0,1 Nukleotiden pro Position im Alignment, was 10 % Unterschied an dieser Position bedeutet.

Um Aussagen über die statistische Signifikanz einer Baumtopologie zu erhalten, wird zusätzlich die Bootstrap-Analyse durchgeführt. Dafür wird das Alignment aus BioEdit in das Programm Seqboot des PHYLIP-Package eingeladen. Dieses erzeugt Pseudodaten indem es die Positionsspalten im Alignment bezüglich ihrer Anordnung randomisiert und in unterschiedlichen Wiederholung einliest. Als Ergebnis der Bootstrapanalyse erhält man einen Konsensusbaum zusammen mit der Information, wie häufig bestimmte Sequenzen in derselben Klade lokalisiert waren. Dadurch wird der Einfluss einzelner Positionen des Alignments auf die Stabilität der Baumstruktur überprüft. Bei einer Anzahl von 1000 Replikaten werden somit 1000 unterschiedliche Alignments (Pseudodaten) erstellt. Die ermittelten Bootstrap-Werte im Konsensusbaum geben an, mit welcher Wahrscheinlichkeit ein Knoten in der Baumtopologie auftritt. Knoten mit einem Bootstrap-Wert ab 70 % gelten als signifikant, da dann eine Übereinstimmung von 70 % in allen 1000 Replikaten vorliegt, sodass man davon ausgehen kann, dass die beobachtete Clusterbildung signifikant ist⁵⁹.

2.2.15 Analyse der rekombinanten Genomstruktur

Um die Rekombinationsstruktur der „URF-*new clade*“ zu ermitteln, werden die beiden frei verfügbare Programme RIP (**R**ecombinant **I**dentification **P**rogram)⁶⁰ der Los Alamos HIV Datenbank und dem *jumping profile Hidden Markov Model (jpHMM)* auf GOBICS (**G**öttingen **B**ioinformatics **C**ompute **S**erver)⁶¹ eingesetzt. Beide Programme dienen der Identifikation von HIV-1-Rekombinanten anhand eines multiplen Alignments aus der zu analysierenden Zielsequenz („*query*“) und einem HIV-1-Referenzsequenzset verschiedener bekannter Subtypen und Rekombinanten. Das RIP-Tool dient der Identifizierung von Rekombinationen im HIV-Genom. Dafür steht ein „Hintergrund“-Alignment aus 12 unterschiedlichen Subtypsequenzen (A1, A2, B-D, F1, F2, G-H, J-K und CRF01_AE) zur Verfügung, mit dem die zu

untersuchende Sequenz, ausgehend von einer Distanzmessungen zwischen Referenzsequenzen und der Zielsequenz, verglichen wird. RIP lässt ein Fenster über dieses Alignment sowie die zu analysierende Sequenz gleiten. Dieses „gleitende Fenster“ besitzt eine definiert Größe (100-400 Positionen) und läuft das Alignment positionsweise ab, um so die Zielsequenz mit dem „Hintergrund“-Alignment zu vergleichen. Dabei wird zwischen den einzelnen Subtypreferenzsequenzen und der Zielsequenz ein Ähnlichkeits-(*s, similarity*)-Distanzplot erstellt, der die größte Übereinstimmung der einzelnen Genregionen wiedergibt. Die gewählte Größe des „gleitenden Fensters“ bestimmt die Sensitivität der Identifikation von HIV-Rekombinanten. Für jedes Fenster wird die Konfidenz mittels (1) z-Testes, unter der Annahme, dass sich jede Position der Sequenz unabhängig, aber im gleichen Prozess entwickelt hat, und (2) einer Hamming-Distanz (*p*-Distanz) bestimmt. Davon ausgehend wird die größte Übereinstimmung innerhalb eines Fensters in Abhängigkeit eines gewählten Konfizendschwellwertes (90-95 %) ausgegeben⁶⁰.

Das *jpHMM* ist eine alternative Möglichkeit zur Identifizierung von HIV-Rekombinanten. Darüber hinaus ist es mit diesem Algorithmus möglich die phylogenetischen Rekombinationsstellen (*breakpoints*) zu bestimmen. Diese Methode nutzt 309 Referenzsequenzen der HIV-Datenbank und ist im Gegensatz zu RIP (12 Referenz-sequenzen) auch variabler. Die Rekombinationsvorhersage basiert ebenfalls auf einer Ähnlichkeit der Zielsequenz mit einem vorkalkulierten multiplen Sequenzalignment aus HIV-1 Subtypen. Dabei ist jeder Subtyp als HMM-Profil modelliert und neben den normalen Transitionen (Übergänge) an definierten Punkten sind hier weitere Transitionen an jeder Stelle im Alignment möglich. Diese Transitionen werden als „*jumps*“ zwischen den HMM-Profilen, also zwischen den Subtypen, bezeichnet. Demzufolge springt das Model in das HMM Profil desjenigen Subtypen, welcher an der jeweiligen Stelle die größte Übereinstimmung mit der Zielsequenz besitzt. Die Rekombinationsvorhersage für eine Zielsequenz wird dann als höchst wahrscheinlichster Weg, der die Zielsequenz mit dem vorliegenden HMM erzeugt, definiert. Die durchgeführten „*jumps*“ beschreiben die Rekombinationsstellen (*breakpoints*)⁶².

Als Ergebnis für die zu analysierende Sequenz wird die vorhergesagte Rekombinations-struktur mit den dazugehörigen Rekombinationsschnittstellen ausgegeben sowie eine grafische Darstellung der Struktur anhand der Referenzsequenz HxB2 (Acc.-Nr.: K03455) erstellt.

3. Ergebnisse

3.1 Virusisolierung aus peripheren mononukleären Blutzellen

Die Vermehrung der Virusisolate (Methode 2.2.2) wurde auf gesunden Spenderzellen (peripheren mononukleären Blutzellen) durchgeführt, um genügend Ausgangsmaterial für die Etablierung und Optimierung einer „ein-Amplikon *full-length* PCR“ zur Verfügung zu haben. Es wurde sich für 2 Patienten entschieden, bei denen, basierend auf vorherigen phylogenetischen Analysen, eine direkte genetische Verknüpfung unwahrscheinlich war und die eine möglichst hohe Viruslast besitzen. Dabei handelte es sich um die Patienten 13-6001 (7.851 cop/mL) und 14-0875 (272.236 cop/mL). Von beiden Plasmaproben war weniger als 500 µL Volumen vorhanden (450 µL 14-0875, 480 µL 13-6001). Für Virusanzuchten werden üblicherweise 4×10^6 Viren/mL eingesetzt. Der Erfolg der Virusanzucht wurde mit dem quantitativen p24-Antigen-ELISA überprüft. Während der 4 Wochen Virusanzucht konnte kein Anstieg der p24-Konzentration gemessen werden (Abb. 4). Ein Großteil der gemessenen Werte war negativ. Lediglich ab Tag 22 p.i. konnte ein minimaler Anstieg der p24-Konzentration beobachtet werden, der jedoch am Tag 26 nicht mehr messbar war (Abb. 4).

Das Ergebnis der Virusanzucht war somit negativ. Es wurde daher auf gelagerte Zellkulturüberstände eines bereits zuvor angezüchteten Isolates zurückgegriffen.

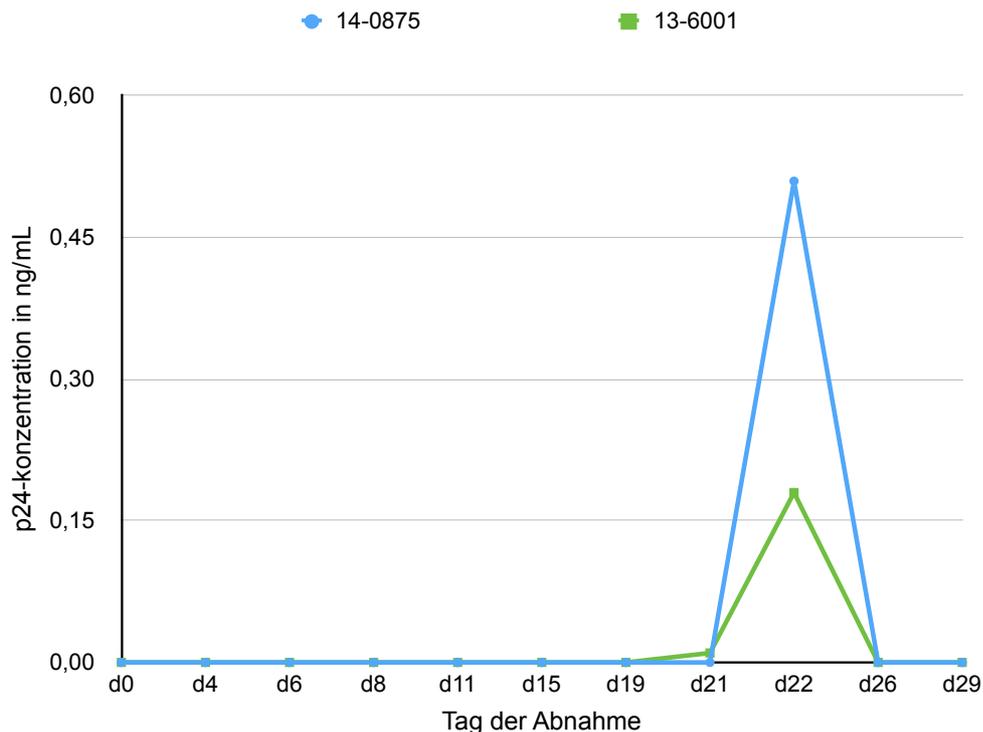


Abb. 4 Anzuchtversuch: p24-Antigen-Konzentration über die Zeit

Die Y-Achse gibt die gemessenen p24-Konzentration und die X-Achse den Tag der Abnahme an. Aufgetragen sind die Messungen für die 24-Lochplatten. Der Konzentrationsverlauf für 14-0875 ist blau und für 13-6001 grün dargestellt.

3.2 Etablierung einer ein-Amplikon HIV-1 *full-length* PCR

Die „vier-Amplikon PCR“ besteht aus 4 überlappenden PCR-Fragmenten. Aufgrund der Variabilität der Quasispezies von HIV ist es daher möglich, dass die erzeugten vier Amplikons von unterschiedlichen RNA-Molekülen stammen können. Da dies die Aussagekraft der Analyse der Rekombinanzstruktur einschränkt, sollte versucht werden, eine *full-length* Amplifikation für HIV-1 ausgehend von der *full-length* cDNA Synthese, die vom 5'-Ende bis zur Polyadenylierungsstelle das komplette Genom abdeckt⁵¹, zu etablieren.

Der *sense* Primer (PAN-HIV-1_1F) des ersten Fragmentes und *antisense* Primer (LTR-full-length-as) des vierten Fragmentes sowie der Long PCR Enzyme Mix wurden beibehalten (Variante 1, Methode 2.2.8). Nach der PCR wurden die Produkte, eine Verdünnungsreihe des Referenzvirus HTLVIII_B, über eine Gelelektrophorese (Methode 2.2.9) sichtbar gemacht (Abb. 5).

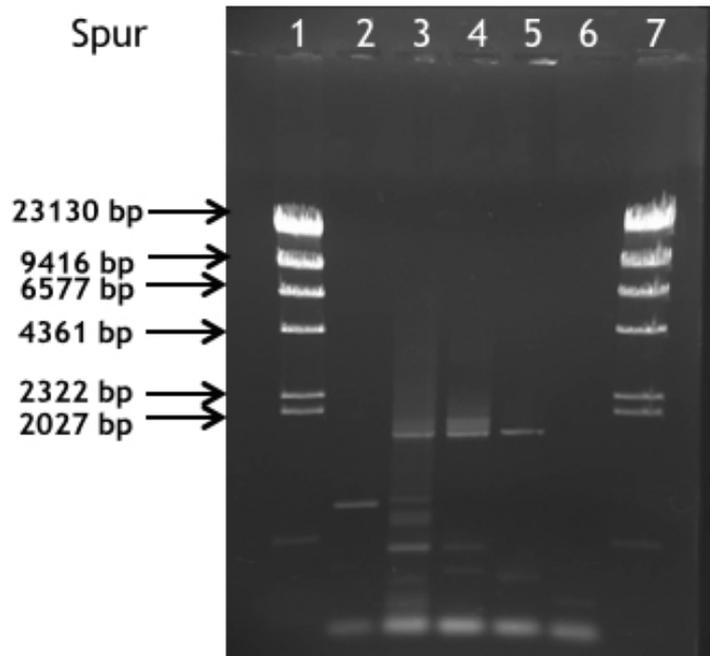


Abb. 5 Qualitative Auswertung der *full-length* PCR Variante 1

Gelelektrophorese zur Darstellung der PCR-Produkte, Agarosegel: 0,6 %, angefärbt mit Ethidiumbromid, Spurbelegung: 1. Lamda DNA/Hind III-Verdau (2,5 µg); 2. HTLVIII B 10³ cop/mL; 3. HTLVIII B 10⁴ cop/mL; 4. HTLVIII B 10⁵ cop/mL; 5. HTLVIII B 10⁶ cop/mL; 6. negative Kontrolle (HIV-1 negatives Blutplasma), 7. Lamda DNA/Hind III-Verdau (2,5 µg)

Ab einer Konzentration von 10⁴ sind schwache Produkte in einer Größe von ca. 2000 Bp zu sehen, aber nicht das erwartete Produkt von ca. 9000 Bp. Es sind weiterhin mehrere kleinere Nebenprodukte zu erkennen.

Daraufhin wurde das PCR-System auf eine *semi-nested* PCR umgestellt, wie es in der Literatur zu finden ist und neue Primer konzipiert (Variante 2). Der Primer URF_full_5'-LTR wurde für beide PCR-Runden (äußere und innere) verwendet. In die äußere PCR wurde zusätzlich der Primer LTR-full-length-as und in die innere PCR der Primer URF_full_3'-LTR eingesetzt. Auch hier konnten nur Produkte mit einer Größe von ca. 2000 Bp amplifiziert werden. Da für die Variante 2 der gleiche Primer in der äußeren PCR benutzt wurde, wie für die *full-length* cDNA-Synthese, wurde dieser gegen einen Primer ausgetauscht, welcher *upstream* vom cDNA-Synthese-Primer liegt, um sicherzugehen, dass dieser auf der cDNA bindet. Weiterhin wurde die eingesetzte Enzymmenge im Rahmen der Herstellerangaben reduziert (Variante 3). Dieser Ansatz wurde mit dem bisher benutzen PCR-Programm sowie mit PCR-Bedingungen aus der Literatur, durchgeführt⁵⁸. Es konnte ebenfalls kein DNA-Produkt amplifiziert werden.

In vorangegangenen Arbeiten zu Etablierung der „vier-Amplikon-PCR“ wurde das Plasmid pNL4.3wt, welches das Komplet-Genom von HIV-1 trägt, verwendet. Durch den Einsatz des DNA-Plasmides kann auf eine vorherige *full-length* cDNA-Synthese verzichtet werden. Des Weiteren können definierte unterschiedliche Molekülmengen in die DNA-PCR eingesetzt werden.

Weiterhin wurde die Menge an eingesetzten Enzym erhöht und damit einhergehend auch zusätzliches MgCl₂ eingesetzt, um ein möglichst großes Fragment zu erhalten. Die Primer der „vier-Amplikon PCR“ wurden für die äußere PCR eingesetzt und für die innere PCR die Primer URF_full_5'-LTR und URF-nef 9140as verwendet. Das PCR-Programm wurde nach Herstellerangaben angepasst (Variante 4).

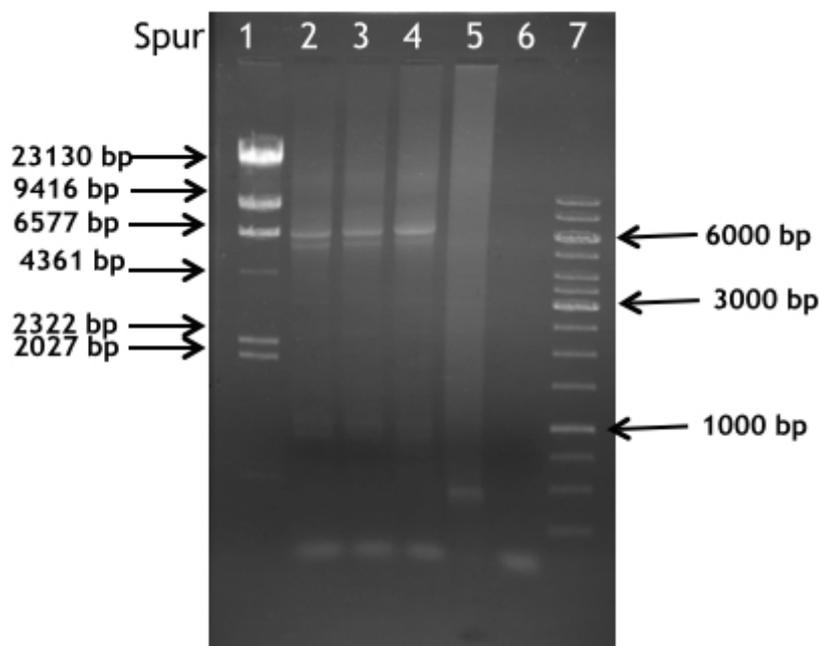


Abb. 6 Qualitative Gelanalyse der *full-length* PCR Variante 4

PCR-Produkte auf 0,6 % Agarosegel dargestellt, angefärbt mit Ethidiumbromid, Spurbelegung: 1. Lamda DNA/Hind III-Verdau (2,5 µg); 2. pNL4.3wt 10¹ cop/mL; 3. pNL4.3wt 10² cop/mL; 4. pNL4.3wt 10³ cop/mL; 5. pNL4.3wt 10⁶ cop/mL; 6. negative Kontrolle (HIV-1 negatives Blutplasma); 7. 1 kb DNA Ladder

Das Gelbild (Abb. 6) zeigt ein Produkt mit einer Größe von ca. 6000 Bp, das als Amplifikation des Vektors erklärt werden kann, da die Primerbindungsstellen der äußeren PCR in beiden *LTR*-Regionen vorliegen (5'-LTR; 3'-LTR). Mit Hilfe der Software „Geneious“ wurden neue Primer konzipiert, welche mit dem Programm „PrimerSelect“ auf ihre Eignung überprüft wurden. Für den Einsatz dieser Primer wurde die Annealing-Temperatur optimiert (58-63 °C, Variante 5, Abb.7).

