



LUNDS UNIVERSITET

octapharma

Master's Thesis

Development and validation of a novel quantitative PCR analysis method for HIV-1

Josefine Sundgren

June 2020

University supervisor: Johan Svensson Bonde, PhD

Supervisor: Rasmus Gustafsson, PhD, Octapharma AB

Examiner: Lei Ye, Professor

Department of Pure and Applied Biochemistry

Faculty of Science

Lund University

SE-221 00 Lund, Sweden

1 Abstract

The major problem when developing primers and probe sets for detection of HIV-1 is its high genetic variability and strong ability to mutate. HIV-1 is divided into four groups, M, N, O, P based on genetic similarities, and there are more than 70 different circulating recombinant forms.

The aim of this master thesis was to develop new primers and probe for a novel quantitative PCR analysis method for detection of a broad range of genotypes of HIV-1. Four target sequences with conserved parts were chosen: the *pol* gene at around nucleotide position 2000 in the reference sequence HXB2, the *pol* gene at around nucleotide position 4000, the *nef* gene at around nucleotide position 9000 and the *gag* gene at around nucleotide position 800.

The primers and probe set targeting the *pol* gene, pol4000, showed best potential. The primer and probe concentrations were optimized, and the best concentration on both primers and probe was 900 nM.

To test the ability to detect a broad range of genotypes of HIV a genotype test was conducted. The result showed that pol4000 could amplify all tested subgroups in group M, but failed to amplify group O and N. Detected replicates in percent from three independent assessor were compared against the average nucleotide mismatches for each genotype and a regression analysis showed that 27.5% of the detected result could be explained by nucleotide mismatches ($p < 0.01$).

Moreover, the results showed that there are potential in the primers and probe set, pol4000, but there is a need for further investigation to be able to determine if pol4000 will function in the new routine analysis by Octapharma AB.

Table of Contents

1	Abstract	2
2	Introduction	5
2.1	Aim and scope.....	6
3	Background	7
3.1	HIV	7
3.1.1	Clinical Disease, AIDS.....	7
3.1.2	HIV-1 pathogenesis.....	7
3.1.3	HIV replication cycle	8
3.1.4	Viral genome.....	9
3.1.5	Viral Groups and subgroups	10
3.1.6	Genetic variation	11
3.2	qPCR.....	12
3.2.1	Real time Taqman qPCR	12
3.2.2	Primer and probe design.....	14
3.2.3	Betaine	15
4	Method	16
4.1	Extraction of HIV-1 RNA.....	16
4.2	Primer and probe design	16
4.3	Primer test.....	17
4.3.1	First evaluation of the P/P sets.....	18
4.3.2	Internal control (IC).....	18
4.3.3	Probe concentration optimization	19
4.3.4	Further primer concentration optimization.....	19
4.3.5	Betaine	19
4.3.6	Double eluate	19
4.4	Genotype test.....	20
4.5	Statistical analysis	20
5	Results.....	22
5.1	Primer design	22
5.2	Primer test.....	24
5.2.1	First evaluation of the P/P sets.....	24
5.2.2	Internal control.....	25
5.2.3	Probe concentration optimization	26
5.2.4	Further primer concentration optimization.....	29

5.2.5	Betaine	29
5.2.6	Double eluate	29
5.3	Tm analysis in OligoAnalyzer	30
5.4	Genotype test.....	31
6	Discussion	34
7	Conclusion.....	37
8	Acknowledgements.....	38
9	Bibliography.....	39

2 Introduction

Octapharma AB produces protein-based pharmaceuticals from human blood plasma and recombinant production. The PCR-laboratory at Octapharma in Stockholm analyse plasma donations from multiple countries to ensure absence of pathologic viruses. Even though it is particularly important that the finished drug not contain any virus particles, plasma donations are tested also before they go into pharmaceutical production. Hence blood plasma is tested by the pharmaceutical company to control that it does not contain Hepatitis A, B, C or E, or parvovirus B19 or Human Immunodeficiency Virus type 1 (HIV-1).

HIV-1 belongs to the genus *Lentivirus* within the family of *Retroviridae* (Murphy & Weaver, 2017). *Lentus* means slow in Latin which is an adjective that describes the deadly disease caused by the virus Acquired Immune Deficiency Syndrome (AIDS) well as it normally takes 8 to 10 years from infection to death if the disease is untreated (Sabin, 2013). The first documented case of HIV infection in humans was in Kinshasa, Democratic Republic of the Congo, in 1959 but the first case of AIDS officially reported, was not until 1981. (Murphy & Weaver, 2017)

In 2018 approximately 37.9 million people around the world were living with AIDS. Every day around 5000 adults and children are infected by HIV. UNAIDS has a goal that by 2020 90% of all HIV infected people will know about their HIV status, and of these 90% will receive antiretroviral therapy and 90% will have viral suppression. UNAIDS work in partnerships with essential stakeholders like national governments, World Health Organization (WHO) and professional medical groups (UNAIDS, 2020). Eradicating AIDS by 2030 is a part of the UN Sustainable Development Goals (Division for Sustainable Development Goals, 2020).

The virus can be divided in two major types based on genetics; HIV-1, the most common type that has its origin from chimpanzees (simian immunodeficiency virus; SIVcpz), and HIV-2 the less common type that comes from sooty mangabeys (SIVsm). HIV-1 and HIV-2 has many similarities, but the main differences are that HIV-2 have a lower transmissibility and a lower risk of progression to AIDS (Nyamweya, et al., 2013). HIV-2 also have the protein vpx instead of the vpr that is present in HIV-1 (Mahdi, et al., 2018). HIV-1 exists worldwide but HIV-2 is mainly present in West Africa (Nyamweya, et al., 2013).

HIV-1 has a large genetic variation with a high polymerase error rate of one error for every 10000 bp (Seitz, 2016). This can be compared to Influenza that has a polymerase error rate of approximately one error for every 13000 bp (Boivin, et al., 2010). HIV-1 can be divided into four groups, M, N, O and P (Eberle & Gurtler, 2012). (See sections 3.1.6 and 4.1.5 for more information about HIV-1 groups and genetic variation).

2.1 Aim and scope

The aim of this master thesis was to develop primers and probe for a novel quantitative PCR analysis method for detection of HIV-1. The method will at a later stage be used in the routine analysis at Octapharma AB for screening of plasma donations prior to pooling and fractionation, and for production pools to ensure absence of the virus.

A specific pedagogical question was how a PCR primers and probe set should be designed to enable detection of a large variety of HIV-1 genomes.

3 Background

3.1 HIV

3.1.1 Clinical Disease, AIDS

In 2018 approximately 37.9 million people around the world were living with AIDS. Every day around 5000 adults and children are infected by HIV (UNAIDS, 2020). HIV can be transmitted through blood, plasma, serum or transplanted organs (Seitz, 2016). Treatments that suppress HIV are available but there is yet no effective vaccine against HIV, or treatment that can eradicate the infection completely. Even though the treatments need to continue for the rest of the patient’s life it has been proven to improve life quality and longevity (Tough & McLaren, 2019).

