



BEUTH HOCHSCHULE
FÜR TECHNIK
BERLIN
University of Applied Sciences

ROBERT KOCH INSTITUT



Nähere Charakterisierung von phylogenetisch nicht klassifizierbaren rekombinanten Formen von HIV-1

Bachelorarbeit

im Bachelor-Studiengang Biotechnologie/Biotechnology der Beuth Hochschule für Technik Berlin – University of Applied Sciences –

zur Erlangung des akademischen Grades eines
Bachelor of Science (B. Sc.)

vorgelegt von

Dominique Seifert

September 2017

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	1
1 Einleitung.....	2
1.1 Epidemiologie und Klassifizierung von HIV	2
1.2 Genom und Replikationszyklus von HIV-1.....	4
1.4 Phylogenie	7
2 Zielstellung	9
3 Material und Methoden.....	10
3.1 RNA-Extraktion mit dem NucliSENS®-easyMag®-Extraktionsautomaten	10
3.1.1 Geräte.....	10
3.1.2 Reagenzien & Materialien	10
3.1.3 Durchführung	11
3.2 PCR-Bedingungen für die Integrase-PCR.....	11
3.2.1 Geräte.....	11
3.2.2 Reagenzien & Materialien	11
3.2.3 Durchführung	12
3.3 PCR-Bedingungen für die Envelope-PCR	13
3.3.1 Geräte.....	13
3.3.2 Reagenzien & Materialien	13
3.3.3 Durchführung	13
3.4 PCR-Bedingungen für die nested Envelope-PCR	14
3.4.1 Geräte.....	14
3.4.2 Reagenzien & Materialien	14
3.4.3 Durchführung	15
3.5 Agarosegelelektrophorese	16
3.5.1 Geräte.....	16
3.5.2 Reagenzien & Materialien	17
3.5.3 Vorbereitung der PCR Produkte.....	17
3.5.4 Durchführung	18
3.6 Aufreinigung der PCR-Produkte mit Agencourt® AMPure® XP Kit.....	19
3.6.1 Geräte.....	19
3.6.2 Reagenzien und Materialien	19
3.6.3 Durchführung	20
3.7 Quantifizierung von PCR-Produkten mit dem Quant-iT™ PicoGreen® dsDNA Assay Kit	20
3.7.1 Geräte.....	21

3.7.2 Reagenzien und Materialien	21
3.7.3 Durchführung	21
3.8 Sequenzierung.....	22
3.8.1 Geräte.....	22
3.8.2 Reagenzien und Materialien	22
3.8.3 int-Sequenzansätze	22
3.8.4 env-Sequenzansätze.....	23
3.8.5 Sanger Sequenzierung.....	24
3.9 Sequenzauswertung	25
3.10 Phylogenie	26
4 Ergebnisse	28
4.1 PCR	28
4.2 Sequenzierung.....	30
4.3 Sequenzauswertung.....	31
4.4 Subtypbestimmung mit Webtools	32
4.5 Maximum-Likelihood-Phylogenie	34
4.6 Vergleich REGA und Maximum-Likelihood-Phylogenie	38
5 Diskussion.....	42
6 Ausblick	47
7 Zusammenfassung.....	48
8 Summary	49
9 Literaturverzeichnis.....	50
Appendix	53
Eidesstattliche Erklärung.....	57
Danksagung	58

Abkürzungsverzeichnis

AIDS	Englisch: acquired immune deficiency syndrome (erworbenes Immundefektsyndrom)
bidest	bidestilliert
bp	Basenpaar
CRF	Englisch: circulating recombinant form of HIV-1 (zirkulierende rekombinante Form von HIV-1)
ddNTPs	Didesoxyribonukleosidtriphosphat
DNA	Desoxyribonukleinsäure
dNTPs	Desoxyribonukleosidtriphosphat
ds	doppelsträngig
EDTA	Ethylendiamintetraacetat
EDTA-Plasma	Plasma aus Blutproben mit EDTA als Gerinnungshemmer
<i>env</i>	Envelope-Genomregion
EtBr	Ethidiumbromid
EtOH	Ethanol
FCV	felines Calicivirus (interne Positivkontrolle)
g	Erdbeschleunigung
geq	Genomäquivalente
HIV	humanes Immundefizienz-Virus
HIV-DB	HIV-Labordatenbank
HTLV-1	humanes T-lymphotropes Virus Typ 1
HTLVIIIB	HIV-1 Laborstamm (interne Positivkontrolle)
HXB2	HIV-1 Subtyp B Referenzsequenz
<i>int</i>	Integrase-Genomregion
kBp	Kilobasenpaare
LTR	Englisch: long terminal repeats (Wiederholungseinheiten)
ML	Englisch: Maximum-Likelihood (maximale Wahrscheinlichkeit)
NGS	Englisch: Next Generation Sequencing (Sequenzierung der nächsten Generation)
PCR	Englisch: polymerase chain reaction (Polymerase-Kettenreaktion)
<i>pol</i>	Genomregion welche die virale Protease, Reverse Transkriptase und Integrase kodiert
<i>prrt</i>	Protease- und Reverse Transkriptase-Genomregion
RNA	Ribonukleinsäure
rpm	Englisch: rounds per minute (Umdrehung pro Minute)
RT-PCR	Reverse Transkriptase und Polymerase-Kettenreaktion
SIV	simianes Immundefizienz-Virus
TAE-Puffer	Tris-Acetat-EDTA-Puffer
TE-Puffer	Tris-HCl-EDTA-Puffer
URF	Englisch: unique recombinant form of HIV-1 (einmalige rekombinante Form von HIV-1)

1 Einleitung

1.1 Epidemiologie und Klassifizierung von HIV

In den letzten 100 Jahren führten soziale, ökonomische und politische Veränderungen zu einer noch nie da gewesenen Verbreitung von Infektionskrankheiten. Die weltweite Reisetätigkeit von Personen, die Einengung der Lebensräume wilder Tiere und der dadurch zunehmend engere Kontakt zwischen Mensch und Tier, die Zunahme des internationalen Warenverkehrs, sowie eine veränderte Tierhaltung und Klimaveränderungen sind die beeinflussenden Faktoren die das Risiko an einer Zoonose zu erkranken erhöhen. Zoonosen sind Infektionskrankheiten die von Erregern verursacht werden, die wechselseitig zwischen Tieren und Menschen übertragbar sind. Weltweit sind etwa zwei Drittel aller Infektionskrankheiten Zoonosen, weshalb insbesondere die Forschung zur Bekämpfung von Zoonosen von enormer gesellschaftlicher, politischer und wirtschaftlicher Bedeutung ist. [1-2]

Eine weit verbreitete virale Epidemie, ursprünglich verursacht durch eine zoonotische Transmission, ist AIDS (erworbenes Immundefektsyndrom, englisch: **a**cquired **i**mmune **d**eficiency **s**yndrome), bedingt durch das HIV (humanes Immundefizienz-Virus, englisch: **h**uman **i**mmunodeficiency **v**irus). [3-4] Das HIV gehört zur Familie der *Retroviridae*, Untergruppe *Orthoretrovirinae*, Gattung *Lentivirus* [5], wobei eine Klassifizierung der zwei Virus-Typen 1 und 2 durchgeführt wird (HIV-1 und HIV-2). Beide Viren resultieren aus artübergreifender Transmission des simianen (Affen) Immundefizienz-Virus (SIV, englisch: **s**imian **i**mmunodeficiency **v**irus), welches innerhalb afrikanischer Primaten zirkuliert. Wann genau die Transmission von nicht-humanen Primaten auf den Menschen stattgefunden hat, lässt sich nur durch mathematische Berechnungen abschätzen. Um weitere Zoonosen zukünftig zu verhindern, ist die Klärung des Übertragungsweges und des genauen Zeitpunkts der Übertragung von großer Bedeutung. [4]

Erstmals wurde AIDS als neues Krankheitsbild im Jahr 1981 identifiziert, als eine stetig wachsende Anzahl von jungen, homosexuellen Männern unüblichen opportunistischen Infektionen und seltenen bösartigen Tumoren erlag. [6] Ist das Immunsystem des Körpers geschwächt, können ansonsten harmlose Mikroorganismen, schwere, sogar tödlich verlaufende Erkrankungen hervorrufen. [5] Diese werden auch als opportunistische Infektionen bezeichnet und nutzen die herabgesetzte Immunabwehr aus, um sich zu vermehren. [7] Das HIV Typ 1, damals noch HTLV-1 (**h**umanes **T**-lymphotropes **V**irus **1**) genannt, wurde in den Jahren 1983 und 1984 von verschiedenen Arbeitsgruppen nach und nach als der Verursacher für die tödliche Krankheit identifiziert. [8-9] Mitte der 1980er Jahre wurde in West Afrika ein neuer Erregerstamm entdeckt, der sich genetisch von dem bis dato bekannten HIV-1 unterschied und als neuer Typ HIV-2 klassifiziert wurde. Der Ursprung der vornehmlich vorherrschenden HIV-Gruppe M (major) und der daraus resultierenden HIV Pandemie ist allerdings auf ein Transmissionsereignis mit einem SIVcpz (SIV eines Schimpansen) in Südost-Kamerun zurückzuführen. [4, 10-11] Ein weiteres, unabhängiges Übertragungsereignis vom Schimpansen auf den Menschen brachte die HIV-1 Gruppe N (non-major) hervor. Des Weiteren entstanden durch unabhängige Transmissionen vom Gorilla auf den Menschen die HIV-1 Gruppen O (outlier) und P (pending). HIV-2 hingegen ist auf eine Übertragung vom Rußmangabes auf den Menschen zurück zu führen. [4] Die Subtypen, auch Clades genannt, A, B, C, D, F, G, H, J und K der epidemisch relevanten HIV-1 Gruppe M unterliegen seit der ersten Transmission auf den Menschen einer Diversifizierung und Rekombination, sodass immer wieder neue zirkulierende rekombinante Formen, auch CRFs genannt (englisch: **c**irculating **r**ecombinant **f**orm), entstehen. Der Subtyp E, zirkulierend in Südost-Asien, ist als solcher nicht mehr existent, da er nur noch als rekombinante Form AE (Abbildung 1) nachzuweisen ist und 1998 als CRF01_AE neu definiert wurde. [12-14] Mittlerweile sind 88 CRFs bekannt. [15]

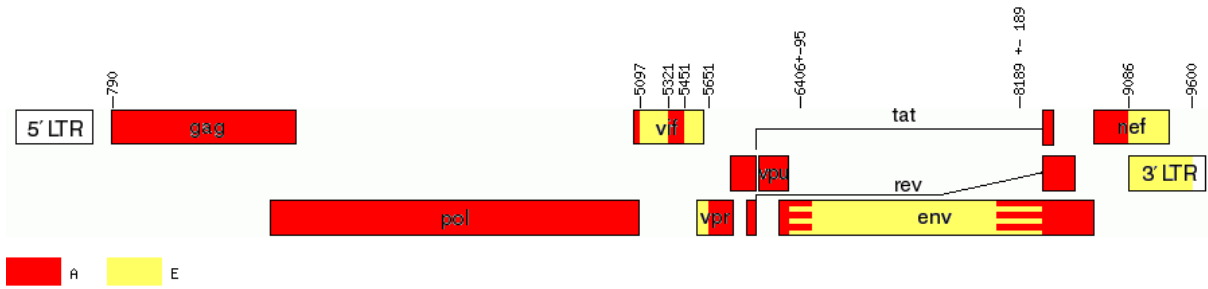


Abbildung 1: Darstellung der verschiedenen Rekombinationspunkte im HIV-1-Genom von CRF01_AE [16]

Unter den CRFs komplexe Formen (Abbildung 2) aus mehreren rekombinanten Formen oder einer rekombinanten Form und einem Subtyp, die als CRF_cpx (cpx für **complex**) bezeichnet werden.

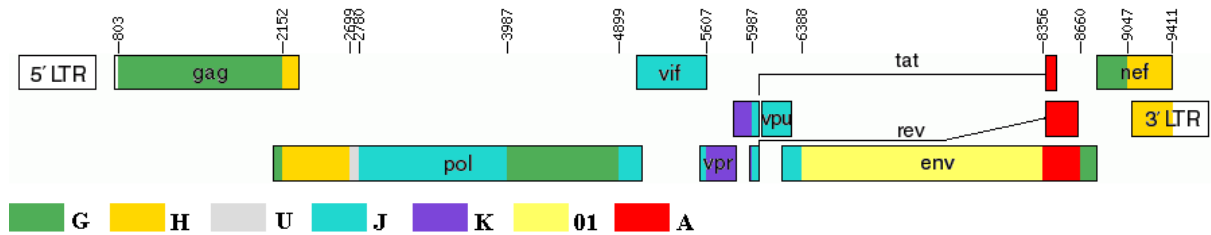


Abbildung 2: Darstellung der verschiedenen Rekombinationspunkte im HIV-1-Genom von CRF27_cpx [16]

Ein hoher Anteil an Rekombinanten kommt einmalig vor (Abbildung 3). Diese werden als singular rekombinante Formen (URF, englisch: **unique recombinant form**) bezeichnet. Treten mehrere, unabhängige Infektionsereignisse mit einer bestimmten URF auf, kann anhand zweier Vollgenom-Sequenzen und einer partiellen Sequenz eine neue CRF definiert werden.

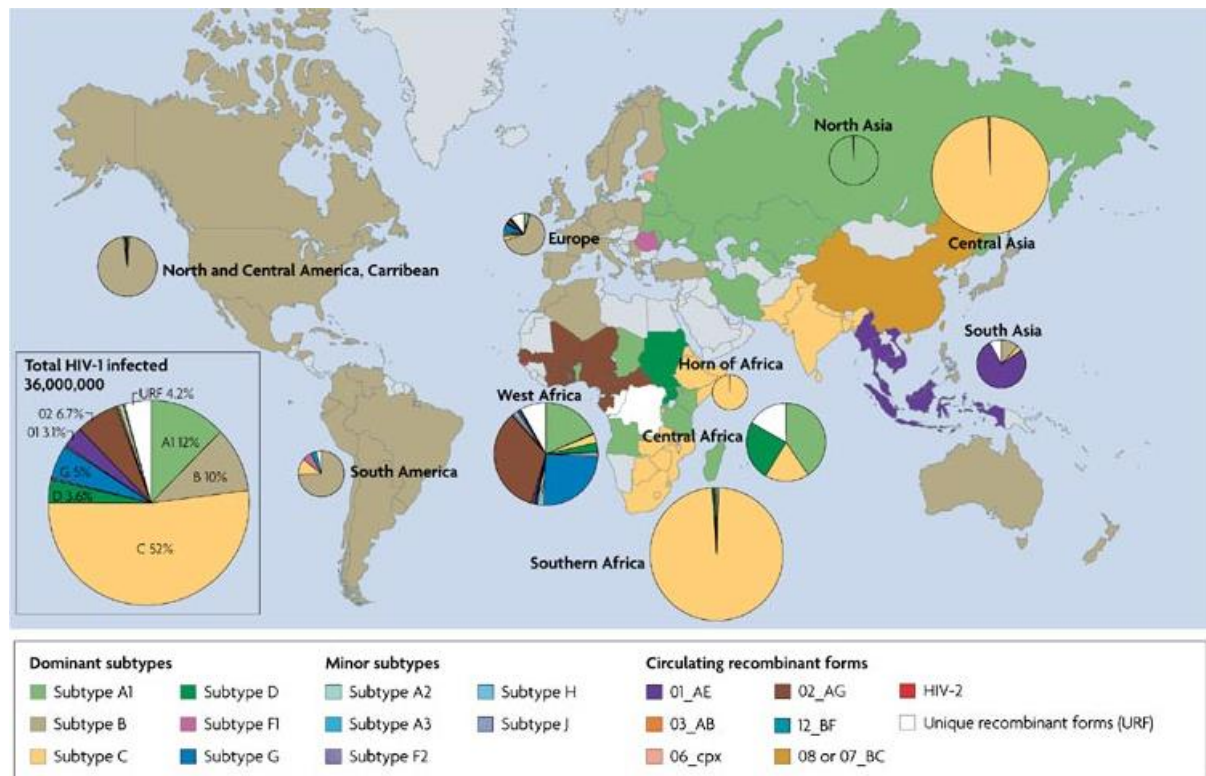


Abbildung 3: Weltweite Verbreitung der HIV-1 Subtypen, CRFs und URFs [19]

Weltweit leben etwa 36,7 Millionen Menschen mit HIV (Stand Ende 2015). [17] Global gesehen kommen Subtyp-C-Infektionen am häufigsten vor, wohingegen in Europa Subtyp B dominant ist. In Deutschland wurde die Zahl der HIV-Neuinfektionen im Jahr 2015 auf 3200 geschätzt. [18] Der Großteil der aktuell in Deutschland infizierten Patienten ist mit HIV-1, Gruppe M, Subtyp B infiziert. Zu den in Deutschland verbreitetsten CRFs gehören der CRF_01AE und der CRF_02AG, die sich aus der Rekombination eines Subtypen A mit E beziehungsweise A mit G gebildet haben. [20] Schätzungen gehen davon aus, dass 2015 zwischen 78.300 bis 91.100 Menschen mit einer HIV-Infektion in Deutschland lebten. [18,21-22]

1.2 Genom und Replikationszyklus von HIV-1

Das RNA-Genom von HIV besteht aus neun Genomregionen mit einer Gesamtgenomgröße von ungefähr 9500 kbp, wobei die Strukturgene *gag*, *pol* und *env* die drei Größten darstellen (Abbildung 4). Das gruppenspezifische Antigen *gag* kodiert die Kern- (*p24*, *p25*), Nukleokapsid- (*p7*, *p6*) und Matrixproteine (*p17*). Die Polymerase (*pol*) kodiert die Enzyme Protease (*pr*), Reverse Transkriptase (*rt*) und Integrase (*int*). Die transmembranen Glykoproteine der Virushülle *gp120* und *gp41* werden von der *env*-Region kodiert. Zusätzlich gibt es noch regulatorische Gene (*tat*, *rev*, *nef*, *vif*, *vpr*, *vpu*) auf die hier jedoch nicht näher eingegangen wird. [23]

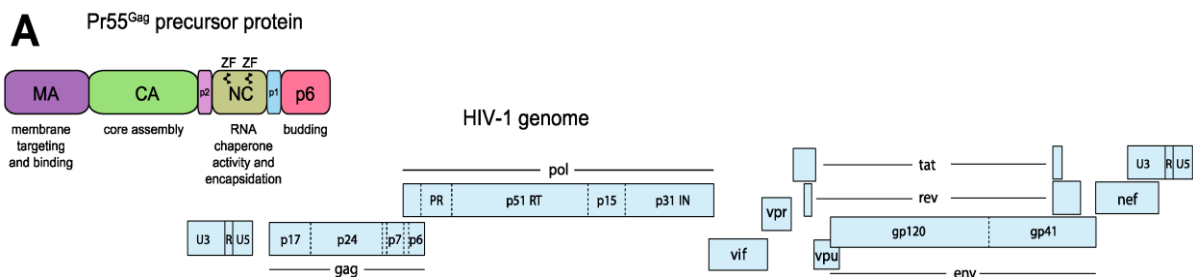


Abbildung 4: Genom vom humanen Immundefizienzvirus und Aufschlüsselung des *gag*-Gens [24]

Aufgrund der überlappenden Leserahmen werden mehrere Proteine von denselben Sequenzabschnitten transkribiert, wodurch die Gesamtgröße des Genoms reduziert wird. An den terminalen Enden des Genoms befinden sich LTRs, sogenannte *long terminal repeats* (DNA-Wiederholungseinheiten), die den Start und das Ende der Transkription steuern. [23-25]

Im strukturellen Aufbau unterschieden sich die beiden Typen HIV-1 und HIV-2 nur wenig voneinander. Das Virus besteht aus einem Kern, zusammengesetzt aus den Kapsid-Proteinen *p24* oder *p25*, welche zwei identische RNA-Einzelstränge enthalten, die mit dem Enzym Reverse Transkriptase und den Nukleokapsid-RNA-Bindungsproteinen *p7* und *p6* verbunden sind. Zusätzlich befinden sich die Integrase, Protease und einige andere akzessorische Proteine ebenfalls im Kapsid. Die äußere Membran des Virus ist eine komplexe Lipiddoppelschicht, die verschiedene Membranproteine und 10 bis 15 „Spikes“ enthält. Bei diesen sogenannten „Spikes“ handelt es sich um *env*-Glykoproteinkomplexe die aus zwei Untereinheiten bestehen: je drei Moleküle des externen Oberflächen-Glykoproteins *gp120* sind nicht-kovalent an drei Moleküle des transmembranen Hüll-Glykoproteins *gp41* gebunden. Das Oberflächen-Glykoprotein *gp120* ist maßgeblich verantwortlich für die Bindung des Virus an die CD4-Rezeptoren der Zielzellen. An der Innenseite der äußeren Membran sind die Matrixproteine *p17* assoziiert (Abbildung 5). [25]

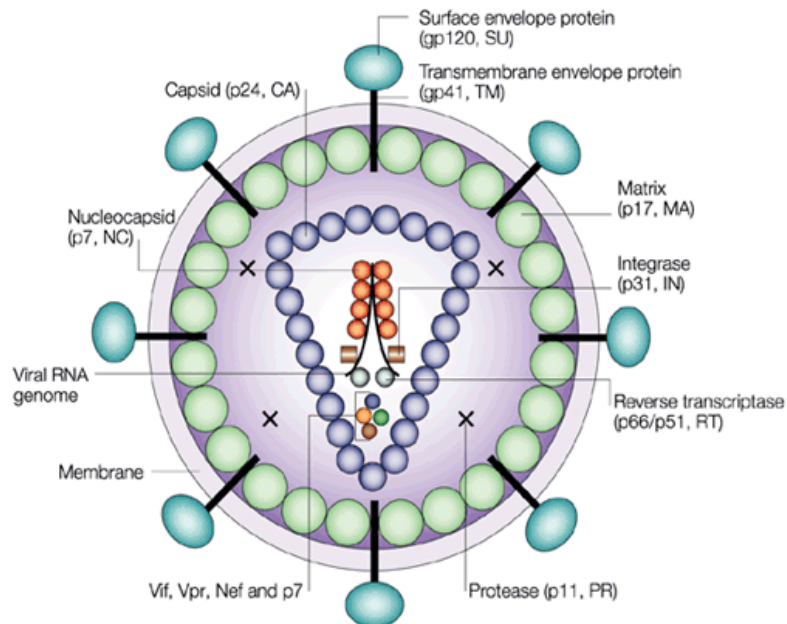


Abbildung 5: Aufbau des humanen Immundefizienzvirus [26]

Der Replikationszyklus von HIV beginnt mit dem Eintritt des HIV-Partikels in den Wirt. Übertragungswege sind neben dem ungeschützten Geschlechtsverkehr, die Gabe von Blut oder Blutbestandteilen, Nadelstichverletzungen mit kontaminierten Nadeln und Spritzen, die Transplantationen von infizierten Organen sowie im Zuge einer Schwangerschaft über die Plazenta während der Entbindung oder über das Stillen. Nachdem die HIV-Partikel im Blutstrom sind benötigen sie eine Wirtszelle zur Replikation, die CD4-Rezeptoren auf der Oberfläche trägt. Dies sind im Besonderen CD4-tragende T-Helferzellen. Mit Hilfe des Glykoproteins *gp120* koppelt das Virus an die CD4-Moleküle der menschlichen Wirtszellen. Nachdem das Virus an der Zielzelle angedockt hat, verliert es seine Hülle und die virale genetische Information wird mit der Reversen Transkriptase in provirale cDNA umgeschrieben und mit Hilfe der Integrase in das Wirtsgenom integriert. Der aktive Zellstoffwechsel führt dann zur Bildung von Virusbestandteilen, die sich innerhalb des Zytoplasmas zu unreifen Virionen formen. Nach dem Abschnüren von der Plasmamembran führt die Protease zu einer abschließenden Reifung der Partikel, welche weitere Wirtszellen infizieren können (Abbildung 6). [18, 27]

The HIV Life Cycle

HIV medicines in six drug classes stop HIV at different stages in the HIV life cycle.

