

Martin-Luther-Universität Halle-Wittenberg

Naturwissenschaftliche Fakultät I
Institut für Biochemie und Biotechnologie



Vollständige Genomanalyse einer neuen rekombinanten Form von HIV-1 aus Oman

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

vorgelegt von

Luise Luckau

angefertigt am

Robert Koch-Institut, Berlin
Fachgebiet 18 „HIV und andere Retroviren“

Erstgutachter:

Prof. Dr. rer. nat. S.- E. Behrens

Zweitgutachter:

PD Dr. rer. nat. N. Bannert

eingereicht am:

25.09.2014

Abkürzungsverzeichnis

AccNo.	Accession Number
AIDS	acquired immunodeficiency syndrome (erworbenes Immundefektsyndrom)
as	antisense
bp	Basenpaare
bzw.	beziehungsweise
CA	Capsidprotein
ca.	circa
cDNA	complementary DNA (komplementäre DNA)
CRF	circulating recombinant form
cpx	complex
d.h.	das heißt
DMSO	Dimethylsulfoxid
DNA	deoxyribonucleic acid (Desoxyribonukleinsäure)
dNTP	Desoxyribonukleosid-5'-triphosphat
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylendiamintetraessigsäure
env	envelope
g	Gramm
gag	group-specific antigen
GOBICS	Göttingen Bioinformatics Compute Server
gp160	Glykoprotein
h	Stunde
HCl	Chlorwasserstoffsäure
HIV	human immunodeficiency virus (humanes Immundefizienzvirus)
HTLV	humanes T-lymphotropes Virus
INT	Integrase
jpHMM	jumping profile Hidden Markov Model
kb	Kilobasen
Konz.	Konzentration
l	Liter
LI	Linkprotein
LTR	long terminal repeat

M	Molar
MA	Matrixprotein
MENA	Middle East and North Africa (Nahost und Nordafrika)
mg	Milligramm
µg	Mikrogramm
MgCl ₂	Magnesiumchlorid
min	Minute
ml	Milliliter
µl	Mikroliter
mM	Millimolar
µM	Mikromolar
n	Anzahl
NaOH	Natriumhydroxid
NC	Nukleocapsid
nef	negative regulatory factor
ng	Nanogramm
nm	Nanometer
Nr.	Nummer
nt	Nukleotid
ORF	open reading frame (offener Leserahmen)
PBMC	peripheral blood mononuclear cell (mononukleare Zellen des peripheren Blutes)
PCR	polymerase chain reaction (Polymerasekettenreaktion)
pol	polymerase
PR	Protease
bzgl.	bezüglich
rev	regulator of expression of virion proteins
RH	RNase H (Ribonuklease H)
RIP	Recombinant Identification Program
RNA	ribonucleic acid (Ribonukleinsäure)
rpm	rotations per minute (Umdrehungen pro Minute)
RT	Reverse Transkriptase (Reverse Transkription)
s	sense
sec	Sekunde
SIV	simian immunodeficiency virus (Immundefizienzvirus der Affen)
TAE	Tris-Acetat-EDTA
tat	transactivator of transcription

Tris	Tris-(hydroxymethyl)-aminomethan
tRNA	Transfer-RNA
U	Unit
UNAIDS	Joint United Nations Program on HIV/AIDS
URF	unique recombinant form (singuläre rekombinante Form)
usw.	und so weiter
UV	ultraviolett
vif	viral infectivity factor
VL	Viruslast
vpr	viral protein r
vpu	viral protein u
z.B.	zum Beispiel

Inhaltsverzeichnis

1	Einleitung	1
1.1	Entdeckung und Herkunft von HIV	1
1.2	Morphologie und Genomstruktur von HIV-1	2
1.3	Genetische Diversität von HIV-1.....	4
1.4	Genetische Klassifizierung von HIV-1.....	5
1.4.1	Subtypeneinteilung von HIV-1 Gruppe M.....	5
1.4.2	Rekombinante Formen in der HIV-1 Gruppe M.....	6
1.5	Epidemiologie von HIV-1	7
1.5.1	Geographische Verteilung der Subtypen und Mosaikviren der HIV-1 Gruppe M	7
1.5.2	Die HIV-1 Epidemie in der MENA-Region.....	8
1.5.3	Eckdaten zur HIV-Epidemie in Oman	8
1.6	Zielsetzung.....	9
2	Material.....	11
2.1	Untersuchungsmaterialien und Patientenproben.....	11
2.2	Referenzmaterialien.....	11
2.3	Reagenzien	12
2.3.1	Chemikalien	12
2.3.2	Puffer für die Agarose-Gelelektrophorese.....	12
2.4	Kommerzielle Kits	14
2.5	Geräte und Verbrauchsmaterialien	14
2.6	Software.....	15
2.7	Oligonukleotide.....	15
3	Methoden.....	17
3.1	Extraktion viraler RNA aus EDTA-Plasma	17
3.2	Full-length cDNA-Synthese von HIV-1	17
3.3	Amplifikation des Kompletengenoms von HIV-1 („vier-Amplikon-PCR“).....	19
3.4	Agarose-Gelelektrophorese zur Qualifizierung und Quantifizierung von DNA.....	21
3.5	Aufreinigung der PCR-Produkte	21
3.6	Sequenzierung der PCR-Produkte.....	22

3.7	Ermittlung der Kompletengenomsequenz	23
3.8	Phylogenetische Analysen.....	23
3.9	Analyse der rekombinanten Genomstruktur	24
3.9.1	RIP (<i>Recombination Identification Program</i>)	25
3.9.2	Springendes Profil-Hidden-Markov-Modell (<i>jumping profile Hidden Markov Model, jpHMM</i>)	25
4	Ergebnisse	27
4.1	Amplifikation des Kompletengenoms einer neuen rekombinanten Form (<i>URF-new clade</i>) von HIV-1 aus Oman („vier-Amplikon-PCR“).....	27
4.1.1	Amplifikation des 3'-Genombereichs (<i>URF-full-3</i> und <i>URF-full-4</i>) von <i>URF-new clade</i> HIV-1 aus Virusisolat.....	27
4.1.2	Amplifikation des Kompletengenoms (<i>URF-full-1</i> bis <i>URF-full-4</i>) von <i>URF-new clade</i> HIV-1 aus EDTA-Plasma.....	28
4.2	Direktsequenzierung der PCR-Fragmente (<i>URF-full-1</i> bis <i>URF-full-4</i>) von <i>URF-new clade</i> HIV-1 Isolaten aus Oman	31
4.2.1	Direktsequenzierung des 3'-Genombereichs (<i>URF-full-3</i> und <i>URF-full-4</i>) von <i>URF-new clade</i> HIV-1 aus Virusisolat.....	33
4.2.2	Direktsequenzierung des Kompletengenoms (<i>URF-full-1</i> bis <i>URF-full-4</i>) von <i>URF-new clade</i> HIV-1 aus EDTA-Plasma.....	33
4.2.3	Sequenzvergleich der überlappenden Regionen zwischen <i>URF-full-1/URF-</i> <i>full-2</i> , <i>URF-full-2/URF-full-3</i> und <i>URF-full-3/URF-full-4</i>	35
4.3	Genetische Klassifizierung von <i>URF-new clade</i> Isolaten von HIV-1 aus Oman anhand der Kompletengenomsequenz	38
4.4	Analyse der rekombinanten Genomstruktur der <i>URF-new clade</i> Isolate aus Oman	40
5	Diskussion	45
5.1	Amplifikation und Sequenzierung des HIV-1 Kompletengenoms einer neuen URF-Variante aus Oman	45
5.2	Genetische Klassifikation einer neu identifizierten URF-Variante aus Oman.....	49
5.3	Ausblick.....	52
6	Zusammenfassung	54
7	Summary	55

8	Anhang	56
8.1	Alignment der Kompletengenomsequenzen der <i>URF-new clade</i> mit der Referenzsequenz HXB2 (K03455)	56
8.2	HIV-1 Referenzsequenzen aus der Los Alamos HIV Sequenzdatenbank, Ausgabe 2010.....	63
8.3	Stammbaum der Kompletengenomsequenzen von <i>URF-new clade</i> HIV-1 aus Oman mit 170 Referenzsequenzen.....	67
9	Literaturverzeichnis.....	69

1 Einleitung

1.1 Entdeckung und Herkunft von HIV

Im Jahre 1981 wurde erstmalig in New York City und Kalifornien von ungewöhnlichen vermehrten Krankheitsfällen berichtet. Es handelte sich um homosexuelle Männer, die an Lungenentzündungen (*Pneumocystis carinii*) in Kombination mit anderen lebensbedrohlichen opportunistischen Erkrankungen und der selten auftretenden malignen Tumorart Kaposi-Sarkom erkrankten (1). Diese Krankheitssymptome wurden später aufgrund eines geschwächten Immunsystems als *erworbenes Immundefektsyndrom (acquired immunodeficiency syndrome, AIDS)* zusammengefasst (1). Das verursachende Retrovirus wurde 1983 von den Wissenschaftlern Luc Montagnier, Françoise Barré-Sinoussi und Robert Gallo aus den Lymphknoten von AIDS-Patienten isoliert und als *lymphadenopathy associated virus (LAV)* oder zunächst auch als *humanes T-lymphotropes Virus 3 (HTLV-III)* bezeichnet (2). 1986 wurde das Virus vom Internationalen Komitee für Virus-Taxonomie als *HIV (human immunodeficiency virus, humanes Immundefizienzvirus)* benannt (3). Im gleichen Jahr wurde aus westafrikanischen AIDS-Patienten ein weiterer HIV-Typ isoliert, der als HIV-2 bezeichnet wurde. Demnach wurde der zuerst entdeckte HIV-Typ als HIV-1 benannt. Trotz der ähnlichen Genomorganisation und Transmissionswege der beiden HIV-Typen, spielen HIV-2-Infektionen in der HIV-Epidemie eine untergeordnete Rolle, da sie vorwiegend auf Westafrika begrenzt sind (4). Dagegen breitete sich HIV-1 pandemisch aus und ist das ätiologische Agens für AIDS (4). Der Erreger kann über Körperflüssigkeiten wie Blut, Muttermilch, Scheidensekret und Sperma übertragen werden (5, 6).

Der Ursprung von HIV-1 geht auf West- und Zentralafrika zurück. Nach dem derzeitigen Kenntnisstand geht man davon aus, dass die Immunschwächeviren der Affen (*simian immunodeficiency virus, SIV*) durch die Jagd und den Verzehr von nicht-humanen Primaten, in den frühen 1900er Jahren durch unabhängige zoonotische Transmissionen auf den Menschen übertragen wurden (7, 8). Durch vier unabhängige Transmissionsereignisse sind innerhalb der menschlichen Population vier genetische Gruppen von HIV-1 entstanden: M, N, O und P, deren Ursprung in Kamerun liegen (9). Die HIV-1 Gruppen M und N stammen direkt aber unabhängig von *SIVcpz* aus dem Schimpansen *Pan troglodytes troglodytes* ab. Die HIV-1 Gruppe O und die neu entdeckte Gruppe P gingen aus *SIVgor* aus Gorillas (*Gorilla gorilla gorilla*) hervor (10). HIV-2 ist dagegen phylogenetisch eng verwandt mit *SIVsm* der Raubmangabes (*Cercocebus atys*) (7).

1.2 Morphologie und Genomstruktur von HIV-1

HIV-1 ist morphologisch ein spherisch geformtes und membranumhülltes Virus mit einem Durchmesser von ca. 100-120 nm, das zwei Moleküle einzelsträngiger RNA als Genom enthält (11, 12). Da das RNA-Genom in DNA revers transkribiert wird, gehört HIV-1 zur Familie der Retroviren und wird dem Genus der Lentiviren zugeordnet (13). Das RNA-Genom von HIV-1 ist ca. 10 Kilobasen (kb) groß und wird an beiden Enden von LTR-Elementen (*long terminal repeat*) flankiert (12, 13). Essentiell für den Aufbau des retroviralen Partikels sind die viralen Strukturproteine, die von den Genomregionen *gag* (*group-specific antigen*) und *env* (*envelope*) kodiert werden sowie die viralen Enzyme, die in der *pol* (*polymerase*)-Genomregion kodiert sind (14). Die Morphologie und die Genomstruktur von HIV-1 sind in Abbildung 1 schematisch dargestellt.

Der *gag*-Genombereich kodiert die Matrixproteine (MA) p17, das Capsidprotein (CA) p24, das Nukleocapsid (NC) p7 und die Linkproteine (LI) p6, die aus dem myristoylierten Vorläuferprotein p55 durch proteolytische Spaltungen der viralen Protease hervorgehen (15). Die Matrixproteine p17 bilden die symmetrisch aufgebaute äußere Matrix, die sich unter der Virushülle befindet und das innere konische Capsid wird aus den Capsidproteinen p24 gebildet (12). Darin enthalten sind das virale RNA-Genom und Enzyme, die mit den Nukleocapsidproteinen p7 einen Komplex bilden können.

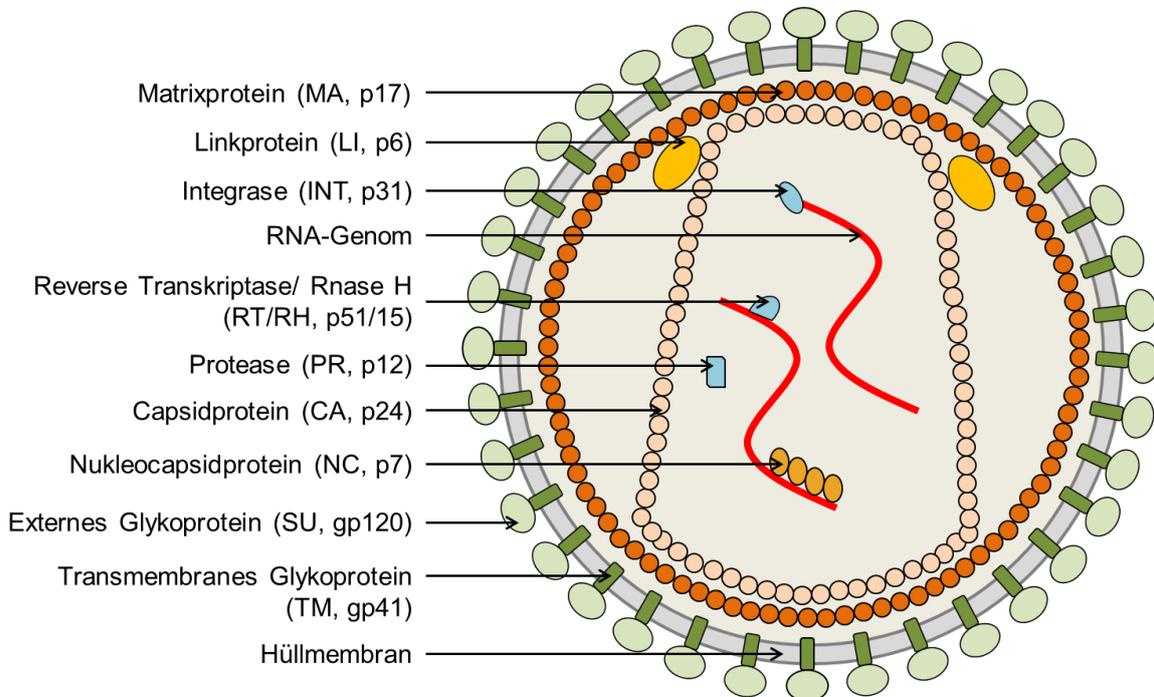
Die Genomregion *pol* kodiert die viralen Enzyme Protease (PR, p12), Reverse Transkriptase mit RNase H (RT/RH, p51/15) und Integrase (INT, p31) (12, 15). Diese Enzyme werden als Gag-Pol Vorläufer-Polypeptid durch eine ribosomale Leserasterverschiebung (*Frameshift*) am 3' Ende von Gag generiert und in funktionale Enzyme durch die Protease prozessiert (16). Die Reverse Transkriptase schreibt das virale RNA-Genom in DNA um, wobei durch die integrale RNase H-Aktivität der Reversen Transkriptase die RNA des RNA-DNA-Hybrides abgebaut wird (13, 17). Die generierte doppelsträngige DNA-Kopie wird anschließend von der Integrase in das Wirtsgenom integriert und wird dann als Provirus bezeichnet (18, 19).

Die *env*-Genomregion kodiert die Hüll-Glykoproteine, die sich aus einer Ectodomäne gp120 (SU) und einer transmembranen Domäne gp41 (TM) zusammensetzen (15). Die Domänen werden proteolytisch aus dem generierten Vorläuferprotein gp160 prozessiert und werden als Trimere an die Hüllmembran transportiert (12). Sie sind verantwortlich für die rezeptorvermittelte Fusion von viraler und zellulärer Membranen, um eine Infektion auszulösen (12, 20, 21).

Das HIV-1 RNA-Genom kodiert zudem mehrere regulatorische und akzessorische Proteine. Die regulatorischen Proteine Tat (*transactivator of transcription*) und Rev (*regulator of expression of virion proteins*) setzen sich jeweils aus zwei Exons zusammen und sind essentiell für die Replikation von HIV (12, 15). Die Proteine Vif (*viral infectivity factor*), Vpr (*viral protein r*) Vpu (*viral protein u*) und Nef (*negative regulatory factor*) werden als

akzessorische Proteine bezeichnet, da sie für die virale Replikation nicht zwingend notwendig sind (15). Sie erhöhen die Infektiosität der Viruspartikel und/ oder unterstützen die HIV-Freisetzung und -Übertragung (12).

A



B

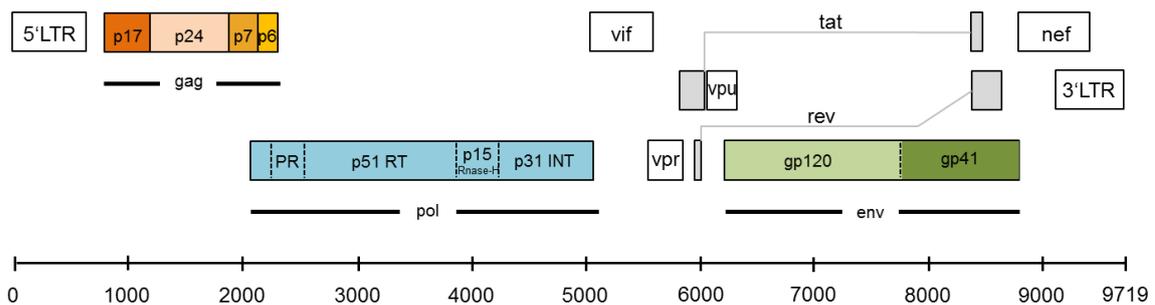


Abbildung 1: Morphologie und Genomstruktur von Hiv-1

A: Schematischer Aufbau eines HIV-1 Partikels. Hüllmembran mit externen (SU, gp120) und transmembranen (TM, gp41) Glykoproteinen, Matrixproteine (MA, p17), konisches Capsid aus Capsidproteinen (CA, p24), Linkproteine (LI, p6), zwei virale RNA-Genome im Komplex mit Nukleocapsidproteinen (NC, p7) und die Enzyme Reverse Transkriptase/ RnaseH (RT/RH, p51/15), Protease (PR, p12) und Integrase (INT, p31)

B: Genomstruktur von HIV-1. Das HIV-1 RNA-Genom mit ca. 10 kb kann in drei Leserastern translatiert werden. Es wird am 5'- und 3'-Ende von den LTR-Elementen (*long terminal repeat*) flankiert. Die Farben der dargestellten offenen Leserastern (ORF, *open reading frame*) entsprechen den Farben der Proteine im HIV-1 Partikel (A). Orange: ORF *gag* (*group specific antigen*), Grün: ORF *env* (*envelope*), Blau: ORF für die Enzyme der *pol* (*polymerase*)-Region, Grau: ORF für die akzessorischen Gene *vif* (*virion infectivity factor*), *vpr* (*viral protein r*), *vpu* (*viral protein u*), *nef* (*negative regulatory factor*) und für die regulatorischen Gene *tat* (*transactivator of transcription*) und *rev* (*regulator of expression of virion proteins*). Die Exons der ORFs *tat* (*tat1*, *tat2*) und *rev* (*rev1*, *rev2*) sind symbolisch jeweils durch Striche miteinander verbunden.

1.3 Genetische Diversität von HIV-1

Zu Beginn einer HIV-Infektion befindet sich im infizierten Organismus eine relativ homogene virale Population (22). Erst im Verlauf einer typischen HIV-Infektion entwickelt sich ein breites Spektrum von unterschiedlichen Virusvarianten innerhalb eines Individuums (22). So können die Nukleotidsequenzen im *env*-Genombereich innerhalb eines HIV-Infizierten um mehr als 10 % voneinander abweichen (23). HIV ist demnach eine Quasispezies, die die Gesamtpopulation von Virusvarianten in einem HIV-Infizierten umfasst (23, 24). Die Diversität der Virusvarianten in einer Quasispezies entwickelt sich durch einen natürlichen Selektionsdruck, der zu einer schnellen Anpassung einer Viruspopulation an den speziellen immunologischen Aktivitäten des Wirts führt sowie zur Entwicklung von Impfstoffversagen und Resistenzen gegenüber antiretroviralen Medikamenten (25).

Das Ausmaß der Sequenzvarianz von HIV ist dabei abhängig von der untersuchten Genomregion. Hochkonservierte Genomregionen sind zum Beispiel Teile der LTR-Region, das aktive Zentrum der Reversen Transkriptase und Integrase sowie das p24 Protein, wohingegen der *env*-Genombereich hypervariable Abschnitte trägt (26, 27). Innerhalb eines HIV-1 Subtyps können sich Aminosäuresequenzen im *env*-Genombereich durchschnittlich um 17 % und im *gag*-Genombereich um 8 % voneinander unterscheiden (23). Zwischen verschiedenen HIV-1 Subtypen, zum Beispiel zwischen Subtyp A und B, können die Aminosäuresequenzen im *env*-Gen um 20-36 % und im *gag*-Gen um 15-22 % voneinander abweichen (23). In der konserviertesten Genomregion *pol* kann der Aminosäuresequenzunterschied zwischen den Subtypen 9-11 % betragen (28).

Durch die sehr schnelle und enorme virale Partikelproduktion (10^{10} HIV-1 Viruspartikel/ Tag) werden Millionen von Virusvarianten in einem infizierten Individuum an einem einzigen Tag generiert (29). Diese charakteristische genetische Diversität des HIV-RNA-Genoms wird vor allem durch die hohe Replikationsrate der fehleranfälligen Reverse Transkriptase verursacht (24, 30). Aufgrund der fehlenden Korrekturfunktion der Reversen Transkriptase werden sehr häufig Basensubstitutionen, Rekombinationsereignisse sowie Insertionen und Deletionen (*indels*) von Genomabschnitten erzeugt, was zu einem schnellen Anstieg von divergenten Virusvarianten führt (23). Die Rate für eine Nukleotid-Substitution beträgt etwa 10^{-4} pro Nukleotid pro Replikationszyklus, was ca. einer Nukleotid-Substitution pro Genom während eines einzigen Replikationszyklus entspricht (31, 32). Charakteristisch für das Reverse Transkriptase-Enzym ist die geringe Bindungsaffinität zum RNA-*Template*, weshalb das intra- und intermolekulare "Springen" der Reversen Transkriptase bei der Genomreplikation möglich ist, was auch als „*Template switching*“ bezeichnet wird (33, 34). Schätzungsweise können 3-12 *Template*-Wechsel pro Genom pro Replikationszyklus stattfinden (35). Das intramolekulare *Template switching* der Reversen Transkriptase auf einem RNA-Molekül verursacht Mutationen wie Insertionen, Deletionen und Duplikationen von Genomabschnitten

(34, 35). Da jedes retrovirale Partikel zwei Kopien einzelsträngiger RNA enthält, ist das intermolekulare „Springen“ der Reversen Transkriptase zwischen den beiden RNA-*Templates* möglich (34). Bei einem co-infizierten Individuum, zum Beispiel mit zwei unterschiedlichen HIV-1 Subtypen können im Organismus heterozygote RNA-Virenpartikel generiert werden (35). Durch das intermolekulare *Template switching* zwischen mindestens zwei unterschiedlichen RNA-Genotypen können neue rekombinante Virenvarianten entstehen, die sehr komplex aus mehreren Subtypanteilen aufgebaut sein können (31, 34, 35).

1.4 Genetische Klassifizierung von HIV-1

Wie zuvor beschrieben sind die hohe Fehlerrate und die Rekombinationseigenschaften verbunden mit einer hohen Replikationsrate der Reversen Transkriptase die Ursache für die stark ausgeprägte genetische Variabilität von HIV-1 (36, 37). Basierend auf phylogenetischen Analysen wird HIV-1 in vier genetische Gruppen eingeteilt: Hauptgruppe M (*major*), Gruppe N, Gruppe O und die Gruppe P, die durch vier unabhängige Transmissionsereignisse von Primaten auf den Menschen entstanden sind (9, 31). Durch die HIV-1 Übertragung von Mensch zu Mensch wurden innerhalb der Gruppe M-Viren verschiedene Subtypen und Mosaikviren generiert, die sich phylogenetisch voneinander unterscheiden (38).

1.4.1 Subtypeneinteilung von HIV-1 Gruppe M

Die pandemisch relevante HIV-1 Gruppe M in der ~ 95 % der weltweiten HIV-1 Infektionen vertreten sind, wird in neun Subtypen (A-D, F-H, J, K) unterteilt (38, 39). Die Subtypen A und F werden weiterhin in die Sub-Subtypen A1 bis A4 und F1 bis F2 differenziert (40). Anfänglich wurden ausschließlich basierend auf *env*-Sequenzen die Subtypen E und I klassifiziert, die allerdings nach der Analyse des Kompletogenoms als Mosaikviren identifiziert wurden, weshalb man diese später in CRF01_AE und CRF04_cpx umbenannte (38, 40-42). Die jeweiligen HIV-1 Subtypen bilden in Stammbaumanalysen eine monophyletische Gruppe (*Clade*), in der Viren mit ähnlicher genetischer Distanz lokalisiert sind (40).

1.4.2 Rekombinante Formen in der HIV-1 Gruppe M

Vor allem in Regionen, in denen mehrere HIV-1 Varianten co-zirkulieren, ist die Wahrscheinlichkeit erhöht, dass sich Menschen mit mehreren verschiedenen HIV-1 Subtypen infizieren (43). Demnach können in dual infizierten Zellen Rekombinationsereignisse zwischen HIV-1 Gruppe M Subtypen auftreten, wodurch inter-Subtyp-rekombinante HIV-1 Varianten generiert werden (40). Der Großteil dieser erzeugten inter-Subtyp-Rekombinanten sind in ihrer rekombinanten Genomstruktur einzigartig und werden als singuläre rekombinante Formen (*unique recombinant form, URF*) bezeichnet (38). Wird in der Bevölkerung eine solche singuläre Form in unabhängigen Infektionen identifiziert, kann sie als zirkulierende rekombinante Variante in der HIV-Epidemie etabliert werden. Aktuell kann eine neue CRF definiert werden, wenn sie in mindestens drei unabhängig voneinander infizierten Individuen nachgewiesen wird. Nach der Analyse von mindestens zwei vollständigen HIV-Genomen und einer partiellen Genomsequenz kann sie als zirkulierende rekombinante Form (*circulating recombinant form, CRF*) definiert und registriert werden (38). Nach Angaben der Los Alamos HIV Datenbank sind derzeit 72 CRFs klassifiziert ((44), 13.08.2014). Neu identifizierte CRFs werden nach der Reihenfolge ihrer Entdeckung fortlaufend nummeriert (CRF01-CRF72) (38, 44). Besteht eine CRF aus zwei Subtypen werden diese an den CRF-Namen angefügt (44). So setzt sich beispielsweise die zweite entdeckte zirkulierende rekombinante Form CRF02_AG aus den reinen Subtypen A und G zusammen (45). Aber auch die Rekombination zwischen Subtypen und CRFs ist möglich, so besteht zum Beispiel die CRF15_01B aus Genomabschnitten der CRF01 und des Subtyps B (46). Setzt sich dagegen eine CRF aus mehr als zwei Subtypen und/oder CRFs zusammen, wird anstelle der Subtypen die Abkürzung „cpx“ (*complex*) an den CRF-Namen gefügt, wie zum Beispiel bei der rekombinanten Form CRF06_cpx, die aus Genomabschnitten der Subtypen A, G, J und K aufgebaut ist oder CRF36_cpx, die sich aus Genomabschnitten der CRF01, CRF02 und der Subtypen A und G zusammensetzt (44, 47, 48).

Wie oben schon beschrieben, müssen mehrere Kriterien erfüllt werden, um letztendlich eine URF als neue CRF definieren zu können (38):

1. Die rekombinante Variante muss in mindestens drei Infizierten nachgewiesen werden, deren Infektionen nicht epidemisch miteinander verknüpft sind,
2. Die Analyse von mindestens zwei Kompletengenomsequenzen sowie einer partiellen Genomsequenz mit jeweils dargestellter Rekombinationsstruktur der rekombinanten Variante von epidemisch unabhängig übertragener Infektionen müssen vorliegen
3. Die genetische Distanz zu anderen Subtypen muss signifikant sein (signifikanter Bootstrap-Wert der Clade). (38)

1.5 Epidemiologie von HIV-1

Seit seiner Entdeckung hat sich HIV pandemisch ausgebreitet. Nach Angaben des jüngsten UNAIDS Global Reports lebten 2012 weltweit ca. 35,3 Millionen Menschen mit HIV, von denen 2,3 Millionen Menschen neu infiziert wurden (49). Im gleichen Jahr starben an den AIDS-Folgen 1,6 Millionen Menschen (49). Insgesamt ist weltweit die Anzahl an Neuinfektionen und AIDS-Toten aufgrund der verstärkten Behandlung mit antiretroviralen Medikamenten und HIV-Schutzmaßnahmen gesunken (49, 50). Die meisten HIV-Infizierten mit 25 Millionen leben in Sub-Sahara Afrika mit der höchsten Prävalenz von 4,7 % (49). Die HIV-1 Gruppe M gilt hierbei als pandemische Form mit ~33 Millionen von ~35 Millionen Infizierten weltweit (9, 51). Die Gruppe O mit weniger als 1% der globalen HIV-1-Infektionen ist sehr viel weniger prävalent und ist weitgehend begrenzt auf Kamerun, Gabun und benachbarte Länder (49, 51). HIV-1 Gruppe N-Infektionen wurden bisher in nur 13 Patienten aus Kamerun identifiziert (50) und die erst kürzlich 2009 entdeckte Gruppe P wurde bisher nur in 2 Personen aus Kamerun nachgewiesen (52).

1.5.1 Geographische Verteilung der Subtypen und Mosaikviren der HIV-1 Gruppe M

Die weltweite geographische Verteilung von verschiedenen Subtypen und Mosaikviren ist sehr heterogen und aufgrund der ansteigenden Mobilität und Migration von Menschen ein dynamischer Prozess (53). Im Zeitraum von 2004-2007 dominiert in der weltweiten Epidemie der HIV-1 Gruppe M Infektionen mit 48 % Subtyp C, gefolgt von Subtyp A (12 %), Subtyp B (11 %), die zirkulierenden rekombinanten Formen CRF02_AG (8 %), CRF01_AE (5 %), Subtyp G (5 %) und Subtyp D (2 %) (8). Die Subtypen F, H, J und K betragen zusammen weniger als 1 % und andere zirkulierende rekombinante Formen nehmen 4 % der weltweiten Infektionen ein (8). Insgesamt repräsentieren Mosaikviren 20 % der globalen HIV-1 Infektionen (8).