3.1.2 HIV-1 pathogenesis

After infection of HIV there is an acute phase where viral RNA levels peak and CD4+ cells decrease dramatically (Figure 1). After the acute phase, there is a short recovery phase before a slow and steady decrease of CD4+ levels. The stage AIDS is defined when CD4+ T cell counts are lower than 200 cells/mm³. Without treatment it normally takes 5 to 10 years from primary infection to AIDS development (Tough & McLaren, 2019). Viral RNA levels can differ between infected individuals and the level is changing depending on stage of the infection.

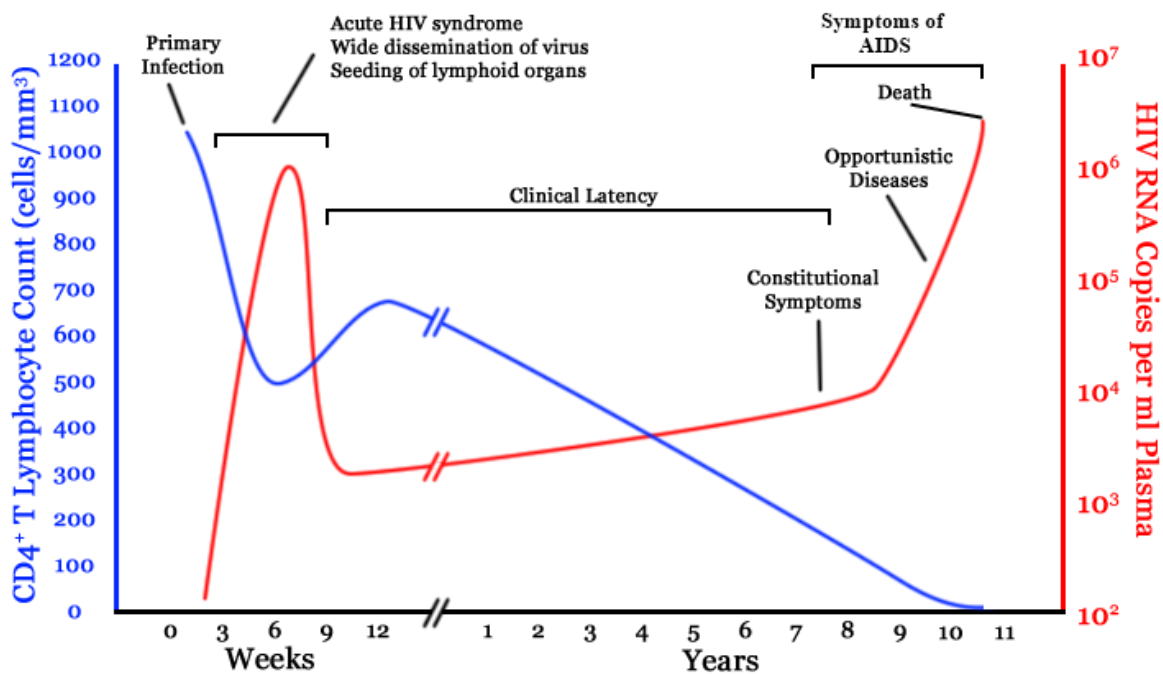


Figure 1. HIV RNA copies and CD4+ counts in a human infected with HIV without treatment. Wikimedia Commons. Jurema Oliveira (<http://creativecommons.org/licenses/by-sa/3.0/>)

3.1.3 HIV replication cycle

The virus is enveloped with 72 knobs that are assembled from env protein trimers. The two proteins encoded by *env* are called glycoprotein (gp)120 and gp41. Gp120 is a surface glycoprotein and gp41 is a transmembrane glycoprotein. The first step in the replication cycle of HIV is attachment to the host cell surface. The gp120 binds to the host cells CD4+ receptor and to the coreceptors CCR5 or CXCR4 (Figure 2). The binding to the coreceptors are mediated through a fragment on the gp120 that is called the V3 loop (Tamamis & Floudas, 2014). A HIV-1 particle without gp120 cannot infect cells (Seitz, 2016). HIV targets primarily three types of immune cells: CD4+ T cells, macrophages, and dendritic cells (Murphy & Weaver, 2017). The second step is a fusion of the viral and cellular membranes. The virus particle is entering the cell via pH dependent endocytosis. Inside the host cell the viral envelope is removed. For the virus to be able to insert the viral genome into the host cell genome, the viral RNA needs to transform into viral DNA. This transformation process is named reverse transcription and mediated by the viral protein reverse transcriptase. The viral DNA is integrated with the host cell genome and from here can get translated into viral RNA and mRNA by the host cell, see Figure 2. The mRNA is translated into the different viral proteins. The new virus is assembled and released from the host cell. This kills the host cell and cause the number of CD4+ cells to drop. The last step of the replication cycle is maturation (Freed, 2015).