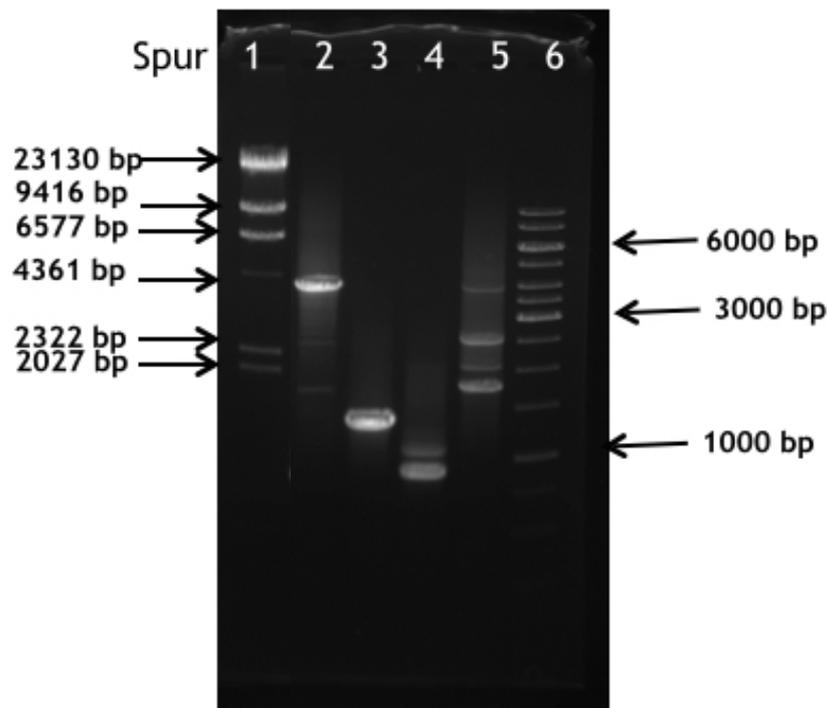


Abb. 7 Ausschnitt der qualitativen Gelelektrophorese der *full-length* PCR bei einer Annealing-Temperatur von 54,5 °C der inneren PCR (Variante 5)

PCR-Produkte auf 0,6 % Agarosegel dargestellt, angefärbt mit Ethidiumbromid, Annealing der äußeren PCR variiert, Spurbelegung: 1. Lambda DNA/Hind III-Verdau (2,5 µg); 2. HTLVIIIB 10⁶ cop/mL (T_{out} 57,9 °C); 3. HTLVIIIB 10⁶ cop/mL (T_{out} 59 °C); 4. HTLVIIIB 10⁶ cop/mL (T_{out} 60,4 °C); 5. HTLVIIIB 10⁶ cop/mL (T_{out} 62,7 °C); 6. 1 kb DNA Ladder, NK auf anderem Gel (negativ)

In Abb. 7 sind Amplifikate unterschiedlicher Größe zu sehen. Mit steigender Annealing-Temperatur konnten längere Produkte erzeugt und weniger zusätzliche Banden beobachtet werden. Die erwartete PCR-Produktgröße von ca. 9000 Bp konnte jedoch auch mit diesen Primern nicht amplifiziert werden. Auch die weitere Erweiterung des Annealing-Temperaturbereiches (55-68 °C) erbrachte nicht das erwartete Produkt (Variante 6).

Für die nächste Variante wurde nochmals der *sense* Primer modifiziert. Der neue Primer (URF_full_5'-LTR_neu) hybridisiert ebenfalls an der konservativen Primerbindungsstelle von HIV und wurde am 5'-Ende um sieben Nukleotide verlängert. Dadurch sollte der *sense* Primer nur noch an der 5'-LTR binden und nicht mehr an die 3'-LTR. Ein Gradient der Annealing-Temperatur wurde eingestellt (52-64 °C, Variante 7, Abb. 8). Als Ergebnis waren für den Ansatz mit pNL4.3wt drei Produkte mit einer Größe von ca. 200 Bp und Artefakte mit unterschiedlichen Größen zu sehen (Abb. 8).

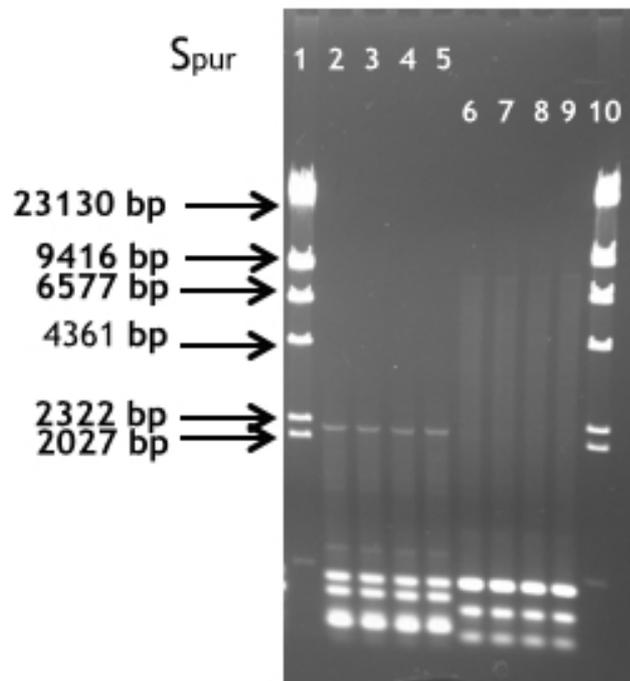


Abb. 8 Ausschnitt der qualitative Gelelektrophorese der *full-length* PCR Variante 7 für pNL4.3wt

PCR-Produkte auf 0,6 % Agarosegel dargestellt, angefärbt mit Ethidiumbromid, Spurbelegung: 1. Lamda DNA/Hind III-Verdau (2,5 µg); 2. pNL4.3 10⁶ cop/mL (T_{out} 61,6 °C, T_{IN} 55,4 °C); 3. pNL4.3 10⁶ cop/mL (T_{out} 61,6 °C, T_{IN} 58,5 °C); 4. pNL4.3 10⁶ cop/mL (T_{out} 61,6 °C, T_{IN} 61,6 °C); 5. pNL4.3 10⁶ cop/mL (T_{out} 61,6 °C, T_{IN} 64 °C); 6. pNL4.3 10⁶ cop/mL (T_{out} 64 °C, T_{IN} 55,4 °C); 7. pNL4.3 10⁶ cop/mL (T_{out} 64 °C, T_{IN} 58,5 °C); 8. pNL4.3 10⁶ cop/mL (T_{out} 64 °C, T_{IN} 61,6 °C); 9. pNL4.3 10⁶ cop/mL (T_{out} 64 °C, T_{IN} 64 °C); 10. Lamda DNA/Hind III-Verdau (2,5 µg)

Daher wurden auch noch die Primer der äußeren PCR variiert, um eine spezifischere Bindung am 3'-Ende zu erlangen (Variante 8). Dafür wurde der Primer URF-nef 9104as in die äußere PCR eingesetzt, da er auf dem *Nef*-ORF und an der 3'-LTR bindet. Die Gelauswertung zeigte Produkte mit einer Größe von ca. 200 Bp.

Die PCR wurde mit einem anderen Enzym durchgeführt, dem Expand Long Template PCR System von der Firma Roche, das in der Literatur zur Amplifikation des HIV-Genoms, über ein Amplikon, verwendet wird⁵⁸. Die PCR-Bedingungen (Mastermix und Zyklusprogramm) wurden nach Herstellerangaben für das Enzym zusammengestellt. Bekannte Primerkombinationen (Variante 9) wurden eingesetzt, zeigten jedoch keinen Erfolg.

Als Nächstes wurde versucht, die Veröffentlichung von Rousseau *et al.* (2006) nach deren Angaben, aber ohne „*wax beads*“, durchzuführen. In parallelen Ansätzen wurden die Primer aus der Publikation und eigene Primer, mit denen bislang die längsten PCR-Produkte erzeugt werden konnten, eingesetzt.

Die Auswertung mittels Gelelektrophorese zeigt schwache Produkte im Bereich von 2000 Bp und kleiner (Abb. 9).

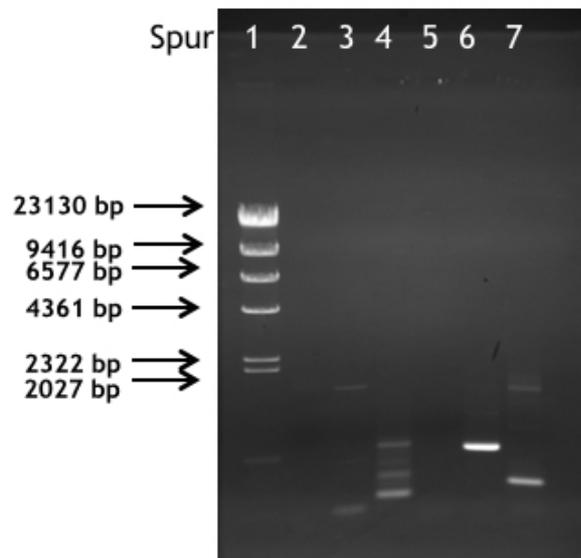


Abb. 9 Ausschnitt der qualitative Gelelektrophorese der *full-length* PCR Variante 10

PCR-Produkte auf 0,6 % Agarosegel dargestellt, angefärbt mit Ethidiumbromid, Spurbelegung: 1. Lamda DNA/Hind III-Verdau (2,5 µg); 2. HTLVIIIB 10⁶ cop/mL 1 µL cDNA in PCR, 3. HTLVIIIB 10⁶ cop/mL 5 µL cDNA in PCR, 4. HTLVIIIB 10⁶ cop/mL 9 µL cDNA in PCR, 5. negative Kontrolle (HIV-1 negatives Blutplasma), 6. HTLVIIIB 10⁶ cop/mL eigene Primer

Um die *“hot start PCR“* auch ohne *„wax beads“* durchzuführen, wurden das PCR-Programm und der Mastermix modifiziert, um eine vorzeitige Exonukleaseaktivität der Polymerase zu verhindern (Variante 11). Um sicherzustellen, dass das Enzym erst bei der spezifischen Annealing-Temperatur zum *Template* gelangt, was durch die *„wax beads“* gewährleistet wird, wurde der Enzym-Mix erst während des PCR-Programms hinzupipettiert. Dazu wurde der Zyklus bei entsprechendem Schritt angehalten. Auch diese Variante zeigt nur kleine Produktbanden auf dem Gel, ähnlich wie die der Variante 10 (Abb. 9).

3.2 Generierung von *full-length* Amplikons mittels *„vier-Amplikon PCR“*

Da mit den ausgetesteten Varianten zur Etablierung einer ein-Amplikon *full-length* PCR keine PCR-Produkte der erwarteten Größe erzeugt wurden, wurden weiterhin PCR-Produkte der *„viel-Amplikon PCR“* verwendet (Methode 2.2.6), um die

Komplett-Genome der erforderlichen drei HIV-1 Isolate der „URF-*calde new*“ zu amplifizieren und zu sequenzieren. Mit Hilfe von dieser PCR wurden vier überlappende Fragmente amplifiziert, im Anschluss aufgereinigt, quantifiziert und sequenziert.

Die vorangestellte *full-length* cDNA-Synthese wurde gegebenenfalls mit Hilfe der inhouse pol PCR überprüft. Dies war nur der Fall, wenn das ursprüngliche Zielvolumen der cDNA-Synthese erhöht wurde, um mehr Material zur Verfügung zu haben. Anschließend wurde die „vier Amplikon PCR“ durchgeführt und deren Erfolg anhand einer Agarose-Gelelektrophorese überprüft (Abb. 10).

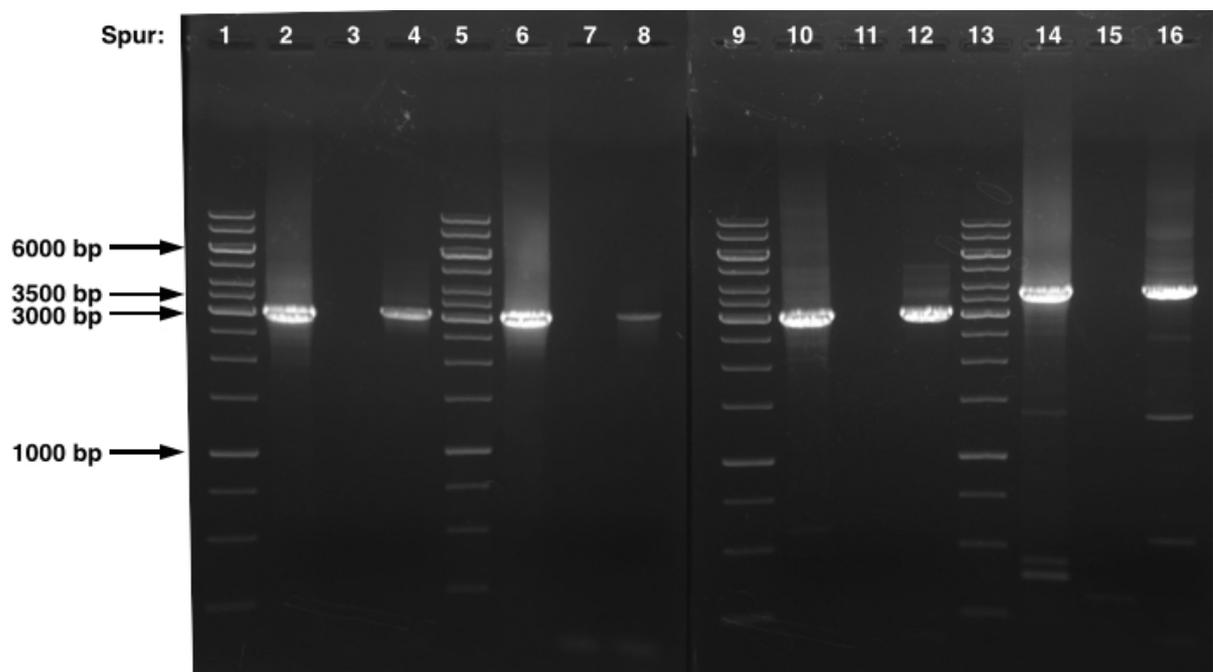


Abb. 10 Qualitative Auswertung der „vier-Amplikon PCR“ über Gelelektrophorese

Darstellung der PCR-Produkte auf einem 1 % Agarosegel, Ethidiumbromidfärbung, Spurbelegung: 1. 1 kb DNA Ladder; 2. 13-0346 Frag 1; 3. NK; 4. PK; 5. 1 kb DNA Ladder; 6. 13-0346 Frag 2; 7. NK; 8. PK; 9. 1 kb DNA Ladder; 10. 13-0346 Frag 3; 11. NK; 12. PK; 13. 1 kb DNA Ladder; 14. 13-0346 Frag 4; 15. NK; 16. PK, NK=negative Kontrolle (HIV-1 negatives Blutplasma); PK=positive Kontrolle (HLTVIII B 10⁶ cop/mL)

Die Fragmente 1-3 haben die erwartete Größe (Fragment1 3053 Bp, Fragment 2 3062 Bp, Fragment 3 3066 Bp). Die Bande des vierten Fragments ist höher gelegen und befindet sich über der 3500 Bp Markerbande. Dies entspricht der erwarteten Größe von 3666 Bp. Die Abbildung 10 zeigt das Ergebnis der „vier-Amplikon PCR“ für das Anzuchtmaterial des Patienten 13-0346.

3.3 Ermittlung der Komplett-Genomsequenz für Isolate der „URF-*new clade*“

Die Komplett-Genomsequenz für das Isolat aus einer vorherigen Virusanzucht (13-0346) lag bereits vor. Von den beiden anderen Patienten (Probennummer 13-5995, 14-0875) waren dagegen nur partielle Sequenzen vorhanden, die durch weitere Sequenzanalysen (Methode 2.2.12) vervollständigt werden sollten. Dadurch gelang es für einen weiteren Patienten (13-5995) eine Komplett-Genomsequenz zu erzeugen sowie von einem dritten Patienten (14-0875) eine nahezu vollständige Genomsequenz zu ermitteln. Da die Sequenzierungsmethode nach Sanger verwendet wurde, stellen die *full-length* Genomsequenzen Populationssequenzen (Konsensussequenz) der viralen Quasispezies dar.