Die Demokratische Republik Kongo zeigt die höchste genetische Diversität bezogen auf co-zirkulierende Subtypen und rekombinante Formen von HIV und gilt als Epicenter von HIV-1 Gruppe M Varianten, die sich von dort aus pandemisch ausgebreitet haben (54, 55). Die globale Verteilung der Subtypen und Mosaikviren von HIV-1 ist in Abbildung 2 graphisch dargestellt. Demnach dominiert der weltweit am häufigsten vorkommende Subtyp C mit fast 100 % in Südafrika, Äthiopien und in Indien (8, 53). In Westafrika gehen 50-80 % der HIV-Infektionen auf CRF02_AG zurück und in Ostafrika leben mehrheitlich Subtyp A infizierte Menschen (8, 53). Subtyp B dominiert in Nordamerika, in der Karibik und Lateinamerika sowie in West- und Zentraleuropa einschließlich Deutschland und in Australien (8). Die Epidemie in Osteuropa und Zentralasien wird von Subtyp A dominiert (8, 53). In Südostasien co-zirkulieren vorwiegend CRF01_AE und Subtyp B (53).

1.6 Zielsetzung

Das HIV-Studienlabor im Fachgebiet „HIV und andere Retroviren“ des Robert Koch Instituts hat seit 2010 eine Kooperation mit dem “Ministry of Health, Central Public Health Laboratories”, in Muscat, Sultanat Oman. In dieser Kooperation wurden Mitarbeiter des Omanischen Zentrallabors in der genotypischen Resistenzbestimmung von HIV-1 geschult. Im Rahmen der Resistenzbestimmung von Omanischen HIV-Patienten, die dort in Krankenhäusern in Behandlung sind, sollten epidemisch relevante HIV-1 Subtypen und CRFs im Oman identifiziert werden. In vorangegangenen Arbeiten wurden 48 Patientenproben in der HIV-1 *pol*-Region (PR/RT) und überwiegend auch zusätzlich in der INT-Region subtypisiert (n = 40). 47 % (19/40) der *pol*-Sequenzen (PR/RT/INT) konnten keinem bekannten Subtyp oder CRF zugeordnet werden und wurden daher zunächst als URF (*unique recombinant form*, singuläre rekombinante Form) klassifiziert. Sechs dieser Isolate bildeten in der phylogenetischen Analyse eine eigenständige monophyletische Gruppe (Clade), die sich signifikant von allen global bekannten Subtypen und CRFs (*pol*-Sequenz) der HIV-Datenbank unterschied, weshalb sie zunächst als „*URF-new clade*“ bezeichnet wurde. Die paarweise Divergenz der *pol*-Sequenzen ergab keinen Hinweis auf eine epidemische Verknüpfung dieser Isolate (Infektketten), sondern sprach dafür, dass es sich bei diesen identifizierten Isolaten um epidemisch unabhängige Infektionen mit einer putativen neuen rekombinanten Form handelt.

Um diese als neue CRF definieren zu können, müssen insgesamt zwei komplette Genomsequenzen dieser *URF-new clade* aus Oman untersucht werden sowie von einer dritten Probe partielle Sequenzen vorliegen, die einen Vergleich der rekombinanten Genomstruktur erlauben. Die Amplifikation des Kompletengenoms aus viraler RNA wurde in der Arbeitsgruppe mit einer *full-length* cDNA-Synthese und anschließender Amplifikation in Form von vier überlappenden PCR-Fragmenten („*vier-Amplikon-Strategie*“) etabliert, um die komplette Genomsequenz daraus ermitteln und zusammensetzen zu können. Die etablierten Methoden wurden erfolgreich für die Amplifikation von viraler RNA aus einem angezüchteten Virusisolat der *URF-new clade* (13-0346) eingesetzt (59). Die Sequenz des 5'-Genombereich (*URF-full-1* und *URF-full-2*, *gag* ORF bis Ende *pol* ORF) dieses Virusisolats (13-0346) konnte bereits ermittelt werden (59). Eine weitere Virusanzucht aus Plasma schlug fehl, daher wurde versucht, die überlappenden PCR-Fragmente aus Plasmaproben (EDTA-Plasma) mit ausreichender Viruslast direkt (ohne Viruskultivierung zur Anreicherung) zu amplifizieren.

Das Ziel dieser Masterarbeit war daher die Sequenz- und Genomstrukturanalyse von Kompletengenomen der *URF-new clade* zu vervollständigen, um sie als neue zirkulierende rekombinante Form (CRF) klassifizieren zu können.

Folgende Arbeitsschritte waren dazu geplant:

1. Amplifikation des 3' Genombereichs (*URF-full-3* und *URF-full-4*) aus RNA des angezüchteten Virusisolates (13-0346),
2. Amplifikation des Komplettenoms in Form von vier überlappenden Fragmenten (*URF-full-1*, -2, -3, -4) der viralen RNA aus EDTA-Plasma (13-5995, 14-0875),
3. Direktsequenzierung von zwei Amplifikaten (13-0346) bzw. vier Amplifikaten (13-05995) nach Sanger inklusive Sequenzanalyse und Assemblierung zur Kompletengenomsequenz,
4. Phylogenetische Analysen der „*URF-new clade*“ Isolate mit aktuellen Referenzsequenzen aus der HIV-Datenbank von Kompletengenomen zur genetischen Klassifizierung,
5. Analyse der rekombinanten HIV-Genomstruktur der „*URF-new clade*“ Isolate zur Identifikation von Rekombinationsstellen bzw. -bereiche

2 Material

2.1 Untersuchungsmaterialien und Patientenproben

In dieser Arbeit wurden eine Verlaufsprobe (VP) eines Patienten aus einem früheren Pilotprojekt und Erstproben (EP) von zwei Patienten des aktuellen Pilotprojektes aus Oman verwendet (Tabelle 1). Bei der Verlaufsprobe 13-0346 handelt es sich um angezüchtetes Virusisolat (aus PBMC einer EDTA-Blutprobe in Co-Kultur angezüchtet) und bei den beiden Erstproben (13-05995, 14-0875) um Plasma aus EDTA-Blut (EDTA-Plasma). Von diesen beiden Erstproben stand nur begrenzt Probenmaterial (ca. 1 ml) zur Verfügung. Der Transport der Patientenproben aus Oman erfolgte mit „FEDEX“ als diagnostisches Material UN 3373 (Ansteckungsgefährliche Stoffe der Kategorie B, 27).

Tabelle 1: HIV-1 Virusisolate, Patientenproben, PCR-Produkte und Sequenzen

Proben-Nr.	Material	Probe ²	VL [Kopien/ml]	Sequenzen	PCR-Produkte
13-0346	Virusisolat ¹	VP	5,51 x 10 ⁶ *	URF-full-1, -2	URF-full-3, -4
13-05995	EDTA-Plasma	EP	9,43 x 10 ⁴ #	-	-
14-0875	EDTA-Plasma	EP	2,72 x 10 ⁵ #	-	-

¹ Tag der Abnahme: Tag 14 (am 31.01.2013)

² VP: Verlaufsprobe, EP: Erstprobe

* Viruslast (VL) des Virusisolates

Viruslast (VL) der Patientenprobe

2.2 Referenzmaterialien

Positivkontrolle

HIV-Laborstamm HTLV_{III B} (6,1 x 10⁶; 6,1 x 10⁵; 6,1 x 10⁴ Kopien/ml) aus Virusstock in H-9 Zellen vermehrt; Tag der Abnahme: Tag 21 (am 19.03.2010), bei -70 °C in Aliquots gelagert

Negativkontrolle

HIV-negatives humanes Plasma, Pool aus 26 Einzelspenden, Paul-Ehrlich-Institut (Frankfurt/Main)

Referenzpanel für phylogenetische Analysen	Sequenzen des Subtypenreferenzpanel, der Los Alamos Datenbank 2010 (n = 170) (60) http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html Zugriff: 19.05.2014 14:22
--	---

2.3 Reagenzien

2.3.1 Chemikalien

Agarose (Ultra Pure™)	Invitrogen GmbH, Karlsruhe
Bromphenolblau 1 %	VWR International GmbH, Darmstadt
dNTP-Mix (10 mM)	Invitrogen GmbH, Karlsruhe
EDTA (Titriplex III)	Carl Roth GmbH + Co. KG, Karlsruhe
Essigsäure 100 %	Merck KGaA, Darmstadt
Ethanol 99,8 %	Carl Roth GmbH + Co. KG, Karlsruhe
Ethidiumbromid (10 mg/ ml)	Carl Roth GmbH + Co. KG, Karlsruhe
Ficoll 400	Sigma-Aldrich GmbH, Steinheim
Gene Ruler™ 1kb DNA-Ladder	Fisher Scientific GmbH, Schwerte
NaOH-Plättchen	Carl Roth GmbH + Co. KG, Karlsruhe
Salzsäure 1 M	Carl Roth GmbH + Co. KG, Karlsruhe
Tris Base	Sigma-Aldrich GmbH, Steinheim

2.3.2 Puffer für die Agarose-Gelelektrophorese

0,5 M EDTA pH 8,0	186,1 g EDTA (Titriplex III) 20 g NaOH-Plättchen lösen in 800 ml Aqua Bidest mit NaOH (1M) auf pH 8,0 nachjustieren ad auf 1 l mit Aqua Bidest
-------------------	--

1 M Tris-HCl pH 8,0	121,1 g Tris Base 562 ml 1 M HCl ad auf 1 l mit Aqua Bidest
50x TAE Puffer	242 g Tris Base 57,1 ml 100 % Essigsäure 100 ml EDTA 0,5 M pH 8,0 ad auf 1 l mit Aqua Bidest
1 % Agarosegele	1x TAE Puffer 1 % Agarose 0,5 µg/ml Ethidiumbromid Aqua Bidest
Laufpuffer	1x TAE Puffer 0,5 µg/ml Ethidiumbromid Aqua Bidest
6x Ladepuffer	0,25 % Bromphenolblau 15 % Ficoll 400 10 mM Tris-HCl pH 8,0 1 mM EDTA pH 8,0

2.4 Kommerzielle Kits

BigDye Terminator v 3.1 Cycle Sequencing Mix	Applied Biosystems, Weiterstadt
GeneJET™ Gel Extraction Kit	Fisher Scientific GmbH, Schwerte
Long PCR Enzyme Mix Kit	Fisher Scientific GmbH, Schwerte
MSB Spin PCRapace	Invitex GmbH, Berlin
QiAmp Viral RNA Mini Kit	Qiagen GmbH, Hilden
Superscript III First-Strand Synthesis Kit	Invitrogen GmbH, Karlsruhe

2.5 Geräte und Verbrauchsmaterialien

Analysenwaage L610 D	Sartorius AG, Göttingen
Bechergläser	DURAN Group GmbH, Wertheim/ Main
Clean Bench Hera Save	Thermo Fisher Scientific, MA, USA
Electrophoresis Power supply ST 305	Gibco BRL, Eggstein
Eppendorf 5417R Zentrifuge	Eppendorf, Hamburg
Erlenmeyerkolben	DURAN Group GmbH, Wertheim/ Main
Geldokumentationssystem E.A.S.Y. Win32	Herolab GmbH Laborgeräte, Wiesloch
Gelelektrophoresekammer Horizon58	Biometra, Göttingen
Heizblock	Biometra, Göttingen
Heraeus Multifuge X3FR Zentrifuge	Thermo Fisher Scientific, MA, USA
Heraeus Pico 17 Zentrifuge	Thermo Fisher Scientific, MA, USA
Magnetrührer mit Heizfunktion	Stuart GmbH, Wuppertal
Magnetrührstäbchen	VWR International GmbH, Darmstadt
Messzylinder	DURAN Group GmbH, Wertheim/ Main
Mikropipetten	Eppendorf, Hamburg
Mikrowelle	Robert Bosch GmbH, Stuttgart
Mikrozentrifuge	Carl Roth GmbH + Co. KG, Karlsruhe
PCR-Reaktionsgefäße (0,2 ml)	Rapidozym, Luckenwalde
pH-Meter QpH 70	VWR International GmbH, Darmstadt
Reaktionsgefäße (0,5 ml, 1,5 ml, 2 ml)	Sarstedt, Nümbrecht

Skalpell, Spatel, Löffel	VWR International GmbH, Darmstadt
ThermoCycler (Mastercycler pro S)	Eppendorf, Hamburg
Thermocycler T3000	Biometra, Göttingen
UV Transilluminator	Herolab GmbH Laborgeräte, Wiesloch
Vortex Mixer	VWR International GmbH, Darmstadt

2.6 Software

BioEdit 7.0.9.	Ibis Biosciences, CA, USA
CorelDRAW Graphics Suite 12	Corel Corporation 2010, USA
Videodokumentation E.A.S.Y. Win32-System	Herolab GmbH Laborgeräte, Wiesloch
DNASTAR-Lasergene 10, Version 10.0.1	DNASTAR, Madison, Winconsin, USA
Geneious 7.1.4	Biomatters Ltd
jpHMM	(61, 62)
Los Alamos HIV-Datenbank	(44)
PHYLIP package 3.6	Joseph Felsenstein
RIP 3.0	(44, 63)
TreeView 1.6.6.	Roderik D.M. Page, UK

2.7 Oligonukleotide

Die für die Amplifikation und Sequenzierung eingesetzten Primer für *URF-full-1* bis *URF-full-4* sind in Tabelle 2 aufgelistet.

Tabelle 2: Primer für die Amplifikation und Sequenzierung des HIV-1 Kompletengenoms.

Primer	URF-full- 1 - 4		Quelle ¹	Sequenz (5' – 3')	Position (nt) *
	PCR	Sequenzierung			
Pan-HIV-1_1F	1	1	(64)	AGCCYGGGAGCTCTCTG	480 - 496
URF_full_5'LTRs		1	diese Arbeit	CTAGCAGTGGCGCCCGAACAGGGACT	629 - 648
URF G20as		1	FG18	TTCTAGCTCCCTGCTTGCCCATAC	915 - 892
Gag 1259as		1	FG18	TTTACCCATGCATTYAAAGTTCTAGGTGA	1259 - 1231
Gag 1322s		1	FG18	ATACCCATGTTTTTCAGCATTATCAGAAGG	1294 - 1322
Pol 2001s	2	1,2	FG18	CTAGRAAAARGGGCTGTT	2012 - 2029
Pol 2036s		1,2	FG18	GTGGAAAGGAAGGACACCAAATGAAAG	2036 - 2062
Pol 2322s		1,2	FG18	TTAGATACAGGAGCAGATGA	2322 - 2341
Pol 2402as		1,2	FG18	AATCCCCCTATCATTITTTGG	2402 - 2382
Pol 2696s		1,2	FG18	ATTGGGCCTGAAAATCCATA	2697 - 2716
Pol 3034s		1,2	FG18	GTAGCATGACAAAATCTTAGAG	3034 - 3056
Pol 3454as		1,2	FG18	TCTGCTTCTTYTGTTAGTGGA	3448 - 3427
Pol 3532as	1	1,2	FG18	TTCTGCTATTAAGTCTTTTGATGGGTCA	3533 - 3506
URF-pol 3626s		2	FG18	TGCCACACTAATGATGT	3626 - 3643
5'INTs		2	FG18	ATTGGAGGAAATGAACAAGT	4173 - 4192
URF-pol 4303as		2	FG18	TCACTAGCCATTGCTCTCCA	4303 - 4284
URF-pol 4560as		2	FG18	TTACTGGCATCTTCCTGC	4560 - 4542
Pan-HIV-1_3F	3	2,3	(64)	TTAAAAGAAAAGGGGGGATTGGG	4783 - 4805
URF 3p31s		3	diese Arbeit	TTGTGTGGCAYGTAGACAGGAT	5066 - 5087
3p31as	2	2,3	FG18	ATCCTGTCTACYTGCCACACAA	5087 - 5066
URF 5435s		3	FG18	GGTGTGAATATCAAGCAGGACA	5435 - 5456
URF 5435as		3	FG18	TGTCCTGCTTGATATTCACACC	5456 - 5435
ENVoutF1		3	FG18	AGARGAYAGATGGAACAAGCCCCAG	5550 - 5574
Pan-HIV-1_4F	4	3,4	(64)	CCTATGGCAGGAAGAAGCG	5967 - 5985
URF 6229as		3	FG18	CTCATTGCCACTGTCTTCTGCT	6229 - 6208
Env 5901s		3	FG18	ATTGTGGGTCACAGTCTATTATGGGGTACCT	6323 - 6353
URF 6543s		3,4	FG18	ACATGGTAGAACAGATGCATGAGG	6520 - 6543
URF 6881as		3,4	FG18	GGCACAATAATGTATGGGAATTGG	6881 - 6858
Env 6537s		3,4	FG18	AATGTCAGCACAGTACAATGTACAC	6945 - 6969
Env 5as		3,4	FG18	TCCTTSGATGGGAGGGGCATACATTGC	7547 - 7521
URF-env 7618s		3,4	diese Arbeit	TGAGACCTTCAGACCTATAGGAGGAG	7619 - 7644
Env 7254asc		3,4	FG18	TCATATCTCCTCCTCCAGGTCTGAA	7650 - 7626
Env 7407asc		3,4	FG18	CATAGTGCTTCTGCTGCTCCYAAGAACC	7814 - 7786
Pan-HIV-1_3R	3	3,4	(64)	TGGCYTGTACCGTCAGCG	7848 - 7831
Gp46 F2		4	FG18	ACAATTATTGTCTGGTATAGTGAACAGCA	7850 - 7879
E180s_SF7		4	FG18	GTCTGGTATAGTGAACAGCA	7859 - 7879
URF-env 8015as		4	diese Arbeit	GCCCCAAATCCCCAGGAGCTGT	8015 - 7994
URF-env 8177s		4	FG18	CCAGCARGAAAAGAATGAACAAG	8177 - 8199
Gp41 R1		4	FG18	AACGACAAAGGTGAGTATCCCTGCCTAA	8374 - 8347
URF-env 8520s		4	FG18	TTCAGCTACCACCGCTTGAGAGA	8520 - 8542
URF-env 8719as		4	FG18	ACTTCTATAACCCTATCTGTCC	8719 - 8698
URF_full_3'LTRas		4	diese Arbeit	GTCATTGGTCTYAAAGGTACYTGTGGTCTGA	9035 - 9005
URF Mlu 13s		4	FG18	TCAGGTACCTTTAAGACCAATGAC	9012 - 9035
URF-nef 9104as		4	FG18	GAGTGAATTAGCCCTTCCAGTCCC	9104 - 9081
Pan-HIV-1_4R		4	(64)	CTTWTATGCAGCWTCTGAGGG	9517 - 9497
LTR-fulllength-as	4	4	FG18	AGCACTCAAGGCAAGCTTTATTGAGGC	9633 - 9607

* Koordinaten entsprechen der Lokalisation in der Referenzsequenz HXB2 (AccNo.: K03455) (44)

¹ Primer aus der Literatur, aus der Primerdatenbank des HIV-Studienlabors (FG18: Fachgebiet 18, HIV und andere Retroviren, Robert Koch-Institut, Berlin)

3 Methoden

3.1 Extraktion viraler RNA aus EDTA-Plasma

Die Extraktion viraler RNA aus Blutplasma erfolgte mit dem QiAmp Viral RNA Mini Kit der Firma Qiagen, das die selektiven Bindungseigenschaften einer Kieselgel-Membran mit der Geschwindigkeit von MicroSpin-Technik kombiniert. Als Positivkontrolle wurde der Laborstamm HTLV_{III}B (1×10^6 Kopien/ml) und als Negativkontrolle HIV-negatives Plasma verwendet (Material, 2.2).

Zur Entfernung von Kryopräzipitaten wurden nach Auftauen 500 μ l der Probe für 10 min bei 4 °C und 5.300 rpm zentrifugiert. Für die anschließende Viruspelletierung wurden 450 μ l des Überstandes für 90 min bei 4 °C und 14.000 rpm zentrifugiert und 310 μ l vom Virusüberstand abgenommen. Das verbleibende Viruspellet mit 140 μ l restlichen Plasmaüberstand wurde mit 560 μ l Lysepuffer inklusive tRNA [10 μ l tRNA (1 mg/ml) pro ml Lysepuffer] versetzt und für 10 min bei Raumtemperatur inkubiert. Anschließend erfolgte die Zugabe von 560 μ l Ethanol (99,8 %, v/w). Die weitere Extraktion wurde nach Herstellerprotokoll durchgeführt. Die Elution der RNA erfolgte mit 60 μ l Elutionspuffer. Die extrahierte RNA wurde zu je 10 μ l aliquotiert und mit einem Tropfen Mineralöl überschichtet. Die Lagerung der RNA erfolgt bei -70 °C für maximal 6 Monate.

3.2 Full-length cDNA-Synthese von HIV-1

Die Methode zur *full-length* cDNA-Synthese wurde in der Arbeitsgruppe etabliert (59). Diese erfolgte in einem Thermocycler T3000 (Biometra) mit dem Superscript III First-Strand Synthesis Kit (Invitrogen). Hierfür wurde zunächst ein Pre-Annealing Mix hergestellt, der anschließend mit der extrahierten RNA versetzt wurde:

Pre-Annealing Mix

Reagenz	Konz. Stock	Konz. final	x1
Primer: LTR-fulllength-as	2 μ M	0,2 μ M	1 μ l
dNTP Mix	2,5 mM	1 mM	4 μ l
Preannealing-Mix			5 μ l + 5 μ l RNA
Gesamtvolumen			10 μ l RNA/ Primer Mix

Anschließend wurde der RNA-Primer-Mix für 5 min bei 65 °C denaturiert, um RNA-Sekundärstrukturen aufzulösen und damit die optimale Primerbindung bzw. cDNA-Synthese zu gewährleisten.

Nach dem Denaturierungsschritt wurde dem Ansatz sofort RT-Mix hinzugegeben, der sich wie folgt zusammensetzt:

RT-Mix

Reagenz	Konz. Stock	Konz. final	x1
10x RT Puffer	10x	1x	2 µl
MgCl ₂	25 mM	4,76 mM	4 µl
DTT	0,1 M	10 mM	2 µl
RNaseOUT	40 U/µl	1,9 U/µl	1 µl
SuperScript III RT	200 U/µl	19 U/µl	2 µl
			11 µl RT Mix
			+ 10 µl RNA/ Primer Mix
Gesamtvolumen			21 µl

Die Reverse Transkription erfolgte anschließend mit folgendem Zyklusprogramm:

PCR-Schritt	Temperatur [°C]	Dauer	Zyklen
Reverse Transkription	50	2 h	1
Termination	85	5 min	1
Hold	4	∞	

Nach erfolgter cDNA-Synthese wurde das RNA-*Template* des cDNA-RNA Hybrides durch Zugabe von 2 µl *E. coli* RNase H abgebaut. Dafür wurde der Ansatz für 20 min bei 37 °C und abschließend für 15 min bei 70 °C inkubiert.

Um aus einem cDNA-Ansatz die Amplifikation der vier Fragmente zu gewährleisten, wurde für alle Schritte der doppelte Ansatz (2x) hergestellt, sodass man letztendlich einen cDNA-Gesamtansatz von 42 µl erhielt.

3.3 Amplifikation des Kompletengenoms von HIV-1 („vier-Amplikon-PCR“)

Die Amplifikation des kompletten HIV-Genoms der *URF-new clade* Isolate erfolgte in Anlehnung an die Literatur in Form von vier überlappenden PCR-Fragmenten (*URF-full-1* bis *URF-full-4*) (64). Die vier subtypgenerischen PCRs waren in einer vorangegangenen Diplomarbeit unabhängig voneinander etabliert worden (59). Die verwendeten Primerpaare für die Amplifikation von *URF-full-1* bis *-4* sind in Tabelle 2 (Material, 2.7) sowie in Tabelle 3 aufgelistet. In Abbildung 3 sind die überlappenden vier Fragmente (*URF-full-1* bis *-4*) und die dazugehörigen Primerpaare bezogen auf das HIV-Genom HXB2 graphisch dargestellt.

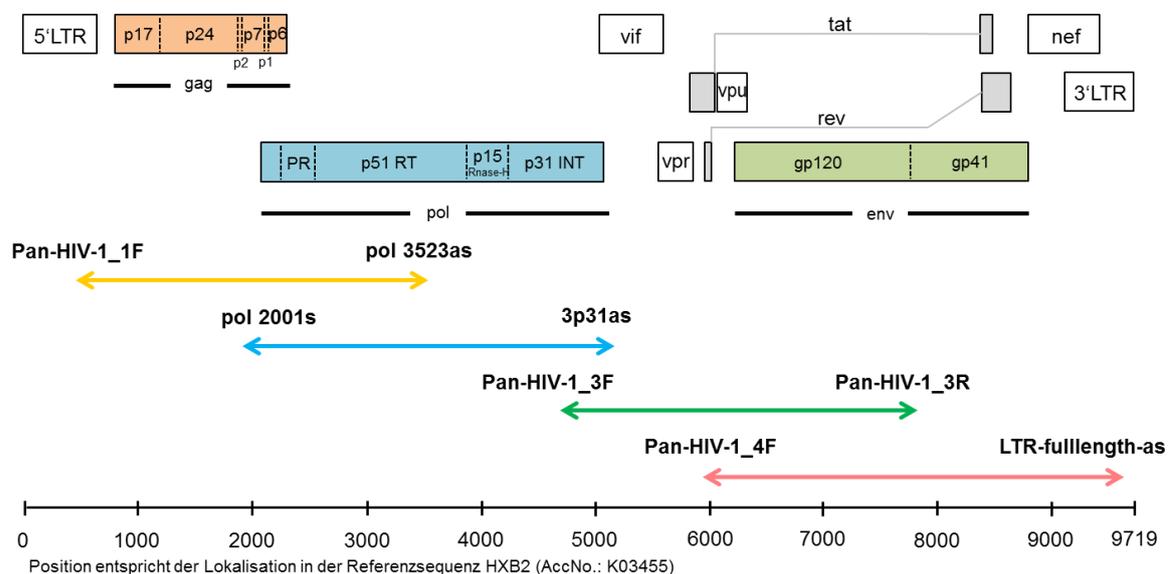


Abbildung 3: Graphisch sind die vier überlappenden PCR-Fragmente (*URF-full-1* bis *-4*) bezogen auf die HXB2-Genomstruktur dargestellt. Die Primerpaare für *URF-full-1* bis *-4* sind am jeweiligen Fragment gezeigt.

Tabelle 3: Primerpaare für die vier PCR-Amplikons *URF-full-1* bis *URF-full-4* und die jeweiligen Fragmentgrößen

Fragment	PCR-Name	Primerpaare	Fragmentgrößen
1	URF-full-1	Pan-HIV-1_1F / pol 3532as	3053 bp
2	URF-full-2	pol 2001s / 3p31as	3062 bp
3	URF-full-3	Pan-HIV-1_3F / Pan-HIV-1_3R	3066 bp
4	URF-full-4	Pan-HIV-1_4F / LTR-fulllength-as	3666 bp

Für einen PCR-Ansatz wurden 9 µl der *full-length* cDNA und 41 µl des PCR-Mix (Long PCR Enzym Mix, Fermentas) eingesetzt, der sich wie folgt zusammensetzt:

Mastermix

Reagenz	Konz. Stock	Konz. final	x1
PCR-Puffer ohne MgCl ₂	10x	1x	5 µl
dNTP Mix	2,5 mM	0,2 mM	4 µl
MgCl ₂	25 mM	1,2 mM	2,4 µl
Primer (sense)	25 µM	0,5 µM	1 µl
Primer (antisense)	25 µM	0,5 µM	1 µl
DMSO	100 %	2 %	1 µl
Long PCR Enzym Mix	5 U/µl	0,04 U/µl	0,4 µl
Nuklease freies H ₂ O			26,2 µl
			41 µl Mastermix
			+ 9 µl cDNA
Gesamtvolumen			50 µl

Da die Länge der jeweiligen vier Amplikons sehr ähnlich ist (zwischen 3053 und 3666 bp), konnten alle vier PCR-Ansätze mit gleichen ZyklerEinstellungen amplifiziert werden:

PCR-Schritt	Temperatur [°C]	Dauer	Zyklen
Denaturierung	96	5 min	1
Denaturierung	96	20 sec	
Annealing	58	30 sec	10
Elongation	68	4 min	
Denaturierung	96	20 sec	
Annealing	58	30 sec	25
Elongation	68	4 min (+10 sec/Zyklus)	
Finale Elongation	68	10 min	1
Hold	4	∞	

3.4 Agarose-Gelelektrophorese zur Qualifizierung und Quantifizierung von DNA

Mit Hilfe der Agarose-Gelelektrophorese werden Nucleinsäuren im elektrischen Feld nach ihrer Größe aufgetrennt, um die Fragmentgröße zu bestimmen oder um die PCR-Produkte quantifizieren zu können. Die elektrophoretische Auftrennung der PCR-Produkte erfolgte mit 1 % Agarosegelen, die mit Ethidiumbromid (0,5 µg/ml), einer DNA-interkalierenden Substanz zur Anfärbung, versetzt wurden. Die fluoreszierenden Eigenschaften von Ethidiumbromid ermöglichen die Detektion der DNA-Fragmente durch UV-Bestrahlung (254 nm - 300 nm) mit einem Transilluminator (Herolab).

Um die Qualität der PCR-Produkte (~ 3000 bp) zu überprüfen, wurden 3 µl der PCR-Ansätze mit 2 µl H₂O und 1 µl 6x Ladepuffer versetzt und auf das Agarosegel aufgetragen. Für die Bestimmung der Fragmentgrößen wurden 3 µl des Größenstandards (GeneRuler™ 1kB DNA-Ladder, 100 ng/µl) mitgeführt. Nach der Produktaufreinigung (Methode 3.5) erfolgte die Quantifizierung der PCR-Produkte ebenfalls mit 1 % Agarosegel. Dazu wurden 2 µl der DNA-Probe mit 2 µl H₂O und 1 µl 6x Ladepuffer versetzt und auf das Gel aufgetragen. 5 µl des Größenstandards (GeneRuler™ 1kB DNA-Ladder, 100 ng/µl) wurden für die Quantifizierung eingesetzt. Nach der Auftrennung der PCR-Fragmente für 120 Minuten bei 70 Volt wurde die DNA anhand einer Eichkurve des Konzentrationsstandards quantifiziert. Die Auswertung der qualitativen sowie quantitativen Gele erfolgte mit dem System E.A.S.Y. RH-3 und der dazugehörigen Software EASY Win32-System.

3.5 Aufreinigung der PCR-Produkte

Bei guter Qualität der aufgetrennten PCR-Fragmente auf dem Agarosegel wurden die PCR-Fragmente direkt aus dem restlichen Volumen (47 µl) der PCR-Reaktion mit Hilfe des MSB Spin PCRapace Kit aufgereinigt. Dabei werden für die Sequenzierung störende Substanzen wie überschüssige Nucleotide, PCR-Primer, Polymerasen und Salze entfernt. Die Aufreinigung wurde nach dem Herstellerprotokoll durchgeführt.