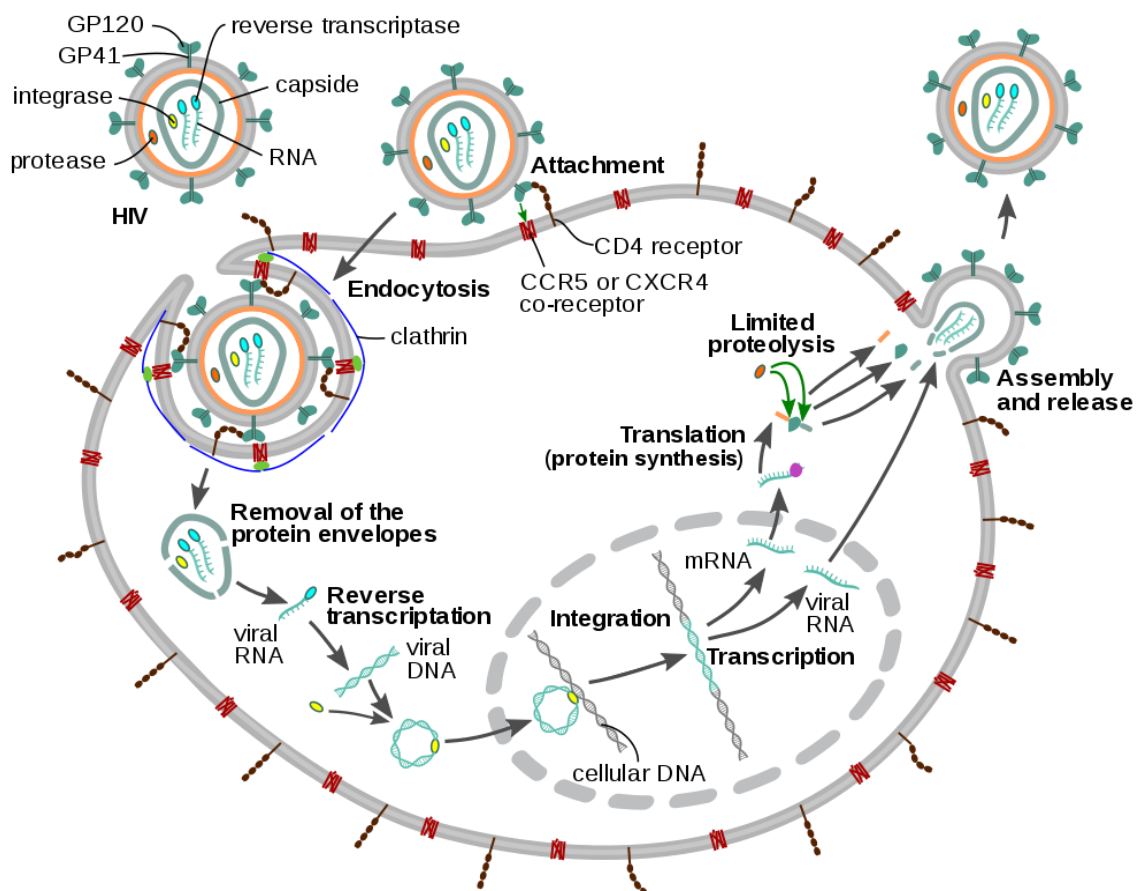


Figure 2. HIV replication cycle. Wikimedia Commons. Jmarchn (<https://commons.wikimedia.org/w/index.php?curid=58188472>)

3.1.4 Viral genome

HIV has a genome of around 10000bp that encodes nine genes in three reading frames (Figure 3). In the virus world this is a quite average size. To mention a few examples for comparisons Influenza A has a genome with a size of 13588bp (Bouvier & Palese, 2008), hepatitis C a genome with 9000bp (Bukh, 2016) and B19V a genome with 5600bp (Guan, et al., 2009). The three largest genes are *gag*, *pol* and *env* (Murphy & Weaver, 2017).

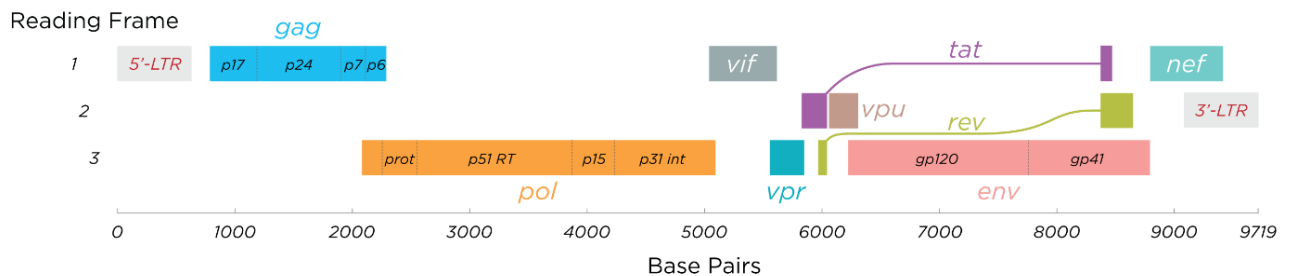


Figure 3. The HIV-1 genome. Wikimedia Commons. Thomas Spletstoeser (www.scistyle.com)

The *gag* gene is located at position 790-2292 on the reference sequence HXB2 (Los Alamos national laboratory, 2014) and encodes the structural proteins in the nucleocapsid core which is the outer core membrane (MA, or p17), the capsid protein (CA, or p24) the nucleocapsid (NC, or p7) and a smaller nucleic acid-stabilising protein (p6) (Freed, 2015). The *pol* gene is located at position 2085-5096 and encodes important viral replication enzymes, enzyme protease (pr, or p12), reverse transcriptase (RT, or p51), RNase H (p15), integrase (IN, or p32) (Murphy & Weaver, 2017). The *env* gene is located at position 6225-8795 and encodes the two viral envelope glycoproteins, gp120 and gp41 (Seitz, 2016).

Nef is a regulatory protein located at position 8797-9417 on HXB2. This protein has a negative impact on the levels of CD4+ on the surface of the cell because its ability to inhibit the transportation of newly synthesized CD4+ molecules to the cell membrane. This makes superinfection of infected cells less possible (Eberle & Gurtler, 2012).

Tat is a protein that can accelerates the access of viral RNA for virus production by a 100-fold. The protein that controls the splicing length of newly formed HIV RNAs is called Rev. Vif, vpr and vpu are additional regulatory proteins. They impact the rate on the production of virus particles (Seitz, 2016).

A mature HIV particle has a diameter of approximately 100nm (Seitz, 2016). Figure 4 displays an illustration of a HIV-1 particle.