Die ermittelte Sequenz des Patienten 13-5995 besitzt eine Länge von 9049 Bp und entspricht einer Lokalisation von 530-9579 Bp (Lokalisation in Bezug auf das Referenzisolat HxB2, Acc.-Nr.: K03455). Es konnte durchgehend eine doppelsträngige Sequenz ermittelt werden, die nur in den letzten 131 Bp einzelsträngig ist, jedoch qualitativ hochwertige Basenzuordnungen aufzeigt.

Die Auswertung der Elektropherogramme des Isolates 13-5995 erwies sich im Bereich von 7450-7575 Bp (Position relativ zu HxB2) als äußerst schwierig. Dieser genomische Bereich umfasst das Ende des V4-Loop bis zum V5-Loop des *envgp120*-Gens von HIV. Wie alle hypervariablen Domänen in dem Oberflächenprotein des viralen Hüllproteins zeigt dieser Bereich ein hohes Maß an Variabilität. Da deutliche Unterpeaks zu erkennen waren, die auf eine minoritäre Sequenzvariante hindeuten, wurden in diesem Bereich nur die Hauptpeaks ausgewertet, das bedeutet, dass minoritären Varianten vernachlässigt wurden und sich für das deutlichste Signal entschieden wurde. Nach der Auswertung wurde die ermittelte Sequenz durch Alignierung mit dem Referenzvirus HxB2 hinsichtlich der Vollständigkeit und Funktionalität überprüft. Es konnten acht Deletionen (AS Pos: 2088, 2186, 2207-2210, 2219-2226, 2384-2385, 2428, 2476 und 2481-2488) und zwei Insertionen (Position: 7601 bp und 8588 bp) lokalisiert werden, die keinen Wechsel des Leserahmens zur Folge haben. In der Abbildung 11 sind zwei Deletionen im V1-Loop exemplarisch dargestellt.

```

1. 13-5995_major          env loop1
Frame 3  TTAAACTGTACT-----AATGTTACTGTAGGCAACAATA YTRRTG-----AAA?
          L N C T ----- N V T V G N N T/I ? ----- E I
2. HxB2 env Bereich
Frame 3  TTAAAGTGCACTGATTTGAAGAATGATACTAATAACCAATAGTAGTAGCGGGAGAAATGATAATGGAGAAAAGGAGAGA?
          L K C T D L K N D T N T N S S S G R M I M E K G E :

```

Abb. 11 Ausschnitt des Sequenzalignment des V1-Loop im Env_{gp120} Protein zur Identifizierung von Deletionen und Insertionen

Mit ClustalW wurden die Referenzsequenz HxB2 (unten) und die Sequenz des Patienten (oben) aligniert. Gelb eingezeichnet ist der Bereich über den sich der V1-Loop erstreckt.

Der Ausschnitt des Sequenzalignments in Abb. 11 zeigt *Env_{gp120}*. Die untere Sequenz zeigt das Referenzvirus und darüber ist die ermittelte Sequenz des Patienten 13-5995 zu sehen. Im Bereich des V1-Loops sind deutlich zwei Deletionen von 3 AS bzw. 8 AS zu sehen. Im gesamten *Env_{gp120}* unterscheidet sich der Patient 13-5995 auf Aminosäureebene um 29,8 % vom Referenzvirus.

Für den dritten Patienten 14-0875 war es nicht möglich eine durchgängige Populationssequenz zu ermitteln. Es konnten jedoch zwei Teile erhalten werden, die die HxB2-Positionen 536-6640 Bp und 6999-9621 Bp besitzen. Der fehlende Sequenzbereich von nur 359 Bp befindet sich am Ende des V1-Loops sowie im kompletten V2-Loop und darüber hinaus. Da auch an dieser Position in den Elektropherogrammen Variantenmischungen beobachtet wurden, wurden die Sequenzen anhand der Hauptvariante ausgewertet.

3.4 Next Generation Sequencing zur Generierung der full-length Sequenz der „URF-new clade“

Mit NGS ist eine bessere Auflösung von Basenmischungen möglich, da minoritäre Virusvarianten in der viralen Quasispezies mit einer sensitiveren Nachweisgrenze detektiert werden können. Die Ergebnisse der Sanger-Sequenzierung sollten daher mit dieser weiteren Sequenzierungsmethode bestätigt werden.

In den gepoolten DNA-Bibliotheken (Methode 2.2.13) wurde eine mittlere Fragmentlänge von 700 Bp ermittelt und die Konzentration des Pools wurde mit 382 pM bestimmt. Für die Sequenzierung wurde der Probenpool auf 18 pM verdünnt. Es konnten im Mittel 706.252 Sequenzen (reads) pro Fragment generiert werden.

Nach der Vorprozessierung mit den Programmen Trimmomatic und Flash konnten im Durchschnitt 80% der Sequenzen für die Kartierung mit Segminator II eingesetzt

werden. Um eine optimale Kartierung zu ermöglichen, wurden für die Proben 13-0346 und 13-5995 die vorher generierten Sanger-Konsensussequenzen des jeweiligen Isolates als Referenzsequenz verwendet. Da für die Probe 14-0875 keine vollständige Sanger-Konsensussequenz vorlag, wurde hier eine Kartierung gegen die Sanger-Konsensussequenzen von 13-0346 und 13-5995 vorgenommen. Jedes Fragment wurde einzeln kartiert und die *full-length* Sanger-Konsensussequenzen in BioEdit entsprechend getrimmt. Es konnte im Durchschnitt eine Abdeckung (coverage) von 29.960 Sequenzen pro Position erzielt werden. Die Nachweisgrenze für minoritäre Varianten liegt bei der Sanger-Sequenzierung etwa bei 20-25%⁶³. Da das Ziel dieser Arbeit eine Bestätigung der Sanger-Sequenzen war und diese Methode in Zukunft für die Anwendung für *full-length* Genome etabliert werden soll, wurden die NGS-Konsensussequenzen der vier Fragmente mit Grenzwerten für Basenmischungen von 20%, 25% und 30% ausgelesen. Anschließend erfolgte in BioEdit ein Vergleich der Sanger-Sequenzen mit den NGS-Sequenzen hinsichtlich der Basenunterschiede sowie eine Auswertung des Segminator II Berichtes zu den Insertionen und Deletionen (Tab. 3). Mit der Software Geneious wurde eine *full-length* NGS-Konsensussequenz aus den vier überlappenden NGS-Konsensussequenzen ausgelesen. Mit Hilfe des *Next Generation Sequencing* konnte für alle drei Patienten (13-0346, 13-5995 und 14-0875) eine Sequenz für jedes der vier Fragmente generiert werden. Es konnte eine Sequenzidentität der einzelnen Fragmente zwischen den Sanger-Sequenzen und den mit NGS erzeugten Sequenzen von mindestens 98,4 % für die Isolate 13-0346 und 13-5995 erreicht werden (Tab. 3). Dabei kam es nur einmal zur Detektion von zwei Transversionen (13-0346 Fragment 1). Die vorhandenen Ambiguitäten in den Sequenzen beider Isolate befinden sich unterhalb von 2,01 % und sind in den Fragmenten 3 und 4 am höchsten, da sich hier der hochvariable Bereich des Hüllglykoproteins von HIV-1 befindet. Sowohl zwischen den Patienten als auch innerhalb der einzelnen Fragmente konnten unterschiedliche Deletionen bzw. Insertionen von Segminator II detektiert werden. Diese sind ebenfalls in der Tab. 3 aufgeführt, dabei ist die Größe der Deletionen bzw. Insertionen mit dem dazugehörigen prozentualen Anteil angegeben. In der mit NGS erzeugten Sequenz wurde jedoch keine der aufgeführten Mutationen beobachtet, obwohl einige mit einem hohen prozentualen Anteil vorkommen (Tab. 3).

Tab. 3 Ergebnisse des NGS für die 4 PCR-Fragmente

13-0346	Frag1	Frag2	Frag3	Frag4
Anteil Ambiguitäten (%)	0,90	0,56	1,29	2,01
Identität mit der Sanger-Sequenz (%)	98,4	98,4	98,4	98,4
Anzahl der unterschiedlichen Basen	2 Transversionen	0	0	0
Deletionen	1 Bp (31 %)	/	/	3 Bp (17 %)
Insertionen	1 Bp (5 %)	1 Bp (5 %)	1 Bp (9 %)	1 Bp (17)
13-5995	Frag1	Frag2	Frag3	Frag4
Anteil Ambiguitäten (%)	0,40	0,06	0,83	0,84
Identität mit der Sanger-Sequenz (%)	99,5	99,9	99,1	99,1
Anzahl der unterschiedlichen Basen	0	0	0	0
Deletionen	3 Bp (17 %) 1 Bp (3 %)	/	3 Bp (19 %)	3 Bp (51 %)
Insertionen	/	/	3 Bp (11 %)	/

Für die beiden Patienten 13-0346 und 13-5995 konnte basierend auf den NGS-Daten eine Komplet-Genomsequenz mit Hilfe der Software Geneious bestimmt werden. Sie wurde mit der Komplet-Genom-Sanger-Sequenz verglichen (Tab. 4). Dabei wurde für das Isolat 13-0346 eine 99,1 %ige und für 13-5995 eine 99,5 %ige Übereinstimmung ermittelt. Für 13-0346 wurde innerhalb der Komplet-Genom-NGS-Sequenz ein höherer Anteil an Ambiguitäten (1,27 %) als in der Komplet-Genom-Sanger-Sequenz (0,98 %) festgestellt. Beim zweiten Isolat (13-5995) ergaben sich Ambiguitäten in vergleichbarer Größenordnung (NGS 0,39 %, Sanger 0,45 %). Mit NGS war ein höherer Ambiguitätenanteil zu erwarten, da mit dieser Methode minoritäre Varianten sensitiver nachzuweisen sind. Bei dem zweiten Isolat 13-5995 ergab bereits die Sanger-Sequenzierung, dass eine Variantenmischung (Anteil >20 %) vorlag. Es wurde in diesem Bereich im Env-Protein nur die majoritäre Virusvariante ausgewertet.

Da die Sanger-Sequenz des Isolates 14-0875 eine Lücke von 359 Bp aufweist, erfolgte die Kartierung dieser Sequenz gegen die beiden anderen Isolate. Beide

Isolate besitzen an unterschiedlichen Positionen Deletionen und Insertionen, deshalb wurden sie zur Kartierung von 14-0875 eingesetzt, um die NGS-Sequenzen optimal zu kartieren.

Da es sich zwar um ein Virus der gleichen URF, es sich jedoch um eine andere Infektion handelt (zwei verschiedenen Quasispezies), sind geringere Sequenzidentitäten zu finden (Tab. 4). So besitzt das Isolat 14-0875 eine geringere Sequenzidentität mit dem Isolat 13-0346 (durchschnittlich 90,4 %) als mit 13-5995 (durchschnittlich 93,1 %). Auch hier fällt auf, dass mehr Sequenzvarianz in den Fragmenten drei und vier, also in der 3'-Genomhälfte vorliegt und der Anteil der Ambiguitäten höher ist, als in den ersten beiden Fragmenten. Dieser Umstand ist auf den hochvariablen Bereich des Hüllglykoproteins zurückzuführen. Da das Isolat nicht gegen seine Sanger-Sequenz kartiert wurde, wurde keine Auswertung der Deletionen und Insertionen durchgeführt, da sich diese auf die jeweilige Referenzsequenz beziehen würden. Nach der Kartierung wurde auch für dieses Isolat versucht eine NGS-Komplett-Genomsequenz zu generieren. Dies ist jedoch nur mit den Sequenzen, welche gegen den Patienten 13-5995 kartiert wurden, gelungen.

Tab. 4 Ergebnisse für die vier PCR-Fragmente des Isolates 14-0875 mit Sanger-Sequenzierung und NGS, kartiert an die beiden anderen Isolate (13-0346 und 13-5995)

14-0875 kartiert an 13-0346	Frag1	Frag2	Frag3	Frag4
Anteil Ambiguitäten (%)	0,40	0,06	0,83	0,84
Identität mit der 13-0346 Sanger-Sequenz (%)	92,4	94,1	87,9	87,2
14-0875 kartiert an 13-5995	Frag1	Frag2	Frag3	Frag4
Anteil Ambiguitäten (%)	0,40	0,06	0,83	0,84
Identität mit der 13-5995 Sanger-Sequenz (%)	95,2	96,4	90,5	90,1

3.5 Threshold-Analyse für den Nachweis minoritärer Varianten mit NGS

Aufgrund der unterschiedlichen Virusvarianten, die in einer HIV-infizierten Person (Quasispezies) vorliegen, kommt es während der Sequenzierung zur Detektion von Basenmischungen an definierten Positionen in der Sequenz. Im Elektropherogramm werden solche Mischungen durch einen Unterpeak unterschiedlicher Höhe sichtbar. In der Sanger-Sequenzierung wird das Vorliegen von Variantenmischungen erkenntlich, wenn diese Variante mindestens mit 20-25 % Anteilen in der Quasispezies vorkommt⁶³. Die Mischpositionen oder Sequenzambiguitäten werden nach der Internationalen Union der Biochemiker (IUB) mit speziellen Buchstaben kodiert (Bsp: IUB-Code Y, wenn T und C vorliegt). Mit mittels NGS erzeugten Daten konnte beim Auslesen über ein speziell in der Arbeitsgruppe entwickeltes Python-Skript (Britta Altmann) ein variabler *Threshold* (Schwellwert) von 25 % und 30 % festgesetzt werden (Tab. 5).

Dabei war zu erkennen, dass mit steigendem Schwellwert weniger Ambiguitäten in NGS detektiert werden konnten (Tab. 5). Um speziell minoritäre Varianten zu detektieren, sind feinere Abstufungen im Schwellwert nötig. Dies könnte man, nach einer erfolgreichen Etablierung der Methode für HIV, mit einer definierten Viruspopulation validieren.

Weiterhin ist im direkten Vergleich zwischen Sanger und NGS bei einem Schwellwert von 20 % zu sehen, dass vor allem bei dem Isolat 13-0346 mehr Ambiguitäten detektiert wurden (Tab. 5). Auch hier ist in den Fragmenten 3 und 4 der 3'-Genomhälfte ein höherer Wert ermittelt worden, was wiederum auf den hochvariablen Bereich des Hüllglykoproteins zurückzuführen ist.

Für das Isolat 14-0875, das gegen die beiden anderen Isolate kartiert wurde, wurden die höchsten Ambiguitäten festgestellt. Dies ist besonders in der 3'-Genomhälfte (Fragment 3 und 4) zu sehen. Hier sind Häufigkeiten von bis zu 4 % detektiert worden (Tab. 5). Dies ist auf die Kartierung gegen ein anderes Isolat (13-5995) zurückzuführen.

Tab. 5 Vergleich der auftretenden Häufigkeiten von Ambiguitäten bei unterschiedlichen Grenzwerten

Häufigkeit der Ambiguitäten in 13-0346 (%)				
Schwellwert	Frag 1	Frag2	Frag3	Frag4
Sanger	0,88	0,27	0,89	1,30
NGS 20 %	0,90	0,56	1,29	2,01
NGS 25 %	0,90	0,56	1,25	1,96
NGS 30 %	0,87	0,53	1,19	1,91
Häufigkeiten der Ambiguitäten in 13-5995 (%)				
Schwellwert	Frag1	Frag2	Frag3	Frag4
Sanger	0,90	0,03	0,91	0,97
20 %	0,40	0,06	0,83	0,84
25 %	0,33	0,03	0,61	0,75
30 %	0,30	0,00	0,40	0,68
Häufigkeiten der Ambiguitäten in 14-0875 kartiert an 13-0346 (%)				
Schwellwert	Frag1	Frag2	Frag3	Frag4
20 %	0,93	1,43	4,04	4,34
25 %	0,76	0,60	2,68	3,99
30 %	0,57	0,40	2,07	2,32
Häufigkeiten der Ambiguitäten in 14-0875 kartiert an 13-0346 (%)				
Schwellwert	Frag1	Frag2	Frag3	Frag4
20 %	1,03	0,92	3,08	2,95
25 %	0,86	0,73	2,35	2,39
30 %	0,67	0,50	1,95	1,86

3.6 Phylogenetische Analysen der „URF-new clade“

In vorherigen Arbeiten wurden phylogenetische Analysen (Methode 2.2.14) anhand partieller Genomsequenzen durchgeführt. Die bislang 13 Isolate, welche als „URF-new clade“ identifiziert werden konnten, clustern in der Baumtopologie mit keinem bisher klassifizierten Subtypen oder rekombinanten Form (CRF) (Abb. 12).