Eine weitere Reinigungsmethode ist die Präparation von DNA-Fragmenten aus dem Agarosegel, um DNA-Fragmente aus einem Gemisch unterschiedlich großer Fragmente zu isolieren und aufzureinigen. Dazu wurde nach qualitativer Auswertung der gesamte PCR-Ansatz von 47 µl mit 8 µl 6x Ladepuffer versetzt und gelelektrophoretisch aufgetrennt. Anschließend wurde das Zielfragment (~3000 bp) unter langwelliger UV-Bestrahlung aus dem Gel geschnitten, das Gelstück gewogen und nach dem Protokoll GeneJET™ Gel Extraction Kit aufgereinigt.

Die DNA wurde in beiden Reinigungsmethoden mit 50 µl Elutionspuffer in ein neues Gefäß eluiert. Die aufgereinigte DNA wurde mittels Agarose-Gelelektrophorese quantifiziert und bei -20 °C gelagert.

3.6 Sequenzierung der PCR-Produkte

Die Sequenzierung der Populationssequenzen erfolgte mithilfe der Sanger-Methode. In Tabelle 2 (Material, 2.7) sind die Sequenzierungsprimer für das jeweilige Amplifikat aufgelistet und die Darstellung der Lokalisationen der Sequenzierungsprimer bezogen auf das HXB2-Genom befindet sich im Anhang (8.1). Dabei handelt es sich um Primer aus der Primerdatenbank des HIV-Studienlabors, aus der Literatur (64) sowie in dieser Arbeit neu designte Primer. Bei dem Primerdesign mit der Software *Geneious* wurde auf den Guanin- und Cytosin-Gehalt (40 % bis 60 %) sowie auf die Schmelztemperatur (55 °C - 70 °C) geachtet. Es sollten zudem möglichst keine Wiederholungen von mehr als drei aufeinander folgenden identischen Nukleotiden auftreten, keine Dimerbildung von Primern mit sich selbst und ein Guanin oder ein Cytosin am 3' Ende vorhanden sein. Die Synthese aller Primer erfolgte durch die Firma Metabion, Martinsried, Deutschland.

Der Sequenzierungsansatz (10 µl) enthielt ~40 ng aufgereinigtes PCR-Produkt, 1 µl 5x Puffer (4 °C), 2 µl BigDye Mix 3.1, 1 µl Primer (5 µM) und Wasser (ultrapur, fluorophorfrees Wasser). Für die Sequenzierung wurde folgendes PCR-Programm verwendet:

PCR-Schritt	Temperatur [°C]	Dauer	Zyklen
Denaturierung	96	2 min	1
Denaturierung	96	10 sec	
Annealing	55	5 sec	25
Elongation	60	4 min	
Hold	4	∞	

Nach der Sequenzierungsreaktion wurden die Proben vom internen Sequenzierungsservice des Robert Koch Instituts über Gelfiltration aufgereinigt und mittels Polyacrylamid-Kapillarelektrophorese aufgetrennt. Die Sequenzrohdaten wurden mit dem Programm *Sequence Analysis Program (ABI)* analysiert und dem Anwender wieder zur Verfügung gestellt.

3.7 Ermittlung der Kompletengenomsequenz

Die Auswertung der Sequenzdaten erfolgte mit dem *SeqMan*-Modul des *DNASTAR Lasergene* Programms (*Version 10.0.1*). Die einzelnen Primersequenzfiles wurden für jedes Amplifikat assembliert und nach manueller Sequenzauswertung eine Konsensussequenz in Form eines *fasta*-Formats erzeugt. Anschließend wurden die Konsensussequenzen von *URF-full-1 bis URF-full-4* mit dem *SeqMan*-Programm zu einer HIV-1 Kompletengenomsequenz assembliert und als *fasta*-Format exportiert.

Die Qualität der Kompletengenomsequenzen der *URF-new clade* Isolate wurde mithilfe des *Quality Control Tools* der HIV-Sequenz-Datenbank überprüft (44), um fehlerhafte Stop-Codons oder Leserahmenwechsel in der Sequenz auszuschließen. Nach positivem Ergebnis der Qualitätskontrolle wurden die Kompletengenomsequenzen der *URF-new clade* Viren für phylogenetische Analysen und für die Analyse der rekombinanten Genomstruktur verwendet.

3.8 Phylogenetische Analysen

Mit Hilfe der Phylogenie (Stammesgeschichte) lassen sich Verwandtschaftsbeziehungen zwischen Populationen oder Individuen, bei denen man einen gemeinsamen Vorfahren vermutet, feststellen. Man spricht von molekularer Phylogenie, wenn molekulare Merkmale (z.B. DNA-Sequenzen, Proteinsequenzen usw.) für Stammbaumanalysen verwendet werden. Phylogenetische Analysen basieren auf der Annahme, dass alle Spezies von einem gemeinsamen Vorfahren abstammen, Veränderungen in der DNA-Sequenz zufällig auftreten und ein Teil davon dauerhaft in den Molekülen erhalten bleibt (65). Anhand phylogenetischer Analysen von partiellen (PR/RT- und INT-Region) oder HIV-Kompletengenomsequenzen kann der Subtyp von HIV-1 Isolaten im Vergleich zu HIV-Subtyp-Referenzsequenzen bestimmt werden.

In dieser Arbeit wurde die genetische Charakterisierung der *URF-new clade* Isolate aus Oman mittels phylogenetischer Analysen der Sequenzen der Isolate im Vergleich mit Referenzsequenzen durchgeführt. Dazu wurden die Nukleotidsequenzen der *URF-new clade* Isolate mit Referenzsequenzen aus dem aktuellen Subtypreferenzpanel der Los Alamos Datenbank verglichen (60) und ein Alignment der Populationssequenz mit der *ClustalW*-Anwendung im Programm *BioEdit* erstellt. Im Alignment stehen untereinander in Spalten (Position) homologe Merkmale von Sequenzen, die in Zeilen ausgeführt sind. Das Alignment wurde mit Hilfe der Software *BioEdit* manuell editiert. Als phylogenetische Methode wurde das distanzbasierte *Neighbor-Joining*-Verfahren gewählt. Hierbei wird zunächst über das *Dnadist*-Programm der *PHYLIP*-Software nach dem *Kimura-2-Parameter-Algorithmus* als Nukleotidsubstitutionsmodell eine Distanzmatrix erstellt, in der alle Positionen im Alignment paarweise verglichen werden. Auf dieser Distanzmatrix basierend wurde mit dem *Neighbor-*

Joining-Algorithmus eine phylogenetische Baumstruktur errechnet, um die Verwandtschaft der Sequenzen abzubilden. Die Topologie des Stammbaumes wurde mit dem Programm *TreeView 1.6.6* dargestellt.

Ein Stammbaum ist aus Ästen und Knoten (Verzweigungspunkten) zusammengesetzt. Die Knoten repräsentieren den letzten gemeinsamen Vorfahren von mindestens zwei sich voneinander abspaltenden Sequenzen. Die Astlängen sind ein Maß für Mutationen, die sich im Verlauf dieser Auftrennung ereignet haben. Je länger ein Ast, desto mehr Mutationsereignisse haben stattgefunden und desto größer ist die evolutionäre Distanz. Um die Richtung der Merkmalsaustausche sowie die Position der Baumwurzel zu erhalten, wird eine Außengruppe, eine homologe Spezies oder Gruppe von Sequenzen benötigt. Mit Hilfe der Außengruppe wird die Wurzel des Baums erstellt, die durch den ersten Knoten, den gemeinsamen Vorfahren der Außengruppe und der zu analysierenden Spezies repräsentiert wird. Von dieser Wurzel zweigen alle anderen Sequenzen ab.

Um Aussagen über die statistische Signifikanz der Baumtopologie (Verzweigungsstruktur) zu treffen, ist eine *Bootstrap-Analyse* erforderlich. Dabei werden Pseudodaten erzeugt, indem die Positionen (Spalten) im Alignment hinsichtlich ihrer Anordnung vertauscht werden. Aus diesen Pseudodaten entstehen entsprechend der Anzahl der Wiederholungen unterschiedliche Bäume, aus denen ein Konsensus-Baum erstellt wird. Die Pseudodaten werden mit der Software *Seqboot* des *PHYLIP*-Paketes erzeugt (erstellt 1000 Pseudoalignments). Mit diesen wird erneut die phylogenetische Analyse durchgeführt (1000 Distanzmatrizes, 1000 *Neighbor-Joining* Bäume). Die ermittelten Bootstrap-Werte im Konsensus-Baum zeigen an, wie stabil die ermittelte Baumtopologie ist und wie signifikant damit die Bildung einer dargestellten monophyletischen Gruppe. Wird zum Beispiel eine monophyletische Gruppe in allen 1000 Pseudodatensätzen gebildet, entspricht dies einem *Bootstrap*-Wert von 100 %. Die über 70 % liegenden *Bootstrap*-Werte werden als signifikant gewertet und an die Knotenpunkte des originalen *Neighbor-Joining* Baums übertragen.

3.9 Analyse der rekombinanten Genomstruktur

Die rekombinante Genomstruktur der *URF-new clade* Isolate aus Oman wurde mit den beiden online verfügbaren Computerprogrammen *RIP (Recombinant Identification Program)* der Los Alamos HIV-Sequenzdatenbank (44, 63) und mithilfe von *GOBICS (Göttingen Bioinformatics Compute Server)* auf Grundlage des *jpHMM (jumping profile Hidden Markov Model)* (61, 62) bestimmt. Beide Programme dienen der Identifikation von HIV-Rekombinanten und erstellen dafür ein multiples Alignment aus eingeladener Nukleotidsequenz mit einem HIV-Referenzsequenzset.

3.9.1 RIP (*Recombination Identification Program*)

RIP ist ein *tool* zur ersten Identifizierung von HIV-Rekombinanten, das hierfür ein bestehendes Alignment aus 12 Referenzsequenzen verwendet, die jeweils einen Subtypen (A1, A2, B-D, F1, F2, G-H, J-K) und die CRF01_AE repräsentieren und vergleicht dieses Alignment mit der eingeladenen zu analysierenden Populationssequenz (Indexsequenz). Ausgehend von Distanzmessungen zwischen den einzelnen Subtyp-Referenzsequenzen und der Indexsequenz wird ein Ähnlichkeits- (*s*, *similarity*)-Distanzplot erstellt, der die beste Übereinstimmung der einzelnen Genomregionen der Indexsequenz mit den Referenzsequenzen wiedergibt. Das Prinzip von RIP beruht auf einem „gleitenden Fenster“, das sich positionsweise von einer Stelle von links nach rechts im Alignment bewegt. Die Wahl der Fenstergröße (100-400 Stellen) bestimmt die Sensitivität der Identifikation von Rekombinanten. Für jedes Fenster im Alignment wird eine Hamming-Distanz (*p*-Distanz) berechnet und davon ausgehend ein Distanzplot erstellt, der zudem die signifikante beste Übereinstimmung innerhalb eines Fensters durch die Wahl eines Konfidenzschwellenwertes (90-95 %) anzeigt. Abhängig davon wie signifikant besser die beste Übereinstimmung zu einer Referenzsequenz gegenüber der zweiten besten Übereinstimmung ist, können auch unbestimmbare Genombereiche in der Rekombinationsanalyse definiert werden.

3.9.2 Springendes Profil-Hidden-Markov-Modell (*jumping profile Hidden Markov Model, jpHMM*)

Auch mit dem jpHMM-Algorithmus können rekombinante Genomstrukturen detektiert werden. Das Prinzip des Modells ist die Bestimmung der Ähnlichkeit einer Anfangssequenz (Indexsequenz) zu der Sequenzfamilie bzw. einzelnen Subtypen mittels eines Alignments der Indexsequenz zu einem multiplen Sequenzalignment der verfügbaren Subtypen. Ein Vorteil dieser Anwendung ist, dass die phylogenetischen Rekombinationsstellen (*Breakpoints*), also die Positionen, an denen die Mosaikstruktur des Genoms von einem Subtyp zu einem anderen wechselt, relativ genau lokalisiert werden können. Die Rekombinationsvorhersage von jpHMM basiert auf einem vorkalkulierten multiplen Sequenzalignment aus 309 HIV-Sequenzen aus der HIV-Sequenzdatenbank, das die neun Subtypen (A1, A2, B-D, F1, F2, G-H, J-K) und die CRF01_AE einschließt (61). Jeder Subtyp im Alignment ist als *profile HMM* modelliert und innerhalb dieser *profile HMMs* sind „Sprünge“ zwischen den verschiedenen *profile HMMs* an fast jeder Position im Alignment erlaubt (66). Das Modell kann also demnach zwischen den verschiedenen Subtypen springen, abhängig davon welcher Subtyp in den unterschiedlichen Genombereichen der Populationssequenz am ähnlichsten ist. Sprünge zwischen verschiedenen Subtypen werden als Rekombinations-*Breakpoints* definiert (66). jpHMM ist allgemein ein statistischer Algorithmus, der statistisch

signifikant falsch vorhergesagte Subtypbereiche in der Populationssequenz als unbestimmbare Regionen definiert und der *Breakpoint*-Intervalle zwischen zwei Subtypen statistisch bestimmt (66). Als Schwellenwert für unbestimmbare Regionen und *Breakpoint* Intervalle wird der Wert 0,99 verwendet (66). Als Ergebnis für die eingeladene Populationssequenz erhält man die vorhergesagte Rekombinationsstruktur mit genauen *Breakpoint* Positionen und *Breakpoint* Intervallen sowie Positionen von undefinierbaren Genomregionen. Zudem wird die Rekombinationsstruktur der Populationssequenz bezogen auf die HXB2-Referenzsequenz graphisch dargestellt.

4 Ergebnisse

4.1 Amplifikation des Kompletengenoms einer neuen rekombinanten Form (*URF-new clade*) von HIV-1 aus Oman („vier-Amplikon-PCR“)

Das Kompletengenom von Vertretern der *URF-new clade* wurde nach erfolgter Extraktion der viralen RNA und der *full-length* cDNA-Synthese in Form von vier überlappenden Fragmenten (*vier-Amplikon-PCR*) amplifiziert (Methoden, 3.3).

4.1.1 Amplifikation des 3'-Genombereichs (*URF-full-3* und *URF-full-4*) von *URF-new clade* HIV-1 aus Virusisolat

Für die Amplifikation des Kompletengenoms aus viraler RNA waren die Methoden zur *full-length* cDNA-Synthese und zur *vier-Amplikon-PCR* anhand des Referenzmaterials HTLV_{III B} und anhand des Vektors pNL4.3 (Acc. Nr. M19921), der ein *full-length* HIV-Genom kodiert, in der Arbeitsgruppe etabliert worden. Die Methode wurde erfolgreich für die Amplifikation viraler genomischer RNA aus dem Virusisolat von Primärprobenmaterial (frisches EDTA-Blut) eines *URF-new clade*-infizierten omanischen Patienten (13-0346) mit Spender PBMC eingesetzt (Ko-Kultur). Der 5'-Genombereich (*URF-full-1* und *-2*, 480-5087 bp bzgl. HXB2, 5'LTR/ *gag* bis *pol*-Ende/*vif*-Start) des Isolates war zuvor bereits sequenziert worden (59). Zur Vervollständigung der Kompletengenomsequenz von 13-0346 erfolgte in dieser Arbeit die Amplifikation und Sequenzierung des 3'-Genombereichs (*URF-full-3* und *-4*, 4783-9633 bp bzgl. HXB2, Ende von *pol* bis 3'LTR einschließlich *env*- und *nef*-Genombereich). In Abbildung 4 sind die amplifizierten PCR-Produkte für *URF-full-3* (3066 bp, 4783-7848 bp bzgl. HXB2) und *URF-full-4* (3666 bp, 5967-9633 bp bzgl. HXB2) anhand eines qualitativen Agarosegels dargestellt. Für die Amplifikation wurde die virale RNA aus dem Primärisolat (13-0346) isoliert. Als Positivkontrolle wurde das Referenzvirus HTLV_{III B} mit $6,1 \times 10^5$ und $6,1 \times 10^4$ Kopien/ml und als Negativkontrolle HIV-1 negatives Plasma eingesetzt. Beide PCRs verliefen erfolgreich und zeigen eindeutige Produktbanden für das Virusisolat. Mit der PCR *URF-full-3* und *URF-full-4* wurden spezifische Produkte ohne Nebenprodukte amplifiziert.

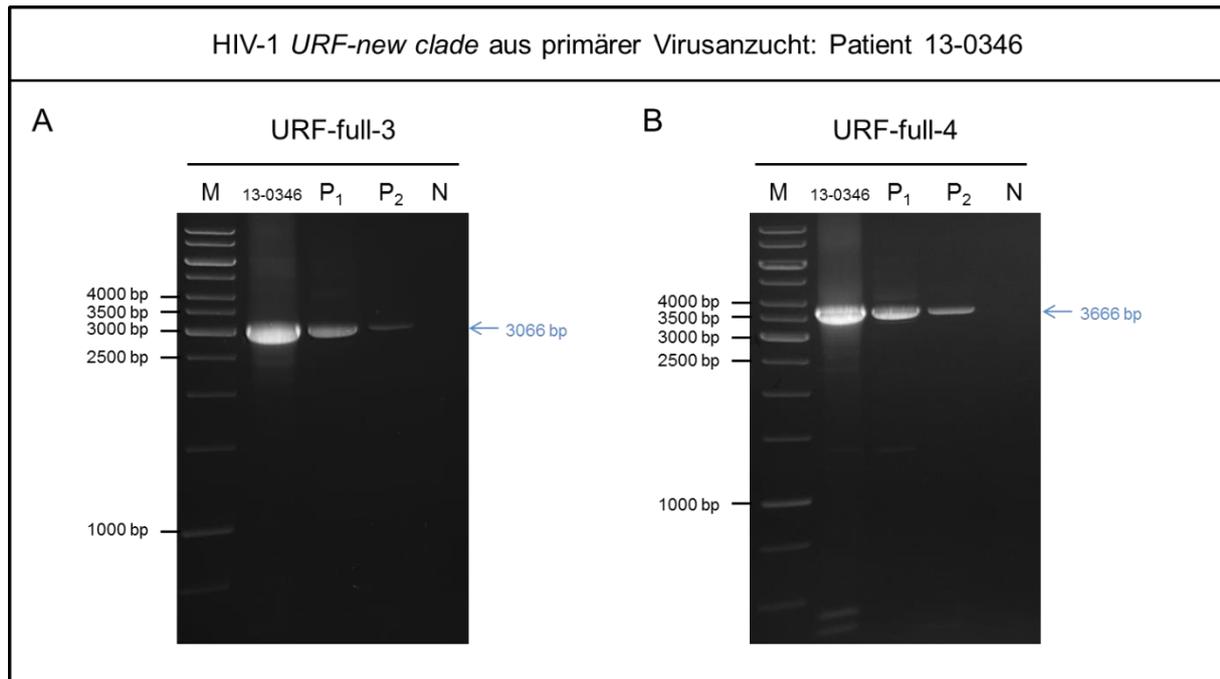


Abbildung 4: Qualitative Auswertung der Amplifikation des 3'-Genombereichs (*URF-full-3* und *URF-full-4*) von *URF-new clade* HIV-1 aus Virusisolat (13-0346). Agarose-Gelelektrophorese zur Darstellung des PCR-Produkts von Fragment 3 (*URF-full-3*) mit einer Größe von 3066 bp (A) sowie des PCR-Produkts von Fragment 4 (*URF-full-4*) mit einer Größe von 3666 bp (B). Agarosegele sind 1 %ig und mit Ethidiumbromid gefärbt. Aufgetragen sind der Marker (M): GeneRuler 1 kb DNA Ladder, das Virusisolat: 13-0346 mit $5,51 \times 10^6$ Kopien/ml, die erste Positivkontrolle (P₁): HTLV_{III B} mit $6,1 \times 10^5$ Kopien/ml, die zweite Positivkontrolle (P₂): HTLV_{III B} mit $6,1 \times 10^4$ Kopien/ml und eine Negativkontrolle (N): HIV-negatives Blutplasma.

4.1.2 Amplifikation des Kompletengenoms (*URF-full-1* bis *URF-full-4*) von *URF-new clade* HIV-1 aus EDTA-Plasma

Das HIV-1 Kompletengenom von zwei weiteren Patientenproben der *URF-new clade* (13-05995 und 14-0875) wurde aus dem Primärprobenmaterial (EDTA-Plasma) amplifiziert (Abbildung 5, Abbildung 6), d.h. für die Amplifikation wurde die virale RNA direkt aus dem EDTA-Plasma der Patienten extrahiert. Für beide Patientenproben konnten die vier PCR-Produkte (*URF-full-1* bis *-4*) erfolgreich amplifiziert werden. Die Qualität der vier Amplifikate für die Patientenprobe 14-0875 ist sehr gut (Abbildung 6), wohingegen bei der Patientenprobe 13-05995 zusätzlich unspezifische Banden bei der Amplifikation der PCR Produkte von *URF-full-3* und *-4* generiert wurden (Abbildung 5, C+D). Deshalb erfolgte die Aufreinigung der Amplikons *URF-full-3* und *URF-full-4* für 13-05995 über ein präparatives Agarosegel mit Extraktion der DNA aus Agarosegel (Methoden, 3.5) für die anschließende Sequenzierung der Fragmente.

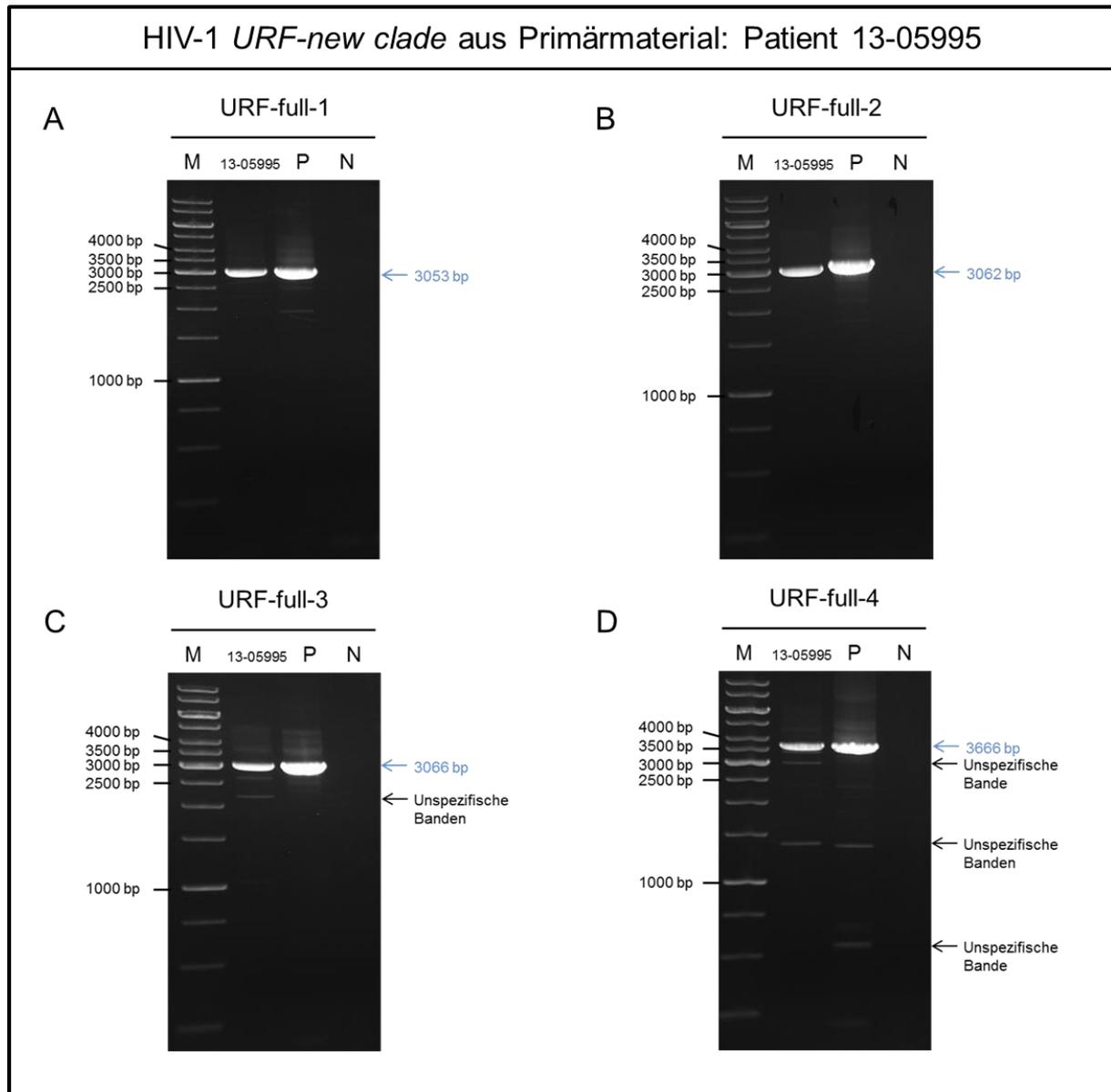


Abbildung 5: Qualitative Auswertung der Amplifikation des Kompletengenoms (*URF-full-1* bis *URF-full-4*) von *URF-new clade* HIV-1 aus EDTA-Plasma (13-05995). Agarose-Gelelektrophorese zur Darstellung des PCR-Produkts von Fragment 1 (*URF-full-1*) mit einer Größe von 3053 bp (A), des PCR-Produkts von Fragment 2 (*URF-full-2*) mit einer Größe von 3062 bp (B), des PCR-Produkts von Fragment 3 (*URF-full-3*) mit einer Größe von 3066 bp (C) sowie des PCR-Produkts von Fragment 4 (*URF-full-4*) mit einer Größe von 3666 bp (D). Agarosegele sind 1 %ig und mit Ethidiumbromid gefärbt. Aufgetragen sind der Marker (M): GeneRuler 1 kb DNA Ladder, die Patientenprobe: 13-05995 mit $9,43 \times 10^4$ Kopien/ml, eine Positivkontrolle (P): HTLV_{III B} mit $6,1 \times 10^6$ Kopien/ml und eine Negativkontrolle (N): HIV-negatives Blutplasma.

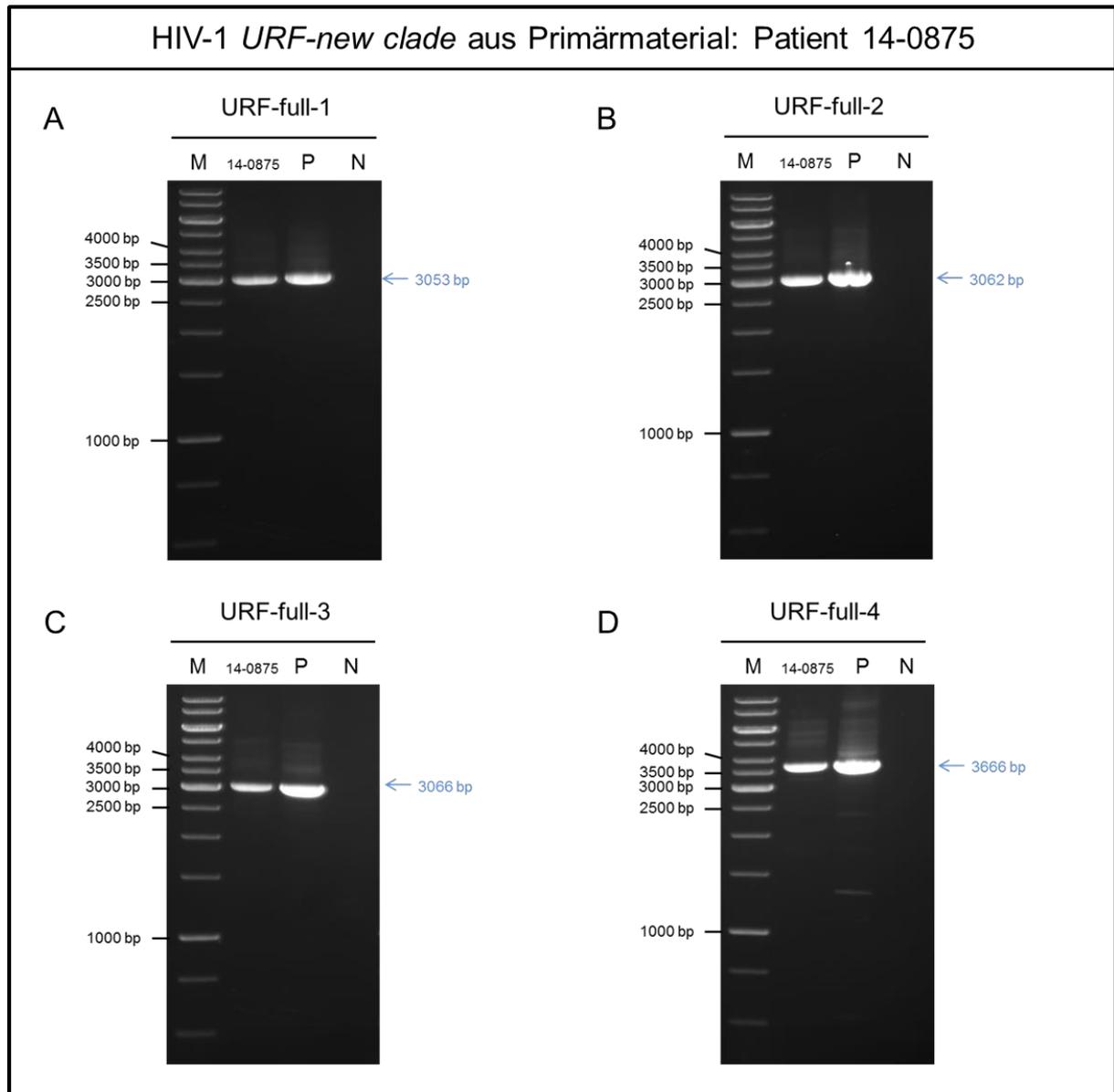


Abbildung 6: Qualitative Auswertung der Amplifikation des Kompletengenoms (*URF-full-1* bis *URF-full-4*) von *URF-new clade* HIV-1 aus EDTA-Plasma (14-0875). Agarose-Gelelektrophorese zur Darstellung des PCR-Produkts von Fragment 1 (*URF-full-1*) mit einer Größe von 3053 bp (A), des PCR-Produkts von Fragment 2 (*URF-full-2*) mit einer Größe von 3062 bp (B), des PCR-Produkts von Fragment 3 (*URF-full-3*) mit einer Größe von 3066 bp (C) sowie des PCR-Produkts von Fragment 4 (*URF-full-4*) mit einer Größe von 3666 bp (D). Agarosegele sind 1 %ig und mit Ethidiumbromid gefärbt. Aufgetragen sind der Marker (M): GeneRuler 1 kb DNA Ladder, die Patientenprobe: 14-0875 mit $2,72 \times 10^5$ Kopien/ml, eine Positivkontrolle (P): HTLV_{III B} mit $6,1 \times 10^5$ Kopien/ml und eine Negativkontrolle (N): HIV-negatives Blutplasma.