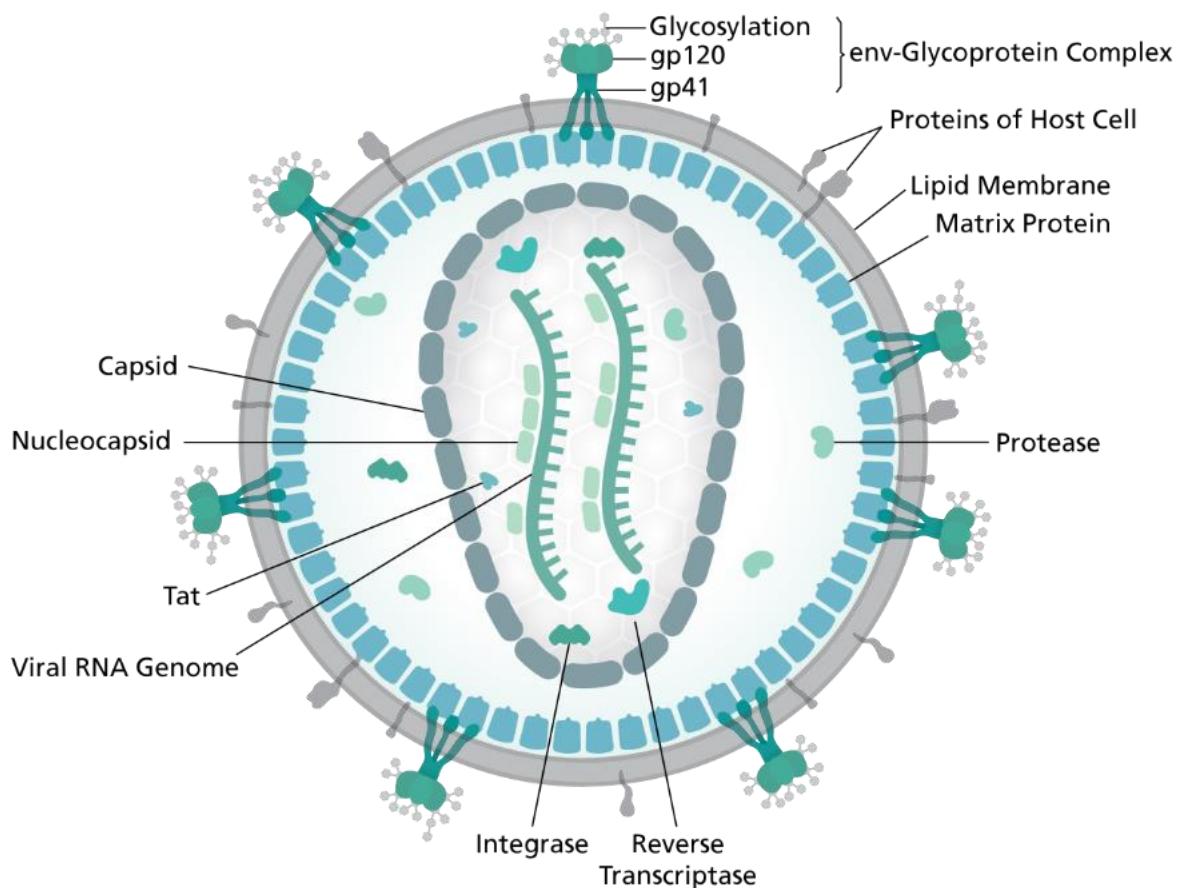


Figure 4. HIV-1 structure. Wikimedia Commons. Thomas Spletstoeser (www.scistyle.com)

3.1.5 Viral Groups and subgroups

HIV-1 is divided into different groups, M, N, O and P based on genetic similarities. Group M is in turn divided into subgroups A, B, C, D, AE, F, G, H, J and K (Eberle & Gurtler, 2012). In Figure 5 there is a Phylogenetic Tree of the SIV and HIV viruses showing the relation between the different groups and subgroups. Recombinant HIV are named CRF (circulating recombinant form) and may originate from a range of different subtypes. In 2016 there were more than 70 different circulating CRFs (Seitz, 2016). HXB2 is commonly used as a reference genome when comparing different subtypes and belongs to group M, subgroup B (NCBI Taxonomy Browser, 2020). Group M is the most common group and is responsible for the majority of the global infections (Tough & McLaren, 2019). In the European population subtype B is the most common subtype (Désiré, et al., 2001). In 2016 there was only approximately 15 cases of HIV-1 group N and 2 cases of HIV-1 group P worldwide (Seitz, 2016).

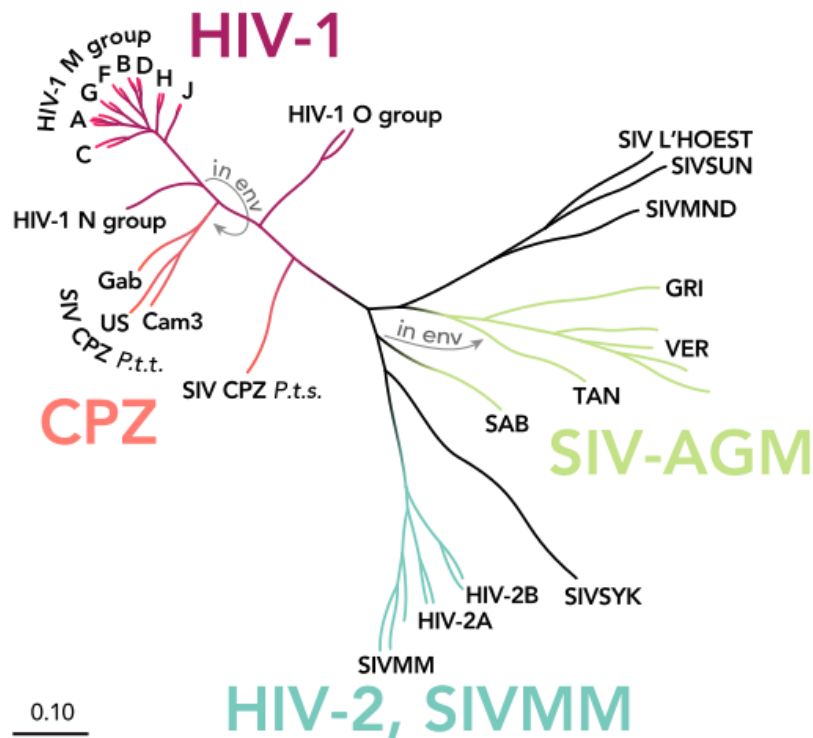


Figure 5. Phylogenetic Tree of the SIV and HIV viruses. Wikimedia Commons. Thomas Spletstoeser (www.scistyle.com)

3.1.6 Genetic variation

HIV has a large genetic variation. The reverse transcription polymerase is error-prone during the transcription from RNA to DNA with an error rate of approximately 1 in 10000 nucleotides. HIV has a rapid replication rate creating mutations fast (Tough & McLaren, 2019). The reverse transcription polymerase also prefers sometimes to insert guanine (G) instead of adenosine (A). Deletion and insertion of DNA is also occurring frequently by the reverse transcription polymerase (Eberle & Gurtler, 2012). An additional common factor that broaden the genetic variation of retroviruses is recombination. Recombination is when the polymerase uses RNA from two genetically different HIV-1 viruses in the infected host cell to create the DNA. The result is that the new HIV-1 particles have a mix of RNA from two genetically different HIV-1 sources. Recombination happens to 1 of 400 HIV particles in a host cell (Eberle & Gurtler, 2012).

To avoid detection of the host cells immune system the virus has some strategies. One strategy is to have a large genetic variation. The *env* gene has the largest genetic variation (Santoro & Perno, 2013). Both gp120 and gp41 have a large degree of mutations, especially the V3 loop, and their functions are not sensitive for mutations. The large genetic variation is needed for the gp120 protein to not be detected by antibodies (Mouquet, et al., 2011). The *pol* gene has lower variability because it encodes the viral proteins that are necessary for the virus (pr and RT). The *gag* gene is more conserved than the *env* gene but less than the *pol* gene. Other examples of highly conserved parts of the HIV genome is the long-terminal repeats (Eberle & Gurtler, 2012).

3.2 qPCR

The first step in a qPCR reaction with HIV is a Reverse Transcript (RT)-step that converts viral RNA to cDNA. This is because the HIV genome is constituted of RNA and RNA has a less stable structure than DNA (Lodish, et al., 2000). After the RT-step denaturation, annealing and extension is run in cycles and the DNA amplifies, see Figure 6.

Denaturation is normally performed at 94-96°C where the DNA's double strands separate because of the high temperature. The primers and probe (P/P) set normally connects to the DNA strands at a temperature of 58-60°C and that is when the DNA polymerase starts to extend the primers (Brown, 2016).

Taq polymerase is a thermostable DNA polymerase that is commonly used in PCR reactions. The Taq polymerase was identified from *Thermus aquaticus* a bacterium extracted from Yellowstone National Park in Montana, USA (Ishino & Ishino, 2014).