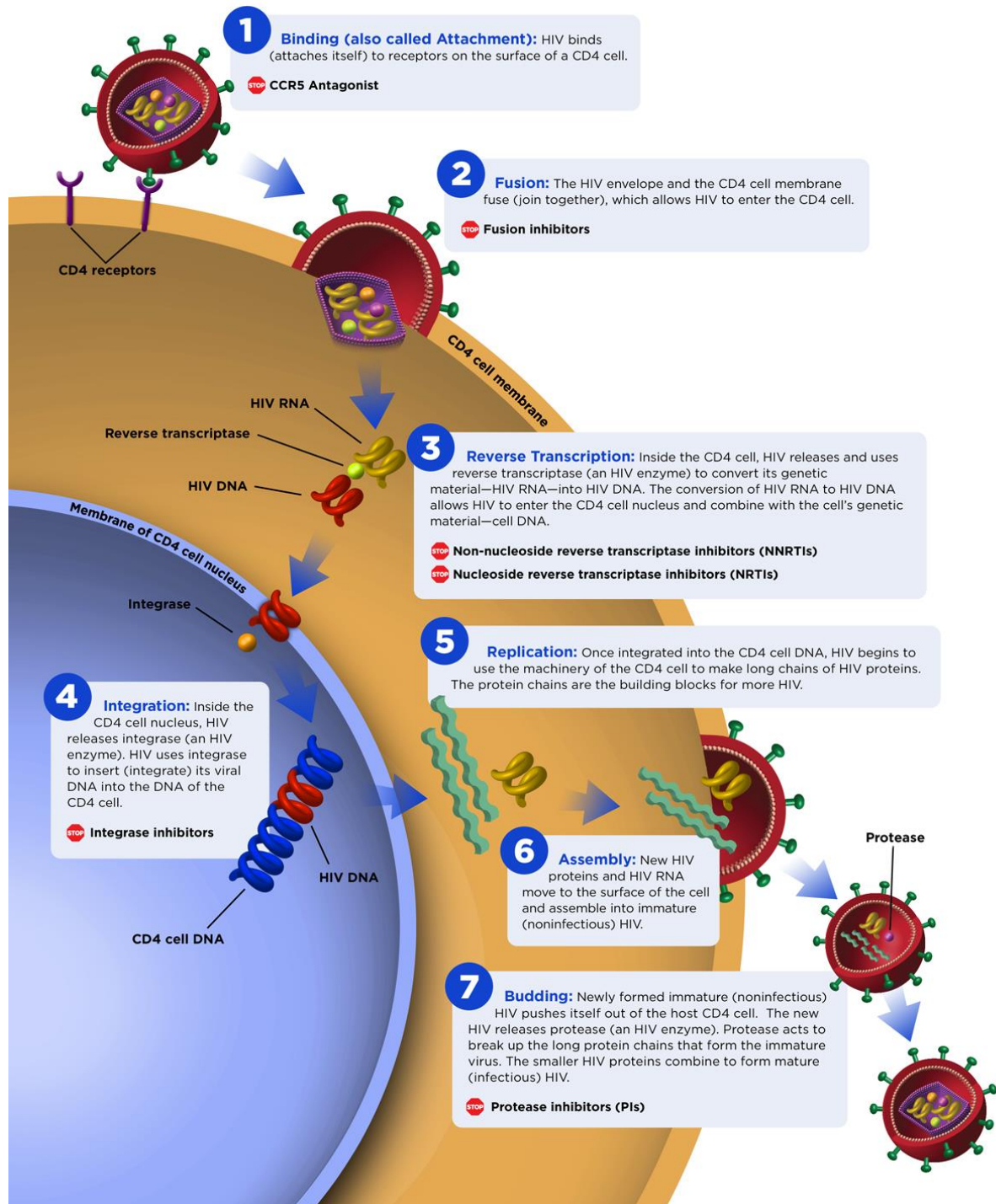


Abbildung 6: Lebenszyklus des humanen Immundefizienzvirus und Angriffsstellen der verschiedenen Wirkstoffklassen [28]

1.3 Genotypische Resistenztestung und Subtypbestimmung

Bei der genotypischen Resistenztestung werden Resistenz-assoziierte Genomregionen von HIV-1 amplifiziert und sequenziert. Mithilfe von Algorithmen, wie dem Webtool der „Stanford University HIV Drug Resistance Database“, werden Resistenzmutationen in der viralen Sequenz identifiziert und die Resistenz des Virus gegenüber verschiedener Inhibitoren vorhergesagt. Ein Nachteil der genotypischen Resistenztestung ist, dass lediglich bekannte Resistenzmuster nachgewiesen werden können. Um neue resistenzassoziierte Mutationen zu entdecken muss eine phänotypische Resistenztestung durchgeführt werden, die sehr zeit- und kostenaufwendig ist. [29-31] Nach den Deutsch-Österreichischen Leitlinien (DAIG 2012) wird eine genotypische Resistenztestung bei Diagnose und gegebenenfalls bei Therapieversagen empfohlen, um ein optimal auf den Patienten zugeschnittenes medikamentöses Therapieschema ausarbeiten zu können. [32]

Des Weiteren kann anhand der erzeugten viralen Sequenz mittels phylogenetischer Analyse oder mit Hilfe von Webtools, wie dem REGA HIV-1 Subtyping Tool [33], der Subtyp des Virus bestimmt werden. Konsequenzen der genetischen Diversität für die Impfstoffentwicklung, die antiretrovirale Therapie, die Sicherheit der Diagnostik, und möglicherweise auch für die Pathogenität (Krankheitsprogression) sind bekannt. Daher ist die nähere Charakterisierung der zirkulierenden Varianten und die Bewertung ihrer epidemischen Relevanz von großer Bedeutung.

1.4 Phylogenie

Die Phylogenie beschäftigt sich mit dem einmaligen historischen Ablauf der Kladogenese, also der Aufspaltung, und der meist damit verbundenen Anagenese, der Höherentwicklung, aber auch der Rückbildung genetischer Eigenschaften von Lebewesen und Organismen im Laufe der Evolution.

Die ersten phylogenetischen Bäume gehen wohl auf den deutschen Zoologen Ernst Haeckel zurück, der ein begeisterter Anhänger von Darwin's Evolutionstheorie war. Ein phylogenetischer Baum besteht aus Ästen beziehungsweise Zweigen (englisch: branches) und Verzweigungen oder Knoten (englisch: nodes), in denen die Zweige zusammenlaufen und die eine Gruppe mit gemeinsamer Abstammung in einer Klade (englisch: clade), oder auch monophyletischen Gruppe, zusammenfassen. Die Position der Wurzel muss erst gefunden werden, wodurch sich die Leserichtung, meist von links nach rechts, ergibt. So können ältere Verzweigungen von späteren Unterschieden werden. Um die Merkmalsausaustausche der Organismengruppen darstellen zu können, wird eine Außengruppe (englisch: outgroup) verwendet, also eine Gruppe die stammesgeschichtlich weiter entfernt von der Innengruppe steht, als alle Gruppen der Innengruppe zueinander (Abbildung 7). [34]

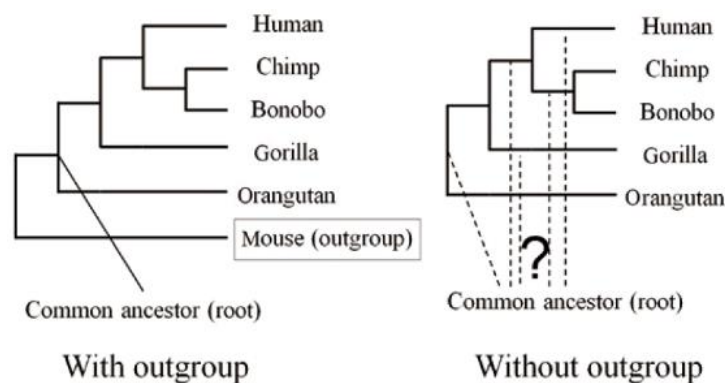


Abbildung 7: Maximum Likelihood Baum, links ohne Außengruppe, rechts mit Außengruppe [35]

Auch die Astlänge ist von Bedeutung. So gibt sie ein Größenmaß für die Veränderungen, beziehungsweise die Unterschiede der verschiedenen Gruppen an. Sind zwei Äste einer Klade gleich lang, so wie im Kladogramm (Abbildung 8 G-I) sind sie sich sehr ähnlich. Unterschiedliche Astlängen zeigen hingegen, dass zwar ein gemeinsamer Vorfahre vorliegt, im direkten Vergleich die zwei Organismengruppen jedoch sehr unterschiedlich sind (Abbildung 8 A-F). [35]

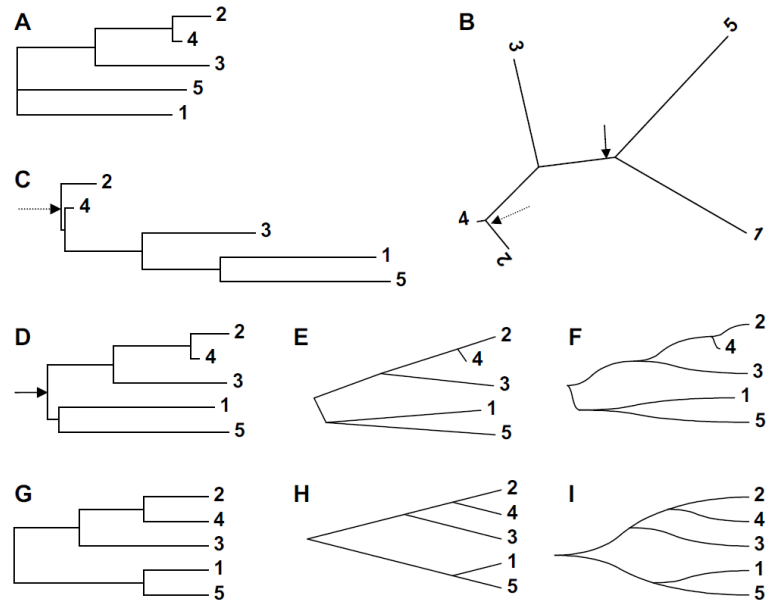


Abbildung 8: Verschiedene graphische Darstellungen einer Stammbaumdatei, **A** als Phylogramm mit basaler Tritomie, **B** als radialer Baum mit zwei Varianten zur Bewurzelung der Topologie mit Pfeil bezeichnet. Die Bewurzelung bei dem gestrichelten Pfeil ergibt das Phylogramm in **C**. In **D** ist das Phylogramm mit Mittelpunktbewurzelung dargestellt, in **E** dasselbe in der „straight“ und in **F** in der „curved“ Optik. **G**, **H** und **I** sind die entsprechenden Darstellungen als Kladogramme [34]

Wurden früher die HIV-1-Isolate noch mit Hilfe der *gag*- und *env*-Sequenzen klassifiziert, da sie von einem gemeinsamen Ursprung ausgingen, werden heute alle Genomregionen für eine Klassifizierung untersucht. Besonders die *env*-Region unterliegt einer extrem hohen Variabilität aufgrund der hohen Mutationsrate des humanen Immundefizienzvirus, welches Replikationsfehler nicht korrigieren kann. Zudem kodiert die *env*-Region die Hüll-Glykoproteine, die nach dem Kapsidprotein p24 zeitig vom Immunsystem als Antikörperziel dienen. Eine dauerhafte, hohe Mutationsrate und Alteration der Antikörperbindungsstellen sorgt dafür, dass das Virus weiterhin frei im Blut zirkulieren kann und weitere CD4-positive Zellen infiziert. Die Unterschiede nur allein zwischen den *env*-Aminosäuresequenzen der verschiedenen Untergruppen seit Beginn der HIV-1-Epidemie und heute liegen bei etwa 25 – 35 %, sodass die Klassifikation des Subtyps immer schwieriger wird und immer mehr rekombinante Formen entstehen. [36-37]

2 Zielstellung

Im HIV-Studienlabor erfolgt im Rahmen von molekularepidemiologischen Studien die Untersuchung von mehreren tausend HIV-positiven Patientenproben jährlich. Ziel dieser Studien ist es, die molekulare Evolution und Ausbreitungsdynamik von resistenten Viren und verschiedenen Subtypen bzw. rekombinanten Formen zu überwachen sowie Transmissionscluster zu identifizieren und deren Aktivität zu beobachten. Anhand dieser Ergebnisse lassen sich gezielte Präventionsmaßnahmen einleiten und Therapieleitlinien anpassen.

Routinemäßig wird im HIV-Studienlabor von Plasmaproben eine genotypische Resistenztestung im Bereich der Protease (*pr*) und Reversen Transkriptase (*rt*) durchgeführt, teilweise auch in der Integrase (*int*). Allerdings sind rekombinante Formen anhand der sehr konservierten *pr*-Sequenz häufig nicht identifizierbar oder nicht eindeutig klassifizierbar, wenn sich Rekombinationsstellen außerhalb des untersuchten Bereichs befinden. Ziel dieser Bachelorarbeit ist daher die zusätzliche Untersuchung der Regionen *int* und *env* (Hüllprotein), um mehr Sequenzinformation für die Subtypisierung zur Verfügung zu haben.

Hierzu sollen von 25 beziehungsweise 45 Patientenproben *int*- beziehungsweise *env*-Sequenzen generiert werden und nachfolgend die Sequenzinformationen aus *pr*, *int* und *env* zur näheren Charakterisierung der Varianten herangezogen werden. Zu diesem Zweck wird die virale RNA aus HIV-positiven Plasma-Proben extrahiert. Anschließend erfolgen RT-PCRs und die Aufreinigung, Konzentrationsbestimmung und Sequenzierung der PCR-Produkte. Nach Auswertung der viralen Sequenzen soll mittels Webtools und eigener Maximum-Likelihood-Phylogenie eine Subtypisierung anhand der Einzelsequenzen und der Gesamtsequenzinformation vorgenommen (Abbildung 9) werden, um die Konkordanz der Algorithmen untereinander und im Vergleich zur eigenen Phylogenie zu bewerten.

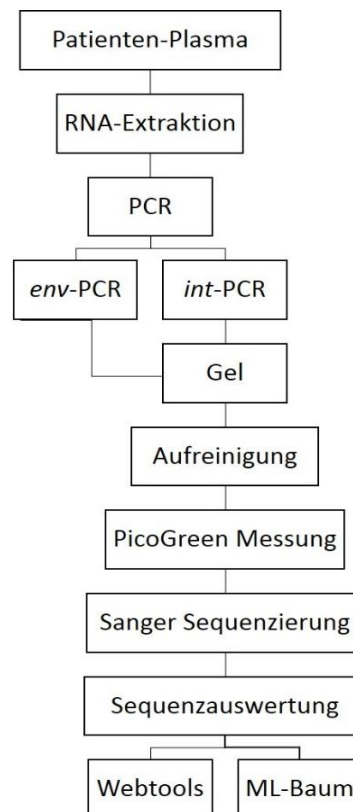


Abbildung 9: Workflow für die Charakterisierung der nicht klassifizierbaren Patientenproben in dieser Arbeit

3 Material und Methoden

3.1 RNA-Extraktion mit dem NucliSENS®-easyMag®-Extraktionsautomaten

Der erste Schritt der genotypischen Resistenztestung ist die Extraktion der viralen genomischen RNA aus zellfreien Viruspartikeln im Blutplasma des Patienten.

Das Prinzip des Untersuchungsverfahrens beruht darauf, dass die Viruspartikel im Plasma durch die Zugabe von Lysispuffer mit Guanidiniumthiocyanat lysieren. Die so freigesetzte RNA wird an magnetische Silika-Kügelchen (magnetic Beads) gebunden, durch Waschschrte gereinigt und wieder von den Silika-Beads eluiert.

3.1.1 Geräte

Sicherheitswerkbänke von BLEYMEHL (PCR-Laminar 1 und 2)

Microzentrifuge Fresco 21 Heraeus

Mikropipetten Eppendorf Research 1000/200/100/10 µl

Vortexer Neolab 7-2020

Kühl-/Gefrierkombination 4 °C/-20 °C von Liebherr

Freezer -80 °C (Plasma) Ultra low VIP Series -86 °C von Sanyo

Freezer -80 °C (RNA) Ultra Low von Sanyo

Kühlelement IsofreezePCR-Rack

Extraktionsautomat NucliSENS® easyMag® (Biomerioux)

3.1.2 Reagenzien & Materialien

Kontrollen:

HIV-positives Plasma aus ViroSeq™ HIV-1 Genotyping System Kit (v2.0 Kit) – Fa. Abbott

HIV-negatives Plasma aus ViroSeq™ HIV-1 Genotyping System Kit (v2.0 Kit) – Fa. Abbott oder HIV-negatives Plasma aus Referenzmaterialliste des FG18

HTLVIIIB Run control (Virus-gespiktes negatives humanes Plasma, interne Positivkontrolle)

FCV (felines Calicivirus, interner Standard für Viruslastbestimmung)

Reaktionsgefäße (SafeSeal) 1,5 ml und 2 ml - Fa. Sarstedt

Einweggefäße – Biomerioux

Extraktionspuffer 1/2/3 - Biomerioux

Magnetisches Silika - Biomerioux

Lysispuffer - Biomerioux

BioHit e1200 (elektrische 8 Kanalpipette (50 – 1200 µl))

BioHit Pipettenspitzen mit Filter

3.1.3 Durchführung

Alle verwendeten Reagenzien werden auf Raumtemperatur gebracht. Sollte der Lysispuffer auskristallisiert sein, muss das ausgefallene Detergenz im Wasserbad bei 37 °C erwärmt und gelöst werden. Die benötigten Plasmaaliquots mit einer Mindestmenge von 500 µl, werden im Kühlblock unter der Sicherheitswerkbank aufgetaut und dürfen keinesfalls über 4 °C erwärmt werden. Nach dem Auftauen werden 500 µl Plasma in ein 1,5 ml Reaktionsgefäß mit Safelock-Deckel unter ständiger Kühlung überführt, der Plasmarest verbleibt im Aliquot-Röhrchen, welches markiert und zurück in die Plasmabank (-80 °C) sortiert wird. Um eventuelle Kryopräzipitate zu entfernen, werden die 500 µl Plasma bei 3.000 g, für 10 Minuten bei 4 °C in der Mikrokühlzentrifuge zentrifugiert. Während der Zentrifugation werden 2 ml Lysispuffer und 3 µl FCV in die Einweggefäße von Biomerieux vorgelegt. Nach der Zentrifugation werden 450 µl des Überstandes (Standard für genotypische Resistenztestung) in die Einweggefäße pipettiert, durch hoch- und runterpipettieren gemischt und für 10 Minuten bei Raumtemperatur inkubiert. Während dieser Inkubation wird der easyMag®-Extraktionsautomat vorbereitet. Nach Ablauf der Inkubation werden 50 µl MagneticBeads dazu pipettiert und mit der BioHit Multikanalpipette gut durchmischt. Die Einweggefäße werden nun in den Extraktionsautomaten eingesetzt und der Extraktionslauf gestartet. Nach Beendigung des Laufs, wird das RNA-Eluat (60 µl) aliquotiert á 10 µl in 0,5 ml Tubes überführt und anschließend bei -80 °C gelagert. Nach anschließender Durchführung der Reinigungsprogramme wird die vom Gerät erstellte Zusammenfassung des Laufes ausgedruckt, abgeheftet und der Automat, sowie der dazugehörige PC abgeschaltet.

Für die hier beschriebenen Versuche wurde mehrfach HTLVIIIB RNA (Acc. No K03455.1, welches als Referenzsequenz für den Subtyp B verwendet wird) als Positivkontrolle extrahiert.

3.2 PCR-Bedingungen für die Integrase-PCR

3.2.1 Geräte

Pipettierbox (Sonderanfertigung)

Dry Block Heating Thermostat TDB-100 - Fa. Boeco Germany

Vortex Genie 2 – Fa. Scientific Industries

Rotilabo®-Mini-Zentrifuge - Fa. Roth

Rotilabo®-Zentrifuge mit Butterfly-Rotor - Fa. Roth

Biometra T3000 Thermocycler – Fa. Biometra

Mastercycler Pro S – Fa. Eppendorf

3.2.2 Reagenzien & Materialien

Qiagen OneStep RT-PCR Kit – Fa. Qiagen

diverse Pipetten – Fa. Eppendorf

diverse Pipettenspitzen – Fa. Thermo, Fa. Eppendorf, Fa. VWR

Kühlelement IsofreezePCR-Rack

Verwendete Primer [38]:

5'INT: 5'- ATT GGA GGA AAT GAA CAA GT -3'

3p31as: 5'- ATC CTG TCT ACY TGC CAC ACA A -3'

3.2.3 Durchführung

Das Protokoll für die Integrase-PCR wurde in der vorher durchgeführten Praxisphase optimiert und hier angewendet. Das Protokoll sieht als ersten Arbeitsschritt die Denaturierung von 10 µl RNA für 10 Minuten bei 65 °C zur Auflösung von Sekundärstrukturen und eine anschließende *hold*-Phase bei 4 °C vor. Zu den 10 µl denaturierter RNA werden 15 µl Mastermix zugegeben, der sich wie folgt zusammensetzt:

Tabelle 1: Zusammensetzung des int-PCR-Mastermixes

Reagenz	Konz. Stock	Konz. Final	µl/Ansatz
H ₂ O	-	-	6
5x Puffer (12,5 mM MgCl ₂)	5x	1x	5
dNTP Mix	10 µM	400 µM (pro dNTP)	1
as-Primer 3p31	25 µM	0,5 µM	0,5
s-Primer 5-INT	25 µM	0,5 µM	0,5
RNasin	40 U/µl	40 U	1
Enzym Mix	200 U/µl	200 U	1
Summe			15

Nach Zugabe des Mastermixes erfolgt die Amplifikation im Zykler mit folgendem Programm:

Tabelle 2: Zyklusprogramm für die int-PCR

Temperatur und Dauer	Wiederholung
50 °C 30 min	
95 °C 15 min	1x
95 °C 30 sek	
65 °C touchdown 2x -5 °C (60 °C) 45 sek	2x
72 °C 3 min	
95 °C 30 sek	
55 °C 45 sek	43x
72 °C 3 min	
72 °C 10 min	1x
4 °C ∞	

Die amplifizierte DNA wurde abschließend über eine Agarosegelelektrophorese auf PCR-Fragmente korrekter Größe ausgewertet.

3.3 PCR-Bedingungen für die Envelope-PCR

3.3.1 Geräte

Pipettierbox (Sonderanfertigung)

Dry Block Heating Thermostat TDB-100 - Fa. Boeco Germany

Vortex Genie 2 – Fa. Scientific Industries

Rotilabo®-Mini-Zentrifuge - Fa. Roth

Rotilabo®-Zentrifuge mit Butterfly-Rotor - Fa. Roth

Biometra T3000 Thermocycler – Fa. Biometra

Mastercycler Pro S – Fa. Eppendorf

3.3.2 Reagenzien & Materialien

Expand High Fidelity PCR System - Fa. Roche

diverse Pipetten – Fa- Eppendorf

diverse Pipettenspitzen – Fa. Thermo, Fa. Eppendorf, Fa. VWR

Kühlelement IsofreezePCR-Rack

Verwendete Primer:

R6: 5'- NNN NNN -3'

6537s: 5'- AAT GTC AGC ACA GTA CAA TGT ACA A -3'

5asc: 5'- TCC TTS GAT GGG AGG GGC ATA CAT TGC -3'

3.3.3 Durchführung

Das Protokoll sieht als ersten Arbeitsschritt die Denaturierung von 10 µl RNA für 3 Minuten bei 65 °C zur Auflösung von Sekundärstrukturen und eine anschließende *hold*-Phase bei 4 °C vor. Zu den 10 µl denaturierter RNA werden 10 µl Mastermix für die cDNA-Synthese zugegeben, der sich wie folgt zusammensetzt:

Tabelle 3: Zusammensetzung des env-PCR-Mastermix (ältere Protokollversion)

Reagenz	Konz. Stock	Konz. Final	µl/Ansatz
H ₂ O	-	-	14,125
PCR-Puffer (mit MgCl ₂)	10x	1x	2,5
dNTP Mix	2,5 mM	200 µM	2
Primer 6537s	25 µM	0,6 µM	0,5
Primer 5asc	25 µM	0,6 µM	0,5
Expand High Fidelity Polymerase	3,5 U/µl	2,6 U	0,375
Summe			20

Nach Zugabe des Mastermixes erfolgt die Amplifikation im Zykler mit folgendem Programm:

Tabelle 4: Zyklusprogramm für die env-PCR (ältere Protokollversion)

Temperatur und Dauer	Wiederholung
94 °C 2 min	1x
94 °C 15 sek	
61,8 °C 30 sek	45x
72 °C 45 sek	
72 °C 7 min	1x
4 °C ∞	

Die amplifizierte DNA wurde abschließend über eine Agarosegelelektrophorese auf PCR-Fragmente korrekter Größe ausgewertet.