4.2 Direktsequenzierung der PCR-Fragmente (*URF-full-1 bis URF-full-4*) von *URF-new clade* HIV-1 Isolaten aus Oman

Die Populationssequenzen der Fragmente *URF-full-1 bis -4* der putativen *URF-new clade* Isolate wurden durch Sanger-Sequenzierung ermittelt. Alle verwendeten Sequenzierungsprimer sind in Tabelle 2 (Material, 2.7) zusammengefasst. Die schematische Darstellung der Lokalisation aller eingesetzten Sequenzierungsprimer auf dem gesamten HIV-1 Genom zeigt Abbildung 7. In der Abbildung sind zudem die in dieser Arbeit neu designten Primer für die Sequenzierung des Kompletengenoms in rot dargestellt. Die Lokalisation der Primer bezogen auf die Referenzsequenz HXB2 befindet sich im Anhang 8.1. und in Tabelle 2. Die Primerlokalisationen wurden dabei so gewählt, dass eine möglichst doppelsträngige Sequenzinformation über die gesamte Länge der Fragmente erhalten wurde (Abbildung 8 und Abbildung 9).

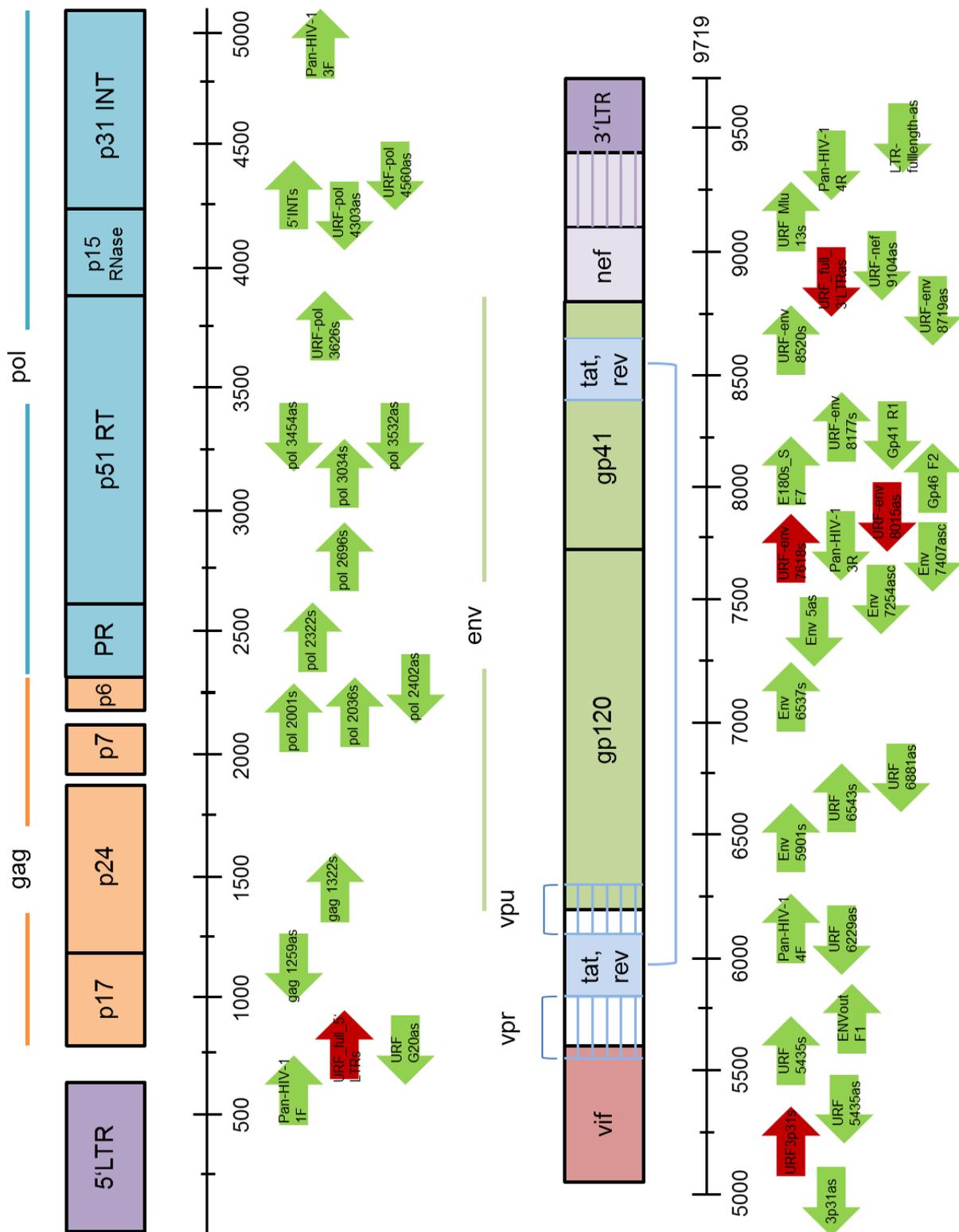


Abbildung 7: Schematische Darstellung der Lokalisation der verwendeten Sequenzierungsprimer bezogen auf das HIV-1 Genom. Grün markierte Primer sind aus der Literatur oder aus der Primerdatenbank des HIV-Studienlabors. Rot markierte Primer wurden für die Sequenzierung in dieser Arbeit neu designt: URF_full_5'LTRs, URF 3p31s, URF-env 7618s, URF-env 8015as, URF_full_3'LTRAs.

4.2.1 Direktsequenzierung des 3'-Genombereichs (*URF-full-3* und *URF-full-4*) von *URF-new clade* HIV-1 aus Virusisolat

Die zwei Fragmente *URF-full-3* und *-4* des Virusisolates 13-0346 (Abbildung 8) konnten erfolgreich doppelsträngig sequenziert werden, wobei für das Fragment *URF-full-4* erst nach Nachsequenzierungen und nach dem Design eines neuen Primers (*URF_full_3'LTRas*) ein vollständiges Contig erzeugt werden konnte (Abbildung 7, Tabelle 2). Das Fragment *URF-full-3* wurde mit sieben sense-Primern und sechs antisense-Primern sequenziert, während die Sequenzierung des Fragments *URF-full-4* mit sieben sense-Primern und mit neun antisense-Primern erfolgte (Abbildung 8). Nach der Assemblierung der Primersequenzfiles für das jeweilige Amplikon und der Sequenzauswertung wurde eine *URF-full-3* und *URF-full-4* Konsensussequenz erzeugt und diese zusammen mit den Konsensussequenzen von *URF-full-1* und *URF-full-2* (59) zu einer Kompletengenomsequenz assembliert. Insgesamt wurde damit eine Kompletengenomsequenz für das *URF-new clade* Isolat 13-0346 von 9113 bp (TAR-Element der 5'LTR bis Poly-A-Signal der 3'LTR, 495-9621 bp bzgl. HXB2) erhalten.

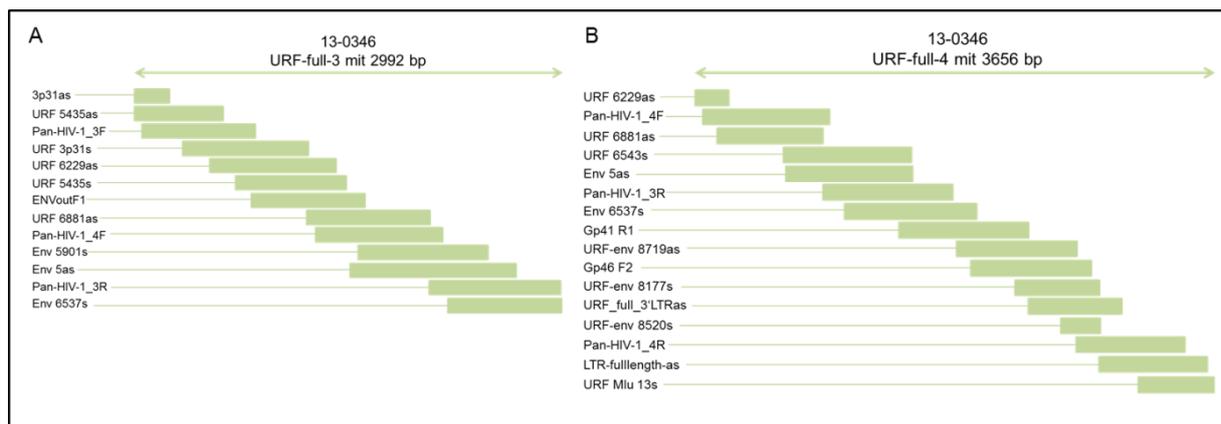


Abbildung 8: Schema zur doppelsträngigen Sequenzierung der Amplikons *URF-full-3* und *-4* des *URF-new clade* Virusisolates 13-0346. (A) *URF-full-3* von 13-0346, (B) *URF-full-4* von 13-0346. Die gelesenen Sequenzen der eingesetzten Sequenzierungsprimer sind graphisch dargestellt.

4.2.2 Direktsequenzierung des Kompletengenoms (*URF-full-1* bis *URF-full-4*) von *URF-new clade* HIV-1 aus EDTA-Plasma

Auch aus dem EDTA-Plasma (13-05995) konnte im Rahmen dieser Masterarbeit die Kompletengenomsequenz ermittelt werden (Abbildung 9). Die Sequenzierung der Fragmente *URF-full-1* bis *-4* aus dem EDTA-Plasma 13-05995 erwies sich als problematischer als für das Virusisolat 13-0346, da die einzelnen Sequenzen teilweise nicht auswertbar waren aufgrund von stärkeren Sequenzüberlagerungen oder Hintergrundrauschen. Durch die Ergänzung des Primersets mit neu entwickelten Primern konnten die Amplikons *URF-full-1* bis *-4* über die jeweils gesamte Länge doppelsträngig sequenziert werden.

Für das *URF-full-1* Fragment aus 13-05995 war zunächst eine einzelsträngige Sequenz im Bereich 1205-1350 bp bezogen auf die HXB2-Referenzsequenz verblieben, die mit dem neu designten Primer *URF_full_5'LTRs* als Doppelstrang nachsequenziert werden konnte (Abbildung 9, A). Unter Verwendung von acht sense-Primern und fünf antisense-Primern wurde eine doppelsträngige *URF-full-1* Sequenz erhalten (Abbildung 9, A). Dagegen konnte das *URF-full-2* Fragment aus dieser Probe problemlos doppelsträngig mit acht sense-Primern und fünf antisense-Primern sequenziert werden (Abbildung 9, B). Für die *URF-full-3* Sequenzierung wurden sieben sense-Primer und sechs antisense-Primer eingesetzt, von denen ein sense-Primer *URF 3p31s* neu designt wurde. Die Sequenzierung von *URF-full-3* war aufwendig und ergab letztendlich keine durchgängig doppelsträngige Sequenzinformation am 3'-Ende von *URF-full-3* (Abbildung 9, C). Diese einzelsträngige Sequenz im Bereich ~6880-7480 bp bezogen auf die HXB2-Referenzsequenz liegt im *env*-Genombereich und wird jedoch vom 5'-Ende des *URF-full-4* Fragments abgedeckt, sodass auch für diesen Bereich letztendlich eine doppelsträngige Sequenzinformation vorliegt. Für die Sequenzierung des Fragmentes *URF-full-4* wurden acht sense-Primer und 13 antisense-Primer verwendet (Abbildung 9, D), die zum Teil in der Arbeitsgruppe zur Verfügung standen. Zusätzlich wurden zwei neue Primer (*URF-env 7618s* und *URF-env 8015as*) designt, um eine doppelsträngige Sequenzinformation sowie ein vollständiges Contig zu erhalten. Nach der Assemblierung der einzelnen Primersequenzfiles und der erfolgten manuellen Sequenzanalyse (DNASTAR-Lasergene 10, Version 10.0.1) für jedes einzelne Fragment wurden die jeweiligen Konsensussequenzen (*URF-full-1* bis *URF-full-4*) zu einer Kompletengenomsequenz von 9019 bp (Poly-A-Signal der 5'LTR bis zum TAR-Element der 3'LTR; 530-9582 bp bzgl. HXB2) assembliert.

Die beiden ermittelten Kompletengenomsequenzen (13-0346, 13-05995) sind in der Form eines Alignments mit der Referenzsequenz HXB2 in Abbildung 14 (siehe Anhang 8.1) dargestellt. Der Anteil an Ambiguitäten in der proteinkodierenden Sequenz ist für das Virus aus EDTA-Plasma (13-05995) mit 0,64 % größer als für das Virusisolat (13-0346) mit 0,51 %. Für das dritte Isolat der *URF-new clade* (14-0875) wurden bereits alle vier Amplikons aus EDTA-Plasma erzeugt (Abbildung 6). Im Rahmen dieser Arbeit konnte für das Plasma-Virus (14-0875) bezüglich der doppelsträngigen Sequenzinformation nur eine partielle Genomsequenz von 5309 bp (*gag p24* bis *env gp120/V1-Loop*, 1329-6638 bp bzgl. HXB2) ermittelt werden (Ergebnisse nicht gezeigt). In weiterführenden Arbeiten wird die Ermittlung der Kompletengenomsequenz erfolgen.

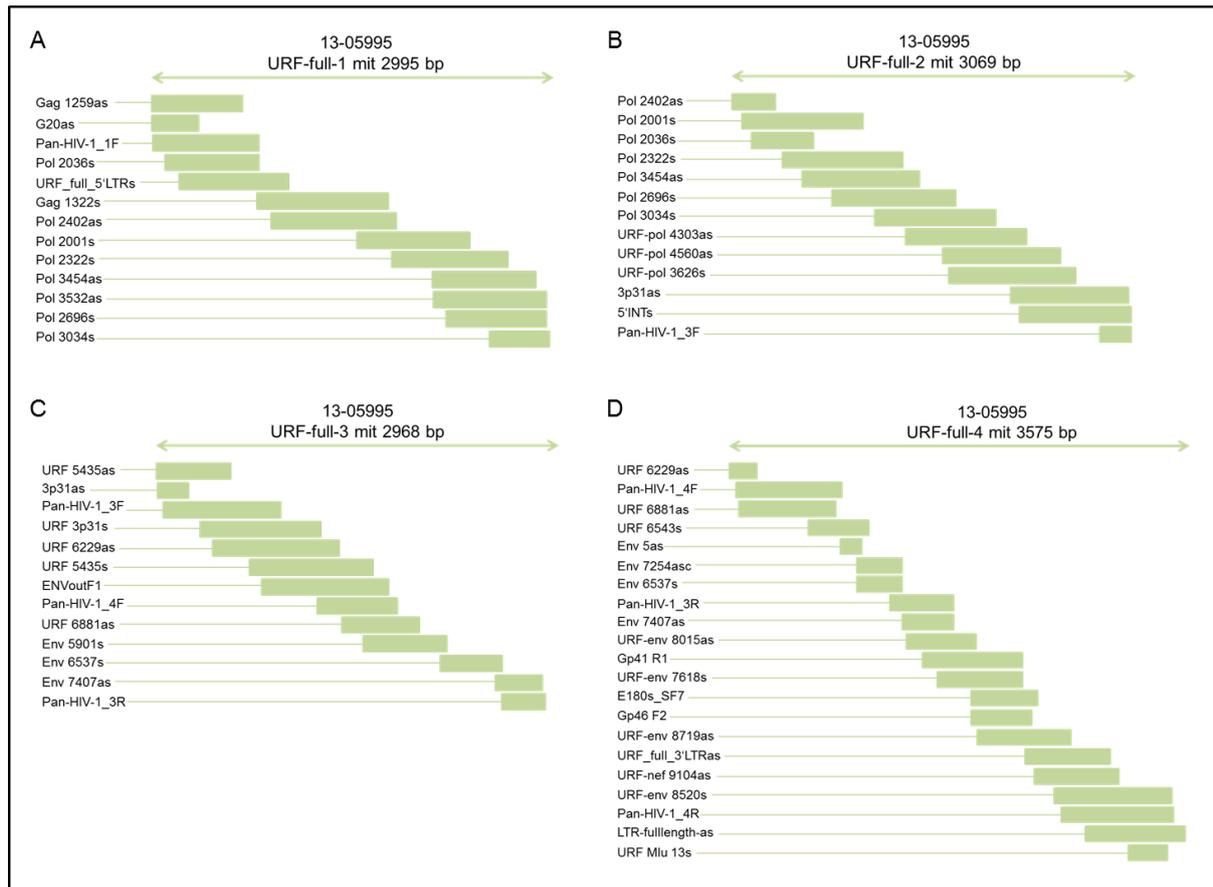


Abbildung 9: Schema zur doppelsträngigen Sequenzierung der Amplikons *URF-full-1* bis *-4* des *URF-new clade* aus EDTA-Plasma (13-05995). (A) *URF-full-1* von 13-05995, (B) *URF-full-2* von 13-05995, (C) *URF-full-3* von 13-05995, (D) *URF-full-4* von 13-05995. Die gelesenen Sequenzen der eingesetzten Sequenzierungsprimere sind graphisch dargestellt.

4.2.3 Sequenzvergleich der überlappenden Regionen zwischen *URF-full-1/URF-full-2*, *URF-full-2/URF-full-3* und *URF-full-3/URF-full-4*

Es werden die überlappenden Regionen der vier Einzelsequenzen hinsichtlich ihrer Übereinstimmung geprüft, da aufgrund der Quasispeziesnatur von HIV mittels der „vier-Amplikon-PCR“ unterschiedliche Varianten der Quasispezies in jedem PCR-Amplikon vorliegen können. Dazu wurden die Konsensussequenzen der einzeln ausgewerteten Amplikonsequenzen assembliert (DNASTAR Lasergene 10, Version 10.0.1) und die überlappenden Bereiche der vier Amplikons verglichen (Abbildung 10). Die überlappende Region von *URF-full-1* und *URF-full-2* (ca. 1500 bp, Lokalisation: 2012-3533 bp bzgl. HXB2) schließt das 5'-Ende der *gag*-Genomregion sowie Teile der *pol*-Genomregion (Protease und Reverse Transkriptase) ein. Der Überlapp zwischen *URF-full-2* und *URF-full-3* (ca. 300 bp, Lokalisation: 4783-5087 bp bzgl. HXB2) deckt den 3'-Genombereich der Integrase ab und der überlappende Bereich zwischen *URF-full-3* und *URF-full-4* (ca. 1800 bp, Lokalisation: 5967-7848 bp bzgl. HXB2) umfasst das erste Exon von *rev*, Teile des *tat*-Exons 1, das *vpu*-Gen, sowie gp120 und den 5'-Bereich von gp41 der *env*-Genomregion (Abbildung 10).

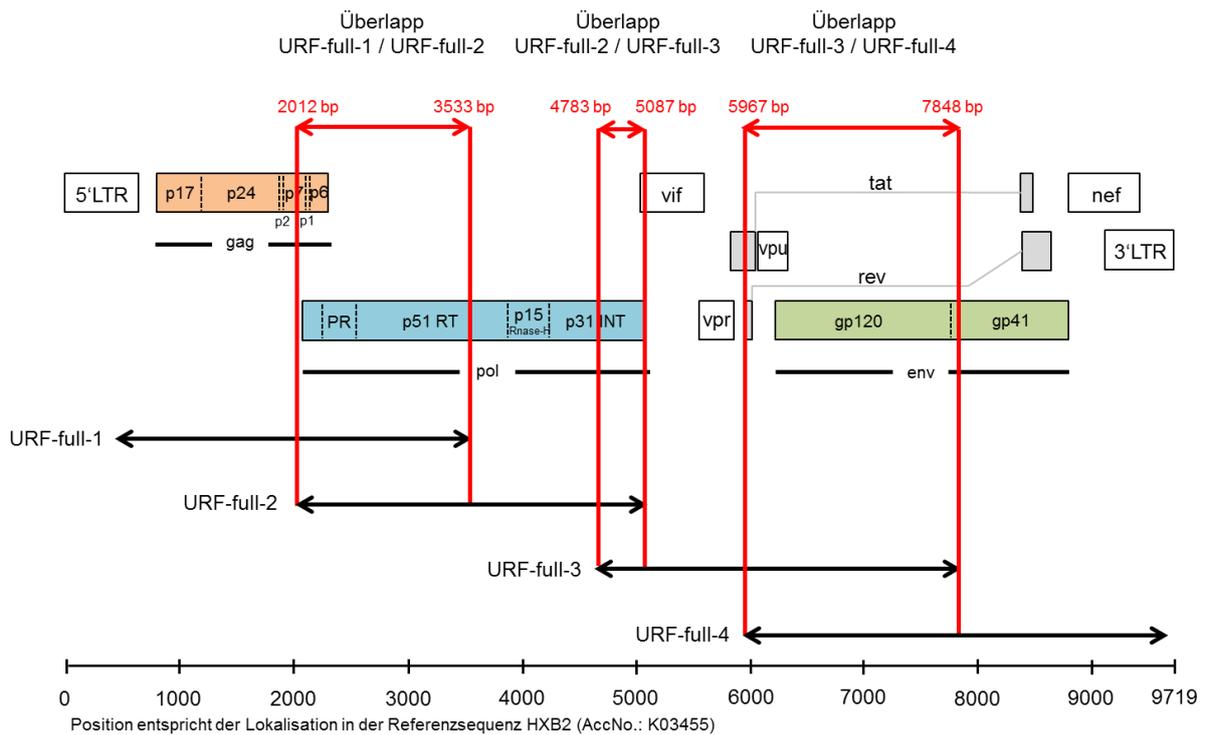


Abbildung 10: Überlappende Bereiche zwischen den Amplikons *URF-full-1* und *URF-full-2*, *URF-full-2* und *URF-full-3* und zwischen *URF-full-3* und *URF-full-4* (rot dargestellt). Die vier Amplikons *URF-full-1* bis -4 sind bezogen auf das HXB2-Genom graphisch dargestellt und zeigen die überlappenden Bereiche auf dem Genom an. Der Überlapp zwischen *URF-full-1* und *URF-full-2* liegt zwischen 2012 bp und 3533 bp, der Überlapp zwischen *URF-full-2* und *URF-full-3* zwischen 4783 bp und 5087 bp und der Überlapp zwischen *URF-full-3* und *URF-full-4* zwischen 5967 bp und 7848 bp bezogen auf HXB2.

Bei Virusisolat 13-0346 traten im überlappenden Bereich zwischen *URF-full-1* und *URF-full-2* mit 1502 bp insgesamt sieben Sequenzunterschiede (0,47 %) auf, von denen sechs Ambiguitäten darstellen, weshalb sich die überlappende Sequenz von Amplikon 1 und 2 letztendlich nur an einer Position (2004 bp bzgl. HXB2) in der Konsensussequenz unterschied. An dieser Position wurde die Nukleotidbase T in *URF-full-1* und die Nukleotidbase C in *URF-full-2* detektiert, weshalb diese Position in der Konsensussequenz als Ambiguität Y (T oder C) erscheint. Dieser Sequenzunterschied führt aber letztendlich nicht zum Aminosäurewechsel, da an dieser Stelle beide möglichen Codons TGC \underline{C} und TGT \underline{C} für die Aminosäure Cystein codieren. Alle anderen Sequenzunterschiede, die auf Ambiguitäten zurückzuführen sind und nicht in der Konsensussequenz auftreten, sind beispielsweise solche, bei denen in *URF-full-1* an einer Position die Ambiguität R (A oder G) und in *URF-full-2* an gleicher Position die Nukleotidbase A auftritt. In diesem Fall erscheint in der Konsensussequenz an dieser Stelle die Nukleotidbase A. In dem überlappenden Bereich zwischen *URF-full-2* und *URF-full-3* mit 292 bp zeigt sich nur ein Sequenzunterschied (0,34 %) und im überlappenden Bereich zwischen *URF-full-3* und *URF-full-4* mit 1808 bp insgesamt sechs Sequenzunterschiede (0,33 %), die ausschließlich Ambiguitäten darstellen

und in der Konsensussequenz nicht als Unterschied gewertet werden. Werden die prozentualen Anteile der Sequenzunterschiede von jedem Überlapp zusammengefasst, beträgt insgesamt der Anteil für das Virusisolat 1,7 %.

Für das Virus aus EDTA-Plasma (13-05995) gibt es im Überlapp zwischen *URF-full-1* und *URF-full-2* (1514 bp) insgesamt 29 Sequenzunterschiede (1,9 %), von denen drei in der Konsensussequenz auftreten, 26 Sequenzunterschiede sind auf Ambiguitäten zurückzuführen und werden in der Konsensussequenz nicht gewertet. Die drei eindeutig gewerteten Sequenzunterschiede (Position: 1481 bp, 1490 bp und 1496 bp bzgl. HXB2) stellen in ihren Codons die dritte Base (Wobble-Basen) dar, weshalb die Sequenzunterschiede letztendlich keinen Einfluss auf die Aminosäuresequenz haben. Zwischen den Amplikons *URF-full-2* und *URF-full-3* mit einem Überlapp von 290 bp wurden gar keine Sequenzunterschiede identifiziert. Zwischen *URF-full-3* und *URF-full-4* mit 1786 bp wurden insgesamt 108 Sequenzunterschiede (6 %) gezählt, von denen 100 Ambiguitäten darstellen und somit acht echte Sequenzunterschiede in der Konsensussequenz identifiziert wurden. Diese acht Sequenzunterschiede liegen im Bereich 6840-7410 bp bezogen auf HXB2 (C2 bis V4-Loop von gp120), von denen sieben zu einem Wechsel der Aminosäure führen. Bei dem Sequenzvergleich zwischen *URF-full-3* und *URF-full-4* fällt auf, dass im Bereich zwischen 7410 und 7650 bp bezogen auf HXB2 (Bereich von 240 bp, V4-Loop bis V5-Loop von gp120) insgesamt 47 Ambiguitäten (ca. 20 %) in der Konsensussequenz auftreten, die nicht auf Sequenzunterschiede zurückzuführen sind, sondern in beiden Amplikons (*URF-full-3* und *URF-full-4*) an gleicher Position detektiert worden sind. Prozentual beträgt der Anteil an Basenheterogenitäten der einzelnen Überlapps zusammen genommen 7,9 % und ist damit für das EDTA-Plasma-Virus höher als für das Virusisolat (13-346) mit 1,7 %. Allgemein bedeutet bei der Sequenzanalyse das Auftreten einer Ambiguität in einer Nukleotidsequenz, dass in der untersuchten Patientenprobe mehrere Virusvarianten (Quasispezies) vorliegen. Zusammenfassend wurde im *URF-full-1* und *URF-full-2* Überlapp (*gag*-Ende bis PR, RT) für das Virusisolat 13-0346 ein Sequenzunterschied gewertet und für das Virus aus EDTA-Plasma 13-05995 drei Sequenzunterschiede, die aber aufgrund der Wobble-Position (3. Nukleotidbase im Codon) letztendlich keinen Einfluss auf die Aminosäuresequenz haben. Im *URF-full-2* und *URF-full-3* Überlapp wurden für beide Isolate keine Sequenzunterschiede identifiziert und innerhalb des *URF-full-3* und *URF-full-4* Überlapps wurden ausschließlich für 13-05995 acht Sequenzunterschiede im hypervariablen gp120 (*env*)-Genombereich detektiert, von denen sieben zu einem Wechsel der kodierenden Aminosäure führen.

4.3 Genetische Klassifizierung von *URF-new clade* Isolaten von HIV-1 aus Oman anhand der Komplet genomsequenz

In vorangegangenen Arbeiten wurde anhand phylogenetischer Analysen von partiellen Genomsequenzen dieser *URF-new clade* Viren gezeigt, dass diese in der Baumtopologie eine eigene monophyletische Gruppe bilden. In dieser Arbeit wurden phylogenetische Stammbaumanalysen mit zwei HIV-1 Komplet genomsequenzen der *URF-new clade* Isolate aus Oman durchgeführt, um zu überprüfen, ob auch die Sequenz des Komplet genomes als eigenständige monophyletische Gruppe in der Baumtopologie lokalisiert ist und sich damit die Annahme bestätigt, dass es sich bei den neu identifizierten URFs aus Oman um eine neue zirkulierende rekombinante Form von HIV-1 handelt.

Die genetische Klassifizierung der *URF-new clade* Isolate erfolgte anhand phylogenetischer Analysen mit 170 vollständig subtypisierten HIV-1 Referenzsequenzen aus dem Subtypen-Panel der HIV-Datenbank. Die verwendeten Referenzsequenzen sind in Tabelle 7 (siehe Anhang 8.2) aufgelistet und der erstellte Stammbaum ist in Abbildung 15 (siehe Anhang 8.3) dargestellt. Für eine bessere Darstellung der Baumtopologie wurde die phylogenetische Analyse zusätzlich mit einem reduzierten Referenzsequenzset von 118 Sequenzen (in Tabelle 7 mit Index versehen) durchgeführt (Abbildung 11). Für die Erstellung der Baumtopologie wurde die Sequenz eines Isolates der Gruppe O (Acc.No.: L20571) als Außengruppe gesetzt (Abbildung 11).

Die zwei *URF-new clade* Isolate aus Oman sind in der *Clade* der G-Subtypen und G-Rekombinanten lokalisiert. Innerhalb dieser G-Subtypen und G-Rekombinanten *Clade* spalten sich die *URF-new clade* Isolate signifikant von den reinen G-Subtyp-Isolaten und G-Rekombinanten ab und bilden damit eine eigene monophyletische Gruppe. Der Bootstrap-Wert von 100 % am Knotenpunkt der beiden Isolate der *URF-new clade* stützt die Annahme, dass beide Patientenisolate nicht epidemisch miteinander verknüpft sind und es sich damit nicht um eine Infektkette handelt. Diese Tatsachen lassen die Interpretation zu, dass es sich bei den *URF-new clade* Isolaten um eine neue zirkulierende rekombinante Form handelt (nicht unter CRF01 bis CRF49, die in der Analyse enthalten waren). Die phylogenetische Analyse wird mit der Komplet genomsequenz des dritten Vertreters (14-0875) in weiterführenden Arbeiten wiederholt werden.