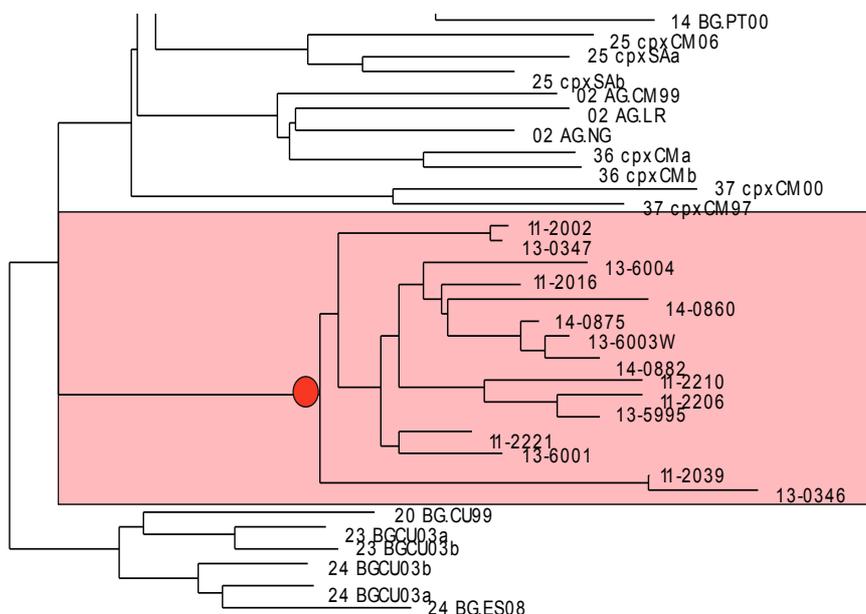


Abb. 12 Phylogenetische *Neighbor-Joining* Analyse mit Isolaten aus Oman (n=13)

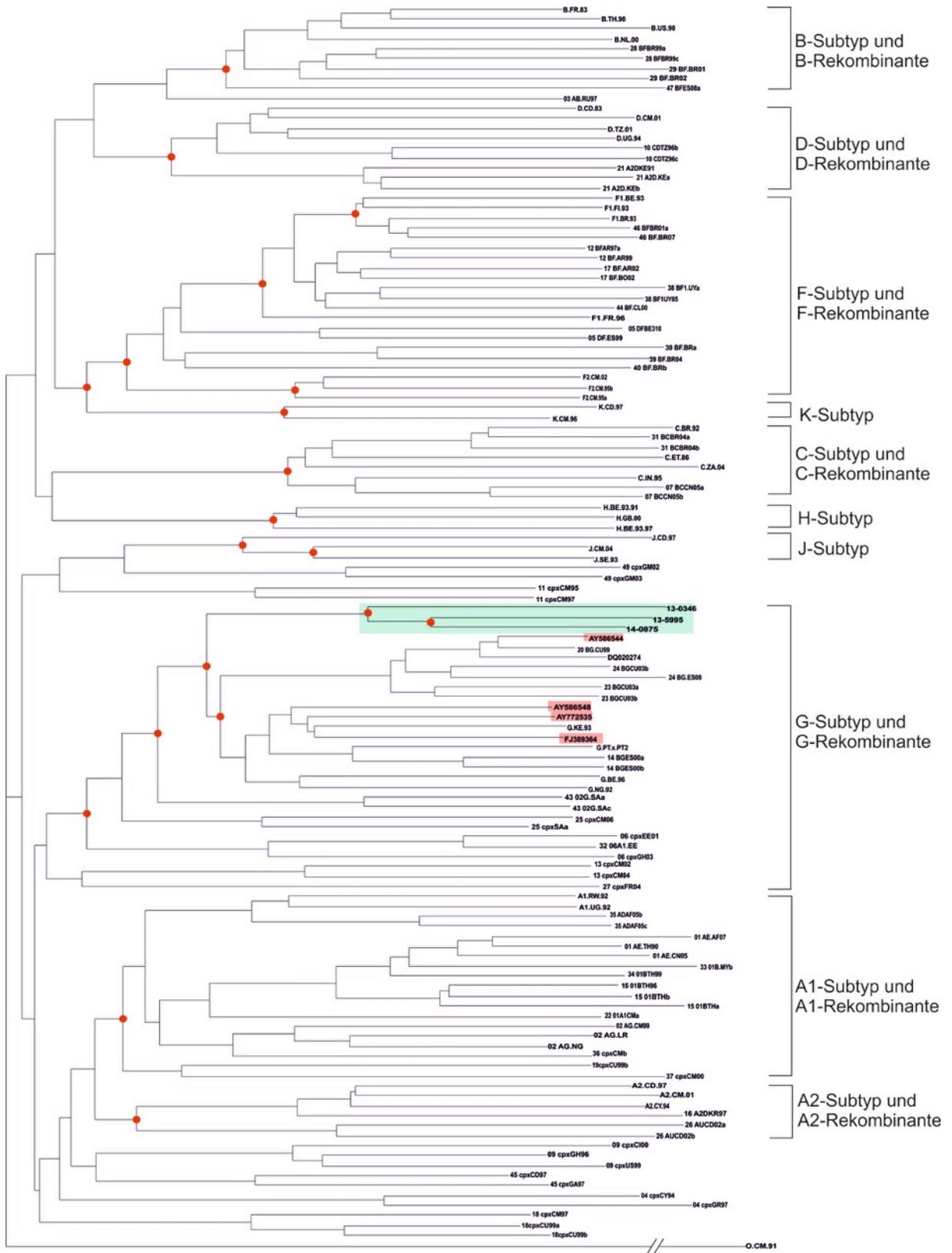
Für diese Analyse wurde ein reduziertes Referenzpanel der *pol*-Genregion der Los Alamos HIV Datenbank verwendet. Als Außengruppe (nicht gezeigt) diente der Subtyp O (O.CM.91, Acc.-Nr.: L20571). Eine *Bootstrap*-Analyse wurde mit 1000 Replikaten durchgeführt und signifikante Knotenpunkte (> 70 %) sind mit einem roten Punkt gekennzeichnet. Die Isolate 11-2002 (Erstprobe) und 13-0347 (Verlaufspröbe) sowie die Isolate 11-2039 (Erstprobe) und 13-0246 (Verlaufspröbe) gehören zum selben Patienten.

Der Ausschnitt des Stammbaumes zeigt die bislang 13 als putativ „URF-new clade“ klassifizierten Isolate. Die derzeit *full-length* analysierten Isolate sind ebenfalls in dieser monophyletischen Gruppe zu finden. Dabei handelt es sich um das Isolat 13-5995 und 14-0875 sowie das Isolat 13-0346, welches ebenfalls im Baum mit der Erstprobe 11-2039 vertreten ist. Die Isolate 11-2002 (Erstprobe) und 13-0347 (Verlaufspröbe) gehören ebenfalls zu einem Patienten. Die drei Isolate (13-0346, 13-5995 und 14-0875) besitzen eine mittlere paarweise Distanz von mehr als 8 % zueinander, was gegen eine epidemiologische Verknüpfung spricht. Aus diesem Grund wurden sie für die *full-length* Sequenzanalyse verwendet.

Die phylogenetischen Analysen wurden mit der Komplet-Genomsequenz mit dem Ziel wiederholt, diese monophyletische Gruppe über das gesamte Genom als zirkulierende rekombinante Form zu klassifizieren (Abb. 13). Die Signifikanz der Baumtopologie wurde mit der Bootstrap-Analyse überprüft (Methode 2.2.14). Zusätzlich zum Subtyppreferenz-panel wurden 5 über eine Blast-Suche ermittelte ähnliche Sequenzen zur Analyse hinzugefügt, um zu überprüfen, ob höchst ähnliche, bereits bekannte Isolate in der gleichen monophyletischen Gruppe der „URF-*new clade*“ clustern. Die Sequenzen mit der größten Ähnlichkeit zu den „URF-*new clade*“ Isolaten aus Oman stammen aus Kamerun und Kuba. Weitere Isolate aus der Blast-Suche, die eine große Ähnlichkeit (87-89 %) zu den Isolaten aus Oman besitzen, stammen aus Nigeria, Spanien und Saudi-Arabien; sie sind jedoch nicht im Stammbaum eingezeichnet.

Die Abbildung 13 zeigt die erhaltene Baumtopologie mit den Komplet-Genomsequenzen. Die Isolate aus Oman sind grün unterlegt. Es ist zu erkennen, dass sie in einer eigenen monophyletischen Gruppe innerhalb der Subtypen G und G-rekombinanten Klade lokalisiert sind. Kein bekanntes Isolat clustert mit dieser neuen Klade. Die Isolate aus der Blast-Suche sind, wie die omanischen Proben, auch in der Klade des Subtypen G und dessen rekombinante Formen zu finden, aber nicht innerhalb der signifikanten Klade der omanischen Isolate (Bootstrap 100 %).

Die phylogenetische Analyse von zwei Komplet-Genomsequenzen und einer dritten annähernd kompletten Genomsequenz von Isolaten, die epidemisch nicht verknüpft sind, bestätigen die Annahme, dass die neu identifizierten Virusvarianten aus Oman eine neue CRF darstellen.



0,1

Abb. 13 Phylogenetische Analyse zur Klassifizierung einer neuen zirkulierenden rekombinanten Form aus Oman

Neighbor-Joining Stammbaumanalyse mit einem reduzierten Referenzpanel der Los Alamos HIV Datenbank (n=118), 3 Isolaten aus Oman (13-0346, 13-5995 und 14-0875, grün) sowie 5, mittels Blast-Suche ermittelte, höchst ähnliche Isolate (AY586544, AY586548, AY772535, DQ020272 und FJ389346, rot). Als Außengruppe wurde ein Isolat der Gruppe O (O.CM.91, Acc.-Nr.: L20571) verwendet. Der Ast der Außengruppe wurde zur besseren Darstellung gekürzt (zwei Striche). Eine Bootstrap-Analyse wurde mit 1000 Datensätzen durchgeführt. Signifikante Knotenpunkte (Bootstrap > 70 %) sind rot markiert. Die Kladen der Subtypen und rekombinante Formen sind durch Klammern markiert. Der Distanzbalken (Scale-Bar) unter dem Baum mit einem Wert von 0,1 beschreibt die Astlänge, die einem Nukleotidaustausch von 10 % pro Position im Alignment entspricht. Das Isolat 14-0875 ist nur partiell vorhanden, dadurch wurde das Referenzpanel auf die Länge von 14-0875 gekürzt (Lokalisation relativ zu HxB2: 536-6640, 6999-9621).

3.7 Analyse der rekombinanten Genomstruktur einer neuen zirkulierenden rekombinanten Form aus Oman

Die phylogenetischen Analysen haben gezeigt, dass die omanischen Isolate in der Klade des Subtypen G und dessen G-rekombinanten Formen lokalisiert sind. Um die genaue Mosaikstruktur der Isolate zu bestimmen, wurden frei zugängliche Programme innerhalb der Los Alamos HIV Datenbank verwendet. Zum einem handelte es sich dabei um das Programm RIP (Methode 2.2.15). Dieses gibt über einen Distanzplot den am ähnlichsten Subtyp zum Isolat an jeder Position im Alignment an. Die Plots der Isolate 13-5995 und 14-0875 zeigen eine sehr ähnliche Mosaikstruktur (Abb.14). Beide Isolate zeigen einen überwiegenden Subtyp G-Anteil im Genom. Dieser wird von drei kurzen Subtyp D-Anteilen unterbrochen. Diese Resultate bestätigen das vorherige Ergebnis, dass die „URF-*new clade*“ eine neue monophyletische Gruppe innerhalb des Subtyp G und G-Rekombinante darstellt.

Die Konsensus-NGS-Sequenz des Patienten 14-0875 wurde in die RIP-Analyse mit einbezogen. Sie weist dieselbe Mosaikstruktur aus Subtyp G- und D-Anteilen auf. In vorherigen Arbeiten wurde bereits das Isolat 13-0346 mit der RIP-Analyse untersucht⁵².

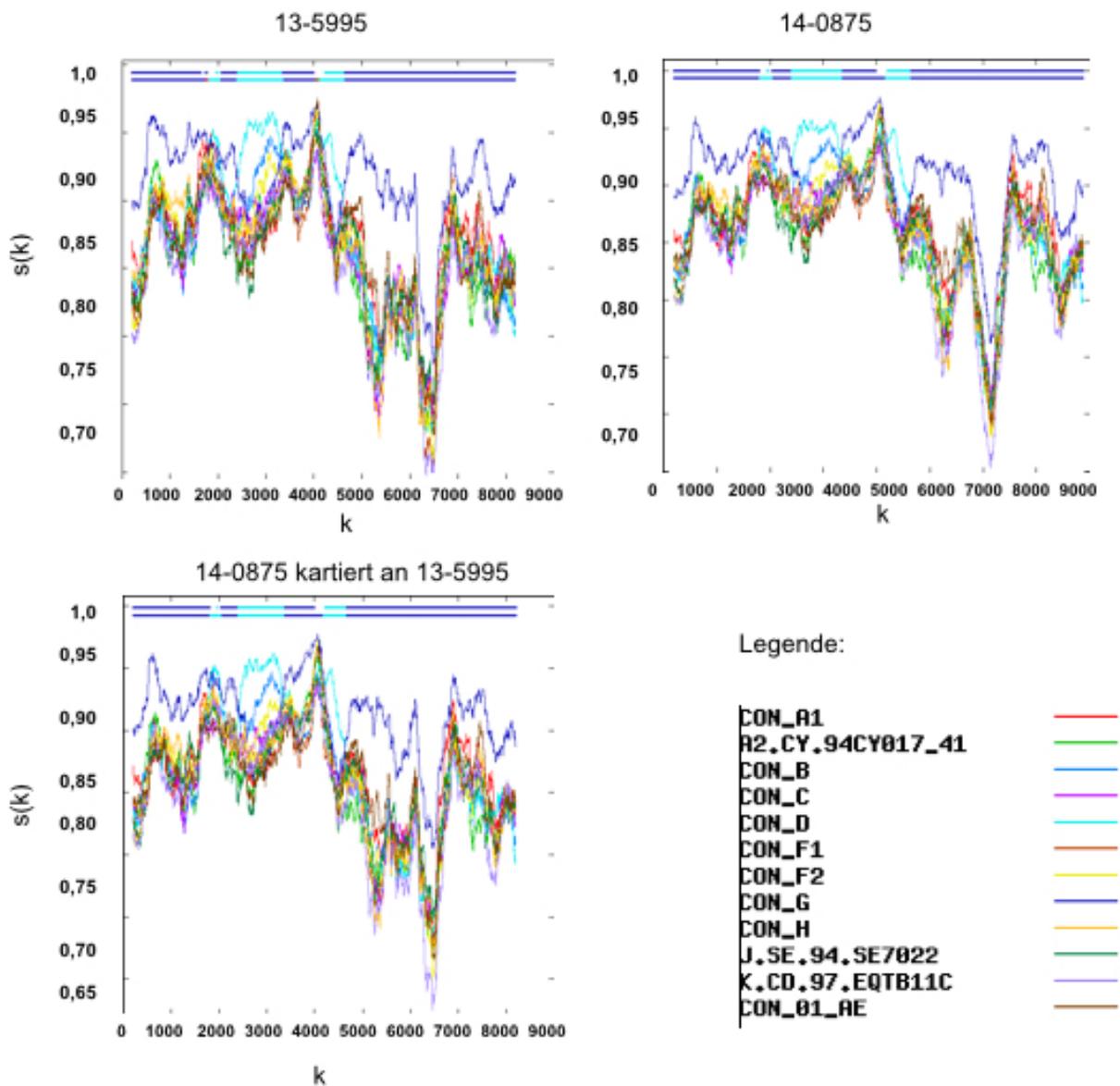


Abb. 14 Ähnlichkeits- (s)-Distanzplot als Ergebnis der RIP-Analyse

Gezeigt sind die Distanzplots der Isolate 13-5995 und 14-0875. Für das Isolat 14-0875 wurden zwei Analysen durchgeführt. Oben rechts mit der generierten Sanger-Sequenz der Komplet-Gemonsequenzierung, die eine Lücke von 359 Nukleotiden enthält. Unten mit der kartierten Sequenz des NGS gegen das Isolat 13-3559. Die Legende zeigt den Farbcode der einzelnen Referenzsubtypen. Als Fenstergröße wurde der Wert 400 genommen. Der Konfidenzschwelligwert beträgt 90 %. Auf der x-Achse (k) ist die Sequenzposition der Isolate im Zentrum des Fensters (Beginn 200) angegeben. Die y-Achse (s(K)) beschreibt die Ähnlichkeit zwischen dem Isolat und der Referenzen. Die zwei Balken über den Plots repräsentieren die Mosaik-Genomstruktur der Isolate. Der untere Balken zeigt die beste Übereinstimmung zwischen dem Isolat und einer Referenz an. Der obere Balken zeigt deren signifikante Übereinstimmung.

In der Tabelle 6 sind die Rekombinationsstellen für die analysierten Isolate 13-5995 und 14-0875 aufgeführt.

Tab. 6 Übersicht der Rekombinationsschnittstellen

	Fragment Position Start-Ende	Breakpoint Intervall Start - Ende	Fragment Subtyp
Position in der Sequenz des Isolates 13-5995	1-298	-	N/A
	279-2760	2753-2798	G
	2761-3708	3602-3741	D
	3709-4601	4555-4611	G
	4602-4873	4864-4893	D
	4874-8862	-	G
	8863-9049	-	N/A
Position in der Sequenz des Isolates 13-5995 bezogen auf das Referenzisolat HxB2	530-789	-	N/A
	790-3283	3276-3321	G
	3284-4231	4125-4264	D
	4232-5124	5078-5134	G
	5125-5396	5387-5416	D
	5397-9409	-	G
	9409-9591	-	N/A
	Fragment Position Start-Ende	Breakpoint Intervall Start - Ende	Fragment Subtyp
Position in der Sequenz des Isolates 14-0875	1-274	-	N/A
	275-2478	2737-2759	G
	2479-3615	3608-3701	D
	3616-4591	4545-4602	FG
	4592-4863	4854-4870	D
	4864-6118	6106-6241	G
	6999-9431	-	G
	9432-9641	-	D
Position in der Sequenz des Isolates 14-0875 bezogen auf das Referenzisolat HxB2	536-789	-	N/A
	790-3281	3270-3292	G
	3282-4148	4141-4234	D
	4149-5124	5078-5135	G
	5125-5396	5387-5403	D
	5397-6641	6711-7121	G
	6999-9411	-	G
	9412-9620	-	N/A

Die Ermittlung der Rekombinationsstellen (Tab. 6) anhand der Referenz HxB2 ermöglicht eine vergleichende Darstellung der rekombinanten Virus-Varianten. In der Abbildung 15 sind die Genkarten mit eingezeichneten *Breakpoints* anhand von HxB2 dargestellt. Sie wurden ebenfalls mit dem Programm *jpHMM* (GOBICS) erhalten. Beide Isolate zeigen eine identische Mosaikstruktur aus Subtyp G und Subtyp D (*pol*, *vif*). Aufgrund einer möglichen Fehldetektion wurden die Teilsequenzen 1 (HxB2-

Lokalisation: 536-6640) und 2 (HxB2-Lokalisation: 6999-9621) des Isolates 14-0875 einzeln in das Programm eingeladen. Zur einfacheren Darstellung wurden die zwei resultierenden Abbildungen übereinander gelegt (Abb. 15). Sie zeigen die gleiche Rekombinationsstruktur wie das Isolat 13-5995. Diese Ergebnisse stimmen mit der Genomstruktur von Isolat 13-0346 überein.