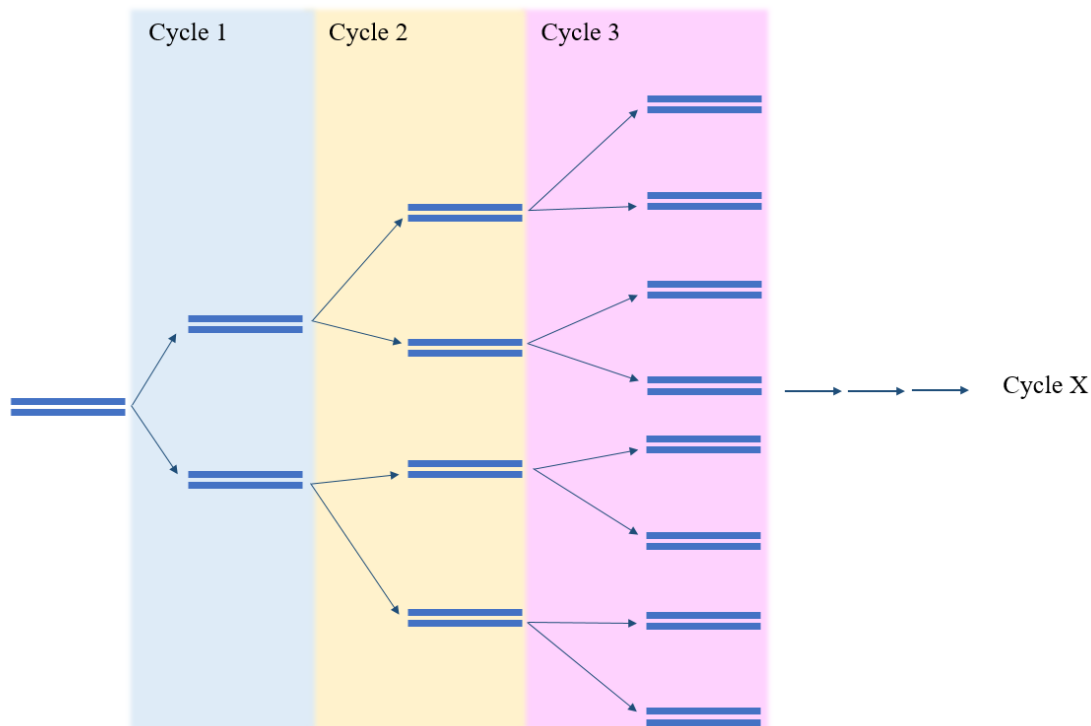


Figure 6. qPCR DNA amplification. The blue lines are the cDNA and after every cycle the amount of DNA is duplicated.

3.2.1 Real time Taqman qPCR

The Taqman qPCR method is the earliest method for monitoring PCR in real time. The method can be used when quantifying RNA or DNA. A good thing with this method is that it only requires one probe. The probe used in a Taqman assays has a fluorophore at the 5' end and a quencher at the 3' end (Heid, et al., 1996).

The Taqman qPCR viral DNA amplification is illustrated in Figure 7. The probe is binding to the upper DNA strand before the two primers binds. The forward primer binds to the upper DNA strand and the reverse primer to the lower strand. Taq polymerase can only add a

nucleotide to a pre-existing 3-‘OH group. The primers have a free 3-‘OH group that the DNA polymerase uses to start the synthesis from (Heid, et al., 1996).

When the probe is intact a quencher prevents the emission of light from the fluorophore of the reporter. During synthesis of the new DNA strands the probe gets hydrolyzed by the Taq polymerase. This is because the Taq polymerase also has an exonuclease property and can remove the probe from the DNA strand by cleaving the probe (Roche, 2001). The reporter is separated from the quencher and is allowed to emit light (Enzo Life Sciences, Inc, 2020).

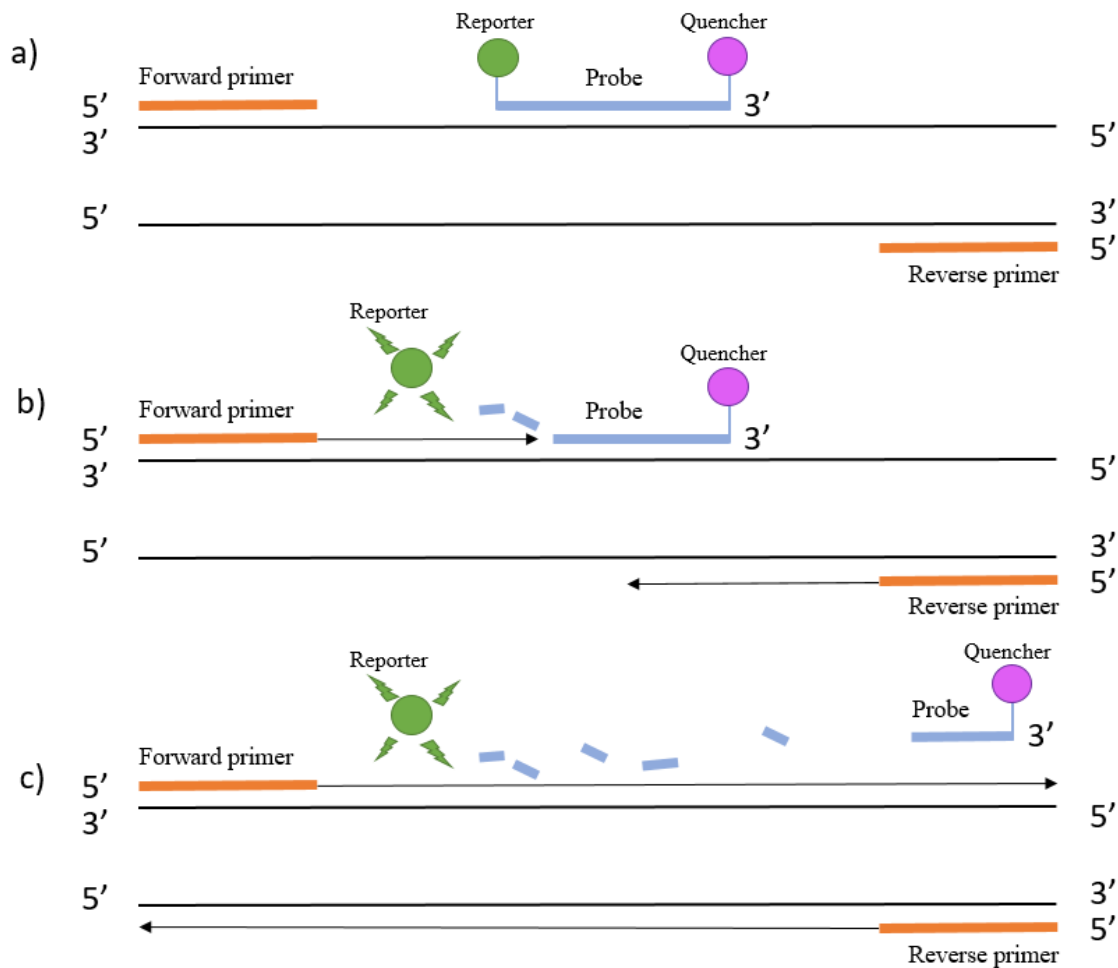


Figure 7. Taqman qPCR reaction. A) The probe binds to the upper strand. The forward primer binds to the upper strand and the reverse primer binds to the lower strand. B) The polymerase starts to produce the new strands and the reporter (fluorophore) is released from the quencher and emits light. C) The polymerase has synthesized two new DNA strands.