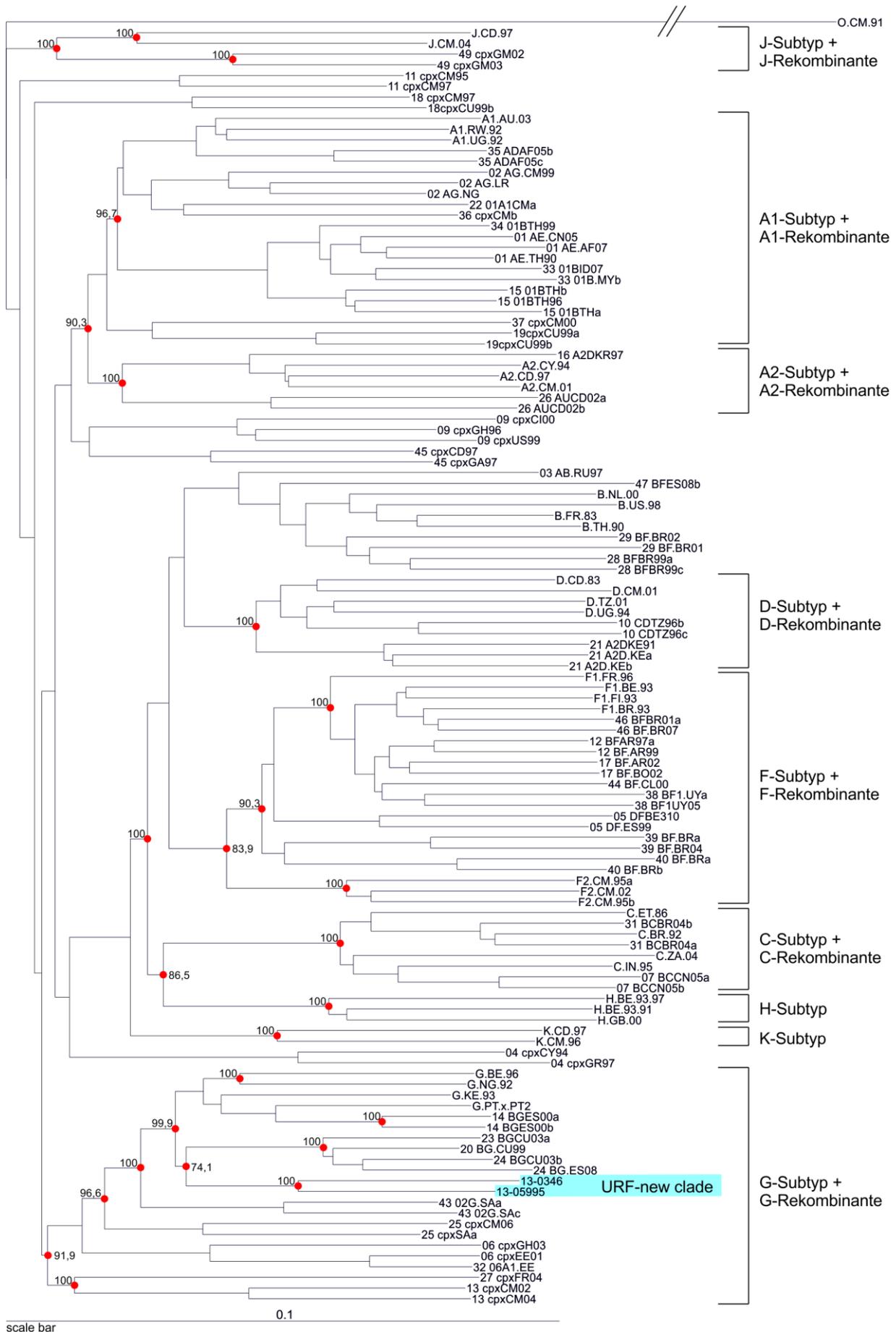


Abbildung 11: Phylogenetischer *Neighbor-Joining* Stammbaum zur Klassifizierung der *URF-new clade* Isolate aus Oman. Phylogenetische Analyse mit Kompletengenomsequenzen von Referenzsequenzen (n=118/170) des HIV-Subtypen-Panels der HIV-Datenbank und von zwei HIV-1 Patientenproben der *URF-new clade* Variante (13-0346, 13-05995). Die Sequenz O.CM.91 (Acc.No.: L20571) wurde als Außengruppe für die Erstellung der Baumtopologie benutzt und deren Astlänge für eine bessere Darstellung der Baumtopologie gekürzt (gekennzeichnet durch zwei Striche). Der Distanzbalken (*scale bar*) mit dem Wert 0,1 zeigt die Astlänge an, die einem Nukleotidaustausch von 10 % pro Position im Alignment entspricht. Die Bootstrap-Analyse wurde mit 1000 Replikaten durchgeführt. Signifikante Knotenpunkte (Bootstrap-Wert: >70 %) für die Einteilung der *Clades* sind rot gekennzeichnet. In Klammern sind die Subtypen und die dazugehörigen rekombinanten Formen zusammengefasst. Die *URF-new clade* Isolate sind blau hervorgehoben.

4.4 Analyse der rekombinanten Genomstruktur der *URF-new clade* Isolate aus Oman

Nachdem die phylogenetische Analyse gezeigt hat, dass die zwei *full-length* Sequenzen in einer monophyletischen Gruppe innerhalb der G/ G-Rekombinanten lokalisiert sind, wurde die rekombinante Genomstruktur analysiert. Zur Identifikation der rekombinanten Genomstruktur der *URF-new clade* Viren wurde das *Recombinant Identification Program* (RIP), das von der HIV-Datenbank zur Verfügung gestellt wird, angewendet. Das Programm erstellt zunächst ein Alignment aus der eingeladenen Kompletengenomsequenz der *URF-new clade* mit 12 Referenzsequenzen, die in Form von Konsensussequenzen jeweils einen Subtypen und die CRF01_AE repräsentieren. Die 12 Referenzsequenzen sind farbig unterschiedlich gekennzeichnet (Abbildung 12, C). Mithilfe von Distanzmessungen zwischen den Referenzsequenzen und der Kompletengenomsequenz der *URF-new clade* Isolate wurde ein Ähnlichkeits-Distanzplot (*similarity*-Distanzplot, s-Distanzplot) erstellt (Abbildung 12, A+B). Die zwei Balken über dem Plot zeigen an, mit welcher Referenzsequenz die Kompletengenomsequenz der *URF-new clade* Isolate die größte Sequenzähnlichkeit zeigt („*best match*“, unterer Balken). Der obere Balken ist an den Positionen farbig markiert, an denen die höchste Ähnlichkeit signifikant besser ist als die zweithöchste Ähnlichkeit (Abbildung 12, A+B). Das Genom der *URF-new clade* setzt sich demnach hauptsächlich aus dem Subtyp G (dunkelblau) mit drei Anteilen von Subtyp D (hellblau) zusammen. Positionen, an denen eine Farblücke besteht, zeigen an, dass hier eine gleichwertig hohe Signifikanz für die Ähnlichkeit mit zwei Referenzen besteht, also keine eindeutige Zuordnung getroffen werden konnte. Die Positionen, an denen im unteren Balken die Farbe wechselt, zeigen an, dass hier die höchste Ähnlichkeit zu einer anderen Referenzsequenz wechselt. Die Lokalisationen der einzelnen Subtypanteile bezogen auf die jeweilige Kompletengenomsequenz der *URF-new clade* Isolate beruht in diesen Distanzplots nur auf Sequenzähnlichkeiten und stellt damit eine Rekombinationsstruktur auf Grundlage einer geschätzten Wahrscheinlichkeit dar.

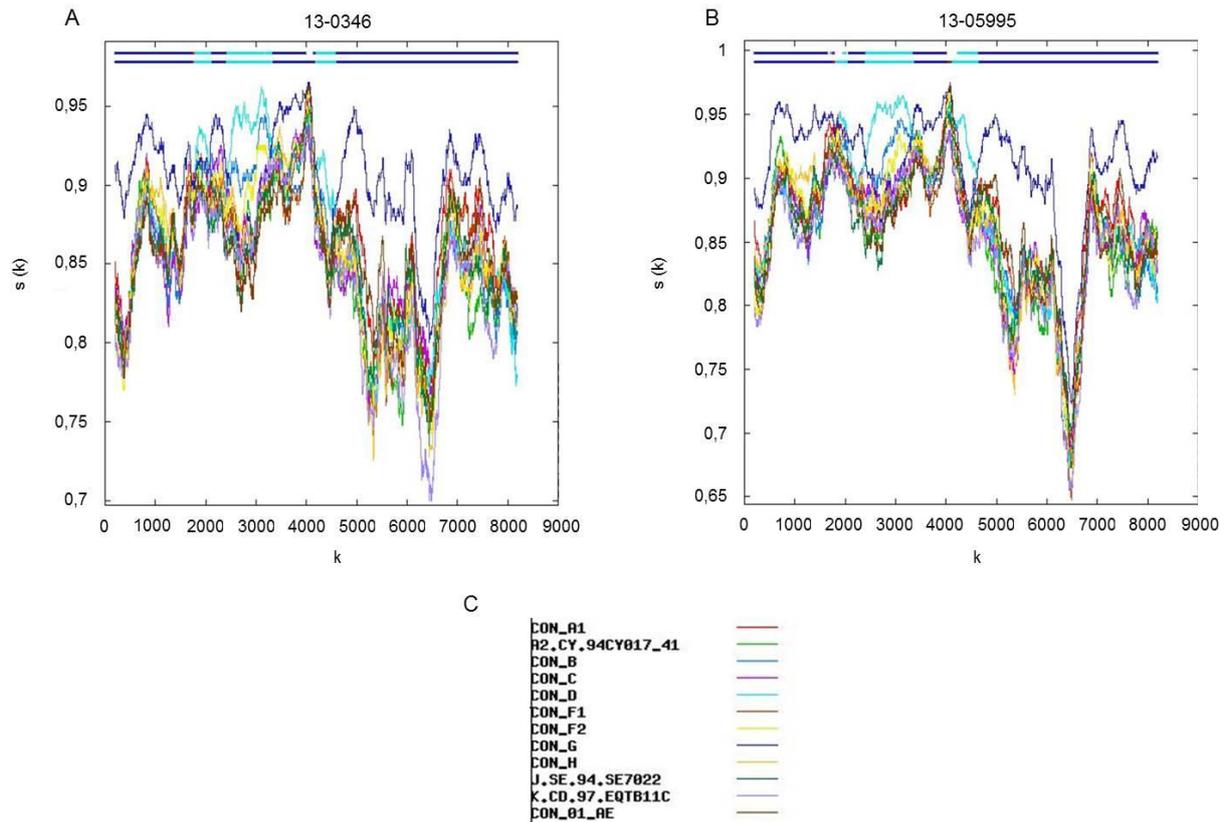


Abbildung 12: Plot der s-Distanz als Ergebnis der RIP-Analyse. Dargestellt sind die Ähnlichkeitsdistanz-Plots (s-Distanzplots) von (A) Virusisolat (13-0346) und (B) full-length Sequenz vom EDTA-Plasma (13-05995) und die dafür verwendeten Referenzsequenzen, die jeweils einen Subtypen und die CRF01_AE darstellen (C). Für die Fenstergröße wurde der Wert 400, für den Konfidenzschwennwert 90 % (*significance threshold* 0,9) gewählt. Die x-Achse (k) repräsentiert die Sequenzposition der *URF-new clade* Isolate im Zentrum des Fensters (also Beginn des Plots bei 200). Die y-Achse s(k) zeigt die Ähnlichkeit (*similarity*) zwischen dem Fenster der *URF-new clade* Sequenzen mit jeder Referenzsequenz an. Die zwei Balken über dem Plot repräsentieren die rekombinante Genomstruktur der *URF-new clade* Viren. Der untere Balken gibt die beste Übereinstimmung der *URF-new clade* Kompletengenomsequenz mit den Referenzsequenzen wieder und der obere Balken die signifikante Übereinstimmung.

Die genaue Bestimmung der phylogenetischen Rekombinationsstellen (*Breakpoints*, Positionen, an denen die Genomstruktur von einem Subtyp zu einem anderen wechselt) erfolgte mithilfe von GOBICS auf Grundlage des *jumping profile Hidden Markov Model* (jpHMM). Dieses Model ermöglicht das „Springen“ der Populationssequenz zwischen den Subtypreferenzsequenzen des vorkalkulierten Alignments und zählt die Sprünge zwischen verschiedenen Subtypen als *Breakpoint*. Auf Basis des statistischen Logarithmus werden zudem *Breakpoint*-Intervalle zwischen zwei Subtypen bestimmt. Die ermittelten Werte für die Positionen der *Breakpoints* bzw. *Breakpoint*-Intervalle sowie die Angaben der Subtypen für die Fragmente zwischen zwei *Breakpoints* sind in Tabelle 4 für das Virusisolat (13-0346) und in Tabelle 5 für das Virus aus EDTA-Plasma (13-05995) zusammengefasst. Die Rekombinationsanalyse mit jpHMM wurde auch mit der partiellen Genomsequenz des Virus aus EDTA-Plasma (14-0875) durchgeführt (Tabelle 6). In den Tabellen sind zudem die jeweiligen *Breakpoint*-Positionen bezogen auf die HXB2-Referenzsequenz angegeben.

Tabelle 4: Ergebnisse der jpHMM-Analyse für das *URF-new clade* Isolat 13-0346.

	Fragment Start- und Endposition (Breakpoints)	Breakpoint Intervall Start - Ende	Fragment Subtyp
Positionen in der Populationssequenz 13-0346	1 - 315	-	N/A
	316 - 2774	2748 - 2785	G
	2775 - 3701	3621 - 3724	D
	3702 - 4612	4576 - 4613	G
	4613 - 4878	4862 - 4910	D
	4879 - 8901	-	G
	8902 - 9113	-	N/A
Positionen der Populationssequenz 13-0346 bezogen auf die HXB2-Genomsequenz	495 - 789	-	N/A
	790 - 3281	3255 - 3292	G
	3282 - 4208	4128 - 4231	D
	4209 - 5119	5083 - 5120	G
	5120 - 5385	5369 - 5417	D
	5386 - 9411	-	G
	9412 - 9621	-	N/A

Tabelle 5: Ergebnisse der jpHMM-Analyse für das *URF-new clade* Isolat 13-05995.

	Fragment Start- und Endposition (Breakpoint)	Breakpoint Intervall Start - Ende	Fragment Subtyp
Positionen in der Populationssequenz 13-05995	1 - 278	-	N/A
	279 - 2760	2753 - 2798	G
	2761 - 3708	3602 - 3741	D
	3709 - 4601	4555 - 4611	G
	4602 - 4873	4864 - 4893	D
	4874 - 6854	-	G
	6855 - 7078	-	B
	7079 - 8862	-	G
	8863 - 9019	-	N/A
Positionen der Populationssequenz 13-05995 bezogen auf die HXB2-Genomsequenz	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 - 7409	-	G
	7410 - 7651	-	B
	7652 - 9409	-	G
9409 - 9560	-	N/A	

Tabelle 6: Ergebnisse der jpHMM-Analyse für das *URF-new clade* Isolat 14-0875.

	Fragment Start- und Endposition (Breakpoints)	Breakpoint Intervall Start - Ende	Fragment Subtyp
Positionen in der Populationssequenz 14-0875	1 – 1946	1936 – 1974	G
	1947 – 2814	2807 – 2900	D
	2815 – 3790	3744 – 3801	G
	3791 – 4062	4053 – 4069	D
	4063 – 5313	-	G
Positionen der Populationssequenz 14-0875 bezogen auf die HXB2-Genomsequenz	1329 – 3281	3271 – 3309	G
	3282 – 4148	4141 – 4234	D
	4149 – 5124	5078 – 5135	G
	5125 – 5396	5387 – 5403	D
	5394 – 6638	-	G

Die Ermittlung der *Breakpoints* einer rekombinanten Populationssequenz bezogen auf die HXB2-Referenzsequenz ist notwendig, um alle rekombinante Formen einheitlich anhand der HXB2-Genommappe darstellen zu können. In Abbildung 13 ist die rekombinante Genomstruktur der *URF-new clade* Viren basierend auf der HXB2-Referenzsequenz gezeigt. Alle drei *URF-new clade* Viren zeigen eine sehr ähnliche rekombinante Genomstruktur, die hauptsächlich aus Subtyp G besteht und mit zwei Fragmentbereichen in der *pol*- und *vif*-Genomregion, die dem Subtyp D zugeordnet werden, kombiniert ist. Das *URF-new clade* Isolat 13-05995 zeigt zusätzlich in der *env*-Genomregion (7410-7651 bp, HXB2) einen Fragmentabschnitt, der dem Subtyp B zugeordnet wird (Abbildung 13, B). Das Programm jpHMM weist aber darauf hin, dass die Detektion von Subtyp B im 3'-Genombereich häufig auf einen technischen Fehler zurückzuführen ist. Die grau hinterlegten Enden der 5'- und 3'LTR der Kompletengenomsequenzen (13-0346, 13-05995) konnten keinem bekannten Subtypen zugeordnet werden (N/A) und stellen den Anfang und das Ende der originalen Populationssequenzen dar. Die partielle Genomsequenz von 14-0875 weist die gleiche rekombinante Genomstruktur auf, bestehend aus Subtyp G und D, wie sie für die Kompletengenomsequenzen ermittelt wurde (Abbildung 13, C). Die jpHMM-Analyse sollte mit der fehlenden *env*-Sequenz von 14-0875 in weiteren Arbeiten wiederholt werden.

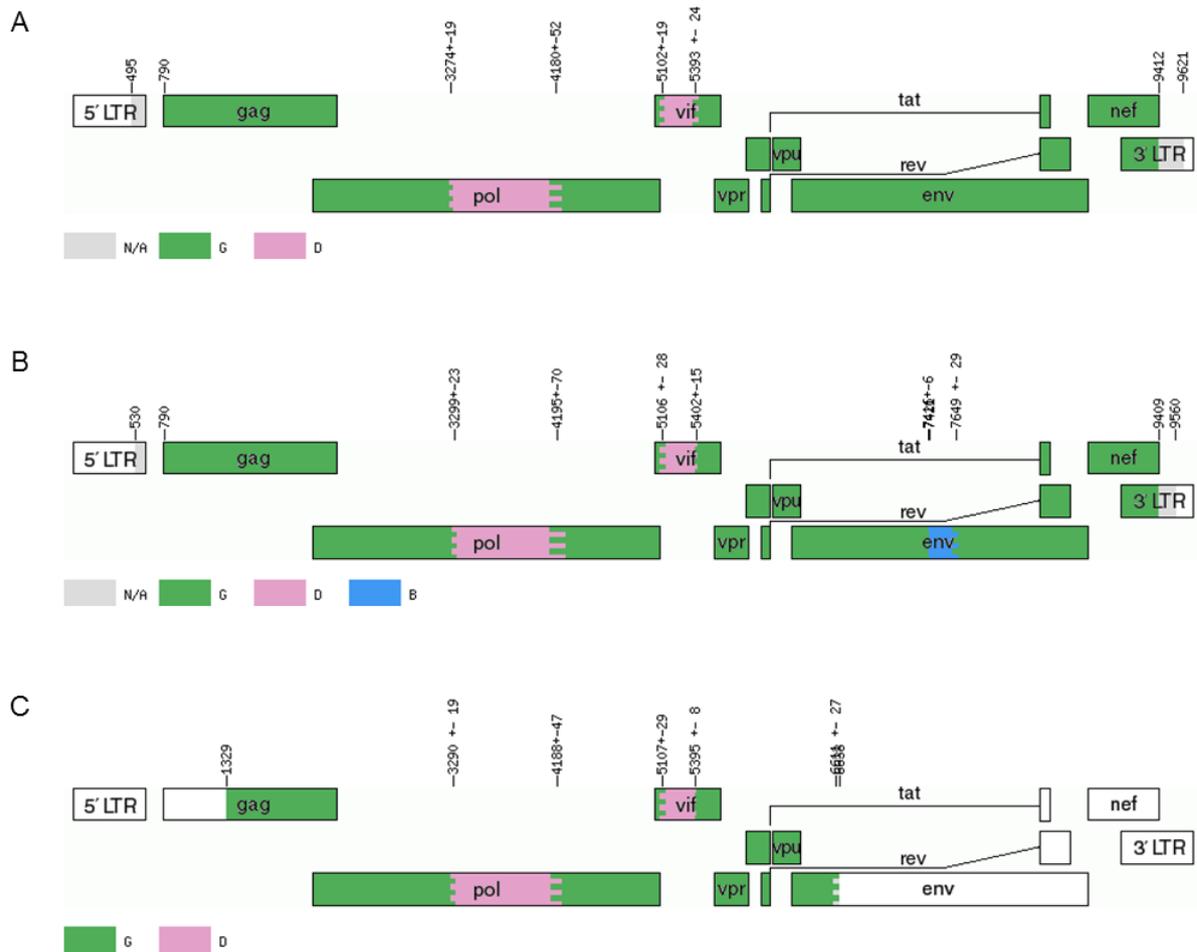


Abbildung 13: jpHMM Analysen der zwei Komplet genomsequenzen (13-0346 und 13-05995) und der partiellen Genomsequenz (14-0875) der URF-new clade Viren. Gezeigt ist die rekonstruierte Genomstruktur der URF-new clade HIV-1 Viren (A) 13-0346, (B) 13-05995 und (C) 14-0875. Die Subtypen G (grün), D (rosa), B (blau) und nicht definierbare Regionen (grau) sind farblich hinterlegt. Die farblos dargestellten 5'- und 3'-Genomenden sind nicht sequenzierte Bereiche der URF-new clade. Die HIV-1 Sequenzen wurden in die online-Anwendung von jpHMM im GOBICS-Server eingeladen. Das Programm nutzt das eigene Subtypen-Referenzalignment und den statistischen jpHMM-Algorithmus für die Bestimmung der Subtypen, *Breakpoints* und *Breakpoint*-Intervalle mit einem Signifikanzwert von 99 %. Die Lokalisationen der *Breakpoints* und *Breakpoint*-Intervalle sind als Zahlenwerte angegeben und entsprechen den Lokalisationen im HXB2-Genom.

5 Diskussion

Im Rahmen der genotypischen Resistenzanalyse bei therapierten und therapie-naiven HIV-1 Patienten aus Oman wurde auch der Subtyp der ermittelten HIV-1 *pol* Sequenzen (PR/RT und INT) analysiert. Ein hoher Anteil von ca. 47 % der *pol*-Sequenzen konnte keinem bekannten Subtyp oder CRF zugeordnet werden und wurden als URF (singuläre rekombinante Form) klassifiziert. Sechs dieser URF-Stämme bildeten in der phylogenetischen Analyse der *pol*-Sequenzen eine eigene monophyletische Gruppe, die sich signifikant von allen global bekannten Subtypen und CRFs der HIV-Datenbank unterschied, weshalb diese zunächst als *URF-new clade* bezeichnet wurden. Anhand phylogenetischer Analysen der Komplet genomsequenzen der neuen URF-Variante mit aktuellen Referenzsequenzen soll geklärt werden, ob es sich hierbei um eine neue zirkulierende Rekombinante (CRF) handelt. Damit wäre die Bestätigung erbracht, dass es sich bei den neu identifizierten Isolaten um HIV-1 Stämme handelt, die für die HIV-Epidemiologie in Oman relevant sind.

Für die Amplifikation des HIV-1 Komplet genom waren in der Arbeitsgruppe die Methoden zur *full-length* cDNA-Synthese und Amplifikation des Komplet genom in Form von vier überlappenden Fragmenten („*vier-Amplikon-PCR*“) etabliert worden. Von einem Virusisolat der *URF-new clade* war der 5'-Genombereich bereits sequenziert worden.

Um die Komplet genomsequenz zu erhalten, wurde in der vorliegenden Arbeit der 3'-Genombereich dieses Isolates amplifiziert und sequenziert. Zwei weitere Vertreter dieser neuen URF-Variante konnten vollständig direkt aus Plasma ohne vorherige Virusanzucht amplifiziert werden und eine weitere vollständige Populationssequenz sowie partielle Sequenzen des dritten Vertreters im Rahmen dieser Arbeit ermittelt werden. Mit den beiden Komplet genomsequenzen der *URF-new clade* Viren konnten diese genetisch über phylogenetische Analysen klassifiziert und ihre rekombinante Genomstruktur analysiert werden.

5.1 Amplifikation und Sequenzierung des HIV-1 Komplet genom einer neuen URF-Variante aus Oman

Die Komplet genomsequenzen der *URF-new clade* Viren wurden durch PCR-Amplifikation und Sequenzierung direkt aus genomischer viraler RNA von freien Viruspartikeln infizierter Patienten erzeugt und repräsentieren damit die zum Zeitpunkt der Blutentnahme zirkulierende und replizierende Viruspopulation im Patienten. Ermittelte Sequenzen aus viraler RNA haben demnach eine höhere pathogene Signifikanz als die Sequenzen, die

durch co-Kultivierung von infizierten Zellen von den integrierten Proviren in Spender-PBMCs vermehrt wurden (67).

Die Amplifikation des Kompletengenoms der *URF-new clade* Viren aus viraler RNA basierte auf der Synthese einer *full-length* cDNA. Die effiziente Synthese von vollständigen cDNA-Molekülen ist eine Herausforderung, da es aufgrund von ausgeprägten Sekundärstrukturen der viralen RNA, wie z.B. dem *Rev-Responsive* Element im *env*-kodierenden ORF, zu Syntheseabbrüchen der Reversen Transkriptase kommen kann (68-70). Aufgrund dessen beruhen viele in der Literatur beschriebenen Methoden zur Kompletengenomsequenzierung auf einer cDNA-Synthese in Form von überlappenden Fragmenten (64). Dadurch wird zwar in der Regel eine höhere cDNA-Ausbeute erzielt, für das Design der anschließenden Amplifikation ist man jedoch auf die einzelnen cDNA-Fragmente beschränkt. Durch die Etablierung eines Protokolls zur effizienten Synthese einer *full-length* cDNA konnte das Design zur Amplifikation des Kompletengenoms flexibel gestaltet werden und verschiedene sequenzspezifische PCR-Systeme konnten von der gleichen cDNA ausgehend verwendet werden. Die effiziente Aufreinigung von intakten RNA-Templates ist die Voraussetzung für die effiziente *full-length* cDNA-Synthese, was wiederum Voraussetzung für die erfolgreiche Amplifikation des Kompletengenoms von HIV ist. Auch die Wahl geeigneter Enzyme ist essentiell. Das für die Reverse Transkription verwendete Enzym Superscript III ist genetisch so modifiziert, dass es eine erhöhte Thermostabilität und keine RNase H-Aktivität aufweist (67). Laut Literatur wird durch die fehlende RNase H-Aktivität die Ausbeute an vollständig synthetisierten cDNA-Molekülen erhöht (71), da die RNA-Moleküle nicht während der cDNA-Synthese bereits abgebaut werden, sondern erst nach erfolgter Reversen Transkription durch nachträglicher Zugabe eines RNase H-Enzyms. Zudem ist bekannt, dass genetisch veränderte Reverse Transkriptase Enzyme ohne RNase H-Aktivität die Wahrscheinlichkeit eines Abbruchs der Reversen Transkriptase durch RNA-Sekundärstrukturen verringert (72). Vermutlich trägt auch die erhöhte Thermostabilität der Reversen Transkriptase zur effizienten Reverse Transkription bei, da das Enzym auch bei sehr hohen Temperaturen wie beim Denaturierungsschritt bei 65 °C und auch während der zweistündigen Inkubationszeit bei 50 °C aktiv bleibt (73). Des Weiteren kann die Ausbeute an spezifischen *full-length* cDNA-Molekülen unter Verwendung von sequenzspezifischen Primern erhöht werden (67), weshalb auch in der etablierten *full-length* cDNA-Synthese ein sequenzspezifischer antisense Primer (LTR-fulllength-as), der in der 3'-LTR bindet, verwendet wurde. Es ist bekannt, dass die Rate an *in vitro*-Rekombinationsereignissen bei langen cDNA-Synthesen hoch ist, wenn vorwiegend partielle cDNA-Moleküle synthetisiert werden aufgrund bereits degradierter RNA-Moleküle oder durch nicht optimale Reaktionsbedingungen der Reversen Transkription (74, 75). In der Literatur ist beschrieben, dass die Rekombinationsrate bei langen cDNA-Synthesen minimiert werden kann, wenn die Inkubationszeit erhöht wird (74,

75). Eine 120 minütige Inkubationszeit wird dabei als geeignet angesehen (75), die auch für die cDNA-Synthese der *URF-new clade* Viren eingesetzt wurde.

Die Amplifikation des HIV-1 Genoms der *URF-new clade* Viren auf der Grundlage von *full-length* cDNA erfolgte in Form von vier überlappenden Fragmenten („*vier-Amplikon-PCR*“). Amplifikationsstrategien über mehrere überlappende PCR-Amplikons (bis zu 10 überlappende Fragmente) ist für virale Genome eine gängige Methode, wie sie auch in mehreren Publikationen beschrieben sind (76-79). Meist wurden sie spezifisch für die untersuchten Isolate etabliert und können nicht immer universell für unterschiedliche HIV-1 Subtypen eingesetzt werden (80). Gall und Kollegen etablierten deshalb eine subtyp-generische *vier-Amplikon-PCR*, die auch im HIV-Studienlabor für die Amplifikation der rekombinanten HIV-1 Variante aus Oman getestet wurde (59, 64). Mit dieser Methode wurden jedoch entweder gar keine Amplifikate generiert oder zusätzlich unspezifische PCR-Produkte amplifiziert (59). Daraufhin wurden die Amplifikationsprimer für drei Fragmente (*URF-full-1*, *URF-full-2* und *URF-full-4*) subtypgenerisch optimiert und konnten erfolgreich für die Amplifikation der Patientenproben aus Oman eingesetzt werden (59, 64). Lediglich das Fragment *URF-full-3* wird mit Primern aus der Literatur generiert (64). Die Sensitivität dieser neu etablierten *vier-Amplikon-PCR* ist jedoch relativ gering, weshalb diese Methode für klinisches Material mit sehr niedrigen Viruslasten nicht geeignet ist. Bei Isolaten mit einer Viruslast unter 10^4 Kopien/ml konnten keine PCR-Amplifikate erzeugt werden. Die vier PCRs des Virusisolates (13-0346) (59) und des Virus aus EDTA-Plasma (14-0875) generierten ausschließlich spezifische PCR-Produkte, weshalb davon ausgegangen werden kann, dass mit der etablierten cDNA-Synthese effizient *full-length* cDNA-Moleküle erzeugt worden waren. Für den Vertreter der *URF-new clade* aus EDTA-Plasma (13-05995) wurden bei der *URF-full-3* und *-4* Amplifikation zusätzlich unspezifische PCR-Produkte detektiert, da dieser Bereich zum einen den hypervariablen *env*-Genombereich umfasst und der Grad der Diversität der Quasispezies im Primärmaterial (EDTA-Plasma) größer ist als im angezüchteten Virusisolat. So ist es möglich, dass speziell in dieser Probe die Amplifikationsprimer an genetisch unterschiedliche Varianten der HIV-Quasispezies hybridisieren und so genetisch unterschiedliche Amplifikationsprodukte generiert wurden. Bei der Amplifikation in mehreren Fragmenten besteht deshalb auch die Gefahr, dass bei co-zirkulierenden Subtypen im Patienten, Amplifikate von mehreren reinen Subtypvarianten generiert und zusammengesetzt werden, was wie bei einer partiellen cDNA-Synthese zu falschen Interpretationen bezüglich inter-Subtyp-Rekombinanten führen kann. Da die Möglichkeit von dualen Infektionen mit mehreren HIV-1 Subtypen besteht, ist es wünschenswert die Anzahl der PCR-Amplikons für die Kompletengenom-Amplifikation so gering wie möglich zu halten bzw. sogar eine *full-length*-Amplifikationsmethode zu etablieren, die es ermöglicht, ein einziges Amplikon über das gesamte HIV-1 Genom zu generieren.

Damit wird die Wahrscheinlichkeit erhöht, dass ein Kompletengenom einer HIV-1 Spezies amplifiziert wird. Ein Nachteil könnte sein, dass die Ausbeute an *full-length*-Amplifikaten gering ausfällt, was nachteilig für die anschließende Sequenzierung des Komplettagensoms sein kann. Die Etablierung einer effizienten *full-length* Amplifikationsmethode erweist sich vermutlich als äußerst schwierig, da in der Literatur nur wenige *full-length* Amplifikationsmethoden beschrieben sind, die zudem subtyp-spezifisch (B, C) etabliert worden sind (81-83).