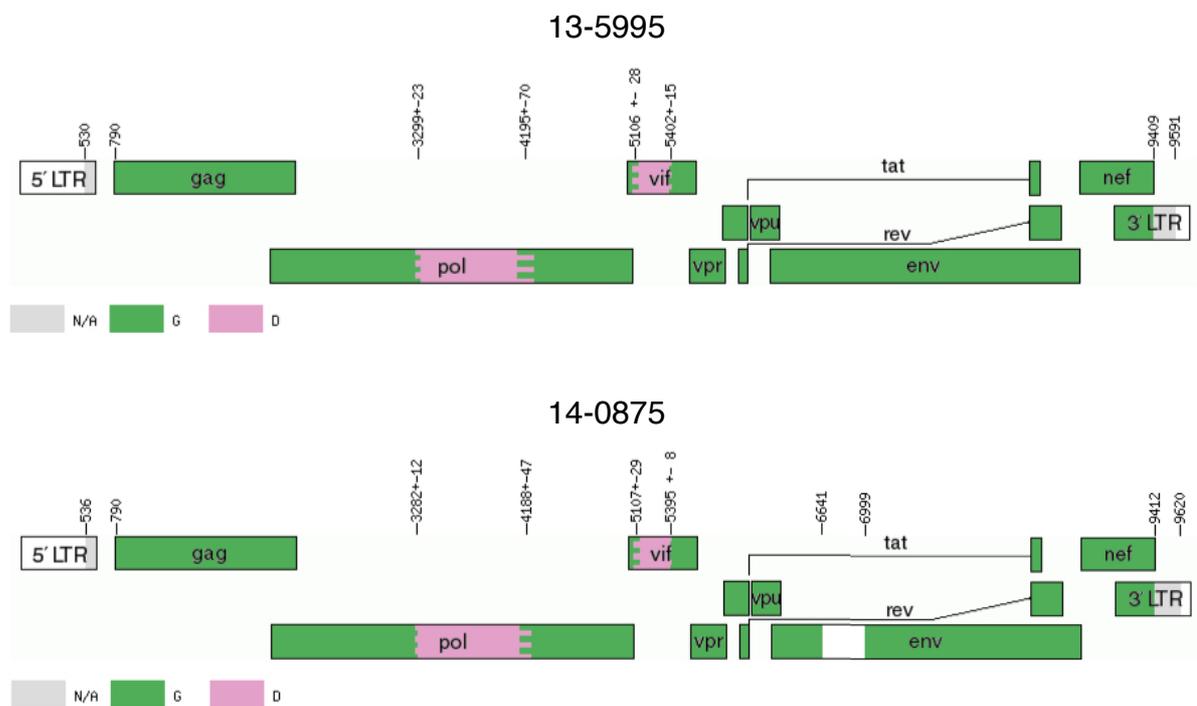


Abb. 15 Grafische Darstellung der *Breakpoints* der analysierten Isolate anhand der Lokalisation auf dem Referenzvirus HxB2 (Acc.-Nr.: K03455)

Dargestellt ist das Genom von HIV-1 mit den drei Leserahmen. Die Farben kodieren für den ermittelten Subtypen und sind in den Legenden beschrieben. Die Koordinaten der Rekombinationsstellen sind über dem jeweiligen Genabschnitt eingezeichnet (bezogen auf HxB2). Oben ist das Isolat 13-5995 dargestellt, 14-0875 unten. Die beiden Ergebnisse der getrennt eingefügten Teilsequenzen von 14-0875 wurden zur besseren Darstellung übereinander gelegt. Der weiße Bereich des Isolates 14-0875 entspricht der Lücke von 359 Bp in der Sequenz.

Erste Analysen der drei Isolate durch die offiziellen Ansprechpartner der internationalen HIV-Sequenz-Datenbank Los Alamos bestätigen, dass es sich bei den neu identifizierten „URF-*new clade*“ Isolaten um eine neue zirkulierende rekombinante Form handelt. Diese neue HIV-1 Virusvariante erhielt die Bezeichnung CRF75_DG. Die hier ermittelten Sequenzen für die CRF75_DG werden erst bei Veröffentlichung der Ergebnisse der Stanford-Datenbank öffentlich zugänglich gemacht (Publikation in Arbeit).

4. Diskussion

Im Rahmen der genotypischen Resistenzanalyse der *pol*-Genomregion (PR/RT und INT) bei omanischen HIV-1-infizierten Patienten mit Therapieversagen wurde der *pol*-Subtyp analysiert. Bisherige Ergebnisse zeigten, dass ein hoher Anteil der untersuchten Proben (47 %) keinem bekannten Subtypen oder einer zirkulierenden rekombinanten Form (CRF) zugeordnet werden konnte und somit als URF (singuläre rekombinante Form) eingestuft wurde. Bisher konnten 13 Isolate, anhand von phylogenetischen Analysen, als eigene monophyletische Gruppe („URF-*new clade*“) klassifiziert werden, da sie sich signifikant von klassifizierten Referenzviren der Los Alamos HIV Datenbank unterscheiden. Durch phylogenetische Untersuchungen der Komplet-Genomsequenz einer Auswahl dieser Isolate sollte geklärt werden, ob es sich bei der im Oman vorkommenden URF um eine neue CRF von HIV-1 und damit um eine HIV-1 Variante mit potentiell epidemischer Relevanz handelt.

Um dieses Ziel zu erreichen, wurde in der Arbeitsgruppe eine *full-length* cDNA-Synthese sowie eine Komplet-Genom-Amplifikationsmethode aus 4 überlappenden Fragmenten („vier-Amplikon PCR“) etabliert. Ein erstes Isolat (13-0346) war über eine Virusanzucht erfolgreich vermehrt worden und mit den oben aufgeführten Methoden ist es gelungen, eine Komplet-Genomsequenz dieses Isolates zu generieren. Für zwei weitere Viren von dieser Klade („URF-*new clade*“) konnten ohne vorherige Anzucht direkt aus Plasmaproben eine Populationssequenz ermittelt werden (13-5995, 14-0875). Von der ersten Probe 13-5995 konnte eine Komplet-Genomsequenz, mit zahlreichen Mischpositionen im hochvariablen *env*-Bereich, ermittelt werden und von der zweiten Infektion (14-0875) wurde eine partielle Genomsequenz erhalten. Analysen mit dem Komplet-Genom- bzw. partiellen Komplet-Genomsequenzen zeigen, dass es sich in voller Länge um eine neue rekombinante Form von HIV-1 handelt.

In der vorliegenden Arbeit wurde der Versuch unternommen, eine „ein-Amplikon *full-length*-RT-PCR“ für HIV-1 zu etablieren, damit die Komplet-Genomsequenz nicht aus vier verschiedenen amplifizierten Genombereichen zusammengesetzt werden muss. Dafür sollten Plasmaproben von zwei omanischen Patienten auf PBMC-Spenderzellen angezüchtet werden, um ausreichend Material für die Etablierung, Optimierung und für spätere Analysen zur Verfügung zu haben. Weiterhin wurden die zu Beginn dieser Arbeit nur partiell vorliegenden Genomsequenzen der Isolate 13-5995 und 14-0875 vervollständigt, um auch von diesen beiden Isolaten jeweils

eine Komplet-Genomsequenz zu erhalten. Mit einer weiteren Sequenzierungsmethode (NGS) sollten für alle drei bislang untersuchten *full-length* Isolate Sequenzen erzeugt und mit den Sanger-Sequenzen verglichen werden, um minoritäre Virusvarianten zu detektieren und um den Zeitaufwand für eine Komplet-Genomsequenzierung zu reduzieren. Von den drei Komplet-Genomsequenzen wurde die Genomstruktur abschließend untersucht, um die Kriterien zur Klassifizierung einer neuen CRF zu erfüllen.

4.1 Virusanzucht eines omanischen Isolates auf PBMC-Spenderzellen

Die Population unterschiedlicher Virusvarianten in einem HIV-infizierten Patienten (Quasispezies) ist zu Beginn der HIV-Infektion relativ homogen, divergiert im Verlauf einer Infektion jedoch zu einer zunehmend heterogenen Population. In Untersuchungen von Primärmaterial eines Patienten ist daher in Abhängigkeit von der Dauer der Infektion bis zur Blutentnahme bei der Sequenzanalyse mit Basenmischungen zu rechnen. Bei einer Virusanzucht kommt es zu einer Selektion der Virusvariante, die sich unter Kulturbedingungen am besten vermehren kann (höchste virale Fitness). Diese Variante besitzt einen Selektionsvorteil und kann sich dadurch in einer Anzucht auf PBMC-Spenderzellen als Hauptvariante behaupten. Somit ist die HIV-Quasispezies in Kulturüberständen aus einer Virusanzucht weniger divergent als die Viruspopulation im Original-Plasma. Aus diesem Grund und auch mit dem Ziel, ausreichend Material für die Etablierung und Optimierung der „ein-Amplikon *full-length* PCR“ sowie für anschließende Untersuchungen zur Verfügung zu haben, wurde versucht, von zwei omanischen Patienten, die der „URF-*new clade*“ angehören, HIV aus EDTA-Plasma in Spender-PBMCs anzuzüchten.

Von diesen Isolaten war zwar nur noch wenig Plasma verfügbar, da aber beide Proben eine hohe Viruslast (13-6001 7.851 cop/mL und 14-0875 272.236 cop/mL) aufweisen, wurden die Anzuchsversuche unternommen, obwohl das standardisierte diagnostische HIV-Anzuchtprotokoll eine Co-Kultur aus PBMCs aus dem infizierten Individuum und gesunden Spender-PBMCs vorsieht. Es konnte nach vierwöchiger Anzucht keine Infektion der Zellen nachgewiesen werden. Gründe dafür können sein, dass die Multiplizität der Infektion für eine erfolgreiche Infektion in Kultur zu gering war. Im Viruslast-Assay wird die Anzahl der RNA-Moleküle bestimmt, unabhängig davon, ob sie von infektiösen oder defekten Viruspartikel stammen. Da die

Plasmaproben bei dem Transport vom CPHL im Oman an das RKI/Berlin auftauten, ist die wahrscheinlichste Erklärung, dass der infektiöse Titer der Plasmaviren für eine Vermehrung auf Spender-PBMCs nicht ausreichend war. Nach der Ankunft wurden die Proben ein weiteres Mal bis zur Bearbeitung eingefroren. Frier bzw. Tauzyklen haben auf das wenig stabile umhüllte Virus einen starken inaktivierenden Effekt. Ob infektiöse Partikel übrig bleiben, hängt stark von der initial vorliegenden Viruslast ab⁶⁴. Ein weiterer Aspekt, den die Arbeitsgruppe um Manak *et al.* (2012) berücksichtigt hat, sind die CD8-Zellen. Sie sind Teil des adaptiven Immunsystems und dienen der Abwehr von Viren⁶⁵. In dieser Arbeit wurde mit Hilfe des CD8 Kits (STEMCELL Technologies) CD8-Zellen eliminiert um eine erfolgreiche Virusanzucht zu ermöglichen.

Probentransport unter Einhaltung der Solltemperatur, Verwendung von EDTA-Blut für die standardisierte Anzucht in Co-Kultur sowie CD8-Depletion der Spender-PBMCs nach Manak *et al.* (2012) könnten in Zukunft zu einer Verbesserung des Anzuchterfolges führen.

4.2 Etablierung einer *full-length*-RT-PCR für HIV-1 in Form einer „ein-Amplikon-PCR“

Durch die Verwendung der „vier-Amplikon PCR“ erhält man eine Komplet-Genomsequenz, die möglicherweise aus vier unterschiedlichen *Templates* zusammengesetzt wurde. Die Etablierung einer subtyp-generischen *full-length* Amplifikationsmethode bietet die Möglichkeit die *in vitro* Rekombinationsrate bei Amplifikationen von Genomen in mehreren Fragmenten zu minimieren, indem ein einziges Amplikon des gesamten HIV-1 Genoms generiert wird. Sollte in den bioinformatischen Analysen der Rekombinationsstruktur eine inter-Subtyp-Rekombination beobachtet werden, könnte dieses Ergebnis auf einer falschen Interpretation beruhen, wenn eine „vier-Amplikon PCR“ verwendet wurde. Auch partiell erzeugte cDNA kann zu falschen Rückschlüssen auf die Genomstruktur führen.

Die Etablierung einer effizienten *full-length* Amplifikationsmethode erweist sich vermutlich als äußerst schwierig, denn auch in der Literatur sind nur wenige *full-length* Amplifikationsmethoden beschrieben^{58,66,67}. Überwiegend werden mehrere PCRs zur Amplifikation des gesamten Genombereichs eingesetzt^{57,68-70}.

Den Ausgangspunkt für die Austestung einer „ein-Amplikon PCR“ bildeten die Primer, die für die äußeren Fragmente in der „vier-Amplikon PCR“ verwendet wurden. Basierend auf publizierten *nested* PCRs⁵⁸ wurden verschiedene PCR-Varianten durchgeführt. In insgesamt 11 Variationen wurden relevante Parameter einer PCR unterschiedlich kombiniert, jedoch konnte mit keiner PCR-Variante ein *full-length* Amplikon synthetisiert werden. Es wurden diverse Primerkombinationen (Variante 2-7) im *nested* und *semi-nested* Aufbau getestet und nach Angaben des Herstellerprotokolls bzw. modifiziert ausgeführt. In 7 Variationen wurde der Long PCR Enzyme Mix der Firma Thermo Scientific/Fermentas erfolglos verwendet. Der Enzym-Mix beinhaltet zwei Enzyme: die *Taq* Polymerase und eine thermostabile DNA-Polymerase mit Korrekturfunktion, welche synergisch Amplikons von bis zu 47 Kilobasen generieren können.

Sowohl die verwendete Enzymcharge als auch die verwendeten Primer führen in anderen PCRs zu positiven Ergebnissen, sodass diese Reagenzien als Ursache für den Misserfolg ausgeschlossen werden können. Die ausreichende Länge der synthetisierten cDNA wurde anhand verschiedener, im Labor etablierter PCRs überprüft⁵¹. Allerdings könnten ausgeprägte Sekundärstrukturen von RNA und DNA von HIV eine Ursache für die negativen *full-length* PCR Ergebnisse sein. Diese Sekundärstrukturen, wie z. B. Haarnadelstrukturen, können einen Syntheseabbruch verursachen. Durch die Erhöhung des DMSO-Anteils im PCR-Ansatz können die Sekundärstrukturen möglicherweise besser aufgelöst werden und der Syntheseabbruch so verhindert werden.

Die PCR-Variante, die exakt nach Literatur eingesetzt wurde, hat möglicherweise nicht funktioniert, da die zugehörigen Primer für den Subtyp C etabliert wurden und die Primer gegebenenfalls nicht optimal für Subtyp B (HTLVIIIB) geeignet sind, da je zwei fehlpaarige Nukleotide vorkommen.

4.3 Amplifikation und Sequenzierung des HIV-1 Komplet-Genoms einer neuen URF-Variante aus Oman

In dieser Arbeit sollten Komplet-Genomsequenzen der „URF-*new clade*“ vervollständigt und in phylogenetischen Analysen untersucht werden, um zu prüfen, ob sie auch über die gesamte Genomlänge eine eigene monophyletische Gruppe

bilden und damit die Annahme bestätigen, dass es sich bei den „URF-*new clade*“ Isolaten aus Oman um eine neue zirkulierende rekombinante Form handelt.

Um die Isolate der „URF-*new clade*“ untersuchen zu können, war es nötig, von beiden Patienten (13-5995 und 14-0875) die Sequenzanalyse zu wiederholen und zu erweitern, da entweder ein hoher Anteil an Sequenzambiguitäten oder zu wenig partielle Sequenz vorlag. Sequenzambiguitäten führen in phylogenetischen Analysen zu überbewerteten Sequenzunterschieden an den Positionen, an denen sie auftreten. Insbesondere Insertionen und/oder Deletionen, die in einzelnen aber nicht in allen Varianten der Quasispezies auftreten, führen zu ganzen Bereichen, deren Sequenz nicht eindeutig bestimmt werden kann.