3.4 PCR-Bedingungen für die nested Envelope-PCR

3.4.1 Geräte

Pipettierbox (Sonderanfertigung)

Dry Block Heating Thermostat TDB-100 - Fa. Boeco Germany

Vortex Genie 2 – Fa. Scientific Industries

Rotilabo®-Mini-Zentrifuge - Fa. Roth

Rotilabo®-Zentrifuge mit Butterfly-Rotor - Fa. Roth

Biometra T3000 Thermocycler – Fa. Biometra

Mastercycler Pro S – Fa. Eppendorf

3.4.2 Reagenzien & Materialien

Expand High Fidelity PCR System - Fa. Roche

diverse Pipetten – Fa. Eppendorf

diverse Pipettenspitzen – Fa. Thermo, Fa. Eppendorf, Fa. VWR

Kühlelement IsofreezePCR-Rack

Verwendete Primer:

R6: 5'- NNN NNN -3'

6537s: 5'- AAT GTC AGC ACA GTA CAA TGT ACA A -3'

7254as: 5'- TCA TAT CTC CTC CTC CAG GTC TGA A -3'

5asc: 5'- TCC TTS GAT GGG AGG GGC ATA CAT TGC -3'

3.4.3 Durchführung

Das Protokoll sieht als ersten Arbeitsschritt die Denaturierung von 10 µl RNA für 10 Minuten bei 65 °C zur Auflösung von Sekundärstrukturen und eine anschließende *hold*-Phase bei 4 °C vor. Zu den 10 µl denaturierter RNA werden 10 µl Mastermix für die cDNA-Synthese zugegeben, der sich wie folgt zusammensetzt:

Tabelle 5: Zusammensetzung Mastermix für die cDNA-Synthese

Reagenz	Konz. Stock	Konz. Final	µl/Ansatz
H ₂ O	-	-	0,1*
RT-Puffer	5x	1x	4
dNTP Mix	2,5 mM	0,5 mM	4
AS Primer R6 (Roche)	20 µM	1 µM	1
DTT	100 mM	5 mM	1
RNasin (Promega)	40 U/µl	20 U	0,4
SuperScript II (Invitrogen)	200 U/µl	100 U	0,5
Summe			11

(* entfällt bei Probenzahl von unter 5)

Nach Zugabe von 10 µl des Mastermixes zu den 10 µl denaturierter RNA erfolgt die cDNA-Synthese im Zykler für 60 min bei 42 °C, die Denaturierung bei 96 °C für 10 min und anschließend folgt eine *hold*-Phase bei 4 °C.

Die Amplifikation sieht folgende Zusammensetzung des Mastermixes vor:

Tabelle 6: Zusammensetzung des äußeren nested env-PCR-Mastermix

Reagenz	Konz. Stock	Konz. Final	µl/Ansatz
H ₂ O	-	-	27,6
PCR-Puffer (ohne MgCl ₂)	10x	1x	5
dNTP Mix	2,5 mM	200 µM	4
MgCl ₂	50 mM	2 mM	1,4
Primer 6537s	25 µM	0,25 µM	0,5
Primer 7254as	25 µM	0,25 µM	0,5
Expand High Fidelity Polymerase	3,5 U/µl	1,75 U	1
Summe			40

Nach Zugabe von 40 µl Mastermix zu 10 µl cDNA erfolgt die Amplifikation im Zykler mit folgendem Programm:

Tabelle 7: Zyklusprogramm für die äußere nested env-PCR

Temperatur und Dauer	Wiederholung
94 °C 2 min	1x
94 °C 1 min	30x
65 °C 2,5 min	
68 °C 10 min	1x
4 °C ∞	

Da es sich um eine nested PCR handelt, wird eine weitere PCR mit anderem antisense-Primer, aber gleichem sense-Primer durchgeführt, um ein kleineres zu amplifizierendes Stück der *env*-Genomregion zu erhalten.

Tabelle 8: Zusammensetzung des semi-nested *env*-PCR-Mastermix

Reagenz	Konz. Stock	Konz. Final	µl/Ansatz
H ₂ O	-	-	33,5
PCR-Puffer (ohne MgCl ₂)	10x	1x	5
dNTP Mix	2,5 mM	200 µM	4
MgCl ₂	50 mM	2 mM	2
Primer 6537s	25 µM	0,25 µM	1
Primer 5asc	25 µM	0,25 µM	1
Expand High Fidelity Polymerase	3,5 U/µl	1,75 U	1
Summe			47,5

Nach Zugabe von 47,5 µl Mastermix zu 2,5 µl PCR-Produkt aus der vorhergehenden PCR erfolgt die Amplifikation im Zykler mit folgendem Programm:

Tabelle 9: Zyklerprogramm für die semi-nested *env*-PCR

Temperatur und Dauer	Wiederholung
94 °C 2 min	1x
94 °C 1 min	30x
65 °C 2,5 min	
68 °C 10 min	1x
4 °C ∞	

Die amplifizierte DNA wurde abschließend über eine Agarosegelelektrophorese auf PCR-Fragmente korrekter Größe ausgewertet.

3.5 Agarosegelelektrophorese

Die Agarosegelelektrophorese dient der Darstellung von PCR-Produkten (DNA-Fragmente). Die doppelsträngigen (ds) PCR-Produkte werden in einer Agarose-Matrix nach ihrer Größe aufgetrennt. Anhand eines Größenstandards kann die Größe des PCR-Produkts bestimmt und somit das Ergebnis der *protrt*-, *int*- oder *env*-PCR positiv oder negativ bewertet werden. Der lineare Trennbereich hängt von der Agarose-Konzentration ab und variiert je nachdem, welches PCR-Produkt analysiert werden soll. Für die Analyse des *protrt*-PCR-Produktes mit einer Größe von 1.079 bp wird ein 1 %iges Agarosegel mit linearem Trennbereich von 400 bp bis 10.000 bp und für die Analyse des *int*-PCR-Produktes mit einer Größe von 915 bp wird ein 1,5 %iges Agarosegel mit linearem Trennbereich von 200 bp bis 4.000 bp eingesetzt. Nach anschließender PCR-Produktaufreinigung über magnetische Beads (Agencourt® AMPure® XP Kit), kann die DNA-Konzentration mit Hilfe eines fluoreszierenden Farbstoffs (PicoGreen®) gemessen werden.

3.5.1 Geräte

pH-Meter QpH 70 – Fa. VWR

Magnetrührer mit Heizfunktion – Fa. Stuart

Sartorius Waage L 610 D

Mikrowelle

Elektrophoresekammer von Whatman HORIZON 58 – Fa. BIOMETRA

8er Kamm mit Zahnbreite von 0,8 mm (maximales Auftragsvolumen 15 µl) – Fa. BIOMETRA

Low Voltage Power Pack P25 – Fa. BIOMETRA

Electrophoresis Power Supply ST 305 – Fa. GIBCO BRL

HEROLAB Gel-Dokumentationssystem mit EasyWin Software – Fa. HEROLAB

3.5.2 Reagenzien & Materialien

Messzylinder – Fa. DURAN Hirschmann

Bechergläser – Fa. Schott DURAN

Glaspipetten – Fa. Brand

500 ml Flaschen – Fa. Schott DURAN

Aktivkohle – Fa. Merck

Ethidiumbromidlösung 10 mg/ml – Fa. Roth

Faltenfilter – Fa. Roth

Pipettenspitzen 0,5 – 10 µl – Fa. Eppendorf

Eppendorf Research 0,5 – 10 µl Pipette

1,5 ml Röhrchen – Fa. Sarstedt

6x DNA Loading Dye – Fa. Thermo

1 kbp DNA Leiter – Fa. Thermo

Low DNA Mass Ladder – Fa. Thermo

Elektrophoresepuffer (2 Liter 1x TAE-Puffer (1x 50 TAE-Puffer 50-fach verdünnt) plus 100 µl Ethidiumbromid 10 mg/ml (Endkonzentration EtBr: 0,5 µg/ml))

Agarose Electrophoresis Grade Pulver – Fa. Thermo

3.5.3 Vorbereitung der PCR Produkte

Bei der Vorbereitung der Produkte ist darauf zu achten, dass die *int*- und *env*-PCR-Produkte nach der PCR bis zur DNA-Gelelektrophorese in der Regel maximal 12 – 16 h bei 4 °C gelagert werden. Eine längere Lagerung muss bei -20 °C erfolgen, da die Nukleinsäuren empfindlich sind und wiederholte Frost-/Tauzyklen oder eine längere Lagerung bei 4 °C zur Degradierung führen kann. Ist die Gelelektrophorese vorbereitet, werden die Proben gekühlt, bei Raumtemperatur pipettiert und anschließend wieder im Kühlschrank gelagert beziehungsweise zur Langzeitlagerung eingefroren.

3.5.4 Durchführung

Zu allererst wird das Agarosegel vorbereitet. Dafür werden 25 ml 1 %iges oder 1,5 %iges Agarosegel in einem Erlenmeyerkolben nach Abnehmen des Parafilms in der Mikrowelle zirka eine Minute bei 600 W erhitzt. Wenn die Agarose vollständig gelöst ist und nicht mehr kocht, werden 10 µl verdünnte EtBr (0,4 µg/ml) hinzu pipettiert, sodass eine anschließende Färbung entfällt. Die hier verwendeten Gele wurden schon vorbereitet und im Kühlraum bei 4 °C gelagert, sodass das Gel nur noch aufgeköcht und mit EtBr versetzt werden musste. Nachfolgend wird die Mini-Whatman Elektrophoresekammer mit Abstandshaltern (englisch: Spacer) vorbereitet, das flüssige Agarosegel hineingegossen (auf ca. 70 °C abkühlen) und ein oder zwei 8er Käbme eingesetzt. Das Gel muss für ca. 20 min erstarren, sodass in der Zwischenzeit die gereinigten *int*- oder *env*-PCR-Produkte vorbereitet werden können. Für qualitative Gele werden 3 µl PCR-Produkt mit 2 µl Aqua bidest und 1 µl 6x Ladepuffer versetzt und gemischt. Das erstarrte Agarosegel wird mit ca. 150 ml 1x TAE-Puffer überschichtet und die Käbme vorsichtig gezogen. Nun können die vorbereiteten Proben in die sichtbaren Geltaschen pipettieren werden, die aufgrund der durch den Ladepuffer erhöhten Dichte, auf den Boden der Geltasche sinken.

Pro Gelhälfte wird eine Spur mit 300 ng (3 µl) der 1 kbp DNA-Leiter (Abbildung 10) als Größenstandard belegt und jeweils eine negative und positive PCR-Kontrolle auf jedem Gel mit analysiert.

GeneRuler 1 kb DNA Ladder

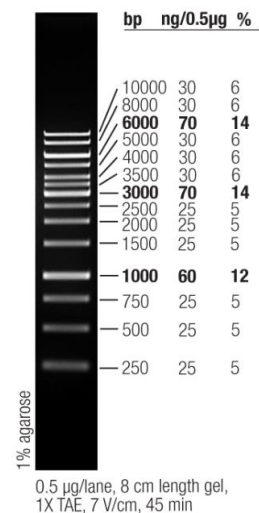


Abbildung 10: 1 kbp DNA Ladder Fa. Thermo [39]

Um die Auftrennung zu starten wird der Deckel der Gelkammer geschlossen und die Kabel von Plus- und Minuspol der Kammer an den entsprechenden Polen des Powerpacks angeschlossen. Die DNA-Fragmente wandern in Richtung Pluspol (Anode). Die Elektrophorese läuft für ungefähr 30 Minuten bei einer Spannung von 90 Volt, wobei das Bromphenolblau als Marker dient und nicht aus dem Gel beziehungsweise in andere Taschen der zweiten Gelhälfte hineinlaufen darf. Das Gelbild wird mit einer Videokamera und dem Gel-Dokumentationssystem aufgenommen. Hierfür muss das Gel luftblasenfrei auf eine Folie auf dem Transluminator gelegt und bei sichtbarem Licht optimal ausgerichtet werden. Nun kann die Rolltür der Kammer verschlossen und das UV-Licht am Transluminator angestellt werden. Durch das UV-Licht wird über das EtBr die DNA zur Fluoreszenz angeregt und die Banden werden sichtbar. Es ist zu beachten, dass die Belichtungsstärke so gewählt wird, dass alle Banden gut sichtbar, aber nicht überbelichtet sind, da diese im Rahmen der Quantifizierung zu verfälschten Konzentrationen führen kann. Die Größe des PCR-Produkts wird

anhand des Größenstandards überprüft beziehungsweise bestimmt. Abschließend werden Kammer, Spacer und Käbme mit Leitungswasser und mit Aqua bidest gespült und sämtliche EtBr-Abfälle (Puffer und Gel) ordnungsgemäß entsorgt. Der TAE-Puffer kann einmal über Aktivkohle gefiltert werden.

3.6 Aufreinigung der PCR-Produkte mit Agencourt® AMPure® XP Kit

Zweck einer Aufreinigung von PCR-Produkten ist das Entfernen von Primern, Nukleotiden, Polymerasen und Salzen, da diese in nachfolgenden Reaktionen stören können. Nach der PCR werden die Reaktionsansätze mit magnetischen Beads versetzt, die an die doppelsträngigen PCR-Produkte binden. Durch das Verwenden eines Magnetblocks wird die DNA über die Beads vom Rest des Reaktionsansatzes separiert, der Primer, Nukleotide, Polymerase, Salze und andere Störfaktoren enthält. Nach mehreren Waschschriften mit Ethanol (EtOH) werden die gereinigten PCR-Produkte von den Beads eluiert. Mit dem Agencourt® AMPure® XP Kit können PCR-Fragmente mit einer Größe zwischen 150 bp und 10 kBp aufgereinigt werden.

3.6.1 Geräte

Pipettierbox (Sonderanfertigung)

diverse Pipetten von Eppendorf und Thermo

serologische Pipetten, diverse Größen – Fa. Sarstedt

3.6.2 Reagenzien und Materialien

Agencourt® AMPure® XP Kit – Fa. Beckman Coulter

EtOH – Fa. Roth

Chromatographie gereinigtes Wasser – Fa. VWR

TE-Puffer – Fa. Promega

Reagenzien-Reservoir – Fa. Beckman Coulter

Nunc 96 microwell plate round bottom – Fa. Sigma

Hard-Shell Low-Profile Thin-Wall 96-Well Skirted PCR Plates – Fa. BioRad

Zentrifugen-Röhrchen 50 ml – Fa. Sarstedt

Microseal B Klebefolien – Fa. BioRad

diverse Pipettenspitzen – Fa. Thermo, Fa. Eppendorf, Fa. VWR

Pipetboy – Fa. Hirschmann

Entsorgungsbeutel – Fa. Geyer

3.6.3 Durchführung

Zuallererst wird ein Belegungsprotokoll mittels Protokollvorlage erstellt, welche auch für die nachfolgende DNA-Quantifizierung mit dem Quant-iT™ PicoGreen® dsDNA Assay Kit gültig ist. Da in Reihe 12 des 96 Well Plattenlayouts bei der nachfolgenden DNA-Quantifizierung die Standardreihe pipettiert wird, werden maximal die Reihen 1 bis 11 mit Proben belegt.

Eine frische 80 %ige EtOH-Lösung wird unter dem Abzug hergestellt, indem 100 %iges EtOH und Wasser einzeln mittels serologischer Pipette abgemessen und in einem zuvor beschrifteten 50 ml Zentrifugenröhrchen nach folgender Formel gemischt werden: 10 ml Wasser + 40 ml EtOH (100 %). Anschließend erfolgt eine Überführung der Lösung in ein Reagenz-Reservoir. Je nach Probenanzahl wird eine entsprechende Menge der AMPure® Beads in ein dafür vorgesehenes Reagenzien-Reservoir pipettiert, die sich wie folgt berechnet: Probenanzahl · 40 µl AMPure® Beads (bei nested *env*-PCR-Produkten 90 µl Beads). Die Beads müssen vor dem Pipettieren gut durchgemischt werden. Eine entsprechende Menge des TE-Puffers (Probenanzahl · 30 µl TE-Puffer) wird in ein Reagenzien-Reservoir pipettiert.

Nachdem die PCR-Produkte in einer DNA-Box aufgetaut wurden, werden je 25 µl PCR-Produkt (bei nested *env*-PCR 50 µl) in das vorgesehene Well gemäß Belegungsplan auf die Nunc-Platte überführt. Anschließend erfolgt die Aufreinigung bei Raumtemperatur, indem jeweils 40 µl AMPure® Beads (bei nested *env*-PCR-Produkten 90 µl Beads) auf die Proben gegeben und durch 10-maliges auf und ab pipettieren gut durchgemischt werden. Es folgt eine fünf minütige Inkubation während der die PCR-Produkte an die Beads binden. Anschließend wird die Nunc-Platte auf den Magneten gestellt und mindestens weitere zwei Minuten inkubiert. In dieser Zeit werden die Beads mit den gebundenen PCR-Produkten von der verbleibenden Lösung separiert. Nun wird der Überstand abgenommen und verworfen und es folgen zwei Waschschriffe mit 80 %igem EtOH währenddessen sich die Platte die gesamte Zeit über auf dem Magneten befindet. Dafür werden jeweils 200 µl EtOH auf die Proben gegeben und für 30 Sekunden inkubiert. Nach dem zweiten Waschschriff wird der gesamte Überstand abgenommen und verworfen, die Platte vom Magneten heruntergenommen und in jedes Well 20 µl TE-Puffer (bei nested *env*-PCR-Produkten 40 µl) pipettiert, um die Beads darin zu resuspendieren. Dabei muss wieder eine gute Durchmischung stattfinden, welches durch 10-maliges auf und ab pipettieren gewährleistet wird. Die Platte wird nun für 10 – 15 Minuten inkubiert, damit das restliche EtOH verdampfen kann. Abschließend wird die Platte erneut auf den Magneten gestellt, für zwei Minuten inkubiert und 20 µl beziehungsweise 50 µl der aufgereinigten, eluierten PCR-Produkte auf eine BioRad-Platte überführt. Die Platte wird mit einer Microseal Klebefolie verschlossen und bis zur DNA-Quantifizierung maximal einen Tag bei 4 °C im Kühlschrank gelagert.

3.7 Quantifizierung von PCR-Produkten mit dem Quant-iT™ PicoGreen® dsDNA Assay Kit

Die Quantifizierung von PCR-Produkten ist für die nachfolgende Sequenzierung von Bedeutung, da nur einheitliche Konzentrationen eingesetzt werden können und die zu sequenzierenden Proben dementsprechend verdünnt werden müssen.

Nach der Aufreinigung der PCR-Produkte wird die DNA-Konzentration mit Hilfe eines fluoreszierenden Farbstoffs (PicoGreen®) gemessen. Durch das Binden des Farbstoffs an doppelsträngige DNA bildet sich ein Komplex, dessen Fluoreszenz mit einem Fluorimeter gemessen werden kann und proportional zur DNA-Konzentration ist. Das Quant-iT™ PicoGreen® dsDNA Assay Kit ermöglicht eine genaue Quantifizierung von einzelsträngiger DNA in einem Bereich zwischen 50 pg und 2 µg.

3.7.1 Geräte

diverse Pipetten von Eppendorf und Thermo

Infinite 200 Reader – Fa. Tecan (inklusive Magellan Software Version 7.1)

3.7.2 Reagenzien und Materialien

Quant-iT™ PicoGreen® dsDNA Assay Kit, darin enthalten: Quant-iT™ PicoGreen® dsDNA reagent (Component A), 20x TE (Component B), Lambda cDNA standard (Component C) – Fa. Invitrogen

Reagenzien-Reservoir – Fa. Beckman Coulter

Hard-Shell Low-Profile Thin-Wall 96-Well Skirted PCR Plates – Fa. BioRad

Microseal B Klebefolien – Fa. BioRad

BRANDplates 96 Well pureGrade schwarz – Fa. Brand

50 ml Zentrifugenröhrchen – Fa. Sarstedt

1,5 ml Reaktionsgefäße – Fa. Sarstedt

Chromatographie gereinigtes Wasser – Fa. VWR

diverse Pipettenspitzen – Fa. Eppendorf, Fa. Thermo, Fa. VWR

3.7.3 Durchführung

Zuallererst wird der Tecan-Reader (Fluorimeter) mit dem dazugehörigen Computer angeschaltet, da der Reader eine Aufwärmzeit von 20 Minuten benötigt, bevor er für die Fluoreszenzmessung verwendet werden kann. Den Lambda cDNA Standard sowie das PicoGreen® Reagenz auf Raumtemperatur bringen, wobei darauf geachtet werden muss das PicoGreen® unbedingt im Dunkeln aufzubewahren, da es lichtempfindlich ist. Nun wird 1x TE-Puffer aus der 20x Stammlösung durch eine 1:20 Verdünnung in einen 50 ml Zentrifugenröhrchen angesetzt. Dabei werden 2 ml 1x TE-Puffer pro Platte für die Verdünnung und Messung der Standardreihe benötigt. Für die Verdünnung und Messung der PCR-Produkte werden inklusive 10 % Pipettierverlust je Probe 220 µl 1x TE-Puffer benötigt. Die benötigte Gesamtmenge des 1x TE-Puffers berechnet sich demnach wie folgt: $220 \mu\text{l} \cdot \text{Anzahl der PCR-Produkte} + 2 \text{ ml für Standardreihe}$. Nun wird die Lambda cDNA Standardreihe hergestellt (1:50, 1:500, 1:5.000, 1:50.000). Pro Platte werden 100 µl Standard pro Verdünnungsstufe benötigt. Dafür werden vier 1,5 ml Reaktionsgefäße vorbereitet und beschriftet, sodass die 1:50 Verdünnung mit 147 µl 1x TE-Puffer und 3 µl Lambda cDNA Standard hergestellt werden kann. Für die weiteren Verdünnungsstufen werden jeweils 180 µl 1x TE-Puffer in die entsprechenden 1,5 ml Reaktionsgefäße vorgelegt und durch die Zugabe von 20 µl der vorherigen Verdünnung hergestellt, wobei vor Entnahme der 20 µl ein kurzes Puls vortexen durchgeführt werden muss. Nachfolgend werden in die Wells A2 bis E12 je 100 µl der Standardreihe von oben nach unten in der Reihenfolge 1:50, 1:500, 1:5.000, 1:50.000 und Blank (1x TE-Puffer) pipettiert. In der DNA-Box werden in einer schwarzen 96 Well Brand-Platte für die Messung der PCR-Produkte pro Well 98 µl 1x TE-Puffer vorgelegt. Anschließend werden je 2 µl PCR-Produkt nach vorherigem 2 – 3-maligem auf und ab pipettieren von der BioRad-Platte in das korrespondierende Well der Brand-Platte überführt. Dies erfolgt unter Zuhilfenahme einer 8-Kanal-Pipette. Nun wird das PicoGreen® 1:200 in einem 50 ml Zentrifugenröhrchen verdünnt. Pro PCR-Produkt werden inklusive 10 % Pipettierverlust 110 µl und für die gesamte Standardreihe zusätzlich 1 ml der PicoGreen®-Verdünnung benötigt. Die benötigte

Gesamtmenge an PicoGreen®-Verdünnung berechnet sich demnach wie folgt: 110 ml · Anzahl PCR-Produkte + 1 ml für die Standardreihe. Unter Zuhilfenahme einer 8-Kanal-Pipette werden je 100 µl der PicoGreen®-Verdünnung in jedes Well mit Probe pipettiert und durch 5 – 6-maliges auf und ab pipettieren gemischt. Danach wird die Platte zügig zur Konzentrationsmessung zum Tecan-Reader gebracht. Am PC wird die Magellan-Software geöffnet, die Platte ohne Deckel in den Reader eingesetzt und in das Gerät gefahren. Das Gerät misst nun die Fluoreszenzemission bei 535 nm bei einer Exzitationswellenlänge von 485 nm. Nach Beendigung der Messung werden die Rohdaten im Excel-Format gespeichert und mit Hilfe einer vorbereiteten Vorlage ausgewertet. Der Messvorgang wird beendet und die Platte aus dem Reader entnommen und verworfen.