A normal thermal cycle profile for Taqman qPCR has a denaturation step at 95°C and annealing step at 60°C (Thermofisher, 2020).

Crossing point (cp)-value is the cycle number at detection threshold. The value is obtained by the LightCycler quantification analysis software (Roche, 2001).

3.2.2 Primer and probe design

The melting temperature (T_m) is where 50% of the nucleotide bases are unpaired (Figure 8).

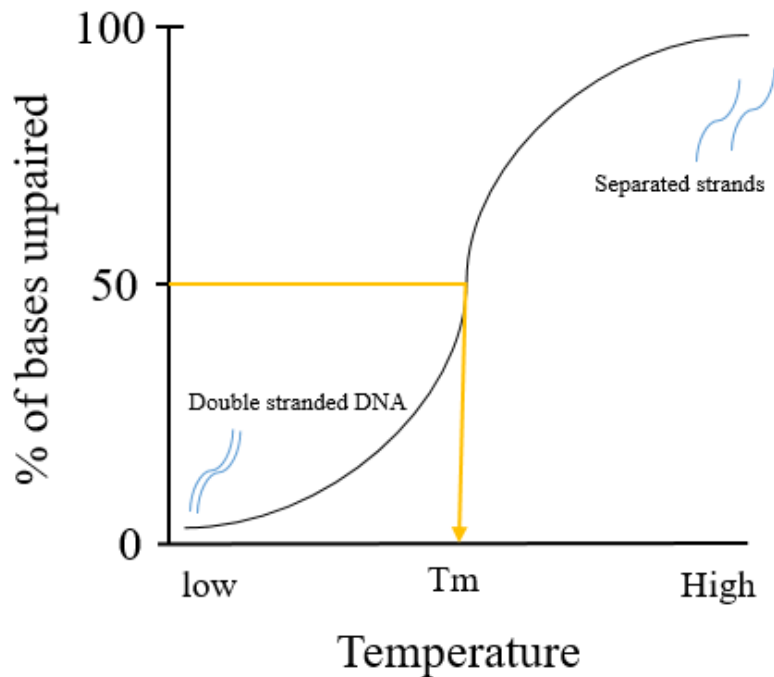


Figure 8. T_m is the temperature where 50% of the bases are unpaired

ThermoFisher recommends that the forward and reverse primers should have a T_m that differs by maximum 1°C. They also recommend that the primers should have a T_m between 58-60°C and a GC content of less than 50% (Thermofisher, 2020). Hydrolysis probes should have a T_m which is 7 to 10°C higher than the primers. The higher T_m ensures that the probe binds to the target before the primers binds in and begin the extension (Roche, 2001).

The GC content affect the T_m of the primer or probe because GC pairs have three hydrogen bonds, whereas AT pairs have two. Hence, the higher content of GC pairs. the higher T_m . T_m is also affected by the length of the primer or probe where a longer primer or probe gives a high T_m and a shorter gives a low T_m (Yuryev, 2007). The primers should have a length of 20 to 24 nucleotides and the probe 20 to 30 nucleotides (Roche, 2001). The length of the primer or probe also effects the specificity. A longer primer or probe have a higher specificity (Integrated DNA Technologies, Inc, 2020).

Primers and probes can have problems with formation of hairpins, self-dimerization, and hetero dimerization. Formation of hairpin occurs when the primer or probe sequence is complementary and can bind to its own other end, forming a loop. Hairpins prevents primers or probes from binding to the target sequence. To avoid problems with formations of hairpins it is important to design the primer or probes with not too many complementary bases in a row at two different sides of the primer or probe. Self-dimers form when two primers of same version are complementary to another copy of the same primer which results in bonding of two identical primers. According to the company Integrated DNA Technologies (IDT) the

calculated delta G (Gibbs free energy) value for the primer or probes ability to form dimers should be lower than the threshold value, -9kcal/mole or there is a risk for problems with self-dimerization. Hetero-dimers forms when two different primers or probes are binding to each other. According to IDT the calculated delta G value should not be lower than -9kcal/mole (Integrated DNA Technologies, Inc, 2020).

It is not recommended to design a primers or probes with more than three G repeats because they can form a tetrad secondary structure (Integrated DNA Technologies, Inc, 2020). Sequences that contain many G repeats forms complex and intrastrand foldings due to the hydrogen bonds between the G repeats at their N-7 ring position (Jensen, et al., 2010). This stops the primer or probe from binding to the DNA target.

3.2.3 Betaine

Betaine is a modified amino acid, which at pH 7, is both negatively and positively charged. The substance helps to equilibrate the different melting temperatures when AT and GC bases are binding (Jensen, et al., 2010). To reduce the risk for dimerization, betaine can be added to the mater mix. The melting temperature of the P/P set may decrease between 1 – 5°C when using betaine (Ampliqon A/S, 2020).

4 Method

4.1 Extraction of HIV-1 RNA

The starting material used in the project was HIV-1 positive plasma aliquotes with 201 international units (IU) HIV-1 RNA/mL (hereafter referred to as Pos A) or 36 IU/mL (hereafter referred to as Pos B). One International Unit can be converted to 0.58 RNA copies (Mossoro-Kpinde, et al., 2016). The Pos B control was extracted by certified personnel in the current routine analyse of HIV-1. Six hundred micro litres of plasma with Pos A was added to each extraction well. In parallel plasma without HIV-1 particles (negative plasma), was extracted and later used as a negative control. Subsequently, 50 μ L of internal control (IC), constituted of Bovine viral diarrhoea virus (BVDV), was added into each well. The BVDV is an enveloped RNA virus like HIV and is therefore suitable as IC. (Khodakaram-Tafti & Farjanikish, 2017). IC was co-extracted together with the positive and negative controls. The RNA was extracted with a Nuclisens easyMAG-instrument (BioMérieux) using elution and lysis buffers from NucliSENS® easyMAG® reagents. The first step was to lys virus particles. After that extraction of RNA with magnetic silica was performed. The RNA was binding to the magnetic beads and the beads was washed in wash buffer before they were dried and added to elution buffer that allowed the RNA to detach from the beads and go into the elution buffer solution. The final volume of eluate from Pos A in each well was 35 μ L and the final volume of eluate from Pos B in each well was 110 μ L. Given the HIV-1 concentration in the two different positive controls before extraction and the concentrating effect in the eluation eluate from Pos A had a concentration of 3.4 IU/ μ L and eluate from Pos B had a concentration of 1.6 IU/ μ L. Maximum 30 min after the extraction the eluate was placed in an eluate well plate and stored at +2 - +8 °C for a maximum of 72 h.