Die Sequenzbewertung der *URF-new clade* Viren war für das Virusisolat (13-0346) einfacher als für das Primärmaterial (13-05995) aufgrund der erhöhten Variabilität der HIV-Varianten aus EDTA-Plasma. Vor allem die Auswertung des hypervariablen *env*-Genombereichs erwies sich durch das Auftreten von Sequenzmischungen teilweise schwierig. Bei dem Virusisolat aus Virusanzucht war die Sequenzanalyse unproblematischer, da die Viren des Patienten (13-0346) mit PBMCs co-kultiviert wurden und die replizierende Hauptvariante der HIV-Quasispezies mit der höchsten viralen Fitness und evolutionären Vorteilen dadurch selektiert wurde, was zu einer weniger divergenten Quasispezies führt. Für beide *URF-new clade* Viren war jedoch die anschließende Assemblierung der vier Einzelsequenzen zu einer Kompletengenomsequenz problemlos. Diese generierten Kompletengenomsequenzen der beiden *URF-new clade* Viren zeigten vereinzelte heterogene Basensubstitutionen. Der Anteil an Ambiguitäten, der zum Wechsel der kodierenden Aminosäure führt, fällt für die Populationssequenz aus EDTA-Plasma-Virus mit 0,64 % größer aus als für die Populationssequenz des Virusisolates mit 0,51 %. Dies liegt wie bereits zuvor erläutert vor allem in der höheren Diversität der Quasispezies aus EDTA-Plasma begründet.

Die Anteile an Sequenzunterschieden in den überlappenden Bereichen der zusammengesetzten Fragmente fällt insgesamt für das (nicht kultivierte) Virus aus EDTA-Plasma mit 7,9 % sehr viel größer aus im Vergleich zum angezüchteten Virusisolat mit 1,7 %. Da insgesamt die Basenheterogenität im Virus aus EDTA-Plasma sehr viel höher ausfällt als im angezüchteten Virusisolat, kann darauf geschlossen werden, dass der Großteil der ambiguitären Sequenzbereiche zwischen den PCR-Amplikons auf die heterogene Quasispezies zurückzuführen ist. Sequenzunterschiede können auch zum Beispiel während der Reversen Transkription durch die fehleranfällige Reverse Transkriptase aufgrund der fehlenden 3'-5' Exonukleaseaktivität (Korrekturfunktion) entstehen (73). Die Fehlerrate der verwendeten SuperScriptIII ist allerdings nicht bekannt (67). Die Mutationsrate während der Amplifikation kann als äußerst gering betrachtet werden, da für die Amplifikation ein kommerzieller Mix aus Polymerasen (Taq-DNA-Polymerase und eine High Fidelity-DNA-Polymerase) eingesetzt wurde. Die High Fidelity DNA Polymerase besitzt eine Korrektur-

funktion und kann falsch eingebaute Nukleotidbasen durch die korrekt komplementären Basen ersetzen.

Die einzelnen Schritte zur Analyse des Kompletengenoms von HIV-1 bauen somit aufeinander auf. Die verwendeten zum Teil in Vorarbeiten etablierten Protokolle sind spezifisch aufeinander abgestimmt und ermöglichen eine ausreichende Ausbeute vollständig synthetisierter cDNA-Produkte für eine erfolgreiche Amplifikation und Sequenzierung vollständiger HIV-1 Genome unterschiedlicher Subtypen und Mosaikviren.

5.2 Genetische Klassifikation einer neu identifizierten URF-Variante aus Oman

Von den weltweit sequenzierten HIV-1 Isolaten entsprechen mindestens 20 % inter-Subtyp-Rekombinanten (8). Man geht davon aus, dass die HIV-Pandemie vermutlich immer stärker von der Ausbreitung rekombinanter Formen beeinflusst wird, da bereits in bestimmten Regionen (z.B. in Malaysia, Brasilien und Westafrika) die HIV-1 Epidemie durch eine graduelle Verdrängung der dominierenden HIV-1 Subtypen aufgrund der ansteigenden Zahlen der epidemiologisch bedeutenden CRFs gekennzeichnet ist (84). Die Identifizierung neuer zirkulierender rekombinanter Formen (CRFs) in der HIV-Epidemie ist essentiell, um Informationen bezüglich der Prävalenz sowie der Verbreitung in einer Population und eventuell in bestimmten Risikogruppen zu erhalten. Zudem können diese rekombinanten Viren, die sich evolutionär in der Bevölkerung etabliert haben, spezielle Eigenschaften bezüglich der Transmissionseffizienz und Pathogenese haben (85). Zudem wird ein Einfluss auf die Virusdiagnostik, Therapiemaßnahmen sowie Impfstoffentwicklung diskutiert (85). Die Identifizierung neuer Mosaikviren und deren genetische Charakterisierung durch Bestimmung der Subtypanteile sowie der Rekombinationsstellen sind essentiell für die molekulare Epidemiologie von HIV-1.

Die phylogenetische Analyse (distanzbasiertes *Neighbor-Joining*-Verfahren) der beiden kompletten Populationssequenzen der *URF-new clade* bestätigt das Ergebnis vorangegangener phylogenetischer Analysen der *pol*-Sequenzen, nämlich dass diese HIV-1 Viren eine eigene monophyletische Gruppe bilden, die sich signifikant von den Subtypreferenzsequenzen (beinhalten alle Subtypen, sowie CRF01-CRF49) abspalten. Die Baumtopologie zeigt eine genetische Verwandtschaft der *URF-new clade* zu Subtyp G und G-Rekombinanten Referenzsequenzen. Die mittlere paarweise genetische Distanz der beiden *URF-new clade* Sequenzen nach der vorliegenden Datenlage und der Bootstrap-Wert von 100 % am Knotenpunkt beider Isolate sprechen dafür, dass es sich um epidemisch nicht verknüpfte HIV-1 Isolate handelt. Das verwendete HIV-Subtypreferenzpanel der Los Alamos Datenbank repräsentiert aber nicht alle bereits registrierten CRFs (bis CRF72, aber nicht alle

Sequenzen sind öffentlich verfügbar) und URF-Sequenzen sind gar nicht im Referenzpanel enthalten. Um die Möglichkeit auszuschließen, dass bereits ähnliche Sequenzen von CRFs und URFs in der HIV-Sequenzdatenbank vorliegen, wurden die beiden ersten Kompletengenomsequenzen der *URF-new clade* der Datenbank zur Sequenzüberprüfung zur Verfügung gestellt. Die Sequenzüberprüfung der Administratoren ergab letztendlich keine signifikante Übereinstimmung mit bereits registrierten CRFs oder URFs in der HIV-Sequenzdatenbank, was darauf hinweist, dass es sich bei den identifizierten *URF-new clade* Viren, um eine neue zirkulierende rekombinante Form handelt.

Um die rekombinante Genomstruktur von Mosaikviren zu bestimmen, werden in der Literatur zahlreiche Rekombinationsidentifizierungs-Programme beschrieben (63, 66, 86, 87). Die hier verwendete und von der HIV-Sequenzdatenbank dafür angebotene Anwendung RIP basiert auf der Bestimmung der besten Sequenzähnlichkeit zu den jeweiligen Subtyp-Referenzsequenzen (63). Diese wird durch abschnittsweise Analyse des Sequenzalignments in überlappenden gleitenden Fenstern (*sliding window*) festgestellt (63). Mit diesem Programm kann eine erste Vorhersage der Rekombinationsstruktur getroffen werden, die sich bei den Populationssequenzen der *URF-new clade* hauptsächlich aus Anteilen von Subtyp G und Anteilen von Subtyp D im Bereich der *pol*- und *vif*-Genomregion zusammensetzt. Mit dem jpHMM-Programm kann die rekombinante Genomstruktur mit einer relativ exakten Bestimmung der phylogenetischen Breakpoints untersucht werden (62, 66). Dieses Programm ist vor allem sensitiver und genauer, da es ein relativ großes Referenzalignment von 309 Sequenzen verwendet im Vergleich zu RIP, das die Ähnlichkeit zu jeweils nur einer Referenzsequenz jeden Subtyps bestimmt (62, 66). Laut Literatur ist das entwickelte jpHMM *tool* in der Bestimmung der Breakpoints effizienter als Methoden, die Breakpoints phylogenetisch detektieren wie z.B. das REGA HIV-1 Subtypisierungs-tool (66). Zudem bestimmt jpHMM sehr signifikant ($p = 0,99$) nicht eindeutig subtypisierbare Regionen und gibt wahrscheinlichkeitkalkulierte Breakpoint-Intervalle an. Dies trägt wesentlich dazu bei, dass keine falschen Schlüsse auf vorhergesagte Subtypanteile gezogen werden (66). Die Ergebnisse der jpHMM-Analyse für die Populationssequenzen der *URF-new clade* sind ähnlich der Ergebnisse mit der RIP-Analyse. Anzumerken ist, dass bei der jpHMM-Analyse bei dem Primärmaterial 13-05995 ein kurzer Genomabschnitt im *env*-Genombereich dem Subtyp B zugeordnet wurde. Auf der Website von GOBICS wird jedoch angegeben, dass detektierte Subtyp B Anteile im 3'-Genombereich methodisch bedingt nicht immer korrekt sind und eine Limitierung des jpHMM *tools* darstellen (88). Ebenso können die selten vorkommenden Subtypen J, K und H oft nicht effizient genug aufgrund der geringen Anzahl vorhandener Kompletengenomsequenzen detektiert werden (66, 88). Der detektierte Subtyp B-Bereich in *env* kann auch auf die Populationssequenz von 13-05995 zurückgeführt werden, da genau in diesem bestimmten Genomabschnitt die Sequenz auffällig viele

Ambiguitäten (~20 %) aufzeigt, die auf beiden Sequenzen von *URF-full-3* und *URF-full-4* detektiert wurden und damit echte Ambiguitäten aus der HIV-Quasispezies des Patienten repräsentieren. Vermutlich berücksichtigt das jpHMM-Programm diese ambiguitären Stellen, was eventuell zum „Sprung“ in den Subtyp B führt. Eine weitere Ursache für die Sequenzambiguitäten können *in vitro* Rekombinationsereignisse sein, die entweder während der cDNA-Synthese oder während der PCR stattgefunden haben.

Insgesamt kann aufgrund der Ergebnisse der RIP und jpHMM-Analyse davon ausgegangen werden, dass es sich bei den *URF-new clade* Viren um G/D-Rekombinante handelt, hauptsächlich bestehend aus Subtyp G mit kleineren Anteilen von Subtyp D in der *pol*- und *vif*-Region des HIV-1 Genoms. Die bestimmten Subtypbereiche im Genom sollten in weiterführenden Arbeiten durch phylogenetische Analysen der einzelnen Bereiche mit Referenzsequenzen verifiziert werden, ebenso der durch jpHMM detektierte B-Subtyp Bereich in *env*, um eindeutig sagen zu können, ob es sich dabei um ein Artefakt handelt, wie es in der Programmdokumentation beschrieben ist.

In der HIV-Sequenzdatenbank gibt es bisher keine publizierten CRFs, die ausschließlich aus Subtyp G und D bestehen, sondern nur in Kombination mit anderen Subtypen wie zum Beispiel CRF02_AG, CRF05_DF, CRF10_CD, CRF14_BG und CRF16_A2D (44, 45, 89-92). Nur eine in Cuba identifizierte komplex aufgebaute CRF19_cpx hat eine rekombinante Genomstruktur aus den Subtypen A1, D und G, die aber keine Ähnlichkeit bezüglich der Genomstruktur der identifizierten G/D-Rekombinanten aus Oman zeigt (93).

Bei der Subtypisierung der *pol*-Sequenzen der untersuchten Patientenkohorte aus Oman (n = 105) wurden keine Isolate identifiziert, die ausschließlich dem reinen Subtyp G oder D HIV-1 zuzuordnen sind (unveröffentlichte Daten, RKI). Dies könnte bedeuten, dass die Anzahl an Subtyp G und D Infektionen in Oman gering ist und Rekombination zwischen beiden Subtypen nicht in Oman, sondern außerhalb in einer Region stattfand, in der diese beiden Subtypen co-zirkulieren. Hinzu kommt, dass die ermittelten Rekombinationsstellen und Rekombinationsintervalle zwischen Subtyp G und D der drei Isolate aus Oman nahezu identisch sind, was eher für Intra-Subtyp-Rekombinationen charakteristisch ist. Die evolutionäre Distanz der analysierten *full-length*-Sequenzen ist relativ hoch und entspricht eher der Distanz wie man sie zwischen Vertretern innerhalb eines Subtyps findet. In Regionen, in denen eine breite Subtypverteilung vorherrscht, ist die Wahrscheinlichkeit erhöht, dass durch Rekombinationsereignisse zwischen verschiedenen Subtypen neue CRFs in der Bevölkerung etabliert werden (43). Bei neu entstandenen CRFs sind die Breakpoint-Positionen und-Verteilungen (Intervalle) zwischen mehreren Isolaten dieser CRF unterschiedlicher und damit auch die Anordnung der Subtypanteile im Genom (94, 95).

All diese Erkenntnisse deuten darauf hin, dass die neu identifizierte G/D-Rekombinante nicht erst kürzlich durch Rekombinationsereignisse in Oman entstanden ist, sondern schon sehr

viel länger in der humanen Population zirkuliert und bisher unentdeckt blieb. Allgemein liegt der Anteil an Subtyp G Infektionen mit 5 % und Subtyp D Infektionen mit 2 % weltweit niedrig (9). Die höchste Subtyp G Präsenz ist in west- und zentralafrikanischen Ländern (35 % und 11 %), wo auch die CRF02_AG dominiert. Subtyp D Infektionen sind mit 50 % in Sudan sehr dominant (96, 97). Als Ursprung der neu identifizierten G/D-Rekombinante aus Oman kommt deshalb Westafrika in Betracht. Infektionen mit Subtyp G und D sind auch in den benachbarten Ländern des Omans vertreten. So sind 16 % einer Kohorte von HIV-Patienten aus Jemen mit Subtyp D HIV-1 infiziert und in einer Studie mit HIV-Patienten aus Saudi Arabien sind 25 % der untersuchten Isolate Subtyp G zuzuordnen (77, 97). Hinweise auf G/D-Rekombinanten aus diesen Ländern gibt es bislang jedoch nicht (77, 97). Um den Ursprung der einzelnen Subtypanteile der hier untersuchten omanischen Isolate zu untersuchen, sollten künftig mit diesen einzelnen Bereichen phylogenetische Analysen mit existierenden G- und D- Sequenzen der HIV-Datenbank durchgeführt werden. Allerdings existieren nicht viele vollständig subtypisierte G- und D-Isolate aus jeder Region (z.B. Isolate aus Sudan, wo 50 % Subtyp D-Infektionen vorliegen) in der Sequenzdatenbank (44), weshalb der Ursprung der einzelnen Subtypanteile der *URF-new clade* vermutlich nicht eindeutig bestimmt werden kann.

Die Resultate der hier vorgelegten Arbeit, sowohl der phylogenetischen Analyse als auch der Rekombinationsstrukturanalysen der *URF-new clade* Isolate aus Oman lassen den Schluss zu, dass es sich bei diesen Viren um eine neue CRF handelt. Die Kriterien zur Definition einer CRF sind erfüllt.

5.3 Ausblick

In der Zukunft sollte für die Verifizierung des Ergebnisses, dass es sich bei den *URF-new clade* Viren um eine neue CRF handelt, die ermittelte partielle Genomsequenz des dritten Vertreters zu einer Kompletengenomsequenz vervollständigt werden, um damit phylogenetische Untersuchungen und Analysen der kompletten rekombinanten Genomstruktur durchführen zu können. Die bereits identifizierten Subtypanteile in der Genomstruktur der *URF-new clade* sollten durch phylogenetische Analysen mit Referenzsequenzen der einzelnen Bereiche bestätigt werden. Darüber hinaus können dann auch Rückschlüsse auf den Ursprung der neuen CRF gezogen werden.

Für eine weitere methodische Optimierung soll in weiterführenden Arbeiten die „*vier-Amplikon-PCR*“ bezüglich der Sensitivität noch verbessert werden und die Effizienz der etablierten subtypgenerischen Amplifikation bei anderen HIV-1 Subtypen überprüft werden, um sie möglichst als universelle Amplifikationsmethode von HIV-1 Kompletengenomen verwenden zu können. Diese Methode könnte auch zu einer effektiven Identifikation und

genetischen Charakterisierung von rekombinanten HIV-1 Varianten, die in Deutschland zirkulieren, eingesetzt werden, was zu einer Identifikation von weiteren rekombinanten HIV-1 Varianten in der Welt beitragen würde. Zudem wäre eine Etablierung einer subtyp-generischen *full-length* Amplifikationsmethode von Vorteil, um die *in vitro* Rekombinationsrate bei Amplifikationen von Genomen in mehreren Fragmenten zu minimieren.

Durch die Sanger-Sequenzierung können vor allem bei hoher Diversität der HIV-Quasispezies in Patienten Konsensussequenzen mit vielen ambiguitären Nukleotiden erzeugt werden, was zu Problemen bei der Untersuchung der rekombinanten Genomstruktur führt. Mit Sequenzierungsmethoden basierend auf *Next Generation Sequencing (NGS)* kann die gesamte HIV-Quasispezies eines Patienten sequenziert werden, wodurch auch Varianten detektiert werden, die nur zu geringen Anteilen in der Quasispezies vorhanden sind. Demzufolge kann auch das Vorhandensein mehrerer HIV-1 Subtypen in einem Patienten identifiziert werden. So kann genauer eingeschätzt werden, ob die identifizierte rekombinante Genomstruktur auf *in vitro* Rekombinationsereignissen beruht oder wirklich der untersuchten Probe entspricht.

6 Zusammenfassung

Die globale HIV-1 Epidemie ist durch die unterschiedliche geographische Verteilung verschiedener HIV-1 Subtypen (A-D, F-H, J, K) und derzeit 72 zirkulierender rekombinanter Formen (CRF) charakterisiert. Um eine neue CRF zu definieren, muss diese in mindestens drei epidemisch nicht verknüpften Infektionen identifiziert werden; zwei Kompletengenomsequenzen sowie eine dritte partielle Genomsequenz, mit der die Rekombinationsstruktur bestätigt werden kann, müssen für die Klassifikation einer CRF vorliegen. Bei omanischen HIV-1 infizierten Patienten mit Therapieversagen wurden im Rahmen der Resistenzbestimmung singuläre rekombinante *pol*-Sequenzen (URFs) identifiziert, die eine eigene monophyletische Gruppe in der Stammbaumanalyse bildeten und zunächst als „*URF-new clade*“ benannt wurden. Ziel der vorliegenden Arbeit war, das vollständige Genom von mindestens zwei inklusive eines partiellen Genoms der putativ neuen Virusvarianten zu analysieren und die phylogenetische Sequenzanalyse dieser HIV-Varianten durchzuführen, um damit eine Klassifikation als neue CRF zu ermöglichen.

Das Kompletengenom der putativ neuen HIV-Variante wurde in einer „*vier-Amplikon-PCR*“ mit vier überlappenden Fragmenten amplifiziert und sequenziert (Sanger). In dieser Arbeit konnte von einem Isolat mit bereits bekannter 5'-Genomsequenz die 3'-Genomsequenz zur Kompletengenomsequenz vervollständigt werden. Zusätzlich konnten von zwei weiteren Isolaten alle vier Fragmente amplifiziert werden. Von diesen wurde eine weitere Kompletengenomsequenz und eine partielle Genomsequenz bestimmt. Die Kompletengenomsequenzen bildeten in der *Neighbor Joining* Stammbaumanalyse wiederum eine eigenständige monophyletische Gruppe. Die relativ hohe paarweise genetische Distanz (> 2,5 %) der viralen Sequenzen bestätigte die epidemiologische Unabhängigkeit der Infektionen. Die Analysen der Genomstruktur zeigen für alle drei Isolate eine sehr ähnliche Rekombinationsstruktur, die sich überwiegend aus Subtyp G und geringen Anteilen an Subtyp D zusammensetzt (*pol*- und *vif*-Region).

Die Ergebnisse der Stammbaumanalyse und der Analyse der Rekombinationsstruktur dieser neu identifizierten omanischen Isolate lassen den Schluss zu, dass es sich bei diesen Viren um eine neue CRF handelt. Die Kriterien zu ihrer Definition sind mit Abschluss dieser Masterarbeit erfüllt. Die offizielle Klassifizierung durch HIV Taxonomie-Experten ist aktuell in Bearbeitung.

7 Summary

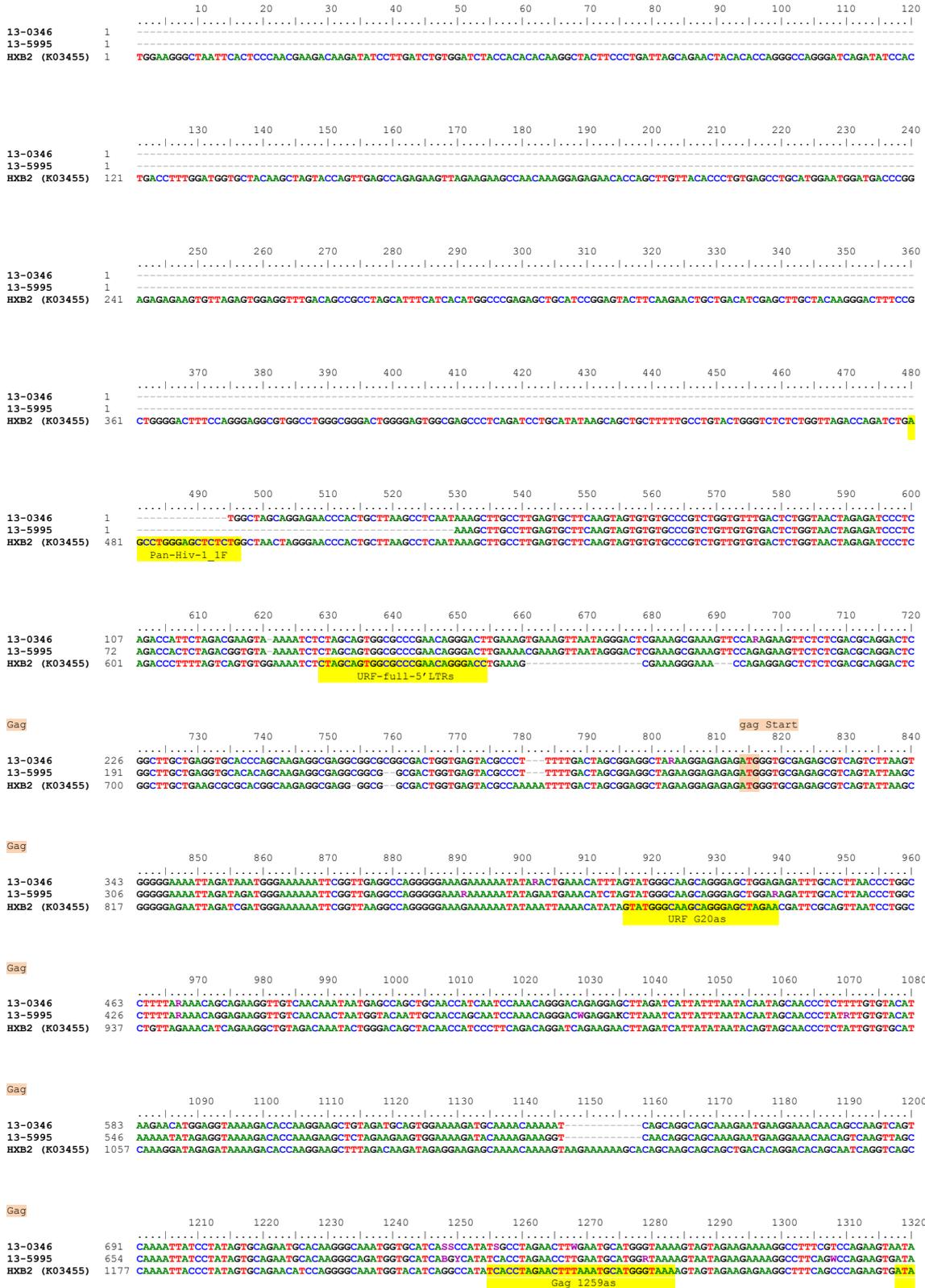
The global HIV-1 epidemic is characterized by the different geographic spread of diverse HIV-1 subtypes (A-D, F-H, J, K) and currently 72 circulating recombinant forms (CRF). To define a new CRF the same virus variant needs to be identified in at least three epidemically unlinked infections; two full-length genome sequences and a third partial genome sequence which confirms the recombination structure of the two full-length genomes have to be presented to classify a CRF. In Omani patients with treatment failure unique recombinant forms (URFs) of *pol*-sequences were identified during resistance testing. These sequences were localized in a distinct clade of the phylogenetic tree (neighbor joining method) and were first named *URF-new clade*. The aim of this study was the analysis of at least two full-length HIV genomes including one partial genome and phylogenetic sequence analysis of the new clade strains to finally enable their future classification as new CRF.

The full-length genome of the putative new HIV variant was amplified with a "four-amplicon-PCR" with four overlapping PCR fragments and sequenced (Sanger method). During this thesis the 3'-genome sequence of a viral isolate of which the 5'-genome was already analyzed was determined and a full-length sequence was produced. In addition, all four PCR fragments could be amplified from two other candidate isolates. Of those, one full-length sequence and one partial genome sequence was determined. In the neighbor joining phylogenetic tree the full-length sequences co-localized in a distinct clade confirming the results obtained with the *pol*-subsequences. The relatively high pairwise genetic distance (> 2,5 %) of the viral sequences confirmed that unlinked HIV infections were identified. Analyses of the recombinant genome structure of all three isolates revealed a very similar pattern of recombination mainly composed of subtype G combined with small parts of subtype D (*pol* and *vif* region).

The results of the phylogenetic analysis and the analysis of the pattern of recombination of these isolates support the view that the newly identified Omani HIV isolates represent indeed a new CRF. The criteria to define a new CRF are full-filled by finalizing this Master thesis. The official classification by HIV taxonomy experts is currently in progress.

8 Anhang

8.1 Alignment der Kompletengenomsequenzen der *URF-new clade* mit der Referenzsequenz HXB2 (K03455)



Gag

1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440

13-0346 811
 13-5995 774
 HXB2 (R03455) 1297 **CCCAATGTTTTCAGATATATCAGAAGGCCACCCACCAAGATTTAAACACCATGCTAAACACAGTGGGGGACATCAAGCAGCCATGCAAAATGTTAAAGAGACCATCAATGAGGAAGCT**
 Gag 1322s

Gag

1450 1460 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560

13-0346 931
 13-5995 894
 HXB2 (R03455) 1417 **GCAGAVTGGGACAGGTTACATCCACCCACAGCCAGGGCCATTCCACAGGCCAGATAAGAGAACCAAGGGGAAGTGATATAGCAGGAAGTACTAGTACCCACAGGAACAAATAACATGG**

Gag

1570 1580 1590 1600 1610 1620 1630 1640 1650 1660 1670 1680

13-0346 1051
 13-5995 1014
 HXB2 (R03455) 1537 **ATGACCAATTAATCCACTATCCAGTAGGAGAAATTTATAAAGATGGATAATCTGGGATTAACAAATAGTAAGAAATGATAGCCCTGTAGCATTTGGACATAGCAAGGGCCCA**

Gag

1690 1700 1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

13-0346 1171
 13-5995 1134
 HXB2 (R03455) 1657 **AAAGAACCCTTTAGAGATTATGTAGACAGGTTCTTTAAACCCCTTAAGAGCTGAGCAAGCTACACAGGAGTAAACAACTGGATGACAGACACCTTTGTTAGTCCAAAATGCAAAACCCAGAT**

Gag

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

13-0346 1291
 13-5995 1254
 HXB2 (R03455) 1777 **TGTAAAGCATTTTAAAGCATTGGGACCCAGCGGCTACACTAGAAAGATGATGACAGCATGTCAAGGAGTAGGAGGACCCGGCCATTAAGCAAGAGTTTGGCTGAAGCAATGAGCCAA**

Gag

1930 1940 1950 1960 1970 1980 1990 2000 2010 2020 2030 2040

13-0346 1411
 13-5995 1374
 HXB2 (R03455) 1897 **GTAAACAAATTCAGCTAC---CATAAATGATGCAGAGAGCCATTTTAAAGAACCAAGAAAGTTGTTAAGTGTTCCTAATTTGTGGCAAGAGGGCCACAGCCAGAAATTCAGGGCCCTC**

Gag Pol

gag-pol -1 ribosomal slip site
 pol-Start

2050 2060 2070 2080 2090 2100 2110 2120 2130 2140 2150 2160

13-0346 1528
 13-5995 1494
 HXB2 (R03455) 2014 **AGGAAAAGGGCTTTGGAAATGTGGAAAGGACGGACATCAATGAAAGACTGCACAGAAAGCAGGCTAAATTTTTAGGGAATCTGGCCTCCAGCAAGGGAGCCAGGGAATTTT**
 Pol 2001s Pol 2036s

Gag Pol

2170 2180 2190 2200 2210 2220 2230 2240 2250 2260 2270 2280

13-0346 1648
 13-5995 1614
 HXB2 (R03455) 2134 **CTTCAGAGCAGACCGAGCCAAACAGCCCAACAGAAAGAGCTTCAGGTCTGGGGTAGAGACAACACTCCCCCTCAGAGCAGGACCGATAGACAAGGACTGTATCTTTAAGCTTCC**

Gag Pol

gag-frame Ende

2290 2300 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

13-0346 1747
 13-5995 1731
 HXB2 (R03455) 2254 **CTCAGTCACTCTTTGGCAACGCCCTCTGTCACAAATAAAGATAGGGGGCACTAAAGAAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAATGAGTTTGCCAGGAA**
 Pol 2322s

Pol

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

13-0346 1867
 13-5995 1851
 HXB2 (R03455) 2374 **GATGAAAACCAAAAATGATAGGGGAAATGGAGTTTATCAAAATGAGACAGTATGATCAGATCTCATAGAAAATCTGGGACATAAAGCTATAGGTACAGTATTAGTAGGACTACAC**
 Pol 2402as

Pol

2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

13-0346 1987
 13-5995 1971
 HXB2 (R03455) 2494 **CTGTCAACATAATTTGGAAGAAATCTGTGACTCAGATTTGGTTGCACTTTAAATTTTCCCAATAGCCCTATTGAGACTGTACCAGTAAAAATAAGCCAGGAATGGATGGCCCAAAAGTTA**

Pol

2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760

13-0346 2107
 13-5995 2091
 HXB2 (R03455) 2614 **AAACAATGGCCATTGACAGAAAGAAAAATAAAGCATTAACAGAAAATTTGTACAGAGCTGGAAAAGGAAAGAAAAATTCAAAAATTTGGCCCTGAAAATCCATACAATCTCCAAATTTG**
 Pol 2696s

Pol
 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880
 13-0346 2227
 13-5995 2211 CCAATAAGAAAAAGACAGTACTAAGTGGAGAAAATAGTAGATTTTCAGAGACTTAATAAAGAAACCCAGACTTTGGGAGTCAACTAGGAATCCCATCCCCAGGGTTAAAAA
 HXB2 (R03455) 2734 CCAATAAGAAAAAGACAGTACTAAGTGGAGAAAATAGTAGATTTTCAGAGACTTAATAAAGAAACCCAGACTTTGGGAGTCAACTAGGAATCCCATCCCCAGGGTTAAAAA