3.8 Sequenzierung

3.8.1 Geräte

diverse Pipetten – Fa. Eppendorf

diverse Pipettenspitzen – Fa. Thermo, Fa. Eppendorf, Fa. VWR

3.8.2 Reagenzien und Materialien

Big Dye Terminator v3.1 Cycle Sequencing Kit – Fa. Thermo Fisher Scientific

Nuklease-freies Wasser

Eisbad – Fa. NeoLab

Verwendete Primer für die Integrase-Sequenzierung:

5'INT: 5'- ATT GGA GGA AAT GAA CAA GT -3'

3p31as: 5'- ATC CTG TCT ACY TGC CAC ACA A -3'

F2s: 5'- TAA GAC AGC AGT ACA AAT GGC AG -3'

F3as: 5'- GCT GTC CCT GTA ATA AAC CCG -3'

Verwendete Primer für die Envelope-Sequenzierung:

C2V5_2s: 5'- TTA ATT ATT GTG GAG GGG AAT TTT TCT -3'

C2V5_2as: 5'- AGA AAA ATT CCC CTC CAC AAT TAA -3'

6537s: 5'- AAT GTC AGC ACA GTA CAA TGT ACA A -3'

5asc: 5'- TCC TTS GAT GGG AGG GGC ATA CAT TGC -3'

3.8.3 *int*-Sequenzansätze

Nachdem die DNA-Konzentration der PCR-Produkte bestimmt wurde, wurde für die Sequenzierung der *int*-PCR-Produkte pro Sequenzierprimer 5'-INT, 3p31, F2s und F3as ein Mix angesetzt in dem 20 ng DNA enthalten sind.

Tabelle 10: Sequenzansatz für die int-PCR-Produkte

	µl/Ansatz
Ansatz	1x
Primer (10 pmol/µl)	0,5 µl
BigDye 3.1	2 µl
5x Puffer	1 µl
H₂O	4,5 µl
Mix	8 µl
DNA	2 µl

Pro Probe werden demnach vier Ansätze mit den vier verschiedenen Sequenzierprimern vorbereitet, pipettiert und mit folgendem Zyklusprogramm amplifiziert:

Tabelle 11. Zyklusprogramm für die int-Sequenzierung

Temperatur und Dauer	Wiederholung
96 °C 2 min	1x
96 °C 10 sek	
55 °C 5 sek	25x
60 °C 4 min	
4 °C ∞	

Nach der Amplifikation wurden die PCR-Produkte für die Sanger-Sequenzierung in das hauseigene Sequenzierlabor versandt.

3.8.4 env-Sequenzansätze

Nachdem die DNA-Konzentration der PCR-Produkte bestimmt wurde, wurde für die Sequenzierung der env-PCR-Produkte pro Sequenzierprimer 2s, 2as, 5asc und 6537s ein Mix angesetzt. Da die Konzentrationen der env-Proben sehr unterschiedlich ausfallen und teilweise unterhalb der einzusetzenden 15 ng DNA lagen, wird das Wasser individuell angepasst, beträgt jedoch in der Regel maximal 5 µl und minimal 2,5 µl für Proben mit negativen DNA-Konzentrationen innerhalb der PicoGreen® Messung.

Tabelle 12: Sequenzansatz für die env-PCR-Produkte

	µl/Ansatz
Ansatz	1x
Primer (10 pmol/µl)	0,5 µl
BigDye 3.1	2 µl
5x Puffer	1 µl
Mix	3,5 µl
DNA	1,5 µl

Pro Probe werden demnach vier Ansätze mit den vier verschiedenen Sequenzierprimern vorbereitet, pipettiert und mit folgendem Zyklusprogramm amplifiziert:

Tabelle 13. Zyklusprogramm für die env-Sequenzierung

Temperatur und Dauer	Wiederholung
96 °C 2 min	1x
96 °C 10 sek	
55 °C 5 sek	25x
60 °C 4 min	
4 °C ∞	

Nach der Amplifikation wurden die PCR-Produkte für die Sanger-Sequenzierung in das hauseigene Sequenzierlabor versandt.

3.8.5 Sanger Sequenzierung

Mit Hilfe der Sanger-Sequenzierung kann die Nukleotidabfolge einer Sequenz bestimmt werden. Dabei wird der zu sequenzierende Doppelstrang durch Denaturierung aufgespalten und eine zugegebene DNA-Polymerase katalysiert die komplementäre Synthese der zu sequenzierenden Einzelstränge. Neben den Sequenzprimern werden, Desoxyribonukleosidtriphosphate (kurz dNTPs) dATP, dTTP, dGTP und dCTP und eine kleine Menge an fluoreszenzfarbstoffmarkierten Didesoxyribonucleosidtriphosphaten (ddNTPs) zugeführt. Die weitere Verlängerung durch die DNA-Polymerase wird durch das Fehlen von 3'-Hydroxylgruppen an den ddNTPs unterbunden, sodass die DNA linear mit unterschiedlicher Länge amplifiziert wird und schließlich Gemische von A-, C-, G- beziehungsweise T-terminierten Elongationsprodukten entstehen. Nach jedem Einbau eines ddNTPs findet ein Abbruch der Elongation statt, weshalb die Methode auch als Kettenabbruchmethode bezeichnet wird. Diese Sequenzfragmente werden anschließend in einem Sequenziergerät kapillarelektrophoretisch nach ihrer Größe aufgetrennt. Die Detektion der Elektropherogramme erfolgt über einen Laser der die Fluoreszenzfarbstoffe anregt und die resultierende Fluoreszenz mit einem Sensor detektiert. Diese entstandenen Fluoreszenzsignale werden anschließend als Elektropherogramm (Abbildung 11) dargestellt, aus dem die Basenabfolge abgelesen werden kann. [40-41]

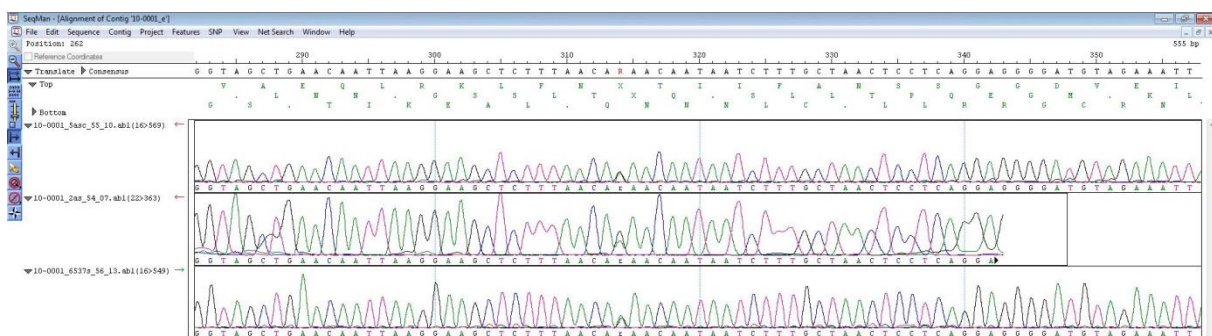


Abbildung 11: Sequenzdarstellung im Programm SeqMan Pro

3.9 Sequenzauswertung

Die Qualität der erhaltenen Sequenz hängt von der Qualität des Ausgangsmaterials ab. Schlecht gelagertes Patientenmaterial, Nebenbanden nach der Amplifikation oder eine unzureichende Aufreinigung können zu schlechten bis gar nicht auswertbaren Sequenzen führen. Da HIV zudem als Quasispezies vorliegt, eine Mischung unterschiedlicher Virusvarianten, kann es an ein und derselben Position im Genom zum Auftreten mehrerer unterschiedlicher Basen kommen (Ambiguitäten). Die manuelle Korrektur der Sequenz ist daher von entscheidender Bedeutung. Für die in dieser Arbeit durchgeführten Sequenzierungen wurden pro Fragment vier Sequenzier-Primer eingesetzt um eine zwei- bis vierfache Sequenzabdeckung in beide Richtungen zu erhalten. Dies ermöglicht eine zuverlässige Auswertung der Sequenz auch bei einem Ausfall einer Sequenzierreaktion. Allerdings sind die Bereiche, die nicht überlappend durch mindestens zwei Primer abgedeckt werden, schwieriger zu beurteilen, da keine weiteren Vergleichssequenzen durch andere Sequenzierreaktionen vorhanden sind. Besonders Ambiguitäten lassen sich in solchen Bereichen nahezu nicht von Sequenzierartefakten unterscheiden (Abbildung 12).

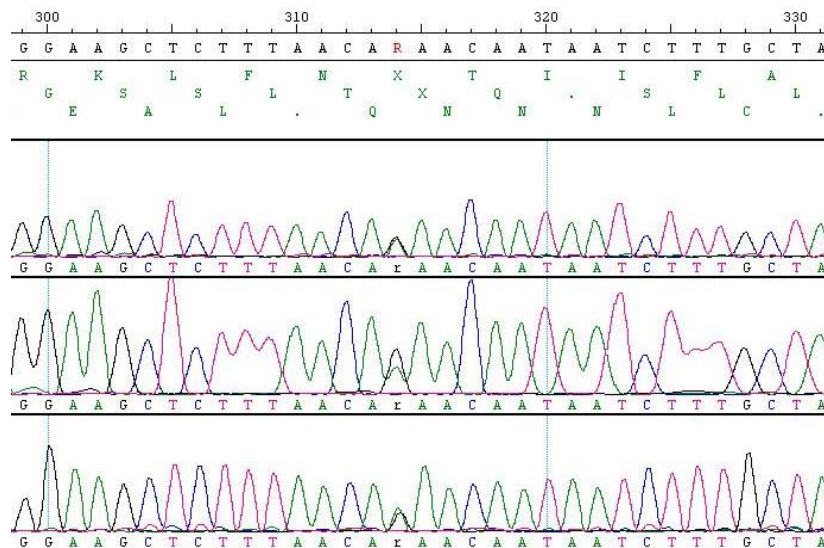


Abbildung 12: Darstellung einer Ambiguität an Basen-Stelle 314

Je nach Kombination der verschiedenen Nucleotide, werden dann sogenannte „Wobble-Basen“ (Abbildung 13) in die Sequenz eingetragen.

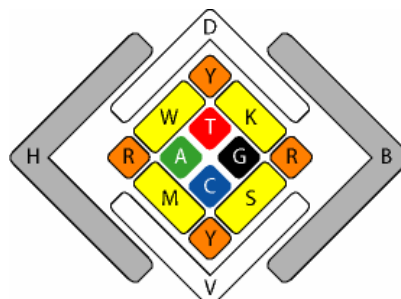


Abbildung 13: IUPAC-IUB Codes für die Sequenzauswertung [43]

Mit Hilfe des geeigneten Programms SeqMan Pro [42] wurden die resultierenden Sequenzen analysiert und ausgewertet (Abbildung 12). Dazu wurden die nach der Sequenzierung erhaltenen Rohdaten zunächst entsprechend der Probenliste mit der dazugehörigen Probennummer und dem verwendeten Sequenzierungsprimer umbeschriftet. Die jeweils vier Einzelsequenzen der *int*- und

env-Sequenzierung wurden in einem Rohdatenordner („Rohdaten“) zusammengeführt und anhand einer Kopie der Rohdaten ausgewertet. Die jeweils vier einzelnen Sequenzen werden mithilfe des Programmes SeqMan Pro zu einem Contig, das die Konsensussequenz aus den einzelnen Sequenzen darstellt, assembliert. Die richtige Orientierung des Alignments wurde anhand der Primerorientierung überprüft, der richtige Leserahmen anhand des Aminosäuremotivs ausgewählt (*int*: DKLV-MAGD; *env*: GIRP-KAMYA) und die Sequenz entsprechend getrimmt. Zur nachfolgend manuellen Bearbeitung der Konsensussequenz wurden folgende Parametereinstellungen vorgenommen: „Consensus Calling“ mit „Compare Trace Evidence“ auf 95 %; „Conflict Split“ mit „Min Inconsistent“ 1 und „Min Coverage“. Im „Strategy View“ wurden über „conflicts“ Konfliktpositionen (Lücken, oder Ambiguitäten) aufgezeigt und im Contig dann manuell bearbeitet. An Positionen, an denen mehrere Basen detektiert wurden diese entsprechend editiert, wenn der Unterpeak mindestens dreimal so groß ist wie das Hintergrundrauschen in der Umgebung dieser Position. Im Falle einer einzelsträngigen Sequenzinformation wurde die entsprechend Mischbase nur dann übernommen, wenn der Unterpeak mindestens 50 % der Höhe des Hauptpeaks aufwies. Nach der Überprüfung aller Konfliktpositionen wurde das Alignment unter der entsprechenden Patientenummer in einem Auswerteordner („Results done“) als *.sqd* abgespeichert. Zur weiteren Auswertung wurde die Konsensussequenz als *.fas* ebenfalls im Auswerteordner abgespeichert.

3.10 Phylogenie

Die erzeugten HIV-Sequenzen wurden mit den bereits vorhandenen Sequenzen pro Genomregion mithilfe der frei zugänglichen Software BioEdit [44] aligniert, getrimmt und mit Hilfe von geeigneter Software, wie zum Beispiel IQ-TREE, in einem phylogenetischen Baum, der die Verwandtschaft der Subtypen zueinander darstellt, zusammengefasst. Dabei werden je nach Programm und Einstellungen unterschiedliche mathematische Algorithmen verwendet, um die bestmögliche Passgenauigkeit der Verwandtschaft zu ermitteln. Eng miteinander verwandte Proben bilden sogenannte Clades aus, die einen hohen Bootstrap aufweisen. Der Bootstrap gibt an, mit welcher Wahrscheinlichkeit sich die Proben bei einer erneuten phylogenetischen Analyse identisch anordnen würden. Es gibt verschiedene Stammbaum-Modelle, wobei in dieser Arbeit Maximum-Likelihood-Bäume (ML-Baum) verwendet werden, die sich mehr an den realen biochemischen Gegebenheiten der Natur orientieren und die größte Wahrscheinlichkeit der Daten wiedergibt. Bei einem ML-Baum wird nach dem Baum mit dem größten Likelihood, also der Baum „der das Alignment unter Annahme eines solchen Modells zur Sequenzevolution am wahrscheinlichsten erscheinen lässt“ gesucht. [45]

Für die Phylogenie wurden die HIV-1 Sequenzen der Patientenproben zusammen mit den bis dato verfügbaren HIV-1 Referenzsequenzen (reine Subtypen und CRFs, n=202) aus der Los Alamos HIV Database mittels dem Programm Mafft [46] aligniert und als FASTA-Dateiformat (*.fas*) gespeichert. Anschließend wurden die verschiedenen Alignments in BioEdit auf eine einheitliche Länge getrimmt. Die *prnt*-Sequenzen wurden so getrimmt, dass der Protease-Start mit dem Aminosäuremotiv „PQITL“ (Position Referenz HXB2: 2.253 bp) den Anfang und das Motiv „KALTE“ (Position Referenz HXB2: 3.425 bp) in der Reversen Transkriptase das Ende bildeten. Die *int*-Sequenzen wurden vom Integrase-Start „FLDGI“ (Position Referenz HXB2: 4.230 bp) bis zum Aminosäuremotiv „MAGDD“ (Position Referenz HXB2: 5.067 bp) getrimmt. Die *env*-Sequenzen wurden von „GIRPV“ (Position Referenz HXB2: 6.972 bp) bis „KAMYA“ (Position Referenz HXB2: 7.553 bp) getrimmt (HXB2 Acc. no: K03455 M38432). Für die Erstellung des Merge-Baumes in dem sich alle drei Sequenzbereiche *prnt*, *int* und *env* befinden, wurden die drei Alignments in BioEdit mit dem Befehl „Merge into Alignment based on a Reference Sequence“ verschmolzen. Da für die Berechnung der Stammbäume das Phylip -Dateiformat (*.phy*) benötigt wird, wurden die Alignments mit dem Programm Seaview [47] im gewünschten Format abgespeichert. Für die getrimmten Alignments im *.phy*-Format wurde nun mit dem Programm IQ-TREE [48] der am besten geeignete Baum nach der Maximum-Likelihood-Methode

berechnet. Folgendes Kommando wurde für die Berechnung mit dem Linux-basierten Programm verwendet:

```
iqtree-omp -s NameDerInputDatei.phy -m TEST -pre NameDerOutputDatei -bb 10000 -nt 6
```

Mit dem ersten Befehl **iqtree-omp** wird das Programm geöffnet. Es folgt die Eingabe des Namens der Ausgangsdatei (**-s NameDerInputDatei.phy**). Mit dem Befehl **-m TEST** wird das am besten geeignete Modell unter 88 Modellen für die Berechnung des Maximum-Likelihood-Baums ermittelt. Der Befehl **-pre NameDerOutputDatei** generiert den Namen der erstellten Dateien. Es wird keine Dateierweiterung am Namen benötigt, da mehrere Dateien mit unterschiedlichen Endungen generiert werden. Das **-bb 10000** bedeutet, dass die Berechnung mit dem Ultrafastbootstrap durchgeführt wird, eine sehr viel schnellere Berechnung als der normale Bootstrap, jedoch auch etwas unsauberer. Aus diesem Grund werden beim Ultrafastbootstrap 10.000 Berechnungen hintereinander durchgeführt, im Gegensatz zum normalen Bootstrap mit nur 100-1.000 Berechnungen. So kann die gleiche Genauigkeit bei einer sehr viel schnelleren Berechnung gewährleistet werden. Der Bootstrap gibt die Wahrscheinlichkeit an, mit der sich die Baumstruktur in dieser Konstellation bei wiederholter Analyse wiederfindet. Der letzte Befehl **-nt 6** gibt an, auf wie vielen CPUs (Englisch: **central processing unit**, Hauptprozessor) des Computers die Berechnung durchgeführt wird. Diese Parallelisierung ermöglicht eine schnellere Berechnung.

Unter den Output-Dateien befindet sich das *.log*-File, in dem die Information über das verwendete best-fit Modell enthalten ist. Die ebenfalls generierte *.treefile*-Datei wurde anschließend mit dem Programm FigTree [49] zur weiteren Auswertung visualisiert. Die HIV-1 Gruppen N, O und P wurden als Außengruppe zum Wurzeln des Baumes definiert. [35, 50]

4 Ergebnisse

4.1 PCR

Von den insgesamt 45 zu untersuchenden Proben, sollte von 25 eine Integrase-Sequenz erzeugt werden. Für die restlichen 20 Proben erfolgte bereits vorher eine Sequenzierung. Die PCR wurde wie unter Punkt 3.2 beschrieben durchgeführt. Als Negativ-Kontrolle diente HIV-negatives Plasma und als Positiv-Kontrolle wurde der Laborreferenzstamm HTLVIIIB mit einer Konzentration von 10^4 Kopien/ml eingesetzt. Exemplarisch ist nachfolgend ein Agarosegel mit den Amplifikaten einer *int*-PCR dargestellt (Abbildung 14).

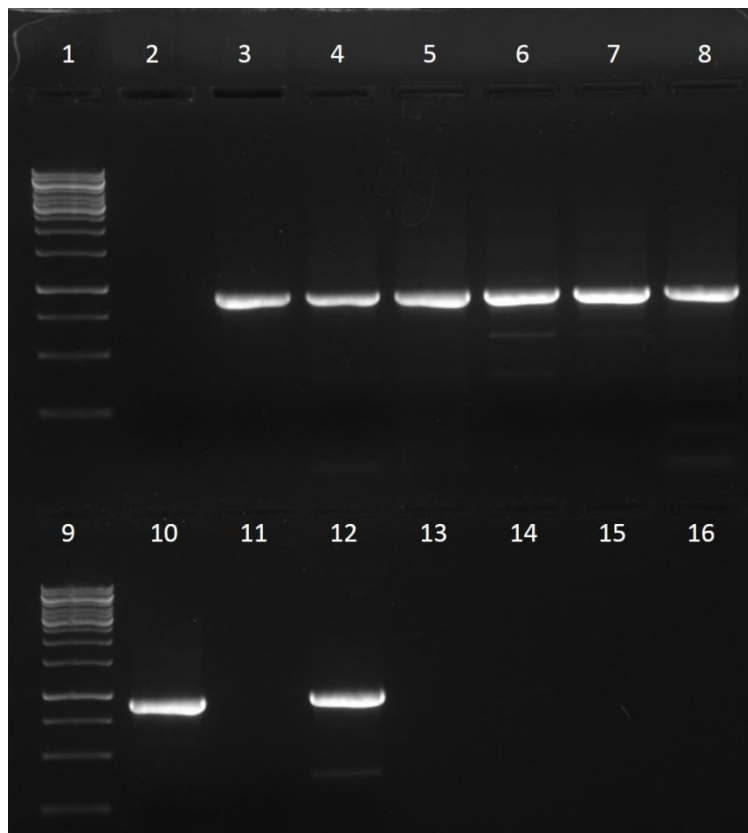


Abbildung 14: Gelbild zur Integrase-PCR 208; Spur (1/9) 1kBp DNA-Ladder, (2) negatives-Plasma (negativ-Kontrolle), (3-11, außer 9) Patientenproben, (12) negativ-Kontrolle PCR

Von den 25 in der *int*-PCR untersuchten Patientenproben waren alle bis auf die Probe 17-00462 (Abbildung 14, Spur 11) PCR-positiv und wiesen ein PCR-Produkt mit der richtigen Größe von ungefähr 900 bp (exakt 915 bp) auf. Für alle 45 Patientenproben wurde zur Untersuchung der *env*-Genomregion die OneStep *env*-PCR wie unter Punkt 3.3 beschrieben durchgeführt. Proben, die in dieser PCR negativ ausfielen, wurden nachfolgend noch einmal mit der nested *env*-PCR wie unter Punkt 3.4 beschrieben untersucht. Exemplarisch werden nachfolgend zwei Agarosegele (Abbildung 15) der OneStep und der nested *env*-PCR gezeigt.

Auf Spur 8 des linken Gelbildes (Abbildung 15) wurde das OneStep *env*-PCR-Produkt der Probe 07-01047 aufgetragen, jedoch ist keine DNA-Bande erkennbar. Aus diesem Grund wurde ausgehend von der gleichen RNA nachfolgend die nested *env*-PCR durchgeführt, deren PCR-Produkt auf dem rechten Gelbild auf Spur 7 (Abbildung 15) aufgetragen wurde. Im Gegensatz zur OneStep *env*-PCR, konnte durch die nested *env*-PCR ein Produkt mit der erwarteten Größe von ungefähr 600 bp (exakt 598 bp) amplifiziert werden.