Für das Isolat 13-5995 konnte in dieser Arbeit eine Komplet-Genomsequenz generiert werden, bei der nur noch ein kurzer Abschnitt von ca. 200 Bp nicht eindeutig ist. In diesem Bereich wurde die Hauptvariante editiert. Ähnliche Probleme gab es auch für den anderen Patienten (14-0875), bei dem aufgrund der hohen Anzahl von Ambiguitäten 359 Nukleotide nicht auswertbar waren, allerdings lag hier keine eindeutig bestimmbare majoritäre Variante vor. Daher wurde für dieses Isolat (14-0875) eine Komplet-Genomsequenz mit einer Lücke von diesen 359 Bp generiert. Die ermittelte Populationssequenz umfasst alle Bereiche der Rekombinationsstellen, die anhand von Untersuchungen zur Rekombinationsstruktur ermittelt werden konnten. Nur eine klonale Sequenzanalyse oder NGS können zur Aufklärung der fehlenden Bereiche verwendet werden.

Für die anschließenden phylogenetischen Analysen wurde im Sequenzalignment der fehlende 359 Bp Abschnitt entsprechend der Sequenz 14-0875 deletiert, um zu verhindern, dass das vorliegende *Gap* als Sequenzunterschied bewertet und daraus eine falsche Distanz zu den anderen Sequenzen abgeleitet wird.

Die Vervollständigung der beiden Komplet-Genomsequenzen der Isolate 13-5995 und 14-0875 und die anschließende phylogenetische Analyse (distanzbasiertes *Neighbor-Joining* Verfahren) der Populationssequenzen der drei „URF-*new clade*“ Isolate bestätigen das Ergebnis der Analysen der kurzen *pol*-Sequenzen. Auch in der phylogenetischen Analyse anhand annähernd vollständiger Genomsequenzen bilden die drei omanischen Isolate eine eigene monophyletische Gruppe, die sich signifikant (*Bootstrap* 100 %) von den Subtypreferenzsequenzen (beinhalten alle Subtypen sowie CRF01-CRF49) abspalten und in der auch keine über Blast-Suche ermittelte Sequenz zu finden ist. Die Baumtopologie zeigte eine genetische Verwandtschaft der „URF-*new clade*“ zu Subtyp G und G-Rekombinante Referenzsequenzen.

Die genetische Distanz von mehr als 8 % der drei „URF- *new clade*“ Sequenzen und ein signifikanter *Bootstrap*-Wert von 100 % am Knotenpunkt der drei Isolate sprechen dafür, dass es sich um epidemisch nicht verknüpfte HIV-1 Isolate handelt. Zur Bestätigung der eigenen Ergebnisse und zum Abgleich mit möglicherweise ähnlichen, aber noch nicht publizierten Sequenzen, wurden die drei Genomsequenzen der „URF-*new clade*“ der Internationalen HIV-Datenbank zur Sequenzüberprüfung übermittelt (<http://www.hiv.lanl.gov/>). In der Überprüfung zeigte sich ebenfalls keine signifikante Übereinstimmung mit bereits registrierten, auch noch nicht publizierten CRF oder URF. Dies führte zu dem Schluss, dass es sich hier bei den identifizierten „URF-*new clade*“ Viren tatsächlich um eine neue zirkulierende rekombinante Form handelt.

4.4 Rekombinationsstruktur der omanischen „URF-*new clade*“ Viren

Die Analysen der Rekombinationsstruktur der beiden neuen nun nahezu vollständigen Genomsequenzen aus Plasma zeigen die gleiche Rekombinationsstruktur wie das schon untersuchte angezüchtete Virusisolat (13-0346) auf⁵². Alle drei Viren bestehen überwiegend aus Subtyp G mit Anteilen von D im *pol*-Genombereich und im offenen Leserahmen des Vif-Proteins. Übereinstimmende Ergebnisse wurden sowohl mit RIP und mit dem *jpHMM* (GOBICS) erhalten. Die RIP-Analyse detektierte jedoch einen weiteren Subtyp D-Anteil in beiden neuen untersuchten Isolaten (13-5995 und 14-0875). Da die RIP-Software nur 12 Referenzsequenzen einbindet und im Gegensatz zum *jpHMM* von GOBICS weniger variabel ist, wird die Aussage der GOBICS-Software als valider eingestuft.

Noch nicht veröffentlichte Daten aus der Kooperation zwischen dem Robert Koch-Institut und dem omanischen CPHL zeigten, dass in Oman der Subtyp C prävalent ist, gefolgt von Subtypen A1 und der rekombinanten Form CRF01_AE. Isolate vom Subtyp G oder andere G-Rekombinanten wurden bislang nicht im Oman identifiziert. Es wird daher angenommen, dass sich dieses CRF nicht im Oman entwickelt hat.

Die mit der Blast-Suche als nächst verwandte identifizierten Isolate zeigen mit ca. 90 % eine für HIV geringe Sequenzidentität und die phylogenetischen Analysen zeigen, dass auch die HIV-Stämme, die größte Ähnlichkeit zur „URF-*new clade*“ besitzen, nicht eng verwandt sind. Diese als nächst ähnlich identifizierten Isolate stammen aus Kuba und Kamerun und besitzen hauptsächlich einen Subtyp G-Anteil. Analysen mit

einer molekularen Uhr oder Analysen von Teilbereichen der untersuchten Isolate können Aufschluss darüber geben, wie lange die omanische Form schon existiert und aus welchen Varianten sie hervorgegangen sein könnte.

4.5 Sequenzierung von *full-length* Genomen mit *Next Generation Sequencing*

Mit Hilfe des *Next Generation Sequencing* (NGS) ist eine bessere Auflösung von Sequenzambiguitäten möglich, wodurch minoritäre Varianten sensitiver in der Quasispezies detektiert werden können. Während die Sanger-Sequenzierung minoritäre Varianten nur ab einem Anteil von 20-25 % detektiert⁶³, ermöglicht die NGS-Methode eine Detektion minoritärer Varianten, die unter 0,5 % in der Quasispezies vorhanden sind⁴⁹.

Eine sensitivere Detektion von minoritären Varianten mit NGS war für das angezüchtete Isolat 13-0346 bei jedem gewählten Schwellwert zu sehen. Für das nicht angezüchtete Isolat aus Plasmavirus (13-5995) wurden überraschenderweise mit NGS weniger Ambiguitäten detektiert. Dies kann an der Tatsache liegen, dass PCR-Amplifikate aus getrennten PCR-Reaktionen für die jeweilige Sequenzierungsmethode genutzt wurden, bei denen Unterschiede im Priming in den ersten Zyklen der PCR zu den beobachteten Sequenzvarianzen führen können. Für das zweite Isolat aus Plasmavirus (14-0875) konnte keine Komplet-Genomsequenz aus der Sanger-Sequenzierung verwendet werden, um die NGS-Sequenzen zu kartieren und somit Aussagen über minoritäre Varianten zu treffen. Derzeit ist es in der Arbeitsgruppe noch nicht möglich, gegen zwei einzelne Teilsequenzen zu kartieren.

In der Arbeitsgruppe wurden anhand dieser Proben erstmalig HIV-Komplet-Genomsequenzen von Patientenproben ermittelt. Die Ergebnisse dieser Arbeit zeigen, dass es über *Next Generation Sequencing* möglich ist, Komplet-Genomsequenzen für HIV-Varianten zu erzeugen, um neue rekombinante Formen zu klassifizieren.

Im Rahmen der Etablierung von NGS in der Arbeitsgruppe sollten die Grenzen der Sensitivität für minoritäre Varianten dieser Methode ausgetestet werden. Daher wurden die Analysen der ermittelten Sequenzen mit unterschiedlichen Schwellwerten durchgeführt.

Mit dem Einsatz eines Grenzwertes von 10 % sollten mehr minoritäre Varianten als mit Sanger-Sequenzierung gefunden werden. Mit Hilfe von speziellen Programmen ist es möglich, diese minoritären Varianten zu rekonstruieren⁷¹ und phylogenetische Analysen sowie Untersuchungen zur Rekombinationsstruktur durchzuführen, um eine möglichst genaue Aufschlüsselung der Quasispeziesvarianten innerhalb eines infizierten Individuums zu erhalten.

Die Kartierung der mit NGS erzeugten *reads* anhand der Sanger-Sequenz war für Isolate, für die eine Sanger-Sequenz erzeugt werden konnte, erfolgreich. Des Weiteren konnten gute Übereinstimmungen der ermittelten Konsensussequenz festgestellt werden. Lediglich an zwei Positionen im Isolat 13-0346 wurden zwei Transversionen im Vergleich zu Sanger detektiert.

Sobald die Etablierung des NGS in der Arbeitsgruppe abgeschlossen ist, bietet es eine sehr gute, kostengünstigere und schnellere Alternative zur herkömmlichen Sanger-Sequenzierung für die Detektion von minoritären Resistenzmutationen.

4.6 Ausblick

In Zukunft könnten noch weitere und detailliertere Untersuchungen mit Isolaten der neuen CRF durchgeführt werden. Neben der Ermittlung weiterer *full-length* Sequenzen können Untersuchungen der CRF durchgeführt werden, anhand derer das Alter dieser CRF über eine molekulare Uhr bestimmt werden könnte, was genauere Informationen zum Ursprung dieser Form liefern kann. Die omanischen Kooperationspartner planen Prävalenzstudien zum Vorkommen von HIV-Subtypen, um die omanische Relevanz dieser Variante in der omanischen HIV-Epidemie beurteilen zu können.

Darüber hinaus kann die vier-Amplikon-PCR optimiert werden, um auch Isolate mit geringeren Viruslasten und weitere Isolate unterschiedlicher Subtypen amplifizieren zu können. Mit dem Fragment URF-*full-2*, welches den *pol*-Bereich beinhaltet, könnte in der Resistenzanalyse ein Amplikon erzeugt werden, das auf Resistenzmutationen untersucht wird, an Stelle der Amplifikation von partiellen Bereichen der *pol*-Region.

Für die Vervollständigung der nicht direkt mit Sangerverfahren sequenzierbaren Bereiche der *env*-Region sollte das URF-*full-3* Amplikon in einen geeigneten Vektor kloniert und in Form klonaler Sequenzen analysiert werden.

Mit Hilfe von *Next Generation Sequencing* Methoden können in kürzerer Zeit mehr Sequenzinformationen erzeugt werden und die Detektion minoritärer Varianten erfolgen. Weitere Etablierungs- und Validierungsschritte können in der Zukunft zur Ablösung der Sanger-Sequenzierung durch NGS-Methoden führen.

5. Zusammenfassung

Die globale HIV-1 Epidemie wird durch die geografisch unterschiedliche Verbreitung von 9 HIV-1 Subtypen (A-D, F-H, J, K) und derzeit 72 zirkulierenden rekombinanten Formen (CRF) charakterisiert. Die genetische Diversität von HIV-1 kann Auswirkungen auf die Sicherheit der HIV-1 Diagnostik, den Krankheitsverlauf, die Impfstoffentwicklung und auf den Erfolg der antiretroviralen Therapie haben.

In einer Kooperation zwischen dem „Central Public Health Laboratory“ des Omans und dem Robert Koch-Institut wurde bei omanischen Patienten mit Therapieversagen neben der genotypischen Resistenzanalyse auch der *pol*-Subtyp analysiert. Ca. 50 % der untersuchten Isolate konnten keinem definierten Subtypen oder einer bekannten zirkulierenden rekombinanten Form zugeordnet werden. Phylogenetische Analysen zeigten, dass 13 Isolate eine eigene monophyletische Gruppe (Clade) im Stammbaum bildeten, die sich in der Baumtopologie signifikant von allen bekannten Referenzviren abspaltet und als putative neue CRF eingestuft wurde. Um eine neue CRF von internationalen Experten definieren zu lassen, müssen von zwei Isolaten Komplet-Genomsequenzen vorliegen, eine dritte Sequenz kann partiell sein.

Ziel dieser Arbeit war es, die *full-length* Sequenzen von drei epidemisch nicht verknüpften Isolaten der putativ neu identifizierten „URF-*new clade*“ zu vervollständigen, um in abschließenden phylogenetischen Sequenzanalysen dieser HIV-Variante eine Klassifizierung als neue CRF durchzuführen.

Um ausreichend Virusmaterial zur Verfügung zu haben, wurden Virusanzuchten aus Plasma auf PBMC-Spenderzellen durchgeführt. Die HIV-Genome wurden mit einer etablierten „vier-Amplikon PCR“ amplifiziert. Zusätzlich wurden verschiedenen Protokolle einer „ein-Amplikon PCR“ ausgetestet, um zuverlässige Aussagen über inter-Subtyp-Rekombinationen treffen zu können. Die erhaltenen Sanger-Populationssequenzen wurden mit Konsensussequenzen anhand der neuen *Next Generation Sequencing* Technologie (NGS, Illumina) verglichen.

Es konnte für zwei Plasmaviren die Komplet-Genomsequenz bzw. eine partielle Sequenz ermittelt werden. Mit einer dritten vorhandenen Komplet-Genomsequenz wurden phylogenetische und Rekombinations-Analysen durchgeführt, die das Vorliegen einer neuen CRF bestätigen.

Der Vergleich der Sanger-Sequenzen mit über NGS ermittelten Sequenzdaten zeigte eine Übereinstimmung von mindestens 98,4 %. Mit NGS konnten mehr Ambiguitäten als mit der Sanger-Methode bestimmt werden. Es sind jedoch noch weitere NGS-

Validierungs-schritte nötig, um einen Grenzwert für die Unterscheidung von Hintergrund zu Ambiguitäten zu definieren, mit dem dann die NGS-Sequenzen der HIV-Quasispezies ausgewertet werden können.

Im Oman sind weitere Untersuchungen zur Ermittlung der Prävalenz und epidemischen Relevanz der neuen Variante in Planung.

6. Summary

The global HIV-1 epidemic is characterised by different geographic distribution of 9 subtypes (A-D, F-H, J, K) and currently 72 circulating recombinant forms (CRF). The genetic diversity of HIV-1 may have an impact on the safety of HIV-1 diagnostics, disease progression, vaccine development and on success of antiretroviral therapy as well.

During a cooperation between the Central Public Health Laboratory in Oman and Robert Koch-Institute Omani patients with therapy failure were analysed regarding *pol* subtype besides genotypic resistance mutations. Around 50 % of the samples couldn't be classified into a defined subtype or know circulating recombinant form. Phylogenetic analyses revealed that 13 of these samples represent an own monophyletic group (clade) in a genealogical tree, which is separated in an significant way from all known HIV-1 reference sequences and therefore was categorised as a putative new CRF. To classify a new CRF by international experts full length genome sequences of 2 samples need to be provided, a third sequence can be partial.

The aim of this study was to complete three full length sequences of the putative new identified „URF-new clade“ and perform final phylogenetic analyses of these HIV variant to classify a new CRF.

To gather enough viral material virus cultivation of plasma samples on PBMC were performed. HIV genomes were amplified with an established „four amplicon PCR“. In addition various protocols of an „one amplicon PCR“ were tested to get reliable information about inter-subtype recombination. Received sequences with Sanger sequencing were compared to consensus sequences with Next Generation Sequencing technology (NGS, Illumina).

Two complete genome sequences could be determined respectively a partial sequence. Together with an sequence determined in work done before phylogenetic analyses were performed to confirm a new CRF.

The comparison between Sanger sequencing and NGS showed a match of more than 98.4 % in sequence. With the use of NGS more ambiguities can be detected compared to Sanger sequencing. More NGS-Validation need to be done to define a threshold for the difference between background and ambiguities to analyse HIV quasi species with NGS.

In Oman more investigations about the prevalence and epidemic relevance of the new CRF are planed.

7. Appendix

7.1 Übersicht der Primerbindungsstellen der *full-length* PCR-Variationen

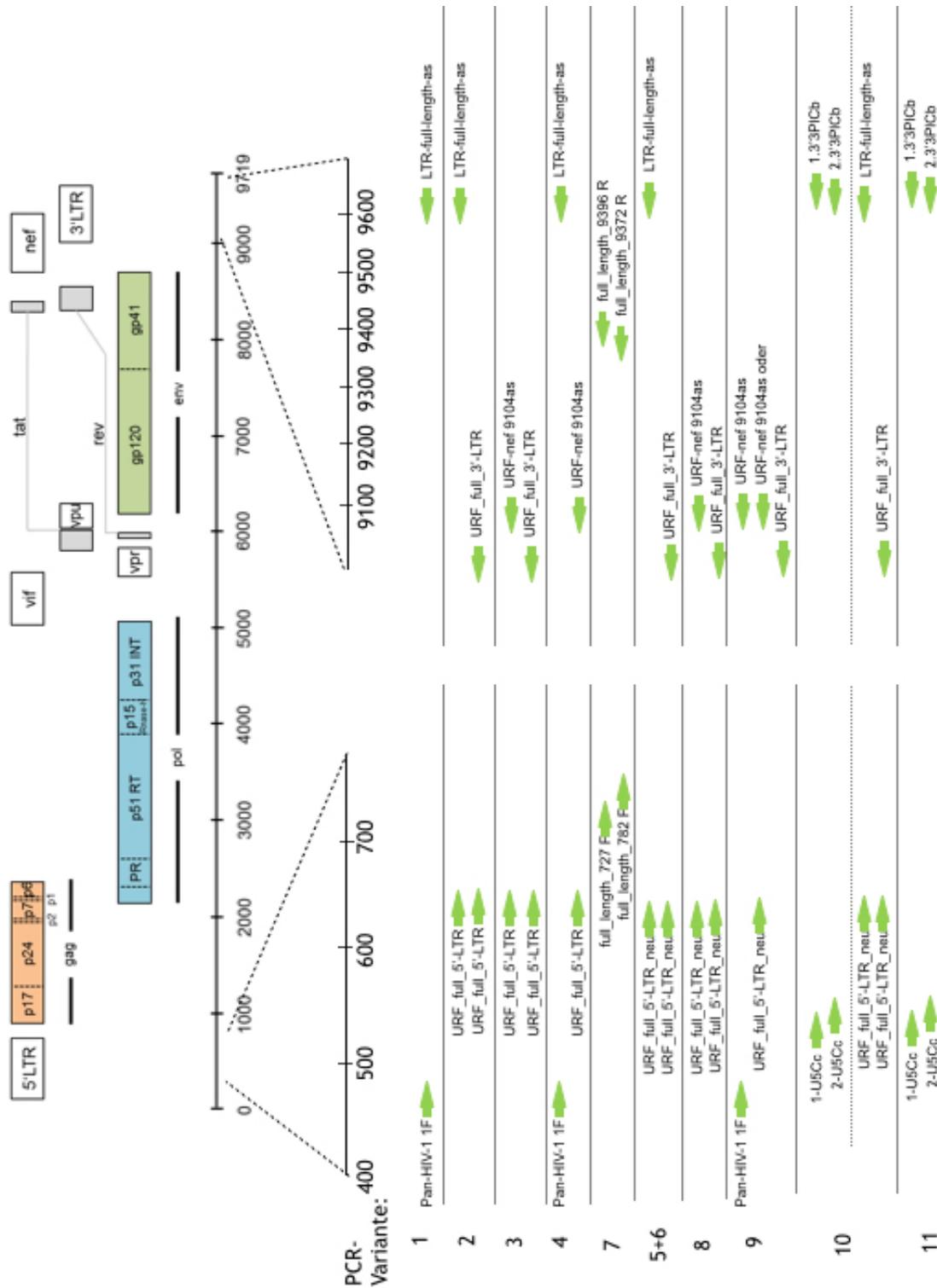


Abb. 16 Übersicht der Primerbindungsstellen für die *full-length* PCR-Varianten HxB2-Genom mit vergrößertem 5'- und 3'-Genombereich, in dem die Primerlokalisationen eingezeichnet sind (grüne Pfeile). Die 11 PCR Variationen sind durch horizontale Striche voneinander getrennt.