4.2 Primer and probe design

A literature study of articles regarding conserved parts of the HIV-1 genome was showing four conserved regions that could be good PCR targets. The poly-purine tract is located on the *nef* gene at around 9000 nucleotides on the reference sequence HXB2 (Los Alamos national laboratory, 2014). This is the integrase attachment site and is highly conserved. The *gag* had some conserved parts around 800 bp because it is the fourth stem loop of encapsulation signal and has a major role in the RNA packaging (Ngandu, et al., 2008). The *pol* gene has two conserved regions, one around 2000 bp and another around 4000 bp. The 2000 bp region is conserved because of the interprotein linker region with multiple helices exist there. The 4000 bp is conserved because there is a central poly-purine tract, splicing sites and overlapping open reading frames (Skitttrall, et al., 2019). The literature study results were combined with looking in the HIV-Sequence Compendium 2013 (Abfalterer, et al., 2013) to find conserved regions to use for the P/P by comparing the HXB2 sequence with example sequences from group M, N and O. The publicly available computer program Primer3 (<http://bioinfo.ut.ee/primer3/>; 2020-02-06) was used to choose location and lengths of the P/P. The objective was to design primers and probes with Tm as described under the Purpose

section and with locations within conserved parts of HIV-1 genome. The publicly available computer program OligoAnalyzer (<https://eu.idtdna.com/calc/analyzer>; 2020-02-06) was used to investigate potential risks of hairpin formation, self-dimer and hetero-dimerization. A comparison with existing P/P from the Hepatitis B and E routine analysis was made in OligoAnalyzer to be able to compare. To investigate if the primer/probe could match with sequences other than HIV-1 the publicly available computer program BLAST (<https://blast.ncbi.nlm.nih.gov/>; 2020-02-06) and the publicly available computer program primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>; 2020-02-06) were used. OligoAnalyzer was used to see how T_m can be affected by variation in the concentrations of Na⁺, Mg⁺⁺ and oligo. The settings in the program were qPCR and DNA.

4.3 Primer test

Master mix containing TaqMan Fast Virus 1-Step Master Mix, one of the designed HIV-1 P/P sets and water was used. The enzyme concentration was kept constant at 5 µL/PCR well in all experiments and the HIV-1 P/P concentrations were varied between experiments. The total reaction volume used was 20 µL in each tested well on a LightCycler® 480 Multiwell Plate, here after referred to as plate, of which 15 µL was master mix and 5 µL was eluate from the extraction. In one experiment with pol4000, 10 µL of master mix and 10 µL of eluate was used. IC P/P set targeting BDVD from routine analysis of hepatitis E virus was added into the master mix. The IC was amplified in parallel with HIV-1 to see if the PCR reaction worked. The plate was sealed and centrifuged for 2 min at 3000 rpm. PCR reactions were run using a light cycler 480 (LC480; Roche) with a PCR program described in Table 1. The excitation was set to 465 nm and the detection was set to 510 nm. In the initial step at 50°C RT converted HIV-1 RNA to cDNA. After the initial step 55 cycles of denaturation, annealing and extension was performed. One cycle contains a denaturation step at 95°C and an annealing and extension step at 60°C. The program ended with a cooling step. The LC480 started to plot the fluorescence data after the 10 first thermo cycles.

Table 1. PCR program used on the Light Cycle 480 instrument.

Step	Target	Acquisition Mode	Hold	Ramp Rate (°C/s)
RT-step	50 °C	None	15 min	4.4
Denaturation/activation	95 °C	None	2 min	4.4
Amplification	55 cycles (10 pre-cycles + 45 cycles)			
Denaturation	95 °C	None	10 sec.	4.4
Annealing/extension	60 °C	Single	35 sec.	2.2
Cooling	37 °C	None	30 sec.	2.2

4.3.1 First evaluation of the P/P sets

The first step was to see if the P/P sets could amplify the HIV-1 positive control, Pos B (8 IU/PCR well). A primer concentration of 300 nM or 600 nM and a probe concentration of 300 nM were used, see Table 2.

Table 2. First evaluation of P/P sets.

P/P set	Genomic region	C _{Primers} (nM)	C _{Probe} (nM)	C _{Positive control} (IU/ml)
pol4000 ¹	<i>pol</i> gene ~4000bp	300	300	8
pol2000	<i>pol</i> gene ~2000bp	300	300	8
nef9000FP1	Nef gene ~9000bp	600	300	8
nef9000FP2	Nef gene ~9000bp	600	300	8

¹ *Pol* is the targeted gene and 4000 is approximately the base pair location for the targeted sequence

After evaluation of the results from the first experiments the two best P/P sets was chosen for optimization. To investigate if the signals could be increased eluate from Pos A was tested. As in previous experiments the same concentrations of P/P were used, 600 nM primers and 300 nM probe.

The primer concentration was increased to 900 nM to assess if the fluorescence would increase. The Pos A control was used. The probe concentration was tested at 300 nM and 450 nM, see Table 3.

Table 3. Optimization of the two best P/P sets.

Region	C _{Primers} (nM)	C _{Probe} (nM)
pol4000	900	300
nef9000FP2	900	300
pol4000	900	450
nef9000FP2	900	450

4.3.2 Internal control (IC)

To evaluate if the P/P set, pol4000, could amplify HIV-1 when IC P/P also was present in the master mix a comparison was made between with and without IC P/P in the master mix. As previously mentioned BVDV was used as the IC. The stock material was produced by the Viral department of Biovitrum and derives from a Madin-Darby bovine kidney (MDBK) cell culture supernatant. The concentration of BVDV was given as 50% tissue culture infectivity dose (TCID₅₀). In each aliquot there was 6.5 x 10³ TCID₅₀/ml. The eluate came from Pos A and the amount of HIV-1 in each PCR well was 17 IU (Table 4). The acceptance criteria of the HEV routine analyse method used by Octapharma AB was a fluorescence higher than 3 and a cp-value lower than 20 cycles.

Table 4. IC experiment.

P/P set	C _{Primers} (nM)	C _{Probe} (nM)	IC P/P set
pol4000	900	450	no
pol4000	900	450	yes

4.3.3 Probe concentration optimization

The best P/P set was chosen for a further study of the optimum probe concentration. The probe concentration was tested at four levels, 150, 450, 600, and 900 nM, see Table 5. The eluate came from Pos A and the amount of HIV-1 in each PCR well was 17 IU.

Table 5. Optimization of probe concentration. The tested probe concentrations was 150nM, 600nM and 900nM.

P/P set	C _{Primers} (nM)	C _{Probe} (nM)
pol4000	900	150
pol4000	900	450
pol4000	900	600
pol4000	900	900

4.3.4 Further primer concentration optimization

To see if the assay could be further improved with higher primer concentrations an experiment with 1350 nM primer and 900 nM probe was performed (Table 6). Pos A was used as the positive control.