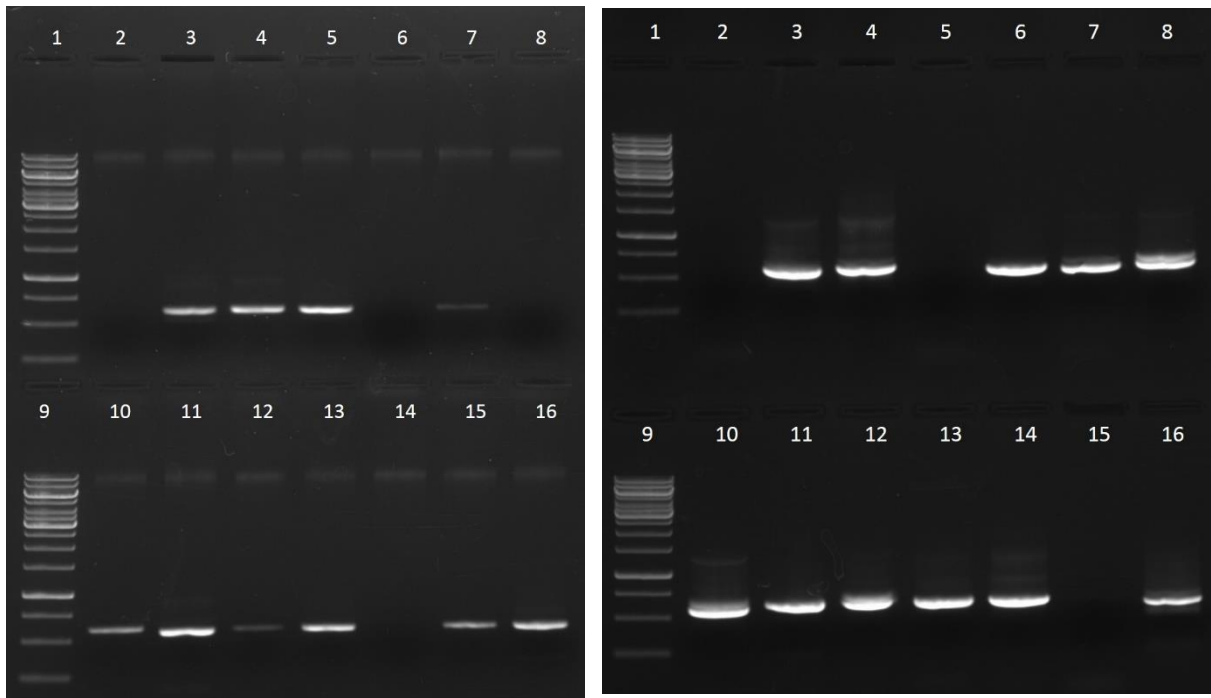


Abbildung 15: links OneStep env-PCR, Spur (1/9) 1 kbp DNA-Ladder, Spur (2-8/10-16) Patientenproben, Spur (8) Probe 07-01047; rechts nested env-PCR, Spur (1/9) 1 kbp DNA-Ladder, Spur (2-8/10-16) Patientenproben, Spur (7) Probe 07-01047

Die Erfolgsrate der OneStep env-PCR lag bei 33,3 % (15/45). Für 30 Patientenproben wurde nachfolgend die nested PCR durchgeführt, mit einer Erfolgsrate von 66,6 % (20/30) Insgesamt konnte für 77,7 % (35/45) der Proben ein positives PCR-Ergebnis für die env-PCR erzielt werden.

Nachfolgend sind in Tabelle 14 alle 45 Proben mit ihrer jeweiligen Probennummer und dem Ergebnis der PCR in den jeweiligen Genomregionen *int* und *env* zusammengefasst. In dieser Tabelle wurden die bereits vorhandenen (o) mit den in dieser Arbeit generierten Ergebnissen kombiniert.

Tabelle 14: Ergebnisse der *int*- und *env*-PCRs (nested *env*-PCRs farbig hinterlegt, +positiv, -negativ, o nicht untersucht)

Probennummer	PCR		Probennummer	PCR	
	<i>int</i>	<i>env</i>		<i>int</i>	<i>env</i>
04-00349	+	+	13-04599	o	+
04-00467	+	+	13-05607	+	+
05-00661	o	-	14-03863	+	+
06-00583	o	+	15-02070	o	-
07-01047	o	+	15-02182	+	+
08-00276	o	+	15-03135	+	+
08-01881	o	+	15-03907	+	+
08-03381	+	+	15-04545	+	+
09-02554	+	+	15-04673	+	+
09-03853	+	+	16-00021	o	+
10-00001	+	+	16-00141	o	+
10-01223	+	+	16-00396	o	-
11-00458	+	+	16-01071	o	-
11-00459	+	-	16-01388	o	+
11-04148	+	+	16-01574	o	-
11-04327	+	-	16-01746	+	+
12-02498	+	+	16-03750	o	+
12-02548	o	+	16-04243	o	-
12-02846	o	+	16-04283	o	+
12-04159	o	+	16-04286	o	-
12-04625	+	-	17-00461	+	+
13-02023	o	+	17-00462	-	+
13-03024	+	+			

4.2 Sequenzierung

Nach der gelelektrophoretischen Auswertung erfolgte die Aufreinigung der *int*- und *env*-PCR-Produkte wie unter Punkt 3.6 beschrieben mit dem Agencourt® AMPure® XP Kit. Anschließend wurde die DNA-Konzentration in ng/µl wie unter Punkt 3.7 beschrieben, durch das Quant-iT™ PicoGreen® dsDNA Assay Kit vermessen.

In der nachfolgenden Tabelle 15 ist der DNA-Gehalt in ng/µl für die jeweiligen PCR-Produkte (*int* oder *env*) aufgelistet.

Tabelle 15: Ergebnisse der DNA-Quantifizierung in ng/μl für die jeweilige Genomregion

Probennummer	<i>env</i>	<i>int</i>	Probennummer	<i>env</i>	<i>int</i>
04-00349	26,4	79,9	13-04599	91,8	-
04-00467	-2,7	62,2	13-05607	29,3	82,3
05-00661	-	-	14-03863	47,0	86,6
06-00583	98,9	-	15-02070	-	-
07-01047	96,8	-	15-02182	37,0	83,2
08-00276	99,2	-	15-03135	-1,6	92,9
08-01881	27,8	-	15-03907	0,7	83,9
08-03381	95,2	53,7	15-04545	40,9	83,8
09-02554	10,5	71,2	15-04673	-	73,6
09-03853	100,0	60,2	16-00021	49,9	-
10-00001	97,6	60,9	16-00141	97,6	-
10-01223	82,0	92,6	16-00396	-	-
11-00458	14,7	96,6	16-01071	-	-
11-00459	-	62,6	16-01388	97,4	-
11-04148	8,8	96,2	16-01574	-	-
11-04327	-3,9	20,7	16-01746	85,3	47,5
12-02498	52,5	96,7	16-03750	95,8	-
12-02548	88,5	-	16-04243	-	20,5
12-02846	19,9	-	16-04283	86,5	-
12-04159	28,4	-	16-04286	-	-
12-04625	-	77,5	17-00461	-2,2	80,9
13-02023	-0,1	-	17-00462	-3,3	-
13-03024	49,0	76,0			

Alle PCR-Produkte konnten in ausreichender Quantität für die Sanger-Sequenzierung generiert werden. Auch die sechs *env*-PCR-Produkte mit negativem DNA-Gehalt wurden in die Sequenzierung eingesetzt, indem die größtmögliche Menge von 4 μl DNA verwendet wurde. Insgesamt wurden somit 35 *env*-PCR-Produkte und 24 *int*-PCR-Produkte sequenziert. Die Sequenzierungen erfolgten wie unter Punkt 3.8 beschrieben.

4.3 Sequenzauswertung

Nachdem die Sequenzansätze im Sequenzierlabor mit Hilfe der Sanger-Methode sequenziert wurden, erfolgte die Auswertung der erhaltenen Sanger-Sequenzen mit SeqMan Pro und BioEdit wie in Punkt 3.9 beschrieben. Insgesamt konnten in dieser Arbeit 24 valide *int*- und 30 valide *env*-Sequenzen generiert werden.

Nachfolgend sind in Tabelle 16 alle 45 Proben mit ihrer jeweiligen Probennummer und dem Ergebnis der Sequenzierung in den jeweiligen Genomregionen *prrt*, *int* und *env* zusammengefasst. Negative oder nicht auswertbare Sequenzen sind mit einem Strich versehen. In dieser Tabelle wurden die bereits vorhandenen (o) mit den in dieser Arbeit generierten Ergebnissen kombiniert.

Tabelle 16: Ergebnisse der Sequenzierung aller zu untersuchenden Proben (+ valide Sequenz, - keine Sequenz, o Sequenz bereits vorhanden)

Probennummer	Sequenz		Probennummer	Sequenz	
	<i>int</i>	<i>env</i>		<i>int</i>	<i>env</i>
04-00349	+	+	13-04599	+	+
04-00467	+	-	13-05607	+	+
05-00661	+	-	14-03863	+	+
06-00583	+	+	15-02070	+	-
07-01047	+	+	15-02182	+	+
08-00276	+	+	15-03135	+	+
08-01881	+	+	15-03907	+	-
08-03381	+	+	15-04545	+	+
09-02554	+	+	15-04673	+	-
09-03853	+	+	16-00021	+	+
10-00001	+	+	16-00141	+	+
10-01223	+	-	16-00396	+	-
11-00458	+	+	16-01071	+	-
11-00459	+	-	16-01388	+	+
11-04148	+	+	16-01574	+	-
11-04327	+	-	16-01746	+	+
12-02498	+	+	16-03750	+	+
12-02548	+	+	16-04243	+	-
12-02846	+	+	16-04283	+	+
12-04159	+	+	16-04286	+	-
12-04625	+	-	17-00461	+	+
13-02023	+	-	17-00462	-	+
13-03024	+	+			

4.4 Subtypbestimmung mit Webtools

Für die Subtypbestimmung wurden auch die 45 bereits vorhandenen *prrt*-Sequenzen und die 20 bereits vorhanden *int*-Sequenzen mit einbezogen. Die Bestimmung des Subtyps erfolgte abhängig von der Genomregion mit verschiedenen Webtools. So wurden für *env*-Sequenzen die Tools Geno2Pheno (G2P), REGA und Comet angewendet. Für *prrt*- und *int*-Sequenzen wurden die Tools REGA und Stanford verwendet. In der folgenden Tabelle 17 sind die Ergebnisse. Fehlende Sequenzen sind mit einem Strich (-) versehen.

Tabelle 17: Ergebnisse der Subtypbestimmung mittels verschiedener Webtools in der jeweiligen Genomregion

Probennummer	<i>prrt</i>		<i>int</i>		<i>env</i>		
	REGA	Stanford	REGA	Stanford	G2P	REGA	Comet
04-00349	02_AG	02_AG	G	02_AG	B	NA (Cluster A1)	A1 (02_AG)
04-00467	02_AG	02_AG	G	02_AG	-	-	-
05-00661	02_AG	02_AG	G	02_AG	-	-	-
06-00583	D	D	D	D	D	D	D
07-01047	12_BF	12_BF	F (F1)	F	B	F (F1)	F1 (12_BF)

Probennummer	<i>prrt</i>		<i>int</i>		<i>env</i>		
	REGA	Stanford	REGA	Stanford	G2P	REGA	Comet
08-00276	02_AG	02_AG	G	02_AG	A/AG	A (A1)	A1
08-01881	B, D	B	B	B	A/AG NA (Cluster A1)	A1 (02_AG)	
08-03381	02_AG	63_02A1	A (A1)	63_02A1	A/AG	A (A1)	A1 (63_02A1)
09-02554	02_AG	02_AG	G	02_AG	C	A (A1)	A1 (02_AG)
09-03853	02_AG	02_AG	G	02_AG	A/AG NA (Cluster A1)	02_AG	
10-00001	02_AG	02_AG	G (02_AG)	02_AG	C	A (A1)	A1 (02_AG)
10-01223	02_AG	02_AG	G	63_02A1	-	-	-
11-00458	02_AG	02_AG	G	02_AG	A/AG	A (A1)	A1
11-00459	02_AG	02_AG	G (02_AG)	02_AG	-	-	-
11-04148	02_AG	02_AG	G	02_AG	A/AG	A (A1)	A1 (02_AG)
11-04327	02_AG	02_AG	G	02_AG	-	-	-
12-02498	02_AG	02_AG	A1, G	A	A/AG	A (A1)	A1 (02_AG)
12-02548	C	C	C (31_BC)	C	A/AG	B	B
12-02846	02_AG	02_AG	G	02_AG	G	NA (Cluster J)	G (06_cpx)
12-04159	11_cpx	11_cpx	A (11_cpx)	A	B	B	B
12-04625	02_AG	02_AG	G	02_AG	-	-	-
13-02023	F (F1)	F	F1, B	F	-	-	-
13-03024	G (02_AG)	02_AG	G	02_AG	A/AG	A (A1)	A1 (02_AG)
13-04599	G (02_AG)	02_AG	G	02_AG	B	B	B
13-05607	02_AG	02_AG	G	02_AG	AE	NA (Cluster C)	01_AE
14-03863	02_AG	02_AG	G	02_AG	B	A (A1)	A1 (02_AG)
15-02070	B, F1	B	B	B	-	-	-
15-02182	02_AG	02_AG	G	02_AG	C	A (A1)	A1 (02_AG)
15-03135	06_cpx	06_cpx	G, 06_cpx	06_cpx	C	C	C
15-03907	02_AG	02_AG	G	09_cpx	-	-	-
15-04545	02_AG	02_AG	G	02_AG	B	A (A1)	A1 (02_AG)
15-04673	C, B	C	C	C	-	-	-
16-00021	D (19_cpx)	D	19_cpx	19_cpx	A/AG NA (Cluster A1)	19_cpx	
16-00141	B	B	B	B	B	B	B
16-00396	B, F1	B	B	B	-	-	-
16-01071	B	B	B	B	-	-	-
16-01388	-	02_AG	G	A, 02_AG	AE	NA (Cluster F1)	AE
16-01574	02_AG-like	02_AG	G	09_cpx	-	-	-
16-01746	F1, A1, K	45_cpx	A (A1)	A	G	NA (Cluster A1)	A1
16-03750	B	B	B	B	C	C	C
16-04243	B, D	B	B	B	-	-	-
16-04283	-	02_AG	G	B	AE	NA (Cluster C)	01_AE
16-04286	02_AG	02_AG	G	02_AG	-	-	-
17-00461	02_AG	02_AG	G	A, 02_AG	B	A (A1)	A1 (02_AG)
17-00462	02_AG	02_AG	-	-	B	A (A1)	A1 (02_AG)

Nachfolgend (Abbildung 16) sind grafisch zusammengefasst die Übereinstimmungsraten der Subtypbestimmung der verschiedenen Webtools dargestellt, wobei eine tendenzielle Übereinstimmung auf Sequenzauswertungen mit Subtypen enger Verwandtschaft wie zum Beispiel A und 02_AG, oder B, F und 12_BF darstellt.

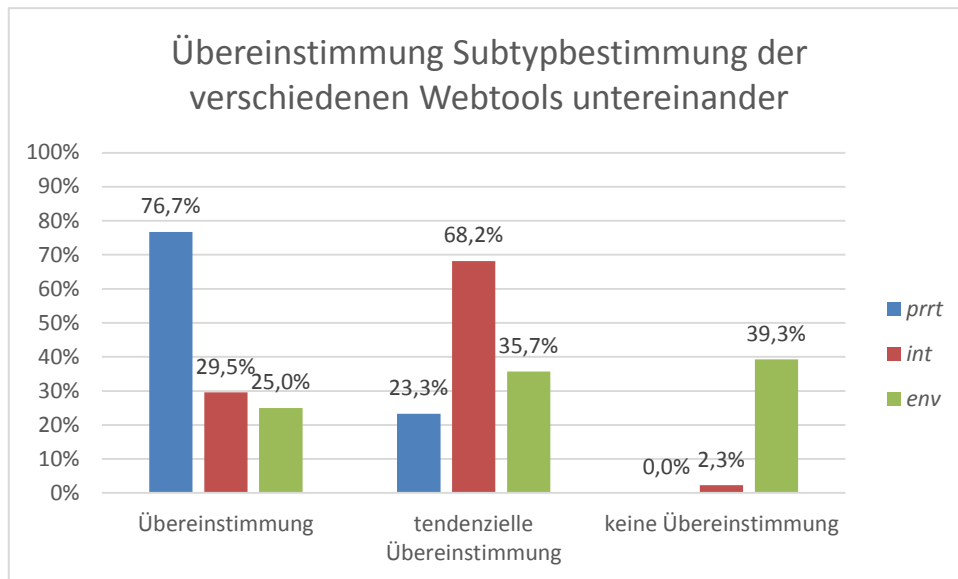


Abbildung 16: Übereinstimmungsrate der Subtypbestimmung der verschiedenen Webtools untereinander

Bei insgesamt 7 (25 %; 7/28) *env*-Sequenzen zeigte sich eine Übereinstimmung zwischen den Tools G2P, REGA und Comet. Für die *int*-Sequenzen konnte zwischen REGA und Stanford für 13 Proben (29,5 %; 13/44) eine Übereinstimmung beobachtet werden. Die größte Übereinstimmung zwischen den Webtool-Bestimmungen wiesen die *prrt*-Sequenzen mit 76,7 % (33/43) auf.

Die größten Diskrepanzen zwischen den Webtools zeigten sich für die *env*-Sequenzen. Hier wurden 39,3 % (11/28) der Sequenzen von den drei Algorithmen unterschiedlich bestimmt und zeigten keinerlei Übereinstimmung. Unter den 25 Proben, für welche *prrt*-, *int*- und *env*-Sequenzen vorlagen, wurden zwei Proben (06-00583, 16-00141) durch REGA in allen drei Bereichen eindeutig und einheitlich klassifiziert. Bei den anderen 23 Proben würde man aufgrund der unterschiedlichen Klassifikation in den drei Bereichen davon ausgehen, dass es sich um URFs handelt.

4.5 Maximum-Likelihood-Phylogenie

Zur Subtypbestimmung mittels einer eigens durchgeführten Maximum-Likelihood-Phylogenie wurden wie unter Punkt 3.10 beschrieben Stammbäume für die einzelnen Sequenzalignments (*prrt*, *int*, *env*) sowie für das *prrt-int-env*-Alignment (Merge) erstellt. Hierbei wurden zwei *prrt*-Sequenzen wegen zu kurzer Sequenzlänge (16-04283) beziehungsweise zu schlechter Sequenzqualität aufgrund eines zu großen Anteils an Ambiguitäten (16-01388) von der Analyse ausgeschlossen. Zudem zeigte sich, dass die *env*-Sequenzen aber nicht die *prrt*- und *int*-Sequenzen der Proben 09-02554 und 10-00001 identisch waren. Dies weist auf eine Probenverwechslung bei der *env*-PCR oder *env*-Sequenzierung hin. Die *env*-Sequenzen der Proben 09-02554 und 10-00001 wurden daher aus der Analyse ausgeschlossen. Somit wurden für den *prrt*-Stammbaum 43 Sequenzen, für den *int*-Stammbaum 44 Sequenzen und für den *env*-Stammbaum 27 Sequenzen eingesetzt werden. Für den Merge-Stammbaum (*prrt*, *int* und *env*) wurde die kleinste gemeinsame Schnittmenge von 25 Sequenzen verwendet.

Die erstellte *.log*-Datei für jeden Stammbaum zeigte, dass unterschiedliche Modelle für die Erstellung genutzt wurden. Für den *prrt*- und den *int*-Baum wurde das GTR+I+G4 Modell, für den *env*-Baum das TIM+I+G4 Modell genutzt und für den Merge-Baum erwies sich ebenfalls das GTR+I+G4 als das am besten geeignete Modell. Nach der Berechnung der Bäume wurden diese grafisch mit FigTree visualisiert und soweit möglich, der Subtyp anhand der Referenzsequenzen und dem Bootstrapwert bestimmt.

Die visualisierten phylogenetischen Bäume befinden sich in Gesamtheit im Anhang (Abbildung 23 bis Abbildung 26) und werden in den nächsten Abbildung 17 bis Abbildung 18 exemplarisch dargestellt, um die Subtypbestimmung nachvollziehen zu können.

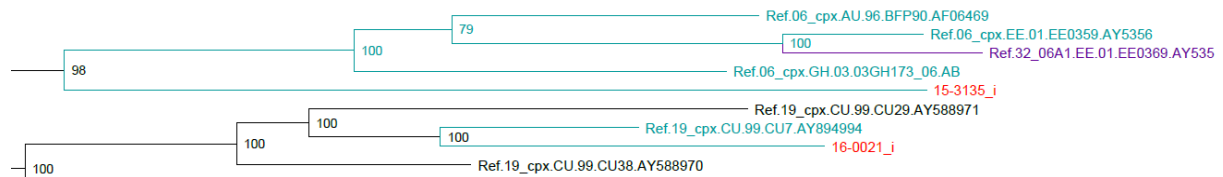


Abbildung 17: Ausschnitt des *int*-Baumes zur Darlegung der Subtypbestimmung

Die rot markierten Taxa (Abbildung 17) sind die untersuchten Patientenproben, in türkis dargestellt eine Klade mit Referenzsequenzen und, in anderen Farben dargestellt, verschiedenen Subtypen. Die Zahlen rechts neben den Knoten stellen den Bootstrapwert dar. Für die Auswertung wurde ein Bootstrap von mindestens 95 % gewählt, wodurch teilweise sehr große Kladen entstanden sind, in denen viele verschiedene Subtypen vorkommen (Abbildung 18, erste Probe 08-03381_e). So kann die Probe mit dem Scount 16-00021_i (Abbildung 17) eindeutig als CRF19_cpx identifiziert werden, wohingegen die Probe 15-03135_i einer Klade angehört, die die Subtypen CRF06_cpx und CRF32_06A1 beinhaltet. Für die oberste Probe mit dem Scount 08-03381_e (Abbildung 18) ist in der *env*-Genomregion daher keine eindeutige Zuordnung zu einem Subtyp möglich, da die Probe mit einer großen weiteren Klade clustert, in der verschiedene weitere Subtypen angesiedelt sind (CRF45_cpx, A1, CRF50_A1D, CRF22_01A1, CRF36_cpx, AD, CRF11_cpx, CRF13_cpx, CRF02_AG, CRF56_cpx, CRF19_cpx, CRF37_cpx und CRF25_cpx). Somit sind für die Proben in Tabelle 18 keine eindeutigen Aussagen zu treffen, wenn mehr als ein Subtyp eingetragen wurde, sowohl für die einzelnen Sequenzbereiche *prnt*, *int* und *env*, als auch für die Gesamtsequenzinformation.

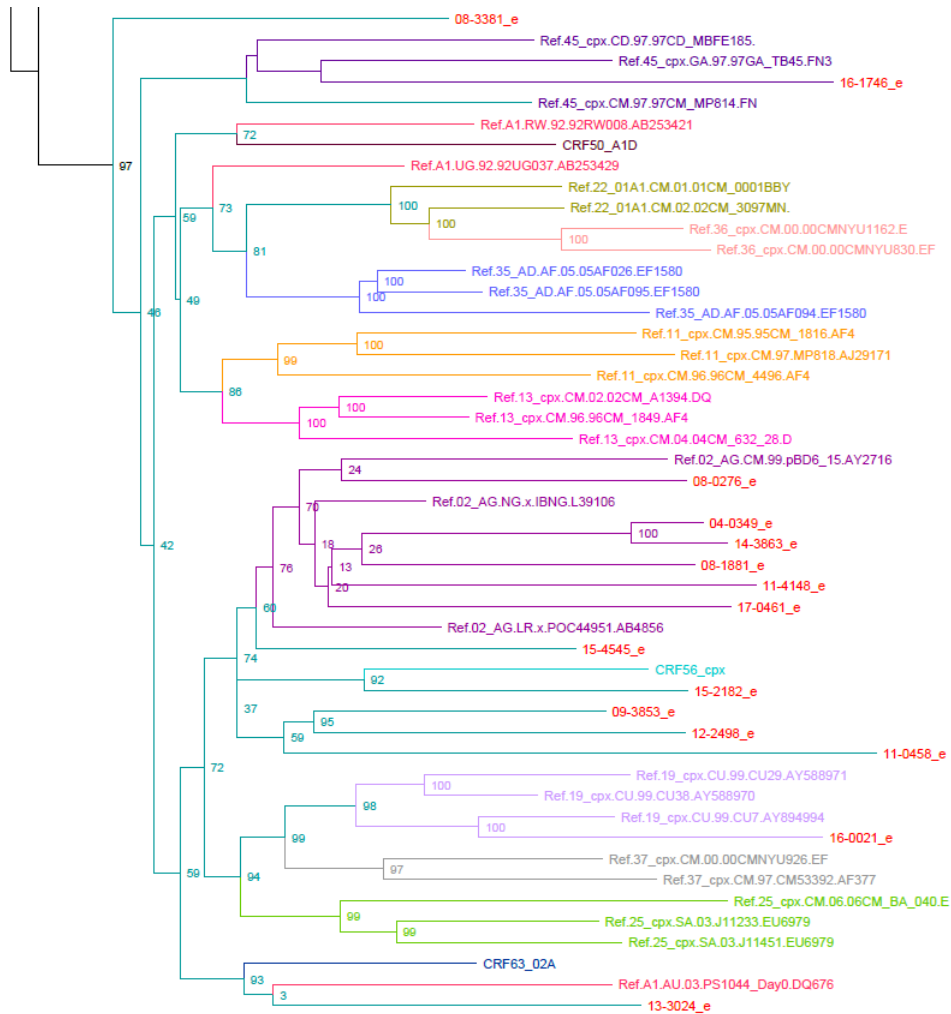


Abbildung 18: Ausschnitt des env-Baumes mit großen Kladen und ihnen einzeln untergeordneter Subtypen

Die Ergebnisse der Klassifizierung aller 45 Proben anhand ihrer verfügbaren Sequenzen sind in Tabelle 18 zusammengefasst.