7.3 Stammbaum mit Genomsequenzen der „URF-new clade“ mit 170 Referenzsequenzen und Ergebnisse der BLAST-Suche

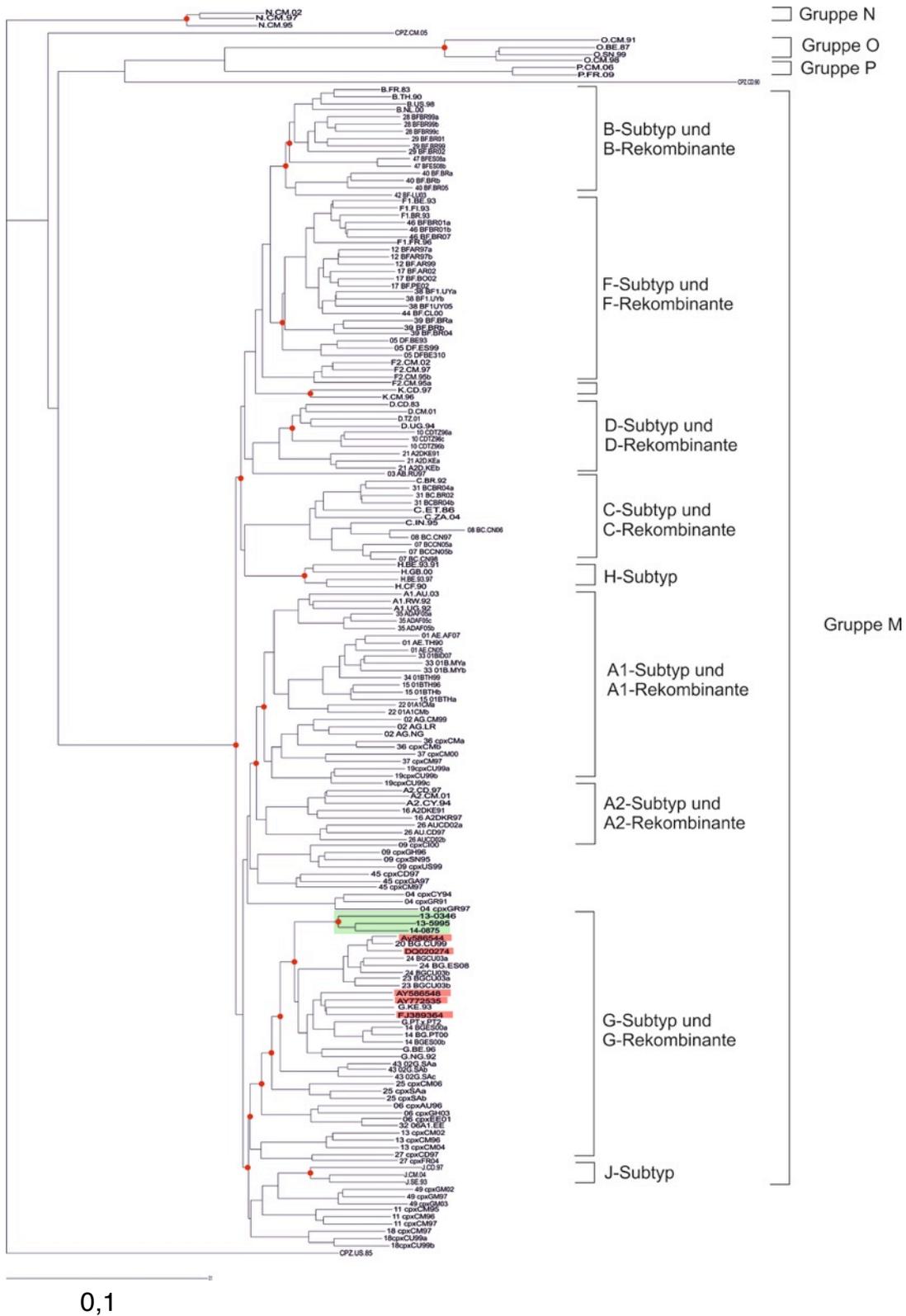


Abb. 18 Phylogenetische Analyse zur Klassifizierung einer neuen rekombinanten Form aus Oman

Neighbor-Joining Analyse mit dem kompletten Referenzpanel der Los Alamos HIV Datenbank (n=170), 3 Isolaten aus Oman (13-0346, 13-5995 und 14-0875, grün) sowie 5, mittels Blast-Suche ermittelte, höchst ähnliche Isolate (AY586544, AY586548, AY772535, DQ020272 und FJ389346, rot). Als Außengruppe wurde das Isolat CPZ.US.85 (Acc.-Nr.: AF103818) verwendet. Eine Bootstrap-Analyse wurde mit 1000 Datensätzen durchgeführt. Signifikante Knotenpunkte (Bootstrap > 70 %), die zur Einteilung der Subtyp-Kladen beitragen, sind rot markiert. Die Kladen der Subtypen und deren rekombinante Formen sind durch Klammern beschrieben. Die Einteilung der HIV-1 Gruppen ist ebenfalls durch Klammern gekennzeichnet. Der Distanzbalken (Scale-Bar) unter dem Baum mit einem Wert von 0,1 beschreibt die Astlänge, die einem Nukleotidaustausch von 10 % pro Position im Alignment entspricht.

Das Isolat 14-0875 ist nur partiell vorhanden, dadurch wurde das Referenzpanel auf die Länge von 14-0875 gekürzt (Lokalisation relativ zu HxB2: 536-6640, 6999-9621).

7.4 HIV-1 Referenzpanel der Los Alamos HIV Datenbank, 2010

Tab. 7 Übersicht der verwendeten Referenzsequenzen des Referenzpanels der Los Alamos HIV Datenbank, 2010

	Isolat	Acc. Nr.	Referenz- Bezeichnung im NJ-Baum	Herkunftsland
HIV-1 Gruppe M				
Subtypen				
A1	PS1044_Day0	DQ676872*	A1.AU.03	Australien
A1	92RW008	AB253421*	A1.RW.92	Rwanda
A1	92UG037	AB253429*	A1.UG.92	Uganda
A2	97CDKTB48	AF286238*	A2.CD.97	Dem. Republik Kongo
A2	01CM_1445MV	GU201516*	A2.CM.01	Kamerun
A2	94CY017_41	AF286237*	A2.CY.94	Zypern
B	HXB2_LAI_IIIB_BRU	K03455*	B.FR.83	Frankreich
B	671_00T36	AY423387*	B.NL.00	Niederlande
B	BK132	AY173951*	B.TH.90	Thailand
B	1058_11	AY331295*	B.US.98	USA
C	BR025_d	U52953*	C.BR.92	Brasilien
C	ETH2220	U46016*	C.ET.86	Äthiopien
C	95IN21068	AF067155*	C.IN.95	Indien
C	04ZASK146	AY772699*	C.ZA.04	Südafrika
D	ELI	K03454*	D.CD.83	Dem. Republik Kongo
D	01CM_4412HAL	AY371157*	D.CM.01	Kamerun
D	A280	AY253311*	D.TZ.01	Tansania
D	94UG114	U88824*	D.UG.94	Uganda
F1	VI850	AF077336*	F1.BE.93	Belgien
F1	93BR020_1	AF005494*	F1.BR.93	Brasilien
F1	FIN9363	AF075703*	F1.FI.93	Finnland
F1	96FR_MP411	AJ249238*	F1.FR.96	Frankreich
F2	02CM_0016BBY	AY371158*	F2.CM.02	Kamerun
F2	95CM_MP255	AJ249236*	F2.CM.95a	Kamerun
F2	95CM_MP257	AJ249237*	F2.CM.95b	Kamerun
F2	CM53657	AF377956	F2.CM.97	Kamerun
G	DRCBL	AF084936*	G.BE.96	Belgien
G	HH8793_12_1	AF061641*	G.KE.93	Kenia
G	92NG083	U88826*	G.NG.92	Nigeria
G	PT2695	AY612637*	G.PT.x.PT2	Portugal
H	VI991	AF190127*	H.BE.93.91	Belgien
H	VI997	AF190128*	H.BE.93.97	Belgien
H	056	AF005496	H.CF.90	Zentralafrikanische Republik
H	00GBAC4001	FJ711703*	H.GB.00	Großbritannien
J	J_97DC_KTB147	EF614151*	J.CD.97	Dem. Republik Kongo
J	04CMU11421	GU237072*	J.CM.04	Kamerun
J	SE9280_7887	AF082394	J.SE.93	Schweden
K	97ZR_EQTB11	AJ249235*	K.CD.97	Dem. Republik Kongo
K	96CM_MP535	AJ249239*	K.CM.96	Kamerun
CRF01_AE	569M	GQ477441*	01_AE.AF07	Afghanistan

	Isolat	Acc. Nr.	Referenz- Bezeichnung im NJ-Baum	Herkunftsland
CRF01_AE	05GX001	GU564221*	01_AE.CN05	China
CRF01_AE	CM240	U54771*	01_AE.TH90	Thailand
CRF02_AG	pBD6_15	AY271690*	02_AG.CM99	Kamerun
CRF02_AG	POC44951	AB485636*	02_AG.LR	Liberia
CRF02_AG	IBNG	L39106*	02_AG.NG	Nigeria
CRF03_AB	KAL153_2	AF193276*	03_AB.RU97	Russland
CRF04_cpx	94CY032_3	AF049337*	04_cpxCY94	Zypern
CRF04_cpx	GR11_97PVCH	AF119820	04_cpxGR91	Griechenland
CRF04_cpx	GR84_97PVMY	AF119819*	04_cpxGR97	Griechenland
CRF05_DF	VI961	AF076998	05_DF.BE93	Belgien
CRF05_DF	VI1310	AF193253*	05_DFBE310	Belgien
CRF05_DF	X492	AY227107*	05_DF.ES99	Spanien
CRF06_cpx	BFP90	AF064699	06_cpxAU96	Australien
CRF06_cpx	EE0359	AY535659*	06_cpxEE01	Estland
CRF06_cpx	03GH173_06	AB286851*	06_cpxGH03	Ghana
CRF07_BC	XJDC6431_2	EF368372*	07_BCCN05a	China
CRF07_BC	XJDC6441	EF368370*	07_BCCN05b	China
CRF07_BC	98CN009	AF286230	07_BC.CN98	China
CRF08_BC	nx2	HM067748	08_BC.CN06	China
CRF08_BC	97CNGX_6F	AY008715	08_BC.CN97	China
CRF09_cpx	00IC_10092	AJ866553*	09_cpxCI00	Elfenbeinküste
CRF09_cpx	96GH2911	AY093605*	09_cpxGH96	Ghana
CRF09_cpx	95SN1795	AY093603	09_cpxSN95	Senegal
CRF09_cpx	99DE4057	AY093607*	09_cpxUS99	USA
CRF10_CD	96TZ_BF061	AF289548	10_CDTZ96a	Tansania
CRF10_CD	96TZ_BF071	AF289549*	10_CDTZ96b	Tansania
CRF10_CD	96TZ_BF110	AF289550*	10_CDTZ96c	Tansania
CRF11_cpx	95CM_1816	AF492624*	11_cpxCM95	Kamerun
CRF11_cpx	96CM_4496	AF492623	11_cpxCM96	Kamerun
CRF11_cpx	MP818	AJ291718*	11_cpxCM97	Kamerun
CRF12_BF	A32879	AF408629*	12_BFAR97a	Argentinien
CRF12_BF	A32989	AF408630	12_BFAR97b	Argentinien
CRF12_BF	ARMA159	AF385936*	12_BF.AR99	Argentinien
CRF13_cpx	02CM_A1394	DQ845388*	13_cpxCM02	Kamerun
CRF13_cpx	04CM_632_28	DQ845387*	13_cpxCM04	Kamerun
CRF13_cpx	96CM_1849	AF460972	13_cpxCM96	Kamerun
CRF14_BG	X605	AF450096*	14_BGES00a	Spanien
CRF14_BG	X623	AF450097*	14_BGES00b	Spanien
CRF14_BG	00PTHDE10	GU230137	14_BG.PT00	Portugal
CRF15_01B	M169	DQ354120*	15_01BTH96	Thailand
CRF15_01B	99TH_MU2079	AF516184*	15_01BTHa	Thailand
CRF15_01B	99TH_R2399	AF530576*	15_01BTHb	Thailand
CRF16_A2D	KNH1271	AY945736	16_A2DKE91	Kenia
CRF16_A2D	97KR004	AF286239*	16_A2DKR97	Südkorea
CRF17_BF	AR02_ARG1139	EU581825*	17_BF.AR02	Argentinien
CRF17_BF	BO02_BOL119	EU581827*	17_BF.BO02	Bolivien

	Isolat	Acc. Nr.	Referenz- Bezeichnung im NJ-Baum	Herkunftsland
CRF17_BF	PE02_PCR0155	EU581828	17_BF.PE02	Peru
CRF18_cpx	CM53379	AF377959*	18_cpxCM97	Kamerun
CRF18_cpx	CU14	AY586541	18cpxCU99a	Kuba
CRF18_cpx	CU68	AY894993*	18cpxCU99b	Kuba
CRF19_cpx	CU29	AY588971*	19cpxCU99a	Kuba
CRF19_cpx	CU38	AY588970*	19cpxCU99b	Kuba
CRF19_cpx	CU7	AY894994	19cpxCU99c	Kuba
CRF20_BG	Cu103	AY586545*	20_BG.CU99	Kuba
CRF21_A2D	KNH1254	AY945737*	21_A2DKE91	Kenia
CRF21_A2D	KER2003	AF457051*	21_A2D.KEa	Kenia
CRF21_A2D	KSM4001	AF457072*	21_A2D.KEb	Kenia
CRF22_01A1	01CM_0001BBY	AY371159*	22_01A1CMa	Kamerun
CRF22_01A1	02CM_3097MN	GQ229529	22_01A1CMb	Kamerun
CRF23_BG	CB118	AY900571*	23_BGCU03a	Kuba
CRF23_BG	CB347	AY900572	23_BGCU03b	Kuba
CRF24_BG	CB378	AY900574	24_BGCU03a	Kuba
CRF24_BG	CB471	AY900575*	24_BGCU03b	Kuba
CRF24_BG	X2456_2	FJ670526*	24_BG.ES08	Spanien
CRF25_cpx	06CM_BA_040	EU693240*	25_cpxCM06	Kamerun
CRF25_cpx	J11233	EU697906*	25_cpxSAa	Saudi-Arabien
CRF25_cpx	J11451	EU697908	25_cpxSAb	Saudi-Arabien
CRF26_AU	02CD_KS069	FM877780*	26_AUCD02a	Dem. Republik Kongo
CRF26_AU	02CD_MBTB047	FM877782*	26_AUCD02b	Dem. Republik Kongo
CRF26_AU	97CD_KTB119	FM877777	26_AU.CD97	Dem. Republik Kongo
CRF27_cpx	97CDKTB49	AJ404325	27_cpxCD97	Dem. Republik Kongo
CRF27_cpx	04CD_FR_KZS	AM851091*	27_cpxFR04	Frankreich
CRF28_BF	BREPM12313	DQ085872*	28_BFBR99a	Brasilien
CRF28_BF	BREPM12609	DQ085873	28_BFBR99b	Brasilien
CRF28_BF	BREPM12817	DQ085874*	28_BFBR99c	Brasilien
CRF29_BF	BREPM16704	DQ085876*	29_BF.BR01	Brasilien
CRF29_BF	BREPM119	AY771590*	29_BF.BR02	Brasilien
CRF29_BF	BREPM11948	DQ085871	29_BF.BR99	Brasilien
CRF31_BC	110PA	EF091932	31_BC.BR02	Brasilien
CRF31_BC	04BR137	AY727526*	31_BCBR04a	Brasilien
CRF31_BC	04BR142	AY727527*	31_BCBR04b	Brasilien
CRF32_06A1	EE0369	AY535660*	32_06A1.EE	Estland
CRF33_01B	JKT194_C	AB547464*	33_01BID07	Indonesien
CRF33_01B	05MYKL007_1	DQ366659	33_01B.MYa	Malaysia
CRF33_01B	05MYKL045_1	DQ366662*	33_01B.MYb	Malaysia
CRF34_01B	OUR2478P	EF165541*	34_01BTH99	Thailand
CRF35_AD	05AF026	EF158043	35_ADAF05a	Afghanistan
CRF35_AD	05AF094	EF158040*	35_ADAF05b	Afghanistan
CRF35_AD	05AF095	EF158041*	35_ADAF05c	Afghanistan
CRF36_cpx	00CMNYU1162	EF087995	36_cpxCMa	Kamerun
CRF36_cpx	00CMNYU830	EF087994*	36_cpxCMb	Kamerun
CRF37_cpx	00CMNYU926	EF116594*	37_cpxCM00	Kamerun