Pol
 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
 13-0346 2347 AGAACAAATCAGTACAGTACTAGATGTGGGGATGCATATTTTCAGTTCCCTTATATGAAGAAATTTAGAAAATATACGCATTCACTATACCTAGTACAATAATGAGACACAGGAA
 13-5995 2331 AGAAAAATCAGTACAGTACTAGATGTGGGGATGCATATTTTCAGTTCCCTTATATGAAGAAATTTAGAAAATATACGCATTCACTATACCTAGTACAATAATGAGACACAGGAA
 HXB2 (R03455) 2854 AGAAAAATCAGTACAGTACTAGATGTGGGGATGCATATTTTCAGTTCCCTTATATGAAGAAATTTAGAAAATATACGCATTCACTATACCTAGTACAATAATGAGACACAGGAA

Pol
 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120
 13-0346 2467 TTAGATATCAGTATAATGTGTTCCACAGGGATGGAAGGATCCACAGCAATATCCAAAGTAGCATGACAAAAATCTTAGAGCCCTTTAGAGCAAAAAATCCAGAAATAGTATATC
 13-5995 2451 TTAGATATCAGTATAATGTGTTCCACAGGGATGGAAGGATCCACAGCAATATCCAAATAGCATGACAAAAATCTTAGAGCCCTTTAGAGCAAAAAATCCAGAAATAGTATATC
 HXB2 (R03455) 2974 TTAGATATCAGTATAATGTGTTCCACAGGGATGGAAGGATCCACAGCAATATCCAAAGTAGCATGACAAAAATCTTAGAGCCCTTTAGAGCAAAAAATCCAGAAATAGTATATC
 Pol 3034s

Pol
 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240
 13-0346 2587 AATAATGGATGACCTTATATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAGCAAAAAATAGAGGAATTAAGAGAACACTTATTGAAATGGGGACTTACCACCCAGATAAAAAATC
 13-5995 2571 AATAATGGATGACCTTATATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAGCAAAAAATAGAGGAATTAAGAGAACACTTATTGAAATGGGGACTTACCACCCAGATAAAAAATC
 HXB2 (R03455) 3094 AATAATGGATGACCTTATATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAGCAAAAAATAGAGGAATTAAGAGAACACTTATTGAAATGGGGACTTACCACCCAGATAAAAAATC

Pol
 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360
 13-0346 2707 AGAAGAACCCCAATCCATTGGATGGGATATGAGCTCCATCCTGACAAATGGACAGTACCAACTATAAAGTTGCCAGAAAAGAAAGCTGGACTGCAATGATATACAGAAGTTAGTAG
 13-5995 2691 AGAAGAACCCCAATCCATTGGATGGGATATGAGCTCCATCCTGACAAATGGACAGTACCAACTATAAAGTTGCCAGAAAAGAAAGCTGGACTGCAATGATATACAGAAGTTAGTAG
 HXB2 (R03455) 3214 AGAAGAACCCCAATCCATTGGATGGGATATGAGCTCCATCCTGACAAATGGACAGTACCAACTATAAAGTTGCCAGAAAAGAAAGCTGGACTGCAATGATATACAGAAGTTAGTAG

Pol
 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480
 13-0346 2827 GAAATTAATTTGGCAAGTCAGATTTATCCAGAAATTAAGTAAAGCAATATGTAATGTTTGGGAAACCCAGCAATGACAGAAATTAATCCACTGACAGTAAGAGCAATTAG
 13-5995 2811 GAAATTAATTTGGCAAGTCAGATTTATCCAGAAATTAAGTAAAGCAATATGTAATGTTTGGGAAACCCAGCAATGACAGAAATTAATCCACTGACAGTAAGAGCAATTAG
 HXB2 (R03455) 3334 GAAATTAATTTGGCAAGTCAGATTTATCCAGAAATTAAGTAAAGCAATATGTAATGTTTGGGAAACCCAGCAATGACAGAAATTAATCCACTGACAGTAAGAGCAATTAG
 Pol 3454as

Pol
 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
 13-0346 2947 AACTGGCAGAAAACAGGAAATTTCTAAAAGCCAGTACATGGAGTGTATTATGACCCATCAAAGACTTAATAGCAGAAAACAGAAAACAGGGCAAGCCAATGGCATATCAAAATTT
 13-5995 2931 AACTGGCAGAAAACAGGAAATTTCTAAAAGCCAGTACATGGAGTGTATTATGACCCATCAAAGACTTAATAGCAGAAAACAGAAAACAGGGCAAGCCAATGGCATATCAAAATTT
 HXB2 (R03455) 3454 AACTGGCAGAAAACAGGAAATTTCTAAAAGCCAGTACATGGAGTGTATTATGACCCATCAAAGACTTAATAGCAGAAAACAGAAAACAGGGCAAGGGCAAGCCAATGGCATATCAAAATTT
 Pol 3532as

Pol
 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720
 13-0346 3067 ATCAAGAACCTTTAAAAATCTGAAAACAGGAAAATATGCAAGATGAGGGGTCGCCACACTAATGATGTAACAACTTAACAGAGGCAGTCCAAAAATAGCTCAGAAAGTATAGTAA
 13-5995 3051 ATCAAGAACCTTTAAAAATCTGAAAACAGGAAAATATGCAAGATGAGGGGTCGCCACACTAATGATGTAACAACTTAACAGAGGCAGTCCAAAAATAGCTCAGAAAGTATAGTAA
 HXB2 (R03455) 3574 ATCAAGAACCTTTAAAAATCTGAAAACAGGAAAATATGCAAGATGAGGGGTCGCCACACTAATGATGTAACAACTTAACAGAGGCAGTCCAAAAATAGCTCAGAAAGTATAGTAA
 URF-pol 3626s

Pol
 3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840
 13-0346 3187 TATGGGAAAAATCTCTAAATTCAGATTAACCAATACAAAAGGAGACATGGGAACTATGGTGGACAGAAATTTGCCAAGCCACTGGATTCTGAGTGGGAGTTTGTCAATACCCCTCCTT
 13-5995 3171 TATGGGAAAAATCTCTAAATTCAGATTAACCAATACAAAAGGAGACATGGGAACTATGGTGGACAGAAATTTGCCAAGCCACTGGATTCTGAGTGGGAGTTTGTCAATACCCCTCCTT
 HXB2 (R03455) 3694 TATGGGAAAAATCTCTAAATTCAGATTAACCAATACAAAAGGAGACATGGGAACTATGGTGGACAGAAATTTGCCAAGCCACTGGATTCTGAGTGGGAGTTTGTCAATACCCCTCCTT

Pol
 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960
 13-0346 3307 TAGTAAATTTGGTTCAGTTAGAAAAGAAACCCATTAATAGGGGCAGAACTTTCTATGATAGTGGGGCAGTAAATAGAGAGACTAAATAGGAAAGCCAGGATATGTTACTGACAGAG
 13-5995 3291 TAGTAAATTTGGTTCAGTTAGAAAAGAAACCCATTAATAGGGGCAGAACTTTCTATGATAGTGGGGCAGTAAATAGAGAGACTAAATAGGAAAGCCAGGATATGTTACTGACAGAG
 HXB2 (R03455) 3814 TAGTAAATTTGGTTCAGTTAGAAAAGAAACCCATTAATAGGGGCAGAACTTTCTATGATAGTGGGGCAGTAAATAGAGAGACTAAATAGGAAAGCCAGGATATGTTACTGACAGAG

Pol
 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080
 13-0346 3427 GAAGCAAAAAGTTGCTCTTTAGCTGACACACAAATCAGAAAGCTGAATTAACAAGCCATTAATCTAGTTTGCAGGATTCAGGATCAGAAGTAAACATAGTAAACAGACTCCCAATATG
 13-5995 3411 GAAGCAAAAAGTTGCTCTTTAGCTGACACACAAATCAGAAAGCTGAATTAACAAGCCATTAATCTAGTTTGCAGGATTCAGGATCAGAAGTAAACATAGTAAACAGACTCCCAATATG
 HXB2 (R03455) 3934 GAAGCAAAAAGTTGCTCACTTAACTGACACACAAATCAGAAAGCTGAGTTACAAGCAATTTATCTAGTTTGCAGGATTCAGGATCAGAAGTAAACATAGTAAACAGACTCCCAATATG

Pol
 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
 13-0346 3547 CATTAGAAATCATTCAGCACACCCAGATTAAGAGTGAATCAGAGTTAGTCAATCAAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTTCACTATCATGGTTACCAGCACAAAGGAA
 13-5995 3531 CATTAGAAATCATTCAGCACACCCAGATTAAGAGTGAATCAGAGTTAGTCAATCAAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTTCACTATCATGGTTACCAGCACAAAGGAA
 HXB2 (R03455) 4054 CATTAGAAATCATTCAGCACACCCAGATTAAGAGTGAATCAGAGTTAGTCAATCAAAATAATAGAGCAGTTAATAAAAAAGGAAAAGGTTCACTATCATGGTTACCAGCACAAAGGAA

Pol
 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320
 13-0346 3667
 13-5995 3651 TTGGAGGAATGAACCAAGTAGATAAAATTAGTCAGTGGAAATCAGGAAGTACTATTTTATAGTGGATAGATAGGCCCAAGTGAACATGGAAGTATCACAGTAATGGAGGCAA
 HXB2 (R03455) 4174 TTGGAGGAATGAACCAAGTAGATAAAATTAGTCAGTGGAAATCAGGAAGTACTATTTTATAGTGGATAGATAGGCCCAAGTGAACATGGAAGTATCACAGTAATGGAGGCAA
 5' INTs URF-pol

Pol
 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440
 13-0346 3787 TGRCTAGTATTAAATACCATCTATAGTAGCAAAAGAGATGGCCAGCTGTGATAAATGCCAGTAAAAGGGGAGCCATGCATGGCCAGTAGACTGTAGTCCAGGAATATGGC
 13-5995 3771 TTGGTAGTATTAAATCTGCCACCTGTAATAGCAAAAGAGATGGCCAGCTGTGATAAATGCCAGTAAAAGGGGAGCCATGCATGGCCAGTAGACTGTAGTCCAGGAATATGGC
 HXB2 (R03455) 4294 TTGGTAGTATTAAATCTGCCACCTGTAATAGCAAAAGAGATGGCCAGCTGTGATAAATGCCAGTAAAAGGGGAGCCATGCATGGCCAGTAGACTGTAGTCCAGGAATATGGC
 4303as

Pol
 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560
 13-0346 3907 AATTAGATTGTACACATTTAGAGGAAAAGTTATCCCTGGTAGCAGTACATGTAGCCAGTGGCTATATAGAACAGAACTTCCAGCAGAAACAGGACAGGAAACAGCATATTTATAT
 13-5995 3891 AATTAGATTGTACACATTTAGAGGAAAAGTTATCCCTGGTAGCAGTACATGTAGCCAGTGGCTATATAGAACAGAACTTCCAGCAGAAACAGGACAGGAAACAGCATATTTATAT
 HXB2 (R03455) 4414 AATTAGATTGTACACATTTAGAGGAAAAGTTATCCCTGGTAGCAGTACATGTAGCCAGTGGCTATATAGAACAGAACTTCCAGCAGAAACAGGACAGGAAACAGCATATTTATAT

Pol
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680
 13-0346 4027 TAAAAATAGCAGGAATGGCCAGTGAAGATAATACATACAGACAATGGTGGCAATTTTCATTAGTGTCCAGTAAAAGGCAGCATGTTGGTGGGCAAAATCACACAAGAAATTTGGAAATTC
 13-5995 4011 TAAAAATAGCAGGAATGGCCAGTGAAGATAATACATACAGACAATGGTGGCAATTTTCATTAGTGTCCAGTAAAAGGCAGCATGTTGGTGGGCAAAATCACACAAGAAATTTGGAAATTC
 HXB2 (R03455) 4534 TAAAAATAGCAGGAATGGCCAGTGAAGATAATACATACAGACAATGGTGGCAATTTTCACCCTGCTAGCGTTAGGCGCCCTGTTGGTGGGCGGAAATCAACAGGAATTTGGAAATTC
 URF-pol 4560as

Pol
 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
 13-0346 4147 CCTACAAATCCCAAGCCAGGAGTAGTAGAATCCATGAATAAAGAAATAAAGAAAATCAATTTGGCCAGTCAAGATCAAGCTGAACACCTTAAGCAGCAGTACAAATGGCAGTATCCA
 13-5995 4131 CCTACAAATCCCAAGCCAGGAGTAGTAGAATCCATGAATAAAGAAATAAAGAAAATCAATTTGGCCAGTCAAGATCAAGCTGAACACCTTAAGCAGCAGTACAAATGGCAGTATCCA
 HXB2 (R03455) 4654 CCTACAAATCCCAAGCCAGGAGTAGTAGAATCCATGAATAAAGAAATAAAGAAAATTTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGCAGCAGTACAAATGGCAGTATCCA

Pol
 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920
 13-0346 4267 TTCACAAATTTAAAAGAAAAGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAAATATAGACATATATAGCATCAGACATACAACTAGGMACTCAAAAACAAATTTCAAAAATTCAAA
 13-5995 4251 TTCACAAATTTAAAAGAAAAGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAAATATAGACATATATAGCATCAGACATACAACTAGGMACTCAAAAACAAATTTCAAAAATTCAAA
 HXB2 (R03455) 4774 TTCACAAATTTAAAAGAAAAGGGGGATTGGGGGGTACAGTGCAGGGGAAAGAAATATAGACATATATAGCATCAGACATACAACTAGGMACTCAAAAACAAATTTCAAAAATTCAAA
 Pan-HIV-1_3F

Pol
 4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040
 13-0346 4387 ATTTTCGGGTTTATTTACAGGACAGCAGAACCCATTGGAAAAGCCAGCAAACTACTCTGAAAAGGTGAAGGGCCAGTAGTAATAAAGCAATACGAAATAAAAGTAGTACCAA
 13-5995 4371 ATTTTCGGGTTTATTTACAGGACAGCAGAACCCATTGGAAAAGCCAGCAAACTACTCTGAAAAGGTGAAGGGCCAGTAGTAATAAAGCAATACGAAATAAAAGTAGTACCAA
 HXB2 (R03455) 4894 ATTTTCGGGTTTATTTACAGGACAGCAGAACCCATTGGAAAAGCCAGCAAACTACTCTGAAAAGGTGAAGGGCCAGTAGTAATAAAGCAATACGAAATAAAAGTAGTACCAA

Pol
 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160
 13-0346 4507 GAAGAAAAGCCAAAGTATTAGGGATTATGAAAACAGATGGCAGGTGATGATTTGGTGGCAGGTAGCAGGATGAGGTTAGAACATGGAACAGTTTGGTAAAACCCATATATATGTT
 13-5995 4491 GAAGAAAAGCCAAAGTATTAGGGATTATGAAAACAGATGGCAGGTGATGATTTGGTGGCAGGTAGCAGGATGAGGTTAGAACATGGAACAGTTTGGTAAAACCCATATATATGTT
 HXB2 (R03455) 5014 GAAGAAAAGCCAAAGTATTAGGGATTATGAAAACAGATGGCAGGTGATGATTTGGTGGCAGGTAGCAGGATGAGGTTAGAACATGGAAGTTTGGTAAAACCCATATATATGTT
 3p31 s/as

Vif
 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280
 13-0346 4627 TCAAGAAAAGCTCAAGGATGGGTATACAGACACCCATTGAATCCCTAAACCCAGAGTAACTTCAGAAAGTACACATCCCATAGGAGAGATAAGCTGATAGTACAAATATTTGGGGT
 13-5995 4611 TCAAGAAAAGCTCAAGGATGGGTATACAGACACCCATTGAATCCCTAAACCCAGAGTAACTTCAGAAAGTACACATCCCATAGGAGAGATAAGCTGATAGTACAAATATTTGGGGT
 HXB2 (R03455) 5134 TCAAGAAAAGCTCAAGGATGGGTATAGACATCCTATGAAAGCCCTCATCCAGAAATAACTTCAGAAAGTACACATCCCATAGGAGATGCTAGATTTGGTAAATACAAATATTTGGGGT

Vif
 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
 13-0346 4747 CTGCTACAGGAGAAAGAAATGGCATCTGGTTCAGGGAGTCTCCATAGAAATGGAGAAAAGAAAGATATAGCACACAAGTAGAACCTGACCTGGCAGACCACTAATTCATATATATAT
 13-5995 4731 CTGCTACAGGAGAAAGAAATGGCATCTGGTTCAGGGAGTCTCCATAGAAATGGAGAAAAGAAAGATATAGCACACAAGTAGAACCTGACCTGGCAGACCACTAATTCATATATATAT
 HXB2 (R03455) 5254 CTGCTACAGGAGAAAGAAATGGCATCTGGTTCAGGGAGTCTCCATAGAAATGGAGAAAAGAAAGATATAGCACACAAGTAGAACCTGACCTGGCAGACCACTAATTCATATATATAT

Vif
 5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520
 13-0346 4867 TTTGATTTTTCAGAACTGCCATAAGGAAAGCCATATAGGAGATATAGTATAGAACCTAGGTGTGARTCAAGCAGGACAYACYAAGGYAGGATCTCTCAAAATTTTGGCCTGAAA
 13-5995 4851 TTTGATTTTTCAGAACTGCCATAAGGAAAGCCATATAGGAGATATAGTATAGAACCTAGGTGTGARTCAAGCAGGACAYACYAAGGYAGGATCTCTCAAAATTTTGGCCTGAAA
 HXB2 (R03455) 5374 TTTGATTTTTCAGAACTGCCATAAGGAAAGCCATATAGGAGATATAGTATAGAACCTAGGTGTGARTCAAGCAGGACAYACYAAGGYAGGATCTCTCAAAATTTTGGCCTGAAA
 URF 5435 s/as

Vif
 Vpr
 5530 5540 5550 5560 5570 5580 5590 5600 5610 5620 5630 5640
 13-0346 4987 GCATTAATTAACCAAAAAGATTAAGCCACCTTTGCTAGTGTATAGAAAATTAACAGAAATAGATGGAACAAGCCCAAGAAAACAGGGGCCACAGAGAGGCCATACATGACTGGA
 13-5995 4971 GCATTAATTAACCAAAAAGATTAAGCCACCTTTGCTAGTGTATAGAAAATTAACAGAAATAGATGGAACAAGCCCAAGAAAACAGGGGCCACAGAGAGGCCATACATGACTGGA
 HXB2 (R03455) 5494 GCATTAATTAACCAAAAAGATTAAGCCACCTTTGCTAGTGTATAGAAAATTAACAGAAATAGATGGAACAAGCCCAAGAAAACAGGGGCCACAGAGAGGCCATACATGACTGGA
 ENVoutF1

Vif
Vpr
vif frame Ende
5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750 5760
13-0346 5107 CATTGGAACTGTAGAGAGCTTAAAAAGAAAGCTTTAGACATTTTCTAGGCCCCCTGGCTCYATGGVTTAGGCARFARNGCTATAACACCTTATGGAGATCTTGGGAGGAGTTGAAAG
13-5995 5091 CATTAGAACTGTAGAGAGCTTAAAAAGAAAGCTTTAGACATTTTCTAGGCCCCCTGGCTCCATGGCTTGGACAGTAGCTATAACACCTTATGGAGATCTTGGGAGGAGTTGAAAG
HXB2 (R03455) 5614 CATTAGAGCTTTTAGAGAGCTTAAAAAGAAAGCTTTAGACATTTTCTAGGATTTGGCTCCATGGCTTAGGCACACATATCTATGAACCTTATGGGATCTTGGCAGAGTTGAAAG

Vpr
Tat exon 1
vpr Ende
tat 1 Start
5770 5780 5790 5800 5810 5820 5830 5840 5850 5860 5870 5880
13-0346 5227 CAATAATAAGAATACTACACACGCTACTGTTTATCCATTT CAGAAATTTGGGTGCCAACAATAGCAGAAATAGGCCATTACTCCACAGAGAAGAGTAAGAAATGGATCCAGTAGATCTTAACCT
13-5995 5211 CAATAATAAGAATACTACACACGCTACTGTTTATCCATTT CAGAAATTTGGGTGCCAACAATAGCAGAAATAGGCCATTACTCCACAGAGAAGAGTAAGAAATGGATCCAGTAGATCTTAACCT
HXB2 (R03455) 5734 CAATAATAAGAATACTACACACGCTACTGTTTATCCATTT CAGAAATTTGGGTGCCAACAATAGCAGAAATAGGCCATTACTCCACAGAGAAGAGTAAGAAATGGATCCAGTAGATCTTAACCT

Tat exon 1
Rev exon 1
5890 5900 5910 5920 5930 5940 5950 5960 5970 5980 5990 6000
13-0346 5346 AAGCCCTGGAAATCATCCGGGAGTCCAGCCTAAAACCTGCTTGAACCACTGGTATTGTAAGTGTGTGGCCATTGCAAGTTTGCTTCTGAACAAAGGCTTAGGCATCTCTAATGG
13-5995 5330 AAGCCCTGGAAATCATCCGGGAGTCCAGCCTAAAACCTGCTTGAACCACTGGTATTGTAAGTGTGTGGCCATTGCAAGTTTGCTTCTGAACAAAGGCTTAGGCATCTCTAATGG
HXB2 (R03455) 5854 AAGCCCTGGAAATCATCCGGGAGTCCAGCCTAAAACCTGCTTGAACCACTGGTATTGTAAGTGTGTGGCTTTCATGCCAAGTTGTTTCTAACAAGGCTTAGGCATCTCTAATGG

Tat exon 1
Rev exon 1
Vpu
tat 1 Ende
rev 1 Ende
vpu Start
6010 6020 6030 6040 6050 6060 6070 6080 6090 6100 6110 6120
13-0346 5466 CAGGAAGAAGCGGAAG---CGACGAGGAACCTCCAGGACGGTGGAGGATCATCAAAATCCGTGACCAAAAGCAGTAAGTAGTAAATAATAGTGTATGTAATGGGGCTTTAGAAATATTT
13-5995 5450 CAGGAAGAAGCGGAAGCGGACGAGGAACCTCCAGGACGGTGGAGGATCATCAAAATCCGTGACCAAAAGCAGTAAGTAGTAAATAATAGTGTATGTAATGGGGCTTTAGAAATATTT
HXB2 (R03455) 5974 CAGGAAGAAGCGGAGACGACGAGGAACCTCCAGGACGGTGGAGGATCATCAAAATCCGTGACCAAAAGCAGTAAGTAGTAAATAATAGTGTATGTAATGGGGCTTTAGAAATATTT
Pan-HIV-1_4P
CATGTAAAGCCACTATACCAATAGTAG

6130 6140 6150 6160 6170 6180 6190 6200 6210 6220 6230 6240
13-0346 5583 CAATAGTAGGCTAGTAGTAGTACGATCTTAGCAGCTATAAATGTTGTGGACTATAGTATATATACAAATATAAGGAAATAGAAACAGAAAAAATAGACAAATTTACTTAAAGAAATAGAG
13-5995 5568 CAATAATAGGATAGTAGTAGGAGCTTAGCAGCTATAAATGTTGTGGACTATAGTATATATACAAATATAAGGAAATAGAAACAGAAAAAATAGACAAATTTACTTAAAGAAATAGAG
HXB2 (R03455) 6084 CAATAGTAGGCTAGTAGTAGTACGATCTTAGCAGCTATAAATGTTGTGGACTATAGTATATATACAAATATAAGGAAATATAAGCAAAAGAAAAATAGACAGGTTAATAGTAGACTAATAG

Vpu
Env
env Start
vpu Ende
6250 6260 6270 6280 6290 6300 6310 6320 6330 6340 6350 6360
13-0346 5703 AAGGACGAGAGCAGTGGAAATGAGAGCGAGGGGAGACAGAGCGAATGGCAGCTCAGG---TGGATGTGGAGATTTTGATCTTTGGTTGGTGAATTTCTAGAGCCCTAATAC
13-5995 5688 AAGGACGAGAGCAGTGGAAATGAGAGCGAGGGGAGACAGAGCGAATGGCAGCTCAGG---TGGAAATGGGAGCTTTGATCCCTGGATGGTGAATTTCTAGAGCTTCAAAATGAC
HXB2 (R03455) 6204 AAGGACGAGAGAGCAGTGGAAATGAGAGTGAAGGAGAAATATAGCAGCTTTGGAGATGGGGTGGAGATGGGCAACATGCTCCTTTGGGATGTTGATGATCTGAGTGTACAGAAAA
URF 6229as

Env
6370 6380 6390 6400 6410 6420 6430 6440 6450 6460 6470 6480
13-0346 5820 TTGTGGCTCAGACTCAATATAGGAGTACCTGTATGGGAAATGCGAGATACCCCTCTAATTTGTGCATCTGATGCTAAGCAACTAGTACTGAAGCCATAATGCTGGGCTACACATGCC
13-5995 5805 TTGTGGCTCAGACTCAATATAGGAGTACCTGTATGGGAAATGCGAGATACCCCTCTAATTTGTGCATCTGATGCTAAGCAACTAGTACTGAAGCCATAATGCTGGGCTACACATGCC
HXB2 (R03455) 6324 TTGTGGCTCAGACTCAATATAGGAGTACCTGTATGGGAAAGCAACCCTCTAATTTGTGCATCTGATGCTAAGCAACTAGTACTGAAGCCATAATGCTGGGCTACACATGCC
Env 5901s

Env
6490 6500 6510 6520 6530 6540 6550 6560 6570 6580 6590 6600
13-0346 5940 TGTGTACCCACAGACCCCAACCCACAGAGATAGATCTGTAAGAAATGTAACAGAAATTTAACATGTGGAAAAATAACTGGTAGAACAAATGCATGAGGATATAATAGTTTATGGGAT
13-5995 5925 TGTGTACCCACAGACCCCAACCCACAGAGATAGATCTGTAAGAAATGTAACAGAAATTTAACATGTGGAAAAATAACTGGTAGAACAAATGCATGAGGATATAATAGTTTATGGGAT
HXB2 (R03455) 6444 TGTGTACCCACAGACCCCAACCCACAGAGATAGATCTGTTGTAAGTGTACAGAAATTTAACATGTGGAAAAATGCACTGGTAGAACAAATGCATGAGGATATAATAGTTTATGGGAT
URF 6543s

Env
6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720
13-0346 6060 GAAAGCTTAAGCCATGTTGTAAGTAAACCCCTCTGTTTACTTTTACAAATGAGTAGAATAAAGCAATAGTAATAATGGAAGCATGAGTAT---TAGTAATGCAACAGG-GAC
6240 GAAAGCTTAAGCCATGTTGTAAGTAAACCCCTCTGTTTACTTTTACAAATGAGTAGAATAAAGCAATAGTAATAATGGAAGCATGAGTAT---TACTGTAGGCAACAATC---GAA
HXB2 (R03455) 6564 CAAGCCCTAAGCCATGTTGTAAGTAAACCCCTCTGTTTACTTTTAAAGTGACATGATTTGAAAGATGATCTAATACCAATAGTAGAGGGGAGAAATGATAATGGAGAAAGGAGG
TRRT

Env
6730 6740 6750 6760 6770 6780 6790 6800 6810 6820 6830 6840
13-0346 6171 ATGAAAACTGCTCTTTCAATGTACCCACAGAAATAGAGATAAAAAGAAAGCAAGAAATAGCGCTTTTCTATAAAATGATATAGTGCCAAATGATATAATAATAATAGTAATAGGACA
13-5995 6132 ATGAAAACTGCTCTTTCAATGTACCCACAGAAATAGAGATAAAAAGAAAGCAAGAAATAGCGCTTTTCTATAAGATGATATAGTACCAATAGATGGGAGTACAGT
HXB2 (R03455) 6684 ATGAAAACTGCTCTTTCAATGTACCCACAGAAATAGAGATAAAAAGAAAGCAAGAAATAGCGATTTTCTATAAACTGATATAATAACCAATAGATAATGATACCTC-----

Env
6850 6860 6870 6880 6890 6900 6910 6920 6930 6940 6950 6960
13-0346 6291 GATTATAGCTAATAAATGTAATGCTCACAAATTAACAGGCTTTGCCAAGGTAATCTTTTGAACCAATTCCTATACATATTTGTGCTCCAGCTGGTTTGGGATTTTAAAGTGTAGG
13-5995 6240 GATTATAGCTAATAAATGTAATGCTCACAAATTAACAGGCTTTGCCAAGGTAATCTTTTGAACCAATTCCTATACATATTTGTGCTCCAGCTGGTTTGGGATTTTAAAGTGTAGG
HXB2 (R03455) 6792 AGCTATAAGTTGCAAGTGTAAACCTCAGCTTACACAGGCTTTGCCAAGGTAATCTTTTGAACCAATTCCTATACATATTTGTGCTCCAGCTGGTTTGGGATTTTAAAGTGTAGG
URF 6881as

Env
6970 6980 6990 7000 7010 7020 7030 7040 7050 7060 7070 7080
13-0346 6411 GATAAGGATTTCAATGGACAGGACCAATGTCACAAATGTCAGTACAGTACAAATGACACATGGTATTAGGCCAGTGGTATCACTCAGTTACTGCTGAATGGCAGTATAGCAGAAG
13-5995 6360 GATAAGGATTTCAATGGACAGGACCAATGTCACAAATGTCAGTACAGTACAAATGACACATGGTATTAGGCCAGTGGTATCACTCAGTTACTGCTGAATGGCAGTATAGCAGAAG
HXB2 (R03455) 6912 AATAAGCCTTCAATGGACAGGACCAATGTCACAAATGTCAGTACAGTACAAATGACACATGGTATTAGGCCAGTGGTATCACTCAGTTACTGCTGAATGGCAGTATAGCAGAAG
Env 6537s

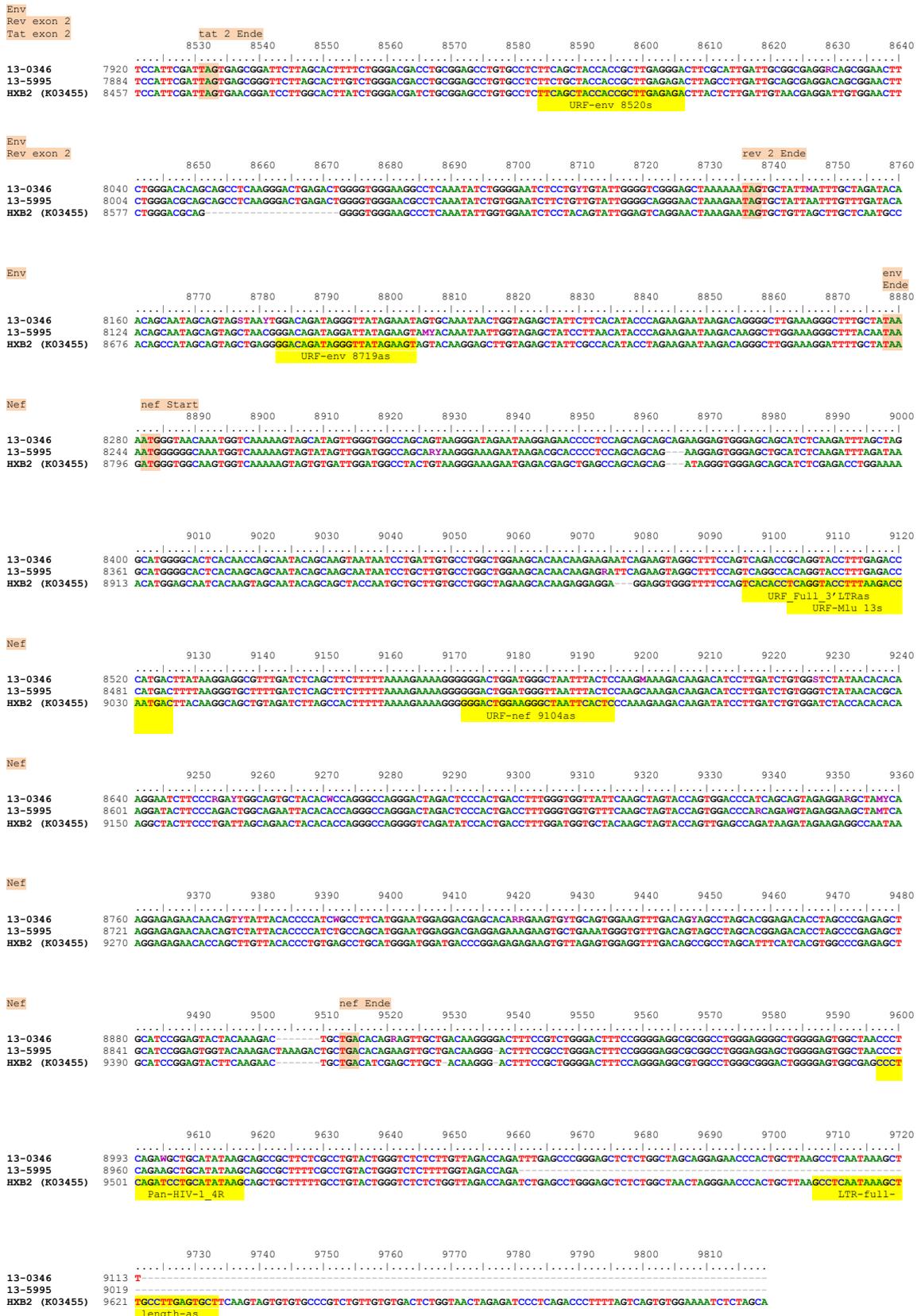


Abbildung 14: Alignment der beiden Kompletengenomsequenzen von 13-0346 und 13-05995 der URF-new clade mit der Referenzsequenz HXB2 (Acc.No.: K03455). Zudem sind die Lokalisationen der verwendeten Sequenzierungsprimer dargestellt, die Positionsangaben stimmen aufgrund der Sequenzverschiebungen im Alignment nicht mit den Positionen bezogen auf die HXB2-Referenzsequenz überein.