Tabelle 18: Ergebnisse der Subtypbestimmung durch die eigens erstellten phylogenetischen Bäume, Fehlende Sequenzen sind mit einem Strich (-) markiert

Probennummer	Subtyp einzeln FigTree			Subtyp Merge FigTree	Subtyp Merge REGA
	<i>prt</i>	<i>int</i>	<i>env</i>		
04-00349	02_AG	02_AG	02_AG	02_AG	02_AG, G
04-00467	02_AG	02_AG	-	-	-
05-00661	02_AG	02_AG	-	-	-
06-00583	D	D	D	D	D
07-01047	12_BF	12_BF	12_BF	12_BF	12_BF, F1, B
08-00276	02_AG	02_AG	02_AG	02_AG	02_AG, G
08-01881	B, 01B, 83_cpx	B, 01B, BF	02_AG	B, BF, 03_AB	B, A1, D
08-03381	02_AG	63_02AG	45_cpx, A1, 50_A1D, 22_01A1, 36_cpx, AD, 11_cpx, 13_cpx, 02_AG, 56_cpx, 19_cpx, 37_cpx, 25_cpx	02_AG, 36_cpx	A1, 02_AG, G

Probennummer	Subtyp einzeln FigTree			Subtyp Merge FigTree	Subtyp Merge REGA
	<i>prrt</i>	<i>int</i>	<i>env</i>		
09-02554	02_AG	02_AG	Sequenz gleich mit 10-0001	-	-
09-03853	02_AG	02_AG	02_AG, 56_cpx	02_AG	02_AG, D, G
10-00001	02_AG	02_AG	Sequenz gleich mit 09-2554	-	-
10-01223	02_AG	02_AG	-	-	-
11-00458	36_cpx	02_AG	02_AG, 56_cpx	02_AG, 36_cpx	02_AG, F1, G
11-00459	36_cpx	02_AG	-	-	-
11-04148	02_AG	02_AG	02_AG	02_AG	02_AG, G
11-04327	02_AG	02_AG	-	-	-
12-02498	02_AG	02_AG, 09_cpx, 13_cpx, 25_cpx, 36_cpx	02_AG, 56_cpx	02_AG	02_AG, A1, G
12-02548	BC	BC, C	B, BF, BG, 15_01B, 03_AB	C, BC	31_BC, C
12-02846	02_AG	02_AG	06_cpx, 32_06A1	G, BG, 43_02G, 06_cpx	02_AG, A1, G
12-04159	11_cpx	11_cpx	B, BF, BG, 15_01B, 03_AB	11_cpx	11_cpx, A1, G
12-04625	02_AG	02_AG	-	-	-
13-02023	F1	F1	-	-	-
13-03024	02_AG	02_AG	63_02A	02_AG	02_AG, G, H
13-04599	02_AG	02_AG	BF	02_AG	02_AG, G
13-05607	02_AG	02_AG	AE, 01B, 78_cpx	02_AG	02_AG, G
14-03863	02_AG	02_AG	02_AG	02_AG	02_AG, G
15-02070	29_BF	B, 01B, BF	-	-	-
15-02182	02_AG	02_AG	56_cpx	02_AG	02_AG, G
15-03135	06_cpx, 32_06A1	06_cpx, 32_06A1	eher C, BC	C, BC, 87_cpx, 65_cpx, 83_cpx	06_cpx, G
15-03907	02_AG	09_cpx	-	-	-
15-04545	02_AG	02_AG	02_AG	02_AG	02_AG, G
15-04673	BC, C, 65_cpx, 87_cpx	BC, C, 49_cpx, 83_cpx	-	-	-
16-00021	19_cpx	19_cpx	19_cpx	19_cpx	19_cpx, H, G, A1
16-00141	B, BF	B, 01B, BF	B, BF, BG, 15_01B	B	B
16-00396	29_BF	B, 01B, BF	-	-	-
16-01071	B, BF	B, 01B, BF	-	-	-
16-01388	zu viele Ambiguitäten	02_AG	AE, 01B, 78_cpx	-	-
16-01574	02_AG	09_cpx	-	-	-
16-01746	45_cpx, 49_cpx	45_cpx	45_cpx	45_cpx	A1, 06_cpx, F1, G
16-03750	B, BF	51_01B	C, BC	B, BF, 03_AB, 01_B	B
16-04243	B, 01B, 83_cpx	B, 01B, BF	-	-	-
16-04283	zu kurz am 3'-Ende	02_AG	AE, 01B, 78_cpx	-	-

Proben- nummer	Subtyp einzeln FigTree			Subtyp Merge FigTree	Subtyp Merge REGA
	<i>prrt</i>	<i>int</i>	<i>env</i>		
16-04286	02_AG	02_AG	-	-	-
17-00461	02_AG	02_AG	02_AG	02_AG	02_AG, G, A1
17-00462	02_AG	-	-	-	-

Unter den 25 Proben, für die *prrt*, *int*- und *env*-Sequenzen vorlagen (farbig hinterlegt), konnten neun Proben in allen drei Bereichen eindeutig und einheitlich klassifiziert werden. Bei weiteren neun Proben konnte bei nicht eindeutiger Zuordnung der Einzelsequenzen noch eine eindeutige Zuordnung anhand der Gesamtsequenzinformation (Merge) erfolgen. Bei den verbleibenden sieben Proben ist aufgrund der vorliegenden Sequenzinformationen davon auszugehen, dass es sich um URFs handelt.

Unter den 20 Proben, für die nicht alle drei Sequenzbereiche vorlagen, konnte bei acht Proben innerhalb der *pol*-Genomregion eine einheitliche Klassifizierung anhand von *prrt* und *int* beobachtet werden. Alle anderen Proben wurden anhand der Einzelsequenzen unterschiedlich klassifiziert oder waren nicht eindeutig klassifizierbar. Leider wurden die *env*-Sequenzen der Proben 09-02554 und 17-00462 nicht für die phylogenetischen Analysen ins Alignment eingefügt, sodass keine Informationen bezüglich des Subtyps vorliegen und insgesamt nur 27 der eingesetzten 30 *env*-Proben (Probenverwechslung 09-02554 und 10-00001) untersucht wurden.

4.6 Vergleich REGA und Maximum-Likelihood-Phylogenie

Nachfolgend wurden die Ergebnisse der Subtypisierung mittels dem REGA HIV-1 Subtyping Tool und der eigenen Maximum-Likelihood-Phylogenie verglichen. Hierzu wurde das REGA Tool auch auf die Gesamtsequenzinformation (Merge) angewendet. In Tabelle 19 sind die Ergebnisse von REGA und der eigenen Phylogenie für die Einzelsequenzen und die Gesamtsequenz gegenübergestellt.

Tabelle 19: Gegenüberstellung der Ergebnisse von REGA und der eigenen Phylogenie für die Einzelsequenzen und die Gesamtsequenz

Proben- nummer	<i>prrt</i>		<i>int</i>		<i>env</i>		Merge	
	REGA	FigTree	REGA	FigTree	REGA	FigTree	REGA	FigTree
04- 00349	02_AG	02_AG	G	02_AG	NA (Cluster A1)	02_AG	02_AG, G	02_AG
04- 00467	02_AG	02_AG	G	02_AG	-	-	-	-
05- 00661	02_AG	02_AG	G	02_AG	-	-	-	-
06- 00583	D	D	D	D	D	D	D	D
07- 01047	12_BF	12_BF	F (F1)	12_BF	F (F1)	12_BF	12_BF, F1, B	12_BF
08- 00276	02_AG	02_AG	G	02_AG	A (A1)	02_AG	02_AG, G	02_AG
08- 01881	B, D	B, 01B, 83_cpx	B	B, 01B, BF	NA (Cluster A1)	02_AG	B, A1, D	B, BF, 03_AB
08- 03381	02_AG	02_AG	A (A1)	63_02AG	A (A1)	Siehe Tabelle 19	A1, 02_AG, G	02_AG, 36_cpx

Proben- nummer	<i>prrt</i>		<i>int</i>		<i>env</i>		Merge	
	REGA	FigTree	REGA	FigTree	REGA	FigTree	REGA	FigTree
09- 02554	02_AG	02_AG	G	02_AG	A (A1)	Sequenz gleich mit 10-0001	-	-
09- 03853	02_AG	02_AG	G	02_AG	NA (Cluster A1)	02_AG, 56_cpx	02_AG, D, G	02_AG
10- 00001	02_AG	02_AG	G (02_AG)	02_AG	A (A1)	Sequenz gleich mit 09-2554	-	-
10- 01223	02_AG	02_AG	G	02_AG	-	-	-	-
11- 00458	02_AG	36_cpx	G	02_AG	A (A1)	02_AG, 56_cpx	02_AG, F1, G	02_AG, 36_cpx
11- 00459	02_AG	36_cpx	G (02_AG)	02_AG	-	-	-	-
11- 04148	02_AG	02_AG	G	02_AG	A (A1)	02_AG	02_AG, G	02_AG
11- 04327	02_AG	02_AG	G	02_AG	-	-	-	-
12- 02498	02_AG	02_AG	A1, G	02_AG, 09_cpx, 13_cpx, 25_cpx, 36_cpx	A (A1)	02_AG, 56_cpx	02_AG, A1, G	02_AG
12- 02548	C	BC	C (31_BC)	BC, C	B	B, BF, BG, 15_01B, 03_AB	31_BC, C	C, BC
12- 02846	02_AG	02_AG	G	02_AG	NA (Cluster J)	06_cpx, 32_06A1	02_AG, A1, G	G, BG, 43_02G, 06_cpx
12- 04159	11_cpx	11_cpx	A (11_cpx)	11_cpx	B	B, BF, BG, 15_01B, 03_AB	11_cpx, A1, G	11_cpx
12- 04625	02_AG	02_AG	G	02_AG	-	-	-	-
13- 02023	F (F1)	F1	F1, B	F1	-	-	-	-
13- 03024	G (02_AG)	02_AG	G	02_AG	A (A1)	63_02A	02_AG, G, H	02_AG
13- 04599	G (02_AG)	02_AG	G	02_AG	B	BF	02_AG, G	02_AG
13- 05607	02_AG	02_AG	G	02_AG	NA (Cluster C)	AE, 01B, 78_cpx	02_AG, G	02_AG
14- 03863	02_AG	02_AG	G	02_AG	A (A1)	02_AG	02_AG, G	02_AG
15- 02070	B, F1	29_BF	B	B, 01B, BF	-	-	-	-

Proben- nummer	<i>prrt</i>		<i>int</i>		<i>env</i>		Merge	
	REGA	FigTree	REGA	FigTree	REGA	FigTree	REGA	FigTree
15- 02182	02_AG	02_AG	G	02_AG	A (A1)	56_cpx	02_AG, G	02_AG
15- 03135	06_cpx	06_cpx, 32_06A1	G, 06_cpx	06_cpx, 32_06A1	C	eher C, BC	06_cpx, G	C, BC, 87_cpx, 65_cpx, 83_cpx
15- 03907	02_AG	02_AG	G	09_cpx	-	-	-	-
15- 04545	02_AG	02_AG	G	02_AG	A (A1)	02_AG	02_AG, G	02_AG
15- 04673	C, B	BC, C, 65_cpx, 87_cpx	C	BC, C, 49_cpx, 83_cpx	-	-	-	-
16- 00021	D (19_cpx)	19_cpx	19_cpx	19_cpx	NA (Cluster A1)	19_cpx	19_cpx, H, G, A1	19_cpx
16- 00141	B	B, BF	B	B, 01B, BF	B	B, BF, BG, 15_01B	B	B
16- 00396	B, F1	29_BF	B	B, 01B, BF	-	-	-	-
16- 01071	B	B, BF	B	B, 01B, BF	-	-	-	-
16- 01388	-	zu viele Ambiguitäten	G	02_AG	NA (Cluster F1)	AE, 01B, 78_cpx	-	-
16- 01574	02_AG-like	02_AG	G	09_cpx	-	-	-	-
16- 01746	F1, A1, K	45_cpx, 49_cpx	A (A1)	45_cpx	NA (Cluster A1)	45_cpx	A1, 06_cpx, F1, G	45_cpx
16- 03750	B	B, BF	B	51_01B	C	C, BC	B	B, BF, 03_AB, 01_B
16- 04243	B, D	B, 01B, 83_cpx	B	B, 01B, BF	-	-	-	-
16- 04283	-	zu kurz am 3'-Ende	G	02_AG	NA (Cluster C)	AE, 01B, 78_cpx	-	-
16- 04286	02_AG	02_AG	G	02_AG	-	-	-	-
17- 00461	02_AG	02_AG	G	02_AG	A (A1)	02_AG	02_AG, G, A1	02_AG
17- 00462	02_AG	02_AG	-	-	A (A1)	-	-	-

In Abbildung 19 sind die Übereinstimmungsraten der Subtypbestimmung mit den eigens in FigTree erstellten phylogenetischen Bäumen im Vergleich mit dem Ergebnis des Tools REGA dargestellt. Eine tendenzielle Übereinstimmung steht für eine Ähnlichkeit.

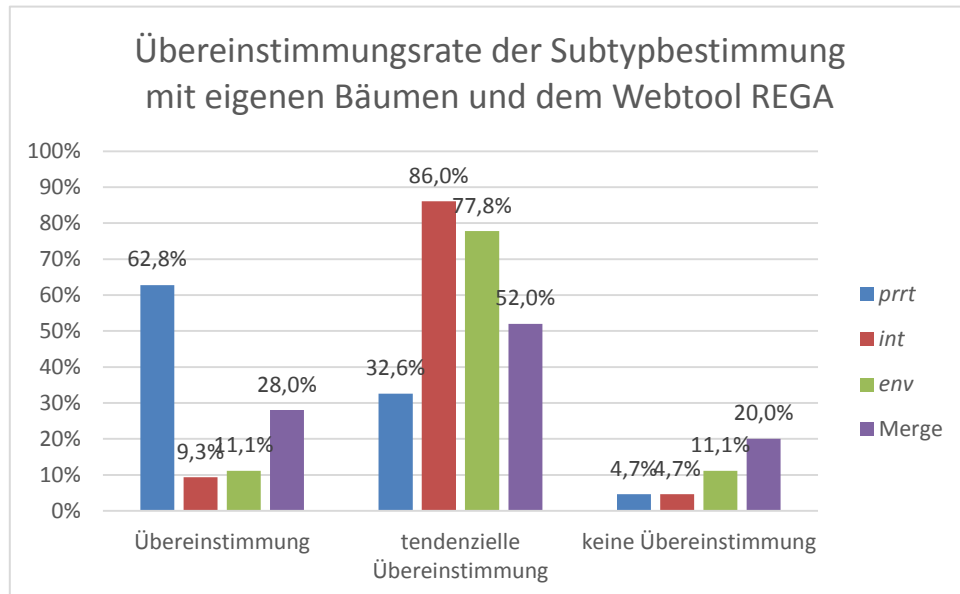


Abbildung 19: Übereinstimmungsrate der Subtypbestimmung mit den eigens erstellten phylogenetischen Bäumen und dem Subtyping Tool REGA

Der Vergleich der Ergebnisse mit dem HIV-1 Subtyping Tool REGA und den eigens erstellten Stammbäumen erzielte eine Übereinstimmung von 62,8 % (33/43) für die *prrt*-Sequenzen. Für die Integrase-Sequenzen wies der Großteil eine tendenzielle Übereinstimmung von 67,4 % (30/44) auf.

Auch für die *env*-Sequenzen fanden sich vorwiegend tendenzielle Übereinstimmungen (77,8 %; 21/27) der Ergebnisse der Webtool-Subtypbestimmung und den eigens erstellten Stammbäumen. Beim Vergleich der Gesamtsequenzinformation aus *prrt*, *int* und *env* konnte über die Hälfte der Proben (52 %; 13/25) ähnlich klassifiziert werden.

5 Diskussion

Ziel dieser Arbeit war die nähere Charakterisierung von HIV-Varianten durch die Analyse mehrerer Genombereiche. Hierdurch können rekombinante Formen erkannt und besser klassifiziert werden. Ausgewählt wurden 45 Proben, bei denen anhand der *prrt*- und gegebenenfalls *int*-Sequenz keine eindeutige Klassifizierung vorgenommen werden konnte.

Hierfür wurde virale RNA aus Plasmaproben von HIV-positiven Patienten gewonnen und mittels PCR verschiedene Genombereiche von HIV-1 amplifiziert. Grundsätzlich kann davon ausgegangen werden, dass die RNA-Extraktionen erfolgreich waren, da bei jedem durchgeführten Protokoll die mitgeführte Positivkontrolle HTLVIIIB in den verschiedenen PCRs ein positives Signal zeigte.

Für die Amplifikation der einzelnen Genombereiche wurden verschiedene RT-PCRs durchgeführt mit spezifischen Primern und als OneStep, TwoStep oder nested PCR. Alle Primer wurden anhand von Referenzsequenzen HIV-1 Gruppe M-generisch entworfen, das heißt sie wurden in konservierten Bereichen lokalisiert und mit Wobble-Basen versehen, um möglichst alle bekannten Varianten der HIV-1 Gruppe M erfassen zu können. Allerdings besitzen diese PCR-Systeme für die Erfassung von unterschiedlichen Subtypen auch eine unterschiedliche Sensitivität.

Die PCRs verliefen zum Großteil positiv. Die Integrase-PCR wurde nach dem zuvor in der Praxisphase optimierten Protokoll durchgeführt und lieferte spezifische PCR-Produkte in ausreichender Menge mit einer Erfolgsrate von 96,0 %. Von den 25 untersuchten Proben konnte für nur eine Probe (Probennummer 17-00462) kein PCR-Produkt generiert werden. Der HIV-RNA-Gehalt in dieser Probe lag mit 27.648 Kopien/ml jedoch über der anhand von Referenzviren experimentell ermittelten subtypgenerischen Nachweisgrenze der *int*-PCR von 10.000 Kopien/ml, sodass die PCR durchaus positiv ausfallen hätte müssen. Womöglich weist das vorliegende Virus Mutationen im Bereich der Hybridisierungsstellen der *int*-Primer auf. Diese Probe muss dahingehend mit einem anderen PCR-System, das diese Primerregionen umspannt, näher untersucht werden.

Aufgrund der hohen Variabilität im Hüllprotein und damit potentiell verbundenen Mutationen in den Primerregionen, lag die Erfolgsrate der OneStep *env*-PCR nur bei 33,3 % (15/45). Die 30 PCR-negativen Proben wurden nachfolgend mit der nested *env*-PCR untersucht. Die nested PCR ist eine PCR-Methode bei der zuerst mit dem primären Primerpaar ein etwas größerer Bereich als der eigentlich Gewünschte amplifiziert wird, um die Sensitivität und Spezifität zu steigern. Dieses PCR-Produkt wird dann noch einmal mit den sekundären Primern amplifiziert, welche intern im zuvor amplifizierten PCR-Produkt hybridisieren. [51] Mit der nested *env*-PCR konnten weitere 22 Proben amplifiziert werden, sodass die Gesamterfolgsrate für die Amplifikation von *env* bei 77,7 % (35/45) lag. Bei vier der zehn PCR-negativen Proben lag die Viruslast unter oder nahe der anhand von Referenzviren experimentell ermittelten subtypgenerischen Nachweisgrenze der nested *env*-PCR von 10.000 Kopien/ml, sodass hier nicht zwangsläufig ein PCR-Produkt zu erwarten war. Für fünf Proben ist die Viruslast nicht bekannt und muss für eine weitere Bewertung erst bestimmt werden. Die verbleibende Probe (Probennummer 16-00396) hat einen HIV-RNA-Gehalt von 29.862 Kopien/ml und sollte daher analog der Probe 17-00462, für die die Integrase nicht amplifiziert werden konnte, mit anderen PCR-Systemen erneut untersucht werden, um mögliche Mutationen im Bereich der Primerbindestellen zu identifizieren. Die in dieser Arbeit untersuchten URFs stellen dahingehend eine besondere Herausforderung dar, da deren Genomsequenz zum Zeitpunkt des Primerdesigns unbekannt waren. Nur durch regelmäßiges Optimieren der PCR-Systeme durch Überprüfen der Passgenauigkeit der Primer auf die aktuell zirkulierenden Virusvarianten, kann die Erfassung neuer rekombinanter Formen gewährleistet werden.

Die Aufreinigung der erhaltenen PCR-Produkte und die darauffolgende Bestimmung des DNA-Gehalts erbrachte teilweise negative Ergebnisse bei den *env*-PCR-Produkten. Nichtsdestotrotz wurden diese Proben unter Einsatz des höchstmöglichen DNA-Volumens sequenziert und es resultierten

auswertbare Sequenzen, was auf eine Unterschätzung des DNA-Gehalts durch die PicoGreen®-Messung hindeutet. Für die Sequenzierung und Sequenzauswertung ist auch die Qualität des PCR-Produkts von großer Bedeutung. Unspezifische Nebenprodukte aus der PCR können zu Rauschen in der Sequenz führen und die Bewertung von Ambiguitäten in der Sequenzauswertung erschweren. Aus diesem Grund werden immer mehrere Sequenzen in beide Orientierungen pro PCR-Fragment generiert. Insgesamt konnten ausgehend von den positiven PCR-Produkten 86,0 % (30/35) valide *env*-Sequenzen und 100 % (24/24) valide *int*-Sequenzen bestimmt werden.