	Isolat	Acc. Nr.	Referenz- Bezeichnung im NJ-Baum	Herkunftsland	
	CRF37_cpx	CM53392	AF377957	37_cpxCM97	Kamerun
	CRF38_BF1	UY04_3987	FJ213781*	38_BF1.UYa	Uruguay
	CRF38_BF1	UY04_4022	FJ213782	38_BF1.UYb	Uruguay
	CRF38_BF1	UY05_4752	FJ213780*	38_BF1UY05	Uruguay
	CRF39_BF	03BRRJ103	EU735534*	39_BF.BRa	Brasilien
	CRF39_BF	03BRRJ327	EU735536	39_BF.BRb	Brasilien
	CRF39_BF	04BRRJ179	EU735535*	39_BF.BR04	Brasilien
	CRF40_BF	04BRRJ115	EU735538*	40_BF.BRa	Brasilien
	CRF40_BF	04BRSQ46	EU735540*	40_BF.BRb	Brasilien
	CRF40_BF	05BRRJ200	EU735539	40_BF.BR05	Brasilien
	CRF42_BF	luBF_05_03	EU170155	42_BF-LU03	Luxemburg
	CRF43_02G	J11223	EU697904*	43_02G.SAa	Saudi-Arabien
	CRF43_02G	J11243	EU697907	43_02G.SAb	Saudi-Arabien
	CRF43_02G	J11456	EU697909*	43_02G.SAc	Saudi-Arabien
	CRF44_BF	CH80	FJ358521*	44_BF.CL00	Chile
	CRF45_cpx	97CD_MBF185	FN392874*	45_cpxCD97	Dem. Republik Kongo
	CRF45_cpx	97CM_MP814	FN392876	45_cpxCM97	Kamerun
	CRF45_cpx	97GA_TB45	FN392877*	45_cpxGA97	Gabun
	CRF46_BF	01BR087	DQ358801*	46_BFBR01a	Brasilien
	CRF46_BF	01BR125	DQ358802	46_BFBR01b	Brasilien
	CRF46_BF	07BR_FPS625	HM026456*	46_BF.BR07	Brasilien
	CRF47_BF	P1942	GQ372987	47_BFES08a	Spanien
	CRF47_BF	X2457_2	FJ670529*	47_BFES08b	Spanien
	CRF49_cpx	N18380	HQ385477*	49_cpxGM02	Gambia
	CRF49_cpx	N26677	HQ385479*	49_cpxGM03	Gambia
	CRF49_cpx	N28353	HQ385478	49_cpxGM97	Gambia
HIV-1 Gruppe O					
	O	ANT70	L20587	O.BE.87	Belgien
	O	MVP5180	L20571*	O.CM.91	Kamerun
	O	98CMU2901	AY169812	O.CM.98	Kamerun
	O	99SE_MP1300	AJ302647	O.SN.99	Senegal
HIV-1 Gruppe N					
	N	DJO0131	AY532635	N.CM.02	Kamerun
	N	YBF30	AJ006022	N.CM.95	Kamerun
	N	YBF106	AJ271370	N.CM.97	Kamerun
HIV-1 Gruppe P					
	P	U14788	HQ179987	P.CM.06	Kamerun
	P	RBF168	GU111555	P.FR.09	Frankreich
SIV-CPZ					
	CPZ	ANT	U42720	CPZ.CD.90	Dem. Republik Kongo
	CPZ	SIVcpzMT145	DQ373066	CPZ.CM.05	Kamerun
	CPZ	US_Marilyn	AF103818	CPZ.US.85	USA

7.5 Ergebnisse der BLAST-Suche

Tab. 8 Übersicht der Suchergebnisse mit BLAST

Accession-Nummer	Beschreibung	Herkunftsland	Identität
AY772535.1	HIV-1 Isolat 96VMABB55, Komplettgenom	Kamerun	89 %
AY586544.2	HIV-1 Isolat R77, Komplettgenom	Kuba	89 %
FJ389364.1	HIV-1 Isolat 314-40, partielles Genom	Kamerun	88 %
DQ020274.2	HIV-1 Isolat CB134, Komplettgenom	Kuba	88 %
AY586548.2	HIV-1 Isolat Cu85, Komplettgenom	Kuba	88 %

8. Literaturverzeichnis

1. Merson, M. H., O'Malley, J., Serwadda, D. & Apisuk, C. The history and challenge of HIV prevention. *Lancet* **372**, 475–488 (2008).
2. Hemelaar, J. The origin and diversity of the HIV-1 pandemic. *Trends in Molecular Medicine* **18**, 182–192 (2012).
3. Sharp, P. M. & Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold Spring Harbor Perspectives in Medicine* **1**, a006841–a006841 (2011).
4. UNAIDS. *Global Report*. (WHO Library Cataloguing-in-Publication Data, 2013).
5. Iyidogan, P. & Anderson, K. S. Current Perspectives on HIV-1 Antiretroviral Drug Resistance. *Viruses* **6**, 4095–4139 (2014).
6. Allers, K. *et al.* Evidence for the cure of HIV infection by CCR5 Δ 32/ Δ 32 stem cell transplantation. *Blood* **117**, 2791–2799 (2011).
7. Hütter, G. *et al.* Long-term control of HIV by CCR5 Δ 32/ Δ 32 stem-cell transplantation. *N. Engl. J. Med.* **360**, 692–698 (2009).
8. Jessen, H., Allen, T. M. & Streeck, H. How a single patient influenced HIV research--15-year follow-up. *N. Engl. J. Med.* **370**, 682–683 (2014).
9. More on Shift of HIV Tropism in Stem-Cell Transplantation with CCR5 Δ 32/ Δ 32 Mutation. *N. Engl. J. Med.* **371**, 2437–2438 (2014).
10. Centers for Disease Control (CDC). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *MMWR Morb. Mortal. Wkly. Rep.* **30**, 305–308 (1981).
11. Centers for Disease Control (CDC). Pneumocystis pneumonia--Los Angeles. *MMWR Morb. Mortal. Wkly. Rep.* **30**, 250–252 (1981).
12. Montagnier, L. 25 years after HIV discovery: Prospects for cure and vaccine. *Virology* **397**, 248–254 (2010).
13. Gallo, R. C. & Montagnier, L. Historical essay. Prospects for the future. *Science* **298**, 1730–1731 (2002).
14. Llano, D. A., Theyel, B. B., Mallik, A. K., Sherman, S. M. & Issa, N. P. Rapid and sensitive mapping of long-range connections in vitro using flavoprotein autofluorescence imaging combined with laser photostimulation. *J. Neurophysiol.* **101**, 3325–3340 (2009).

15. Barré-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868–871 (1983).
16. Barré-Sinoussi, F. HIV: A discovery opening the road to novel scientific knowledge and global health improvement. *Virology* **397**, 255–259 (2010).
17. Gallo, R. C. HISTORICAL ESSAY: The Early Years of HIV/AIDS. *Science* **298**, 1728–1730 (2002).
18. Case, K. *Nomenclature Human Immunodeficiency Virus*. (American College of Physicians, 1986).
19. Ratner, L. *et al.* Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**, 277–284 (1985).
20. Barré-Sinoussi, F. HIV as the cause of AIDS. *Lancet* **348**, 31–35 (1996).
21. Demeulemeester, J., Tintori, C., Botta, M., Debyser, Z. & Christ, F. Development of an AlphaScreen-Based HIV-1 Integrase Dimerization Assay for Discovery of Novel Allosteric Inhibitors. *Journal of Biomolecular Screening* **17**, 618–628 (2012).
22. Campbell, S. & Rein, A. In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *Journal of Virology* **73**, 2270–2279 (1999).
23. Delelis, O., Carayon, K., Saïb, A., Deprez, E. & Mouscadet, J.-F. Integrase and integration: biochemical activities of HIV-1 integrase. *Retrovirology* **5**, 114 (2008).
24. Piralla, A., Paolucci, S., Gulminetti, R., Comolli, G. & Baldanti, F. HIV integrase variability and genetic barrier in antiretroviral naïve and experienced patients. *Virology Journal* **8**, 149 (2011).
25. Ceccherini-Silberstein, F. *et al.* Characterization and structural analysis of HIV-1 integrase conservation. *AIDS Rev* **11**, 17–29 (2009).
26. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).
27. Kaplan, A. H., Manchester, M. & Swanstrom, R. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *Journal of Virology* **68**, 6782–6786 (1994).
28. Haseltine, W. A. Molecular biology of the human immunodeficiency virus type 1. *FASEB J.* **5**, 2349–2360 (1991).

29. Kirchhoff, F. Immune Evasion and Counteraction of Restriction Factors by HIV-1 and Other Primate Lentiviruses. *Cell Host and Microbe* **8**, 55–67 (2010).
30. Korber, B. *et al.* Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* **58**, 19–42 (2001).
31. Eberle, J. & Gürtler, L. HIV types, groups, subtypes and recombinant forms: errors in replication, selection pressure and quasispecies. *Intervirology* **55**, 79– 83 (2012).
32. Pathak, V. K. & Hu, W.-S. “Might as Well Jump!” Template Switching by Retroviral Reverse Transcriptase, Defective Genome Formation, and Recombination. *Seminars in VIROLOGY* **8**, 141–150 (1998).
33. Santos, A. F. & Soares, M. A. HIV Genetic Diversity and Drug Resistance. *Viruses* **2**, 503–531 (2010).
34. Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E. & Hammer, S. M. The challenge of HIV-1 subtype diversity. *N. Engl. J. Med.* **358**, 1590–1602 (2008).
35. Smyth, R. P., Davenport, M. P. & Mak, J. Virus Research. *Virus Research* **169**, 415–429 (2012).
36. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996).
37. Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168–1171 (1988).
38. Lau, K. A. & Wong, J. J. L. Current Trends of HIV Recombination Worldwide. *Infect Dis Rep* **5**, e4 (2013).
39. Delviks-Frankenberry, K. *et al.* Mechanisms and Factors that Influence High Frequency Retroviral Recombination. *Viruses* **3**, 1650–1680 (2011).
40. Sousa, J. D. de, Müller, V., Lemey, P. & Vandamme, A.-M. High GUD Incidence in the Early 20th Century Created a Particularly Permissive Time Window for the Origin and Initial Spread of Epidemic HIV Strains. *PLoS ONE* **5**, e9936 (2010).
41. Peeters, M., Jung, M. & Ayoub, A. The origin and molecular epidemiology of HIV. *Expert Rev Anti Infect Ther* **11**, 885–896 (2013).
42. Van Heuverswyn, F. *et al.* Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **444**, 164 (2006).
43. Wertheim, J. O. & Worobey, M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* **5**, e1000377 (2009).

44. Plantier, J.-C. *et al.* BRIEF COMMUNICATIONS. *Nature Publishing Group* 1–2 (2009). doi:10.1038/nm.2016
45. Vallari, A. *et al.* Confirmation of putative HIV-1 group P in Cameroon. *Journal of Virology* **85**, 1403–1407 (2011).
46. Perrin, L., Kaiser, L. & Yerly, S. Travel and the spread of HIV-1 genetic variants. *Lancet Infect Dis* **3**, 22–27 (2003).
47. Powell, R. L. R., Urbanski, M. M., Burda, S., Kinge, T. & Nyambi, P. N. High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. *J. Acquir. Immune Defic. Syndr.* **50**, 84–92 (2009).
48. Quiñones-Mateu, M. E., Avila, S., Reyes-Teran, G. & Martinez, M. A. Deep sequencing: becoming a critical tool in clinical virology. *J. Clin. Virol.* **61**, 9–19 (2014).
49. Archer, J. *et al.* Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. *PLoS ONE* **7**, e49602 (2012).
50. Bresk, A. Untersuchung von HIV-1 Subtypen und rekombinanten Formen (CRFs) in HIV-infizierten Patienten aus dem Oman .
51. Kenfack Guepi, E. Vollständige Genomanalyse neu identifizierter singulärer HIV-1 Rekombinanten (URF) aus Oman.
52. Luckau, L. Vollständige Genomanalyse einer neuen rekombinanten Form von HIV-1 aus Oman.
53. Hall, T. A. BioEdit a user-friendly biological sequence alignment editor and analysis program for Windows 95_98_NT. *Nucleic Acid Symposium Series No. 41* 95–98 (1999).
54. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).
55. Archer, J. *et al.* Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics* **13**, 47 (2012).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
57. Gall, A. *et al.* Universal Amplification, Next-Generation Sequencing, and Assembly of HIV-1 Genomes. *Journal of Clinical Microbiology* **50**, 3838–3844 (2012).

58. Rousseau, C. M. *et al.* Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *Journal of Virological Methods* **136**, 118–125 (2006).
59. Knoop, V. *Gene und Stammbäume.* (2009).
60. Siepel, A. C., Halpern, A. L., Macken, C. & Korber, B. T. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Research and Human Retroviruses* **11**, 1413–1416 (1995).
61. Zhang, M. *et al.* jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res.* **34**, W463–5 (2006).
62. Schultz, A.-K. *et al.* jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res.* **37**, W647–51 (2009).
63. Palmer, S. *et al.* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *Journal of Clinical Microbiology* **43**, 406–413 (2005).
64. Manak, M. *et al.* Pilot Studies for Development of an HIV Subtype Panel for Surveillance of Global Diversity. *AIDS Research and Human Retroviruses* **28**, 594–606 (2012).
65. Vanders, R. L., Murphy, V. E., Gibson, P. G., Hansbro, P. M. & Wark, P. A. B. CD8 T cells and dendritic cells: key players in the attenuated maternal immune response to influenza infection. *J Reprod Immunol* (2014). doi:10.1016/j.jri. 2014.09.051
66. Wei, H. *et al.* Near full-length genomic characterization of a novel HIV type 1 CRF07_BC/01_AE recombinant in men who have sex with men from Sichuan, China. *AIDS Research and Human Retroviruses* **29**, 1173–1176 (2013).
67. Li, Z. *et al.* Tracing the origin and history of HIV-1 subtype B' epidemic by near full-length genome analyses. *AIDS* **26**, 877–884 (2012).
68. Thomson, M. M., Casado, G., Posada, D., Sierra, M. & Najera, R. Identification of a novel HIV-1 complex circulating recombinant form (CRF18_cpx) of Central African origin in Cuba. *AIDS* **19**, 1155–1163 (2005).
69. Casado, G., Thomson, M. M., Sierra, M. & Najera, R. Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. *J. Acquir. Immune Defic. Syndr.* **40**, 532–537 (2005).
70. Yamaguchi, J. *et al.* Identification of new CRF43_02G and CRF25_cpx in Saudi Arabia based on full genome sequence analysis of six HIV type 1 isolates. *AIDS Research and Human Retroviruses* **24**, 1327–1335 (2008).

71. Jayasundara, D. *et al.* ViQuaS: An improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu754

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die Masterarbeit ohne fremde Hilfe angefertigt und nur die im Literaturverzeichnis angeführten Quellen und Hilfsmittel benutzt habe.

....., den

Ort

Datum

Unterschrift

Danksagung

Herrn PD Dr. rer. nat. Norbert Bannert und Frau Dr. rer. nat. Claudia Kücherer danke ich für die Möglichkeit meine Masterarbeit im Fachgebiet 18 „HIV und andere Retroviren“ anfertigen zu können. Claudia danke ich ebenfalls für die Anfertigung des Gutachtens dieser Arbeit.

Herrn Prof. Dr. Bernd Müller-Röber danke ich für die Übernahme der internen Betreuung meiner Masterarbeit.

Bei Dr. rer. nat. Sybille Somogyi bedanke ich mich für die Betreuung während meiner gesamten Zeit am Robert Koch-Institut und für die Betreuung meiner Masterarbeit.

Fr. Dr. Karolin Meixenberger und Britta Altmann danke ich für die Unterstützung in der Thematik des Next Generation Sequencing.

Ganz lieber Dank geht an das gesamte HIV-Studienlabor für die Unterstützung im Labor. Danke: Katrin Arndt, Hanno von Spreckelsen und Luise Luckau für die angenehme Atmosphäre und den Spaß den wir zusammen hatten.

Hiermit möchte ich auch meinen Freunden danken, die viel Verständnis in dieser Zeit für mich aufgebracht haben.

Als letztes möchte ich mich bei Gilbert bedanken, danke für deine Unterstützung und dein

Verständnis. Danke, dass du mein Anker bist ;)