8.2 HIV-1 Referenzsequenzen aus der Los Alamos HIV Sequenzdatenbank, Ausgabe 2010

Tabelle 7: Referenzsequenzen des Subtypenpanels der Los Alamos Datenbank, Ausgabe 2010 (n=170)

Isolat	Acc. No.	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
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
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
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

* ausgewählte Isolate für eine veranschaulichte Darstellung der phylogenetischen Analyse der *URF-new clade* Viren (13-0346, 13-05995)

8.3 Stammbaum der Komplet genomsequenzen von *URF-new clade* HIV-1 aus Oman mit 170 Referenzsequenzen

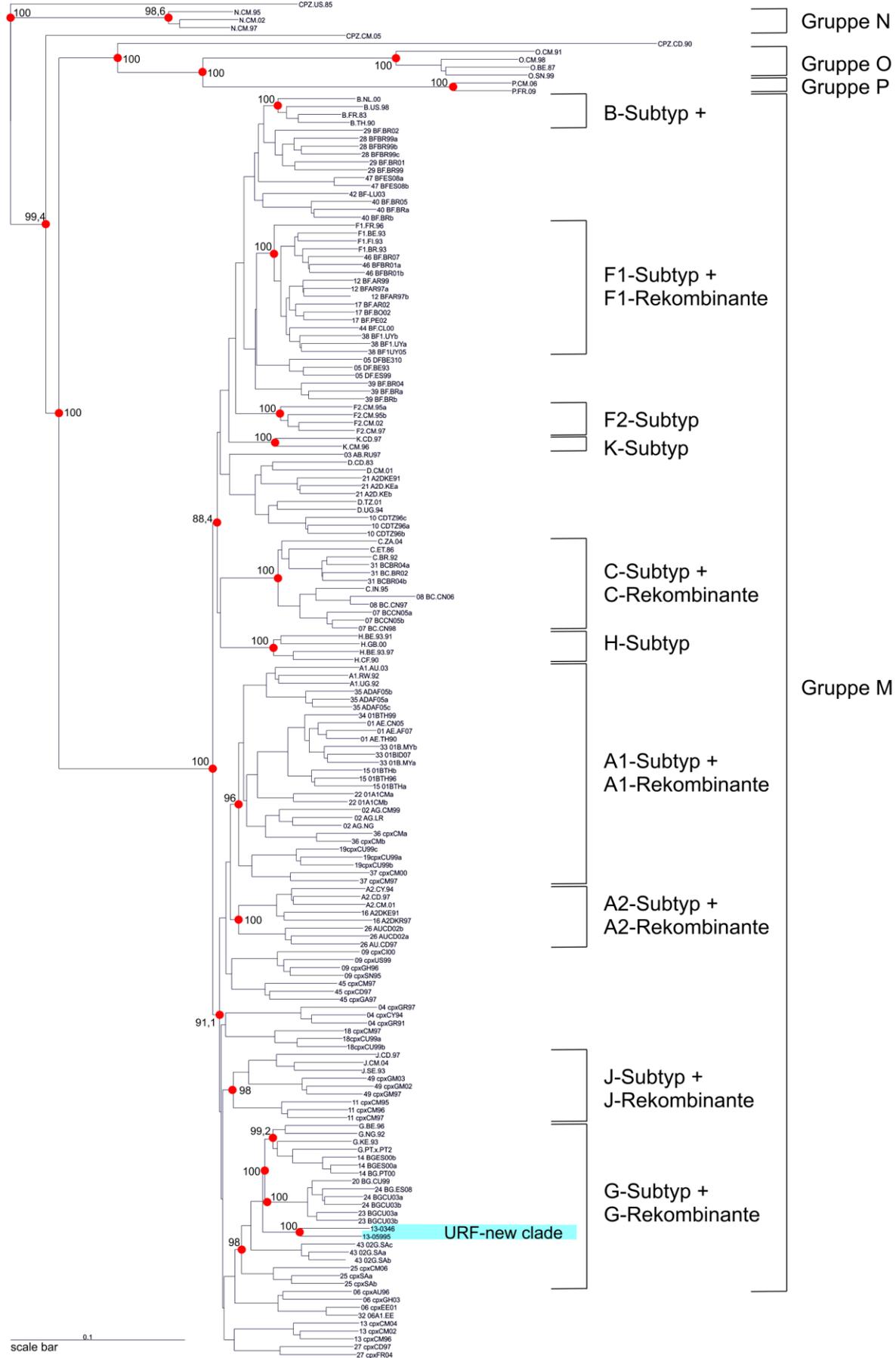


Abbildung 15: Phylogenetischer *Neighbor-Joining* Stammbaum zur Klassifizierung der *URF-new clade* Isolate aus Oman. Phylogenetische Analyse mit Kompletengenomsequenzen von 170 Referenzsequenzen des HIV-Subtypen-Panels der HIV-Sequenzdatenbank und von zwei HIV-1 Patientenproben der *URF-new clade* Variante (13-0346, 13-05995). Die Sequenz CPZ.US.85 (Acc.No.: AF103818) wurde als Außengruppe für die Erstellung der Baumtopologie benutzt. Der Distanzbalken (*scale bar*) mit dem Wert 0,1 zeigt die Astlänge an, die einem Nukleotidaustausch von 10 % pro Position im Alignment entspricht. Die Bootstrap-Analyse wurde mit 1000 Replikaten durchgeführt. Signifikante Knotenpunkte (Bootstrap-Wert: >70 %) für die Einteilung der *Clades* sind rot gekennzeichnet. In Klammern sind die Subtypen und die dazugehörigen rekombinanten Formen zusammengefasst. Die *URF-new clade* Isolate sind blau hervorgehoben.

9 Literaturverzeichnis

1. Friedman-Kien, A., Laubenstein, L., Marmor, M., Hymes, K., Green, J., Ragaz, A., Gottleib, J., Muggia, F., Demopoulos, R., and Weintraub, M. (1981) Kaposi sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California, *MMWR. Morbidity and Mortality Weekly Report* 30, 305-308.
2. Barre-Sinoussi, F. (2010) HIV: a discovery opening the road to novel scientific knowledge and global health improvement, *Virology* 397, 255-259.
3. CASE, K. (1986) Nomenclature: Human immunodeficiency virus, *Ann. Intern. Med.* 105, 133-133.
4. Reeves, J. D., and Doms, R. W. (2002) Human immunodeficiency virus type 2, *J. Gen. Virol.* 83, 1253-1265.
5. Prameela, K. K. (2012) HIV transmission through breastmilk: the science behind the understanding of current trends and future research, *Med. J. Malaysia* 67, 644-651.
6. (2012) Assessing the risk of HIV infection after an isolated exposure incident, *Prescrire Int.* 21, 102-103.
7. Hahn, B. H. (2000) AIDS as a Zoonosis: Scientific and Public Health Implications, *Science* 287, 607-614.
8. Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S., Isolation, W.-U. N. f. H., and Characterisation. (2011) Global trends in molecular epidemiology of HIV-1 during 2000-2007, *AIDS* 25, 679-689.
9. Hemelaar, J. (2012) The origin and diversity of the HIV-1 pandemic, *Trends Mol. Med.* 18, 182-192.
10. Plantier, J.-C., Leoz, M., Dickerson, J. E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D. L., and Simon, F. (2009) A new human immunodeficiency virus derived from gorillas, *Nat. Med.* 15, 871-872.
11. Campbell, S., and Rein, A. (1999) In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain, *J. Virol.* 73, 2270-2279.
12. Haseltine, W. A. (1991) Molecular biology of the human immunodeficiency virus type 1, *FASEB J.* 5, 2349-2360.
13. Gonda, M. A. (1988) Molecular genetics and structure of the human immunodeficiency virus, *J. Electron Microsc. Tech.* 8, 17-40.
14. Gelderblom, H. R., Ozel, M., and Pauli, G. (1989) Morphogenesis and morphology of HIV. Structure-function relations, *Arch. Virol.* 106, 1-13.
15. Freed, E. O. (2001) HIV-1 replication, *Somat. Cell Mol. Genet.* 26, 13-33.
16. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression, *Nature* 331, 280-283.
17. DeStefano, J. J., Buiser, R. G., Mallaber, L. M., Myers, T. W., Bambara, R. A., and Fay, P. J. (1991) Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled, *J. Biol. Chem.* 266, 7423-7431.
18. Varmus, H. E., Swanstrom, R. (1985) *Replication of retroviruses.*, R. Weiss, N. Teich, H. Varmus, J. Coffin (Eds.), New York.
19. Bushman, F. D., and Craigie, R. (1991) Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA, *Proc. Natl. Acad. Sci. U. S. A.* 88, 1339-1343.
20. Stein, B. S., Gowda, S. D., Lifson, J. D., Penhallow, R. C., Bensch, K. G., and Engleman, E. G. (1987) pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane, *Cell* 49, 659-668.
21. Lifson, J. D., Feinberg, M. B., Reyes, G. R., Rabin, L., Banapour, B., Chakrabarti, S., Moss, B., Wong-Staal, F., Steimer, K. S., and Engleman, E. G. (1986) Induction of

- CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein, *Nature* 323, 725-728.
22. Wolinsky, S. M., Korber, B. T., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D., and Safrin, J. T. (1996) Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection, *Science* 272, 537-542.
 23. Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001) Evolutionary and immunological implications of contemporary HIV-1 variation, *Br. Med. Bull.* 58, 19-42.
 24. Domingo, E., Sheldon, J., and Perales, C. (2012) Viral quasispecies evolution, *Microbiol. Mol. Biol. Rev.* 76, 159-216.
 25. Llauro, A. S., and Andino, R. (2010) Quasispecies theory and the behavior of RNA viruses, *PLoS Pathog.* 6, e1001005.
 26. Eberle, J., and Gurtler, L. (2012) HIV types, groups, subtypes and recombinant forms: errors in replication, selection pressure and quasispecies, *Intervirology* 55, 79-83.
 27. Johnson, P. R., and Hirsch, V. M. (1992) Genetic variation of simian immunodeficiency viruses in nonhuman primates, *AIDS Res. Hum. Retroviruses* 8, 367-372.
 28. Santos, A. F., and Soares, M. A. (2010) HIV Genetic Diversity and Drug Resistance, *Viruses* 2, 503-531.
 29. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M., and Ho, D. D. (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time, *Science* 271, 1582-1586.
 30. Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. (1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection, *Nature* 373, 123-126.
 31. Santoro, M. M., and Perno, C. F. (2013) HIV-1 Genetic Variability and Clinical Implications, *ISRN microbiology* 2013, 481314.
 32. Nowak, M. (1990) HIV mutation rate, *Nature* 347, 522.
 33. Temin, H. M. (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation, *Proc. Natl. Acad. Sci. U. S. A.* 90, 6900-6903.
 34. Pathak, V. K., and Hu, W.-S. (1997) "Might as Well Jump!" Template Switching by Retroviral Reverse Transcriptase, Defective Genome Formation, and Recombination, *Semin. Virol.* 8, 141-150.
 35. Delviks-Frankenberry, K. G., Andrea ; Nikolaitchik, Olga ; Mens, Helene ; Pathak, Vinay K. ; Hu, Wei-Shau (2011) Mechanisms and Factors that Influence High Frequency Retroviral Recombination, *Viruses* 3, 1650-1680.
 36. Zhuang, J., Jetzt, A. E., Sun, G., Yu, H., Klarmann, G., Ron, Y., Preston, B. D., and Dougherty, J. P. (2002) Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots, *J. Virol.* 76, 11273-11282.
 37. Ramirez, B. C., Simon-Loriere, E., Galetto, R., and Negroni, M. (2008) Implications of recombination for HIV diversity, *Virus Res.* 134, 64-73.
 38. Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S., and Korber, B. (2000) HIV-1 nomenclature proposal, *Science* 288, 55-56.
 39. Delviks-Frankenberry, K., Galli, A., Nikolaitchik, O., Mens, H., Pathak, V. K., and Hu, W. S. (2011) Mechanisms and factors that influence high frequency retroviral recombination, *Viruses* 3, 1650-1680.
 40. Taylor, B. S., Sobieszczyk, M. E., McCutchan, F. E., and Hammer, S. M. (2008) The challenge of HIV-1 subtype diversity, *N. Engl. J. Med.* 358, 1590-1602.
 41. Carr, J. K., Salminen, M. O., Koch, C., Gotte, D., Artenstein, A. W., Hegerich, P. A., St Louis, D., Burke, D. S., and McCutchan, F. E. (1996) Full-length sequence and

- mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand, *J. Virol.* **70**, 5935-5943.
42. Paraskevis, D., Magiorkinis, M., Vandamme, A. M., Kostrikis, L. G., and Hatzakis, A. (2001) Re-analysis of human immunodeficiency virus type 1 isolates from Cyprus and Greece, initially designated 'subtype I', reveals a unique complex A/G/H/K/? mosaic pattern, *J. Gen. Virol.* **82**, 575-580.
 43. Powell, R. L., Urbanski, M. M., Burda, S., Kinge, T., and Nyambi, P. N. (2009) High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa, *J. Acquir. Immune Defic. Syndr.* **50**, 84-92.
 44. <http://www.hiv.lanl.gov/>.
 45. Carr, J. K., Salminen, M. O., Albert, J., Sanders-Buell, E., Gotte, D., Birx, D. L., and McCutchan, F. E. (1998) Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants, *Virology* **247**, 22-31.
 46. Tovanabuttra, S., Watanaveeradej, V., Viputtikul, K., De Souza, M., Razak, M. H., Suriyanon, V., Jittiwutikarn, J., Sriplienchan, S., Nitayaphan, S., Benenson, M. W., Sirisopana, N., Renzullo, P. O., Brown, A. E., Robb, M. L., Beyrer, C., Celentano, D. D., McNeil, J. G., Birx, D. L., Carr, J. K., and McCutchan, F. E. (2003) A new circulating recombinant form, CRF15_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand, *AIDS Res. Hum. Retroviruses* **19**, 561-567.
 47. Montavon, C., Toure-Kane, C., Nkengasong, J. N., Vergne, L., Hertogs, K., Mboup, S., Delaporte, E., and Peeters, M. (2002) CRF06-cpx: a new circulating recombinant form of HIV-1 in West Africa involving subtypes A, G, K, and J, *J. Acquir. Immune Defic. Syndr.* **29**, 522-530.
 48. Powell, R. L., Zhao, J., Konings, F. A., Tang, S., Nanfack, A., Burda, S., Urbanski, M. M., Saa, D. R., Hewlett, I., and Nyambi, P. N. (2007) Identification of a novel circulating recombinant form (CRF) 36_cpx in Cameroon that combines two CRFs (01_AE and 02_AG) with ancestral lineages of subtypes A and G, *AIDS Res. Hum. Retroviruses* **23**, 1008-1019.
 49. Peeters, M., Gueye, A., Mboup, S., Bibollet-Ruche, F., Ekaza, E., Mulanga, C., Ouedrago, R., Gandji, R., Mpele, P., and Dibanga, G. (1997) Geographical distribution of HIV-1 group O viruses in Africa, *AIDS* **11**, 493-498.
 50. Vallari, A., Bodelle, P., Ngansop, C., Makamche, F., Ndembi, N., Mbanya, D., Kaptué, L., Gürtler, L. G., McArthur, C. P., and Devare, S. G. (2010) Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low, *AIDS Res. Hum. Retroviruses* **26**, 109-115.
 51. Sharp, P. M., and Hahn, B. H. (2011) Origins of HIV and the AIDS pandemic, *Cold Spring Harb. Perspect. Med.* **1**, a006841.
 52. Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembi, N., and Gürtler, L. (2011) Confirmation of putative HIV-1 group P in Cameroon, *J. Virol.* **85**, 1403-1407.
 53. Peeters, M., Jung, M., and Ayouba, A. (2013) The origin and molecular epidemiology of HIV, *Expert Rev. Anti Infect. Ther.* **11**, 885-896.
 54. Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B., and Delaporte, E. (2000) Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa, *J. Virol.* **74**, 10498-10507.
 55. Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic, *Nature* **391**, 594-597.
 56. (2013) *Global report: UNAIDS report on the global AIDS epidemic 2013.*, WHO Library Cataloguing-in-Publication Data.
 57. Bozicevic, I., Riedner, G., and Calleja, J. M. G. (2013) HIV surveillance in MENA: recent developments and results, *Sex. Transm. Infect.* **89**, iii11-iii16.
 58. (2012) Global AIDS Response Progress Report, Country Progress Report Sultanate of Oman

59. Kenfack Guepi, E. (2013) Vollständige Genomanalyse neu identifizierter singulärer HIV-1 Rekombinanten (URF) aus Oman.
60. Kuiken C, F. B., Leitner T, Apetrei C, Hahn B, Mizrachi I, Mullins J, Rambaut A, Wolinsky S, and Korber B, Eds. . HIV Sequence Compendium 2010, *Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 10-03684*.
61. Zhang, M., Schultz, A. K., Calef, C., Kuiken, C., Leitner, T., Korber, B., Morgenstern, B., and Stanke, M. (2006) jpHMM at GOBICS: a web server to detect genomic recombinations in HIV-1, *Nucleic Acids Res.* 34, W463-465.
62. Schultz, A. K., Zhang, M., Leitner, T., Kuiken, C., Korber, B., Morgenstern, B., and Stanke, M. (2006) A jumping profile Hidden Markov Model and applications to recombination sites in HIV and HCV genomes, *BMC Bioinformatics* 7, 265.
63. Siepel AC, H. A., Macken C, Korber BT. (1995) A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences., in *AIDS Res. Hum. Retroviruses*, pp 1413-1416.
64. Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., Berry, N., Pillay, D., and Kellam, P. (2012) Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes, *Journal of clinical microbiology* 50, 3838-3844.
65. Volker Knoop, K. M. (2006) *Gene und Stammbäume: ein Handbuch zur molekularen Phylogenetik.* , Spektrum Akademischer Verlag
66. Schultz, A. K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., and Stanke, M. (2009) jpHMM: improving the reliability of recombination prediction in HIV-1, *Nucleic Acids Res.* 37, W647-651.
67. Nadai, Y., Eyzaguirre, L. M., Constantine, N. T., Sill, A. M., Cleghorn, F., Blattner, W. A., and Carr, J. K. (2008) Protocol for nearly full-length sequencing of HIV-1 RNA from plasma, *PLoS One* 3, e1420.
68. Klarmann, G. J., Schaubert, C. A., and Preston, B. D. (1993) Template-directed pausing of DNA synthesis by HIV-1 reverse transcriptase during polymerization of HIV-1 sequences in vitro, *J. Biol. Chem.* 268, 9793-9802.
69. Harrison, G. P., Mayo, M. S., Hunter, E., and Lever, A. M. (1998) Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site, *Nucleic Acids Res.* 26, 3433-3442.
70. Simon-Lorriere, E., Martin, D. P., Weeks, K. M., and Negroni, M. (2010) RNA structures facilitate recombination-mediated gene swapping in HIV-1, *J. Virol.* 84, 12675-12682.
71. Villahermosa, M. L., Thomson, M., Vazquez de Parga, E., Cuevas, M. T., Contreras, G., Perez-Alvarez, L., Delgado, E., Manjon, N., Medrano, L., and Najera, R. (2000) Improved conditions for extraction and amplification of human immunodeficiency virus type 1 RNA from plasma samples with low viral load, *J. Hum. Virol.* 3, 27-34.
72. Das, M., Harvey, I., Chu, L. L., Sinha, M., and Pelletier, J. (2001) Full-length cDNAs: more than just reaching the ends, *Physiological genomics* 6, 57-80.
73. Malboeuf, C. M., Isaacs, S. J., Tran, N. H., and Kim, B. (2001) Thermal effects on reverse transcription: improvement of accuracy and processivity in cDNA synthesis, *Biotechniques* 30, 1074-1078, 1080, 1082, passim.
74. Yu, W., Rusterholtz, K. J., Krummel, A. T., and Lehman, N. (2006) Detection of high levels of recombination generated during PCR amplification of RNA templates, *Biotechniques* 40, 499-507.
75. Fang, G., Zhu, G., Burger, H., Keithly, J. S., and Weiser, B. (1998) Minimizing DNA recombination during long RT-PCR, *J. Virol. Methods* 76, 139-148.
76. Luk, K. C., Holzmayer, V., Ndembu, N., Swanson, P., Brennan, C. A., Ngansop, C., Mbanya, D., Kaptue, L., Gurtler, L., Devare, S. G., and Hackett, J. (2008) Near full-length genome characterization of an HIV type 1 CRF25_cpx strain from Cameroon, *AIDS Res. Hum. Retroviruses* 24, 1309-1314.
77. Yamaguchi, J., Badreddine, S., Swanson, P., Bodelle, P., Devare, S. G., and Brennan, C. A. (2008) Identification of new CRF43_02G and CRF25_cpx in Saudi

- Arabia based on full genome sequence analysis of six HIV type 1 isolates, *AIDS Res. Hum. Retroviruses* 24, 1327-1335.
78. Swanson, P., Devare, S. G., and Hackett, J., Jr. (2003) Molecular characterization of 39 HIV isolates representing group M (subtypes A-G) and group O: sequence analysis of gag p24, pol integrase, and env gp41, *AIDS Res. Hum. Retroviruses* 19, 625-629.
79. Pessoa, R., Carneiro Proietti, A. B., Busch, M. P., and Sanabani, S. S. (2014) Identification of a Novel HIV-1 Circulating Recombinant Form (CRF72_BF1) in Deep Sequencing Data from Blood Donors in Southeastern Brazil, *Genome announcements* 2.
80. Sanabani, S., Neto, W. K., de Sa Filho, D. J., Diaz, R. S., Munerato, P., Janini, L. M., and Sabino, E. C. (2006) Full-length genome analysis of human immunodeficiency virus type 1 subtype C in Brazil, *AIDS Res. Hum. Retroviruses* 22, 171-176.
81. Li, Z., He, X., Wang, Z., Xing, H., Li, F., Yang, Y., Wang, Q., Takebe, Y., and Shao, Y. (2012) Tracing the origin and history of HIV-1 subtype B' epidemic by near full-length genome analyses, *AIDS* 26, 877-884.
82. Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., Moore, S. W., Shindo, N., Learn, G. H., Korber, B. T., Brander, C., Goulder, P. J., Kiepiela, P., Walker, B. D., and Mullins, J. I. (2006) Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes, *J. Virol. Methods* 136, 118-125.
83. Wei, H., Su, L., Feng, Y., He, X., His, J., Liang, S., and Shao, Y. (2013) 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 Res. Hum. Retroviruses* 29, 1173-1176.
84. Lau, K. A., and Wong, J. J. (2013) Current trends of HIV recombination worldwide, *Infect. Dis. Rep.* 5.
85. Ramirez, B. C., Simon-Loriere, E., Galetto, R., and Negroni, M. (2008) Implications of recombination for HIV diversity, *Virus Res.* 134, 64-73.
86. Etherington, G. J., Dicks, J., and Roberts, I. N. (2005) Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination, *Bioinformatics* 21, 278-281.
87. de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E. J., Wensing, A. M., van de Vijver, D. A., Boucher, C. A., Camacho, R., and Vandamme, A. M. (2005) An automated genotyping system for analysis of HIV-1 and other microbial sequences, *Bioinformatics* 21, 3797-3800.
88. <http://jphmm.gobics.de/>.
89. Laukkanen, T., Carr, J. K., Janssens, W., Liitsola, K., Gotte, D., McCutchan, F. E., Op de Coul, E., Cornelissen, M., Heyndrickx, L., van der Groen, G., and Salminen, M. O. (2000) Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America, *Virology* 269, 95-104.
90. Koulinska, I. N., Ndung'u, T., Mwakagile, D., Msamanga, G., Kagoma, C., Fawzi, W., Essex, M., and Renjifo, B. (2001) A new human immunodeficiency virus type 1 circulating recombinant form from Tanzania, *AIDS Res. Hum. Retroviruses* 17, 423-431.
91. Delgado, E., Thomson, M. M., Villahermosa, M. L., Sierra, M., Ocampo, A., Miralles, C., Rodriguez-Perez, R., Diz-Aren, J., Ojea-de Castro, R., Losada, E., Cuevas, M. T., Vazquez-de Parga, E., Carmona, R., Perez-Alvarez, L., Medrano, L., Cuevas, L., Taboada, J. A., and Najera, R. (2002) Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure, *J. Acquir. Immune Defic. Syndr.* 29, 536-543.
92. Gomez-Carrillo, M., Quarleri, J. F., Rubio, A. E., Carobene, M. G., Dileria, D., Carr, J. K., and Salomon, H. (2004) Drug resistance testing provides evidence of the globalization of HIV type 1: a new circulating recombinant form, *AIDS Res. Hum. Retroviruses* 20, 885-888.

93. Casado, G., Thomson, M. M., Sierra, M., and Najera, R. (2005) Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba, *J. Acquir. Immune Defic. Syndr.* 40, 532-537.
94. De Sa Filho, D. J., Sucupira, M. C., Caseiro, M. M., Sabino, E. C., Diaz, R. S., and Janini, L. M. (2006) Identification of two HIV type 1 circulating recombinant forms in Brazil, *AIDS Res. Hum. Retroviruses* 22, 1-13.
95. Foster, G. M., Ambrose, J. C., Hue, S., Delpech, V. C., Fearnhill, E., Abecasis, A. B., Leigh Brown, A. J., and Geretti, A. M. (2014) Novel HIV-1 recombinants spreading across multiple risk groups in the United Kingdom: the identification and phylogeography of Circulating Recombinant Form (CRF) 50_A1D, *PLoS One* 9, e83337.
96. Abecasis, A. B., Lemey, P., Vidal, N., de Oliveira, T., Peeters, M., Camacho, R., Shapiro, B., Rambaut, A., and Vandamme, A. M. (2007) Recombination confounds the early evolutionary history of human immunodeficiency virus type 1: subtype G is a circulating recombinant form, *J. Virol.* 81, 8543-8551.
97. Hierholzer, M., Graham, R. R., El Khidir, I., Tasker, S., Darwish, M., Chapman, G. D., Fagbami, A. H., Soliman, A., Birx, D. L., McCutchan, F., and Carr, J. K. (2002) HIV type 1 strains from East and West Africa are intermixed in Sudan, *AIDS Res. Hum. Retroviruses* 18, 1163-1166.

Danksagung

An dieser Stelle möchte ich allen großen Dank aussprechen, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Arbeit beigetragen haben.

Mein erster Dank gilt Dr. rer. nat. Sybille Somogyi für die Überlassung dieser interessanten Thematik sowie für die kontinuierliche und außerordentliche Betreuung bei der Anfertigung dieser Arbeit.

Herrn PD Dr. rer. nat. Bannert danke ich für die Möglichkeit, diese Masterarbeit im Fachgebiet „HIV und andere Retroviren“ anfertigen zu können sowie für die Übernahme des Gutachtens der Arbeit.

Ich möchte mich auch bei Herrn Prof. Dr. rer. nat. Behrens für die Übernahme des Amtes als mein institutsinterner Betreuer und mit der von ihm geschaffenen Möglichkeit, diese Masterarbeit extern schreiben zu dürfen, herzlichst bedanken.

Ein weiterer Dank gilt Frau Dr. rer. nat. Claudia Kücherer für die konstruktiven Ratschläge während der schriftlichen Anfertigung dieser Arbeit.

Dem gesamten HIV-Studienlabor möchte ich für die hilfreiche Unterstützung und Einarbeitung im Labor danken. Ganz besonders danke ich Anika Bresk, Katrin Arndt und Hanno von Spreckelsen für die freundschaftliche Arbeitsatmosphäre und für die vielen erheiternden fachlichen und nicht-fachlichen Diskussionen zwischendurch.

Zudem möchte ich auch meinen Freunden für die schöne Zeit während meines Studiums danken. Insbesondere danke ich Christian, der stets und ständig für mich da war und bei Problemen immer eine passende Lösung parat hatte.

Ein besonders herzlicher Dank gilt meinen Eltern, die mich während meines gesamten Studiums in meinen Vorhaben unterstützt und ermutigt haben. Vielen Dank für euer Vertrauen in mich, das mir immer den nötigen Rückhalt gegeben hat.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die eingereichte schriftliche Fassung der Arbeit entspricht der auf dem elektronischen Speichermedium. Weiterhin versichere ich, dass die vorliegende Arbeit noch nicht als Abschlussarbeit an anderer Stelle eingereicht wurde.

Halle (Saale), den 25. September 2014

Luise Luckau