Anhand der Einzelsequenzen *prrt*, *int* und *env* wurde für jede Probe versucht den Subtyp mittels verschiedener webbasierter Subtypisierungs-Tools zu bestimmen, die unterschiedliche Algorithmen und Referenzsequenzen nutzen. Hierbei zeigten sich deutliche Unterschiede zwischen den Tools, insbesondere bei der Klassifizierung anhand der variablen *env*-Sequenzen. Somit lässt sich sagen, dass eine Subtypbestimmung mit Webtools für die *env*-Region nicht besonders aussagekräftig und verlässlich ist. Bei der Klassifizierung anhand der stärker konservierten *pol*-Genomregion (*prrt* und *int*) stimmten die Ergebnisse zwischen den Webtools zumindest tendenziell überein. Ein Grund hierfür kann sein, dass aufgrund eines hohen Anteils an Ambiguitäten in den kurzen *env*-Sequenzen im Vergleich zu den längeren und konservierteren *prrt*- und *int*-Sequenzen die Zuordnung zu einer Referenzsequenz deutlich schwieriger ist und daher zwischen den Tools stärker variiert. Für die *int*-Sequenzen konnte größtenteils eine tendenzielle Übereinstimmung beobachtet werden, während für die *prrt*-Sequenzen eine hohe absolute Übereinstimmung zwischen den Webtools beobachtet werden konnte. Die ist darauf zurückzuführen, dass sich die Subtypen 02_AG und G besonders im Integrase-Bereich (Position Referenz HXB2: 4.230 bp bis 5.067 bp) aufgrund der Breakpoints im Genom (Abbildung 20) schwer differenzieren lassen und das untersuchte Probenet einen hohen Anteil CRF02_AG beinhaltet. Auch im Vergleich von REGA und der eigenen Phylogenie fand sich die höchste Übereinstimmung für die *prrt*-Sequenzen, die Zuordnung der *int*- und *env*-Sequenzen stimmte zumindest tendenziell.

```

1 789 G
790 2155 A
2156 2625 A/G
2626 3274 G
3275 4174 A
4175 4874 G
4875 6224 A/G
6225 8311 A
8312 8588 A/G
8589 9089 A/G
9090 9719 G

```

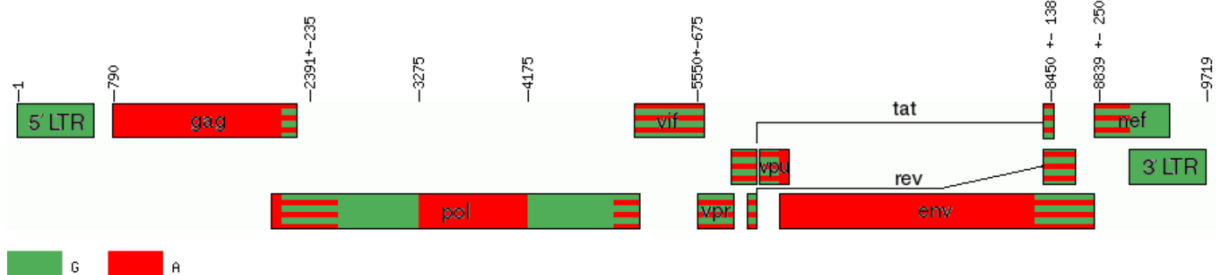


Abbildung 20: HIV-1 Genomkarte mit eingetragenen Breakpoints der CRF02_AG [16]

Für die 25 Proben, für die *prrt*, *int*- und *env*-Sequenzen vorhanden waren, wurde des Weiteren die Klassifizierung der einzelnen Bereiche mit dem Subtypisierungstool REGA analysiert. Hierbei zeigte sich bei 23 Proben keine Übereinstimmung zwischen den Genomregionen, sodass man aufgrund der REGA-Analysen von URFs ausgehen würde. Im Vergleich dazu konnte bei der eigenen Maximum-Likelihood-Phylogenie anhand der drei Sequenzbereiche eine eindeutige Klassifizierung bei neun der 25 Proben erfolgen. Abschließend wurde noch ein Vergleich der Subtypisierung zwischen REGA und der eigenen Phylogenie für die Gesamtsequenzinformation (Merge) durchgeführt. Anhand der

Merge-Sequenz ließen sich mit REGA zusätzlich nur eine Probe, mit der eigenen Phylogenie jedoch zusätzlich neun Proben eindeutig klassifizieren. Insgesamt zeigte sich, dass bei einer Subtypisierung auf Basis der konservierten *prt*-Sequenz das REGA HIV-1 Subtyping Tool ganz gute Ergebnisse liefert. Für eine umfassendere Klassifizierung anhand mehrerer Genomregionen ist die eigene phylogenetische Analyse dem Tool allerdings überlegen. Ein Grund hierfür könnte sein, dass im eigenen Stammbaum die Referenzsequenzen der Los Alamos HIV-1 Database bis zur dato aktuellsten verfügbaren Referenz CRF88_BC genutzt wurde, wohingegen das REGA-Tool weit weniger Referenzsequenzen nutzt.

Aufgrund der Mosaikstruktur bei rekombinanten Formen ist eine eindeutige Klassifizierung anhand nur eines Genomabschnittes kaum möglich, da besonders bei den komplexen Formen viele verschiedene Subtypen im Genom auftreten. Beispielhaft hierfür ist die Probe 08-03381. Im Envelope-Bereich clusterte die Probe im phylogenetischen Stammbaum mit 13 verschiedenen Subtypen (45_cpx, A1, 50_A1D, 22_01A1, 36_cpx, AD, 11_cpx, 13_cpx, 02_AG, 56_cpx, 19_cpx, 37_cpx, 25_cpx) und ist somit definitiv nicht klassifizierbar. Auch im stark konservierten *prt*-Bereich gibt es teilweise nicht eindeutig klassifizierbare Proben, so zum Beispiel die Probe 16-01746. Im eigenen Stammbaum clustert die Probe mit der CRF45_cpx (Abbildung 21) und der CRF49_cpx (Abbildung 22). Der Bereich der *prt*-Region umfasst die Position 2.253 bp bis 3.425 bp, sodass nur durch das Betrachten dieser einen Genomregion keine Klassifizierung erfolgen kann. Besonders Subtypen mit enger Verwandtschaft wie 02_AG und G der Gruppe M lassen auch in konservierten Bereichen wie der Integrase keine eindeutige Zuordnung zu einem Subtyp zu, da teilweise keine signifikante Abspaltung einzelner Subtypen oder CRFs stattfindet. Dies ist auch auf die Kürze des Fragments zurückzuführen

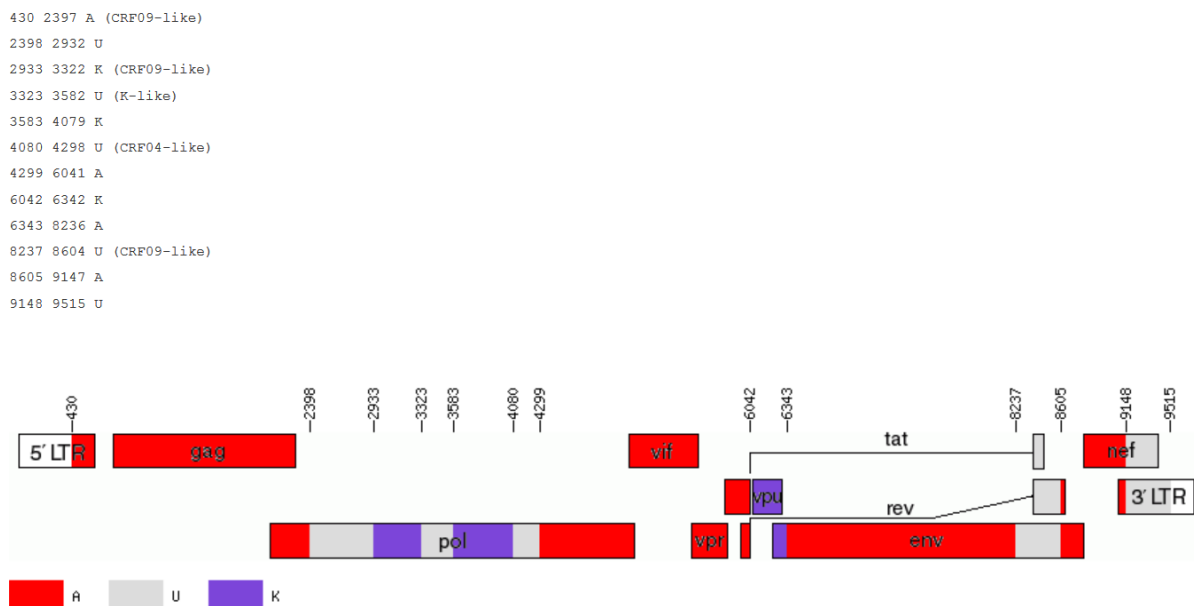


Abbildung 21: HIV-1 Genomkarte mit eingetragenen Breakpoints der CRF45_cpx [16]

556	789	U
790	2385	A
2386	2395	A/J
2396	2631	J
2632	2652	J/U
2653	2909	U
2910	3001	U/K
3002	3364	K
3365	3381	K/J
3382	3728	J
3729	3758	J/A
3759	4056	A
4057	4108	A/C
4109	5572	C
5573	5730	C/J
5731	9078	J

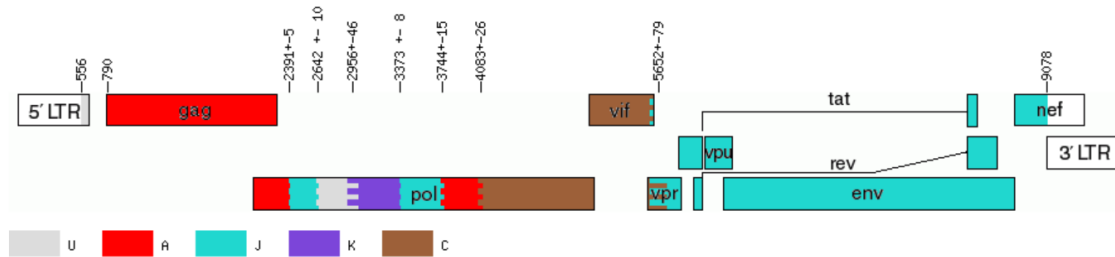


Abbildung 22: HIV-1 Genomkarte mit eingetragenen Breakpoints der CRF49_cpx [16]

Ein Vorteil des fragmentierten Sequenzierens der Genombereiche ist, dass kürzere Produkte eine höhere Erfolgsquote bei der Amplifikation zeigen. Allerdings werden so nur Bereiche mit einer Größe von etwa 1.800 bp (*prf1*), 600 bp (*env*) und 900 bp (*int*) abgedeckt, das Genom von HIV-1 hingegen besitzt rund 9.500 bp. Dementsprechend wird mit ungefähr 3.300 bp gerade mal ein Drittel des kompletten Genoms abgedeckt, wodurch die Überlegung einer *full length* Sequenzierung in den Vordergrund tritt. Je nach Protokoll werden hier mehrere Kilobasenpaar-große Fragmente (7.000 bp [52], 8.400 bp [53] bis über 9.000 bp [54]) amplifiziert, wobei sich überlappende Fragmente die Gesamtheit des Genoms sichern. Dementsprechend stehen längere Sequenzen für eine exaktere Subtypbestimmung zur Verfügung, was wiederum besonders bei der Subtypbestimmung mit den Webtools Vorteile liefert: je länger die Sequenz, desto exakter kann sie mit vorhandenen Referenzsequenzen verglichen werden und würde eine eindeutige Zuordnung zu einem bekannten Subtyp ermöglichen, als zuvor mit den kurzen Fragmenten. Bei einer Untersuchung zu Subtypen, CRFs und URFs ist die Auswahl der Referenzsequenzen von entscheidender Bedeutung. Die verfügbaren Webtools beruhen auf unterschiedlichen Datenbanken und nutzen jeweils andere mathematische Modelle zur Berechnung des Subtyps. Große Bedeutung für die Subtypisierung ist eine aktuelle Datenbank welche in regelmäßigen Abständen aktualisiert wird. Diese Problematik zeigt deutlich der Anstieg der Anzahl an CRFs. 2010 waren 44 CRFs bekannt, für das Jahr 2017 konnten 88 unterschiedliche CRFs in den Datenbanken gefunden werden. Mit den neuen Sequenzieretechnologien und den sinkenden Kosten für Ganzgenomsequenzierung ist mit weiteren neuen CRFs zu rechnen. Auch die Länge der zu untersuchenden Referenzen ist bei der Webtool-Subtypbestimmung von Bedeutung, da eventuell zu kurze oder zu Ambiguitäten-reiche Sequenzen kein clustern der Probe in der korrekten Subtyp-Klade ermöglichen. Besonders für den zusammengefassten Stammbaum der das *prf1-int-env* Alignment beinhaltet, wird eine genau passende Referenz benötigt, um den Subtyp, beziehungsweise die rekombinante Form, deklarieren zu können. Hierbei muss erwähnt werden, dass bei der Alignierung der einzelnen Genomregionen, Sequenz-Bereiche in denen keine Aminosäureinformationen vorhanden war, entfernt wurden, um keine Mismatches mit den Referenzsequenzen zu erzeugen.

Alternativ zur verwendeten Sanger-Sequenzierung, kann auch ein Next Generation Sequencing (NGS) durchgeführt werden. Bei der Untersuchung von mehreren großen Genombereichen stellt NGS eine schnellere, kostengünstigere und der Sanger-Sequenzierung qualitativ überlegen Alternative dar. Ebenfalls von Vorteil ist die gesteigerte Sequenztiefe um minoritäre Varianten zu detektieren,

insbesondere wenn es sich hierbei um resistenzassoziierte Mutationen handelt. [55] Auch Deletionen und Insertionen können besser detektiert werden. Somit wird mit dem NGS im Vergleich zur Sanger-Sequenzierung eine deutlich umfassendere und genauere Analyse gewährleistet, im Vergleich zur konventionellen Einzelgen-Analyse. [55-56] Zusätzlich können Resistenzmutationen erfasst werden, die mit der Sanger-Sequenzierung nicht detektiert werden. [55] Im Bereich des NGS gibt es verschiedene Methoden. Marktführend ist hier die Firma Illumina. Die Amplikon-Sequenzierung auf deren Gerät MiSeq wird im HIV-Studienlabor bereits durchgeführt und stellt für das hohe Probenaufkommen in der molekularen Surveillance eine Zeit- und Kostenersparnis dar.

Eine weitere, preislich vergleichbare Alternative wäre die Sequenzierung mit dem portablen MinION-Sequenziergerät. Mit diesem Gerät können Sequenzen mit einer Länge von bis zu 200 kbp generiert werden, währenddessen die Sequenzen in *real time* auf dem Computer ausgewertet werden können. [57] Es werden allerdings qualitativ hochwertige PCR-Produkte benötigt, die bei dem vorliegenden Probenmaterial jedoch nicht immer gegeben sind. Auch die *single molecular real time* (SMRT) Sequenzierung von Pacific Biosciences [58] erfordert Probenmaterial besonders hoher Güte. Zudem ist diese NGS-Methode ein sehr kostenintensives Verfahren, bei dem ungefähr 2.000 € pro Sequenz eingeplant werden müssen und welches daher nicht für den Hochdurchsatz in der molekularen Surveillance geeignet ist.

Zusammenfassend konnten 18 von 25 Viren mittels eigener phylogenetischer Analyse auf Basis der *prnt-int*- und *env*-Sequenzen eindeutig einem Subtyp oder einer CRF zugeordnet werden, was einer Erfolgsrate von 72 % entspricht. Mit dem REGA HIV-1 Subtyping Tool waren nur drei der 25 Viren eindeutig klassifizierbar. Somit lässt sich schlussfolgern, dass die Analyse weiterer viraler Genomregionen neben *prnt* und die Subtypisierung mit eigens erstellten Stammbäumen für eine umfassendere Charakterisierung der Viren sinnvoll ist, um die aktuell zirkulierenden HIV-Varianten zu erfassen und deren epidemische Relevanz bewerten zu können.

6 Ausblick

Die Subtypbestimmung von HIV-Proben ist ein nicht endender Prozess, da regelmäßig neue CRFs entdeckt werden, aufgrund der regen Genommutationen des humanen Immundefizienzvirus. Neue Methoden werden entwickelt, um die Sequenzierung noch genauer zu gestalten. Nichtsdestotrotz ist es von Vorteil das gesamte Genom zu sequenzieren (*full length* sequencing), um möglichst viele Informationen für die Analyse zur Verfügung zu haben. Die Verbreitung und Bildung von neuen URFs wird weiterhin bestehen bleiben, da die weltweite Reisetätigkeit von Personen mindestens gleichbleibt, wenn nicht sogar zunimmt. Die molekulare Surveillance, also die Untersuchung der viralen Eigenschaften von HIV, wird auch in Zukunft ein wichtiger Bestandteil der Forschungsarbeiten in Zusammenhang mit HIV bleiben, um die aktuell zirkulierenden HIV-1-Subtypen, die übertragenen HIV-Resistenzen und die phylogenetischen Transmissionscluster bestimmen und analysieren zu können.

7 Zusammenfassung

Das humane Immundefizienzvirus (HIV), ursprünglich entstanden durch eine zoonotische Transmission, ist der Erreger des erworbenen Immunschwächesyndroms (AIDS). Hierbei verursacht HIV-1 Gruppe M eine weltweite Epidemie, wobei unter den neun Subtypen und bis dato 88 bekannten rekombinanten Formen (CRFs) global der Subtyp C am verbreitetsten ist. Durch Mutation und Rekombination unterliegen die Viren jedoch einer stetigen Diversifizierung, was zur Entstehung immer neuer, vornehmlich singulärer rekombinanter Formen (URFs) führt. Da diese genetische Diversität Konsequenzen für die antiretrovirale Therapie, die Sicherheit der Diagnostik, die Impfstoffentwicklung und möglicherweise auch für die Pathogenität hat, ist die Beobachtung der epidemischen Verbreitung verschiedener HIV-1 Subtypen und rekombinanter Formen von großer Public Health Bedeutung. Vor Initiierung einer antiretroviralen Therapie oder bei Therapieversagen wird zur Identifikation von resistenz-assoziierten Mutationen eine genotypische Resistenztestung anhand der Sequenz der viralen Protease (*pr*) und Reversen Transkriptase (*rt*), teilweise auch der Integrase (*int*), vorgenommen. Die erzeugten Virussequenzen können zudem für die Bestimmung des HIV-Subtyps herangezogen werden. Jedoch ist anhand der sehr konservierten *pr*-Sequenz eine eindeutige Klassifizierung nicht immer möglich, da markante Rekombinationspunkte häufig außerhalb dieses Bereichs liegen. Aus diesem Grund wurden in der vorliegenden Arbeit 45 Proben von HIV-1 positiven Patienten näher untersucht durch die zusätzliche Betrachtung der viralen *int*- und *env* (Hüllprotein)-Sequenzen. Insgesamt konnte für 24 von 25 untersuchten Proben eine valide *int*-Sequenz und für 30 von 45 untersuchten Proben eine valide *env*-Sequenz generiert werden. Für die Subtypisierung anhand der Einzelsequenzen sowie der vorliegenden Gesamtsequenzinformation wurden verschiedene Webtools und eine eigene Maximum-Likelihood-Phylogenie eingesetzt und die Ergebnisse auf Konkordanz verglichen. Aufgrund der unterschiedlichen Algorithmen, auf denen die webbasierten Subtypisierungstools basieren, zeigten sich vor allem in dem hochvariablen *env*-Bereich mit 39,9 % deutliche Unterschiede bei der Klassifizierung, wohingegen die Übereinstimmung in der recht konservierten *pr*-Region mit 76,7 % relativ hoch war. Von insgesamt 25 der 45 Proben lag eine Sequenz in allen drei untersuchten Genombereichen vor. Anhand dieser umfassenden Sequenzinformation konnten mit dem REGA HIV-1 Subtyping Tool dennoch nur drei Proben eindeutig klassifiziert werden, wohingegen mittels der eigenen Phylogenie 18 Proben eindeutig zugeordnet werden konnten. Bei den verbleibenden sieben Proben könnte es sich um URFs handeln. Insgesamt lässt sich schlussfolgern, dass für eine korrekte Subtypisierung eine eigene phylogenetische Analyse anhand mehrerer Genomregionen erfolgen muss und das für die nähere Charakterisierung von URFs und die Bewertung ihrer epidemischen Relevanz eine Gesamtgenom-Sequenzierung notwendig ist.

8 Summary

The human immune deficiency virus (HIV) causes the acquired immune deficiency syndrome (AIDS) and was originally developed through a zoonotic transmission. The epidemic relevant type is HIV-1 group M comprising nine subtypes and up to date 88 circulating recombinant forms (CRFs). Globally, subtype C is most common. HIV is subjected to continuous diversification and recombination, steadily giving rise of unique recombinant forms (URFs) of which some become epidemic. The molecular surveillance of currently circulating HIV variants is an important public health aspect as the genetic diversity of HIV has consequences on antiretroviral therapy, diagnostic, vaccine development and potentially pathogenicity. HIV sequences of viral protease (*pr*) and reverse transcriptase (*rt*), in some cases also integrase (*int*), determined by genotypic resistance testing can be used to identify the HIV subtype. However, the *prrt*-region is quite conserved and breakpoints of recombination are often located outside this region. The aim of this bachelor thesis was to further characterize the subtype of 45 viruses using *int*- and *env* (envelope)-sequences in addition to *prrt*-sequences. To this end, 25 and 45 plasma samples of HIV-1 positive patients were analyzed for viral *int* and *env*, respectively. In total, 24/25 valid *int*-sequences and 30/45 valid *env*-sequences could be determined. Several webtools using different algorithms for subtyping as well as an own maximum-likelihood-phylogeny were applied to classify the viruses based on their respective sequence information. When comparing the outcomes, the biggest difference between the webtool interpretations was observed in the highly variable *env*-region, whereas the results for the conserved *prrt*-region were quite concordant. For 25 samples, the full sequence information on *prrt*, *int*, and *env* was available. Only three of the 25 samples could be clearly classified using the REGA HIV-1 Subtyping Tool, whereas 18 samples could be subtyped by the maximum-likelihood-phylogeny. The remaining seven samples are supposed to be URFs. In conclusion, analysis of several genomic regions by maximum-likelihood phylogeny is necessary for correct determination of HIV-1 subtypes. Full length-sequencing is needed for further characterization of URFs.

9 Literaturverzeichnis

- [1] Erneuerung der Forschungsvereinbarung zu zwischen Tieren und Menschen übertragbaren Krankheiten (Zoonosen); Bundesministerium für Bildung und Forschung; 2016
- [2] High GUD Incidence in the Early 20th Century Created a Particularly Permissive Time Window for the Origin and Initial Spread of Epidemic HIV Strains; João Dinis de Sousa, Viktor Müller, Philippe Lemey, Anne-Mieke Vandamme; PLoS ONE, April 2010, Volume 5, Issue 4
- [3] Origins of HIV and the AIDS Pandemic; Paul M. Sharp, Beatrice H. Hahn; Cold Spring Harbor Perspectives in Medicine; 2011 Sep; Volume 1, Issue 1
- [4] Origins of HIV; David M. Hillis; Science; 09 June 2000; Vol. 288, Issue 5472, pp. 1757-1759
- [5] HIV and AIDS – Medical Perspectives Series; K. E. Nye, J. M. Parkin; Bios Scientific Publishers Limited 1994; pp.1 ff.
- [6] A history of AIDS: Looking back to see ahead; Greene WC; 2007 European Journal of Immunology 37 (Suppl. 1): p 94 – p102
- [7] Med-Info – Medizinische Informationen zu HIV und AIDS; Deutsche AIDS-Hilfe; Aktualisierte Neuauflage Nr. 74
- [8] Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS); Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, et al. 1983; Science Volume 220, Issue 4599; pp. 868–871
- [9] Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS; Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, Palker TJ, Redfield R, Oleske J, Safai B, et al. 1984; Science, Volume 224, Issue 4648, pp. 500–503
- [10] The evolution of HIV-1 and the origin of AIDS; Paul M. Sharp, Beatrice H. Hahn; Philosophical Transactions of the Royal Society B, Biological Sciences; 2010 Aug 27; 365(1552): 2487–2494
- [11] Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*; Feng Gao, Elizabeth Bailes, David L. Robertson, Yalu Chen, Cynthia M. Rodenburg, Scott F. Michael, Larry B. Cummins, Larry O. Arthur, Martine Peeters, George M. Shaw, Paul M. Sharp, Beatrice H. Hahn; Nature 397, 436-441 (4 February 1999)
- [12] HIV sequence compendium 2005; Leitner T., Foley B., Hahn B. et al; Los Alamos, NM: Theoretical Biology and Biophysics Group; Los Alamos National Laboratory; 2005
- [13] Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand; Carr JK, Salminen MO, Koch C, et al; Journal of Virology, 1996; Volume 70, No. 9, pp. 5935–43
- [14] The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin; Gao F, Robertson DL, Morrison SG, et al; Journal of Virology 1996, Volume 70, No. 10, pp. 7013–7029
- [15] HIV CRFs; <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/breakpoints.html> - 09.08.17
- [16] Breakpoints of CRFs; <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/breakpoints.html> - 21.07.17
- [17] HIV Statistics; <http://www.unaids.org/en/resources/fact-sheet> - 20.07.17

-
- [18] Epidemiologisches Bulletin Nr. 45; Schätzung der Zahl der HIV-Neuinfektionen und der Gesamtzahl von Menschen mit HIV in Deutschland; Robert Koch Institut; 14. November 2016; S. 2 ff.
- [19] Is HIV-1 evolving to a less virulent form in humans?; Kevin K. Ariën, Guido Vanham, Eric J. Arts; Nature Reviews Microbiology March 2007
- [20] HIV-Subtypen; <http://www.hivleitfaden.de/cms/index.asp?inst=hivleitfaden&snr=2173> – 21.07.17
- [21] HIV and AIDS – Medical Perspectives Series; K. E. Nye, J. M. Parkin; Bios Scientific Publishers Limited 1994; pp.1 ff.
- [22] Biological Resource Centers - Knowledge Hubs for the Life Sciences; Scott Stern; Brookings Institution Press 2005; p. 44
- [23] HIV; <http://www.spektrum.de/lexikon/biologie/hiv/32099> - 24.07.17
- [24] The Life-Cycle of the HIV-1 Gag–RNA Complex; Figure 2: Human immunodeficiency virus type 1 (HIV-1) genome and Gag organization; Elodie Mailler, Serena Bernacchi, Roland Marquet, Jean-Christophe Paillart, Valérie Vivet-Boudou, Redmond P. Smyth; Viruses 2016, Volume 8(9), p. 248
- [25] HIV and AIDS – Medical Perspectives Series; K. E. Nye, J. M. Parkin; Bios Scientific Publishers Limited 1994; pp.1 ff.
- [26] New hope for an AIDS vaccine; Harriet L. Robinson; Nature Reviews Immunology 2 (April 2002), pp. 239-250
- [27] HIV und AIDS: Ein Leitfaden für Ärzte, Apotheker, Helfer und Betroffene; HIV-Arbeitskreis Südwest; 5. Auflage Springer Verlag Berlin-Heidelberg; S. 4 ff.
- [28] HIV lifecycle; <https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/73/the-hiv-life-cycle> - 24.07.17
- [29] HIV und AIDS: Ein Leitfaden für Ärzte, Apotheker, Helfer und Betroffene; HIV-Arbeitskreis Südwest; 5. Auflage Springer Verlag Berlin-Heidelberg; S. 75
- [30] Die Infektiologie; D. Adam, H. W. Doerr, H. Link, H. Lode; Springer Verlag Berlin-Heidelberg 2004; S. 590
- [31] LaborMedizin: Indikationen, Methodik und Laborwerte - Pathophysiologie und Klinik; Hans D. Bruhn, Ulrich R. Fölsch, Heiner Schäfer; 2. Auflage Schattauer Stuttgart; S. 307
- [32] Genotypische Resistenztestung;
<http://www.hivleitfaden.de/cms/index.asp?inst=hivleitfaden&snr=2342> – 25.07.17
- [33] REGA Subtyping Tool; Developed by Tulio de Oliveira, Koen Deforche, Sharon Cassol, Andrew Rambaut and Anne-Mieke Vandamme; Revised 25th of March 2014
- [34] Gene und Stammbäume – Ein Handbuch zur molekularen Phylogenetik; Volker Knoop, Kai Müller; 2. Auflage Spektrum Akademischer Verlag 2009; S. 58 ff.
- [35] An introduction to molecular phylogenetic analysis; Tokumasa Horiike; Reviews in Agricultural Science 2016, Issue 4, pp. 36-45
- [36] Genetik; Jochen Graw; 4. Auflage Springer Verlag Berlin-Heidelberg 2006; S. 360 ff.
- [37] Phylogenie; <http://www.spektrum.de/lexikon/biologie/phylogenie/51553> - 28.07.17
- [38] Evolution of raltegravir resistance during therapy; Nadine Sichtig, Saleta Sierra, Rolf Kaiser, Martin Däumer, Stefan Reuter, Eugen Schülter, Andre Altmann, Gerd Fätkenheuer, Ulf Dittmer, Herbert Pfister, Stefan Esser; Journal of Antimicrobial Chemotherapy 1st July 2009, Volume 64, Issue 1, pp. 25–32

-
- [39] User guide GeneRuler 1kBp DNA-Ladder;
<https://www.thermofisher.com/order/catalog/product/SM0312> - 31.07.17
- [40] Sanger-Sequenzierung; <http://www.spektrum.de/lexikon/biochemie/sanger-sequenzierung/5530> - 26.07.17
- [41] Biologie für Mediziner; Werner Buselmaier; 13. Auflage Springer Verlag 2005; S. 251
- [42] SeqMan Pro™ (©1988-2017); DNASTAR, Madison, Wisconsin
- [43] IUPAC-IUB codes; http://www.gendx.com/SBTengine/Help_220/hs310.htm - 28.07.17
- [44] BioEdit; Version 7.2.6.1 (2017/Juni); Tom Hall
- [45] Gene und Stammbäume – Ein Handbuch zur molekularen Phylogenetik; Volker Knoop, Kai Müller; 2. Auflage Spektrum Akademischer Verlag 2009; S. 114
- [46] MAFFT; Version 7.310 (2017/März); © 2006 Kazutaka Katoh
- [47] Seaview; Gouy M, Guindon S, Gascuel O (2010); SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building; *Molecular Biology and Evolution*, Volume 27, Issue 2, pp. 221-224
- [48] IQ-TREE; Version 1.5.5 (2017/Juni); Modelfinder: S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermini; ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 2017, Volume 14, pp. 587–589; Ultrafastbootstrap: B.Q. Minh, M.A.T. Nguyen, and A. von Haeseler; Ultrafast approximation for phylogenetic bootstrap; *Molecular Biology and Evolution*, 2013, Volume 30, pp.1188-1195
- [49] FigTree; Version 1.4.3 (2016/Okttober)
- [50] The Phylogenetic Handbook, Philippe Lemey, Marco Salemi, Anne-Mieke Vandamme; 2nd Edition Cambridge University Press 2009; pp. 181 ff
- [51] Principles and Technical Aspects of PCR Amplification; Elizabeth van Pelt-Verkuil, Alex van Belkum, John P. Hays; Springer Science & Business 2008; p. 190
- [52] Quasispecies Analyses of the HIV-1 Near-full-length Genome With Illumina MiSeq; Hirotaka Ode, Masakazu Matsuda, Kazuhiro Matsuoka, Atsuko Hachiya, Junko Hattori, Yumiko Kito, Yoshiyuki Yokomaku, Yasumasa Iwatani, Wataru Sugiura; *Frontiers in Microbiology*, 12 November 2015
- [53] Protocol for Nearly Full-Length Sequencing of HIV-1 RNA from Plasma; Yuka Nadai, Lindsay M. Eyzaguirre, Niel T. Constantine, Anne M. Sill, Farley Cleghorn, William A. Blattner, Jean K. Carr; *PLoS ONE* 2008, Volume 3, Issue 1
- [54] Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management; Sebastian Grossmann, Piotr Nowak, Ujjwal Neogi; *Journal of the International AIDS Society* 2015, Volume 18, pp. 20035
- [55] Next-Generation Sequencing to Help Monitor Patients Infected with HIV: Ready for Clinical Use?; Richard M. Gibson, Christine L. Schmotzer, Miguel E. Quiñones-Mateu; *Current Infectious Disease Reports*, April 2014, Volume 16, pp. 401
- [56] Next Generation Sequencing-based HIV-1 Drug Resistance Monitoring System; E. Rakhmanaliev, P. Ariyaratne, P. Nimitsuntiwong, C. Watthphan, Z. Rui, E. Passomsub, W. Chantratita, G. Michel; Scientific Poster B-063 presented at: AACC Annual Scientific Meeting. July 31-Aug. 4, 2016; Philadelphia
- [57] MinION; <https://nanoporetech.com/products/minion> - 30.08.17
- [58] SMRT Sequencing; <http://www.pacb.com/smrt-science/smrt-sequencing/> - 30.08.17

Appendix

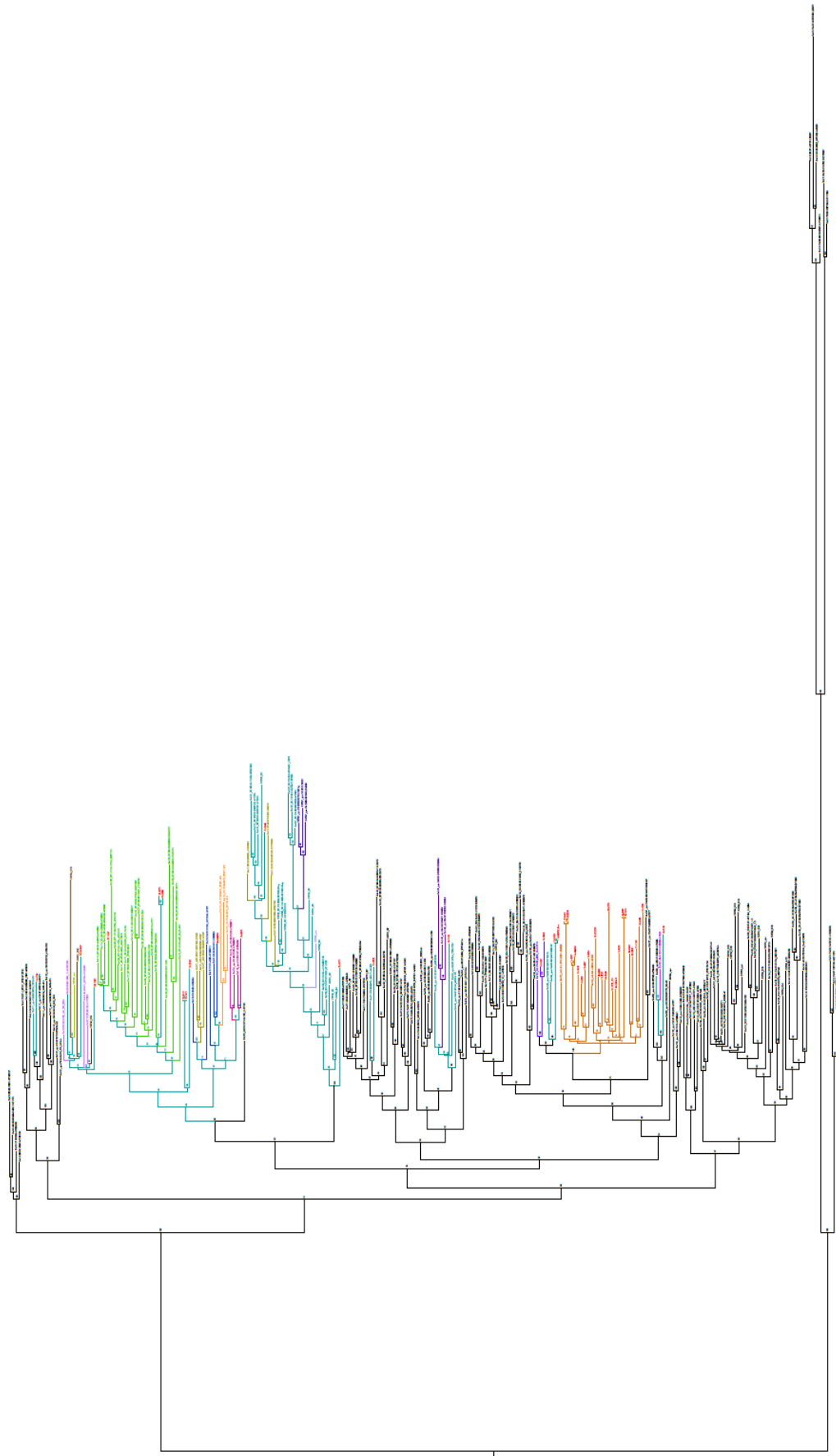


Abbildung 23: phylogenetischer Stammbaum der prrt-Genomregion, ganz rechts die Außengruppe der Untergruppen N, O und P

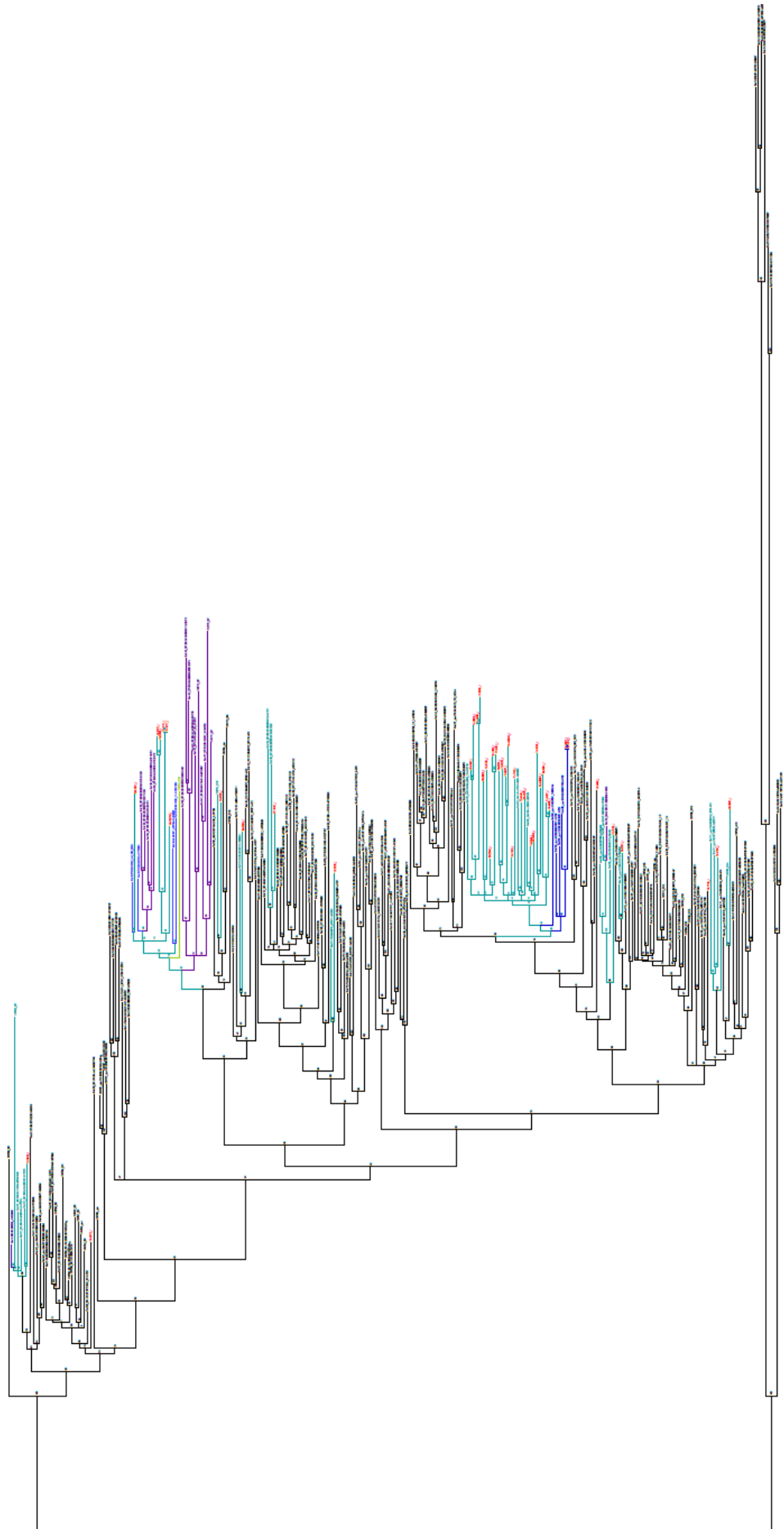


Abbildung 24: phylogenetischer Stammbaum der int-Genomregion, ganz rechts die Außengruppe der Untergruppen N, O und P

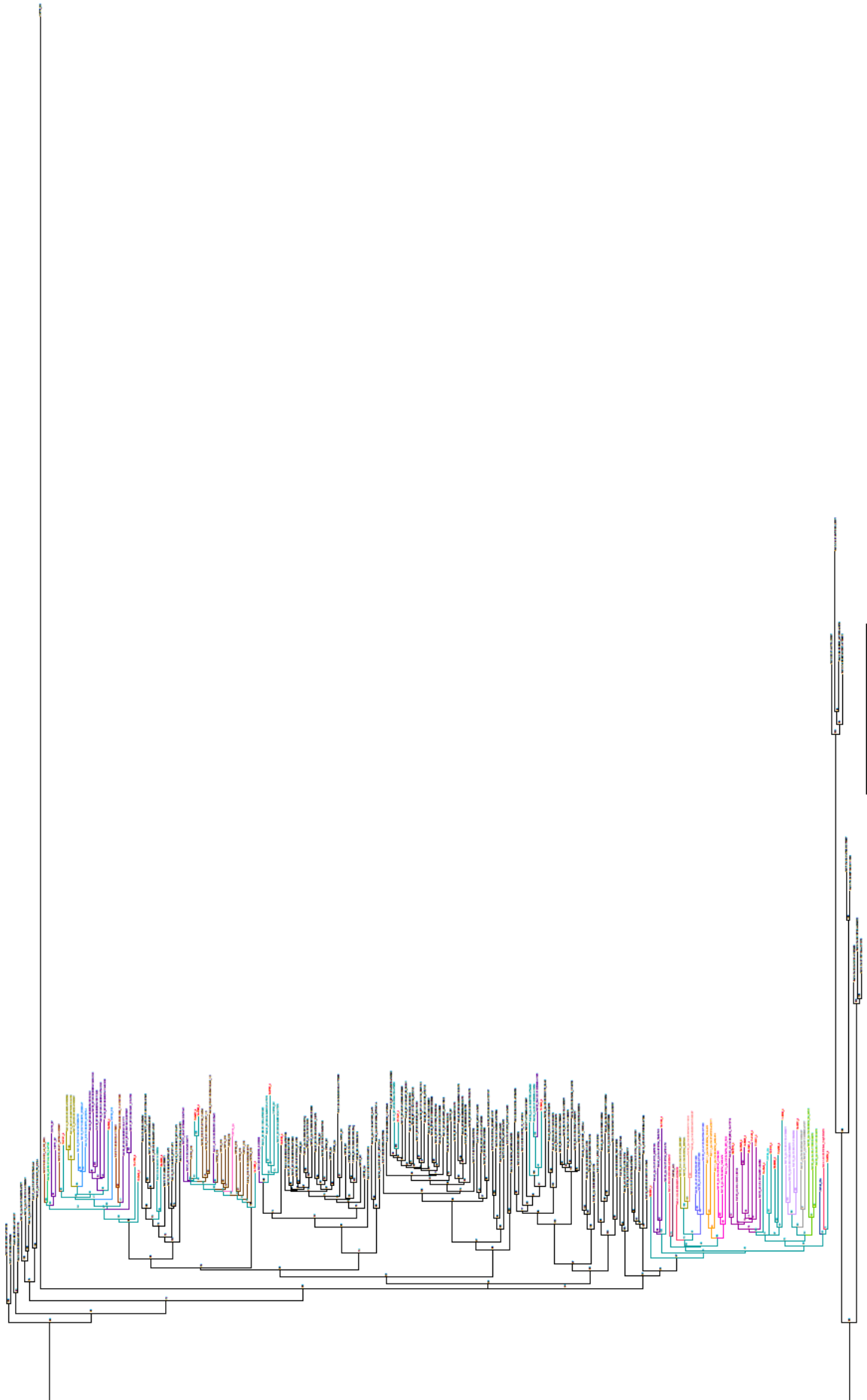


Abbildung 25: phylogenetischer Stammbaum der env-Genomregion, ganz rechts die Außengruppe der Untergruppen N, O und P

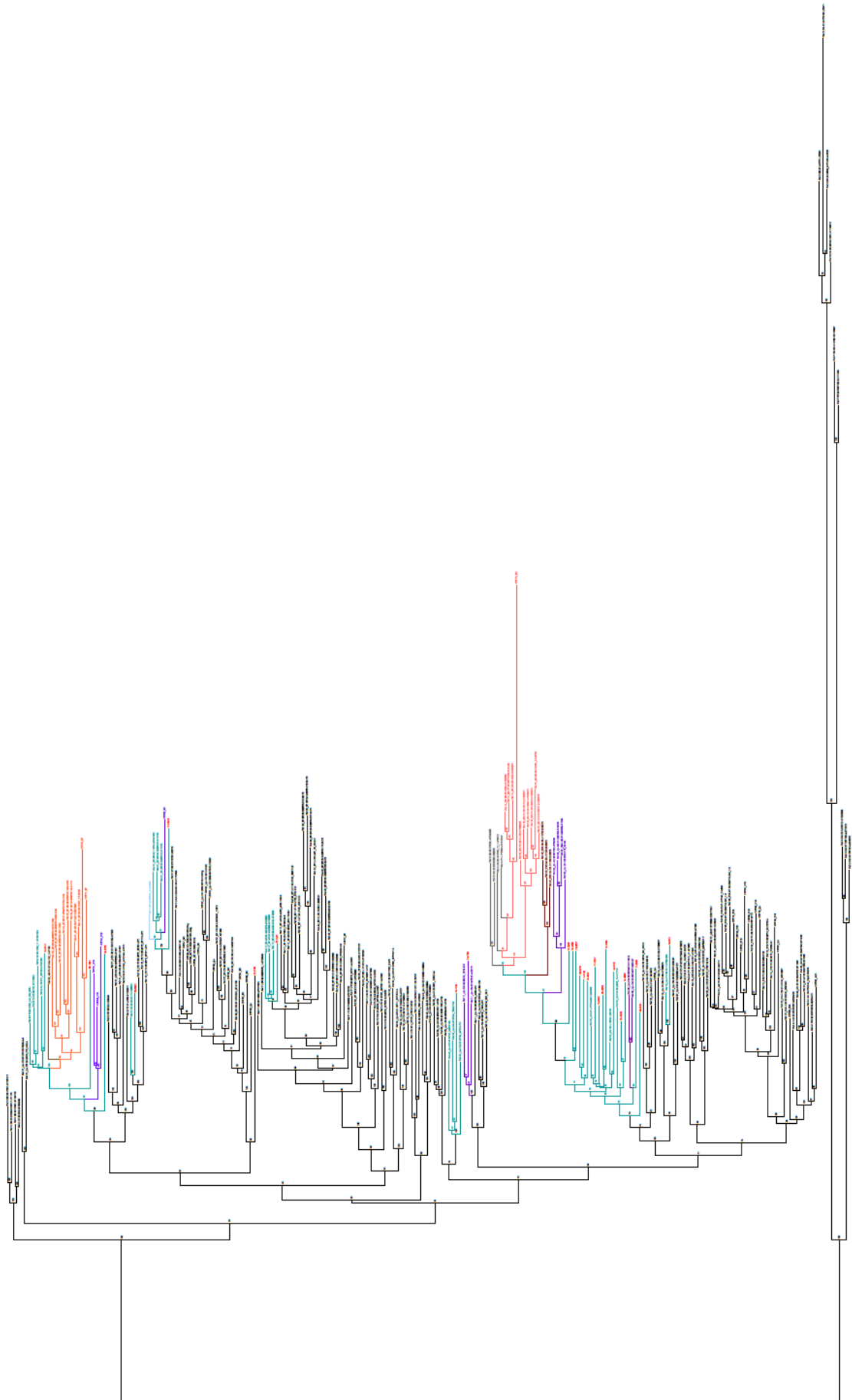


Abbildung 27: phylogenetischer Stammbaum der prrt-, int- und env-Genomregion, ganz rechts die Außengruppe der Untergruppen N, O und P

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst habe. Es wurden keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt. Die wörtlichen oder sinngemäß übernommenen Zitate wurden als solche kenntlich gemacht.

Berlin, den 05.09.2017

Dominique Seifert

Danksagung

Mein Dank gilt Prof. Dr. Norbert Bannert für die Möglichkeit meine Bachelorarbeit am Robert Koch-Institut in der Fachgruppe 18 für HIV und andere Retroviren zu schreiben, sowie Prof. Dr. Steffen Prowe für die Betreuung meiner Arbeit.

Besonders erwähnen möchte ich Dr. Karolin Meixenberger und Dr. Patrycja Machnowska, die mich intensiv betreut und mir mit Rat und Tat zur Seite gestanden haben.

Vielen Dank auch an Julian und Patrycja für eine tolle gemeinsame Zeit mit vielen unterhaltsamen Argumentationen und dem regen Austausch von nicht-wissenschaftlichen Publikationen.

Für jegliche Hinweise und Denkanstöße möchte ich mich bei Karolin, Patrycja und Stefan bedanken, die mir sehr weiter geholfen haben.

Dem gesamten Fachbereich 18 danke ich für die netten Gespräche und das große Spektrum an weiterführenden Informationen durch Vorträge und Seminare.

Vielen Dank an das Sequenzierlabor.

Mein ganz besonderer Dank gilt meiner Familie, die mich immer unterstützt, jeglicher Widrigkeiten zum Trotz!