Table 6. Further optimization of the P/P set pol4000.

P/P set	C _{Primers} (nM)	C _{Probe} (nM)
pol4000	1350	900

4.3.5 Betaine

To assess if betaine could enhance the pol4000 P/P set performance, betaine was added to the master mix in one experiment. Three micro litres Betaine per PCR reaction with a concentration of 5 M was added to the master mix in one experiment (Table 7). Eluate from Pos B was used.

Table 7. Betaine test

P/P set	C _{Primers} (nM)	C _{Probe} (nM)	Betaine 5M (µl)
pol4000	600	300	3

4.3.6 Double eluate

To evaluate if adding more eluate to the PCR wells would give stronger fluorescence signals double amount of eluate was added into each PCR well. Ten micro litres of eluate from Pos B was added in each PCR well instead of 5 µL. Hence, the amount of HIV-1 in each PCR well doubled to 16 IU. The primer and probe concentration was set to 600 nM and 300 nM respectively, see Table 8. The amount of added water in the master mix was reduced by 5 µL

to achieve a volume of 10 μL master mix in each PCR well. The total volume in each PCR well was 20 μL , 10 μL master mix and 10 μL eluate.

Table 8. Double eluate from PosQ612A

P/P set	C _{Primers} (nM)	C _{Probe} (nM)	Eluate
pol4000	600	300	10 μL

4.4 Genotype test

To assess if the po4000 P/P set could detect a broad range of groups and subtypes of HIV-1 a genotype panel of the groups O, N and M were tested, and of the M group subtypes A, B, C, D, AE, F, G, and H were tested. The primer concentration was set to 900 nM and probe concentration to 900 nM based on the previous optimization experiments. Prior to the extraction the genotype samples had a HIV-1 concentration of 730 IU/mL. Two samples of each genotype were extracted and then the eluates of the duplicates were pooled together in pairs. Given the HIV-1 concentration in the sample before extraction and the concentrating effect in the eluation the eluates had a concentration of 12.5 IU/mL. After 5 μL of eluate was added to each PCR well the total amount of HIV-1 in each PCR well was 63 IU. Two Pos A and two negative controls were extracted in the same run as the samples and pooled in pairs. The amount of HIV-1 in each Pos A PCR well was 17 IU.

All genotypes and the positive control were analysed with PCR in two runs in quadruplicates and in duplicates.

To evaluate how many mismatches the different genotypes had with the P/P set a comparison was made between the nucleotide sequences of the P/P set and each of the genotypes tested. Sequences from the different HIV-1 groups and subtype used were taken from HIV Sequence Compendium 2013 (Abfalterer, et al., 2013). An average mismatch value was calculated for every genotype by adding the total number of mismatches for the investigated genotype and divide them by the number of sequence that the genotype has. Three independent assessors were deciding how many detected positive replicates there were on each genotype.

4.5 Statistical analysis

All statistical analyses were made in Excel (Microsoft).

ANOVA was used in comparisons between several groups at a time and t-test for comparisons of two groups at a time. The variance in the group should not be larger than the variance between the groups for a significant result from an ANOVA test (Fraiman & Fraiman, 2018).

The equation for calculating the t-test with unequal sample sizes and unequal variances can be seen in equation 1. X is the average value for the group, S is the variance for the group and N is the number of samples in the group.

$$t = \frac{X_1 - X_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (\text{Equation 1})$$

The result from the t test was compared to the critical t-value and the degrees of freedom to estimate the p-value (Kim, 2015).

Regression analysis was performed on the results from the genotype experiment to see how much the independent variable affected the dependent variable. The R^2 value was used to see how much the independent variable could explain the results. The p-value was used to see if the effect of the independent variable was statistically significant (Schneider, et al., 2010).

5 Results

5.1 Primer design

Four target sequences with conserved parts were chosen: the *pol* gene at around nucleotide position 2000 in the reference sequence HXB2 (Los Alamos national laboratory, 2014), the *pol* gene at around nucleotide position 4000, the *nef* gene at around nucleotide position 9000 and the *gag* gene at around nucleotide position 800, hereafter referred to as pol2000, pol4000, nef9000 and gag800 respectively.

The selected P/P can be found in Table 9. Within the target sequence in the *gag* gene at the desired position for the probe for gag800 the various subtypes contained a high degree of mismatches, compared to the HXB2 sequence. The pol2000 probe could not obtain a T_m around 65-68°C like the other probes because of the low GC content in the region. The pol4000 probe and the nef9000 probe had the highest temperature regarding hairpins. The hairpin temperature is the T_m of the most stable hairpin structure that the primer or probe can create.

Table 9. Potential primers created with primer3. Conservation rate (%) is calculated in Excel by comparing the selected primer or probe sequence with the HXB2 sequence in the same region, data from the HIV Sequence Compendium 2013. The hairpin temperature is the T_m of the most stable hairpin structure that the primer or probe can form.

P/P set	Type	T _m (°C)	Conservation rate (%)	Hairpin (°C)
pol2000	Forward primer	58.65	90	0
	Reverse primer	58.64	95	0
	Probe	59.37	83	0
pol4000	Forward primer	59.31	90	0
	Reverse primer	59.23	89	0
	Probe	67.37	70	41.72
nef9000FP1	Forward primer	58.46	95	0
	Reverse primer	59.24	89	0
	Probe	65.68	93	37.36
gag800	Forward primer	58.51	50	0
	Reverse primer	58.87	95	0
	Probe	60.24	100 to subtype B	36.05

Further analysis with OligoAnalyzer showed that the reverse primer in the *nef* gene, nef9000, had a self-dimerization value below the threshold ($\Delta G < -9\text{kcal/mole}$), see Table 10. The lower the value, the higher is the risk for self-dimerization. (Integrated DNA Technologies, Inc, 2020)

According to IDT (Integrated DNA Technologies, Inc, 2020) the ΔG value should be above a threshold of -9kcal/mole or else there is a risk for problems with hetero-dimerization. The primers in the *pol* gene, pol4000 had a ΔG value of -15.12 and -10.81 which is lower than the threshold. Therefore, there is a risk for problems with hetero-dimers when using pol4000.

Table 10. Results from OligoAnalyzer. A low value (delta G) indicate a risk for hairpin, self-dimer and hetero-dimer to occur.

P/P set	Type	Hairpin (kcal/mole)	Self-dimer (kcal/mole)	Hetero-dimer (kcal/mole)
pol2000	Forward primer	- 0.48	- 4.67	- 4.67
	Reverse primer	- 0.30	- 3.07	- 7.44
	Probe	- 0.08	- 3.14	- 4.67
pol4000	Forward primer	0.48	- 7.05	- 10.81 ¹
	Reverse primer	0.18	- 3.14	- 15.15
	Probe	- 3.32	- 7.44	- 5.49
nef9000FP1	Forward primer	- 0.08	- 1.6	- 7.74
	Reverse primer	0.87	- 9.28	- 7.93
	Probe	- 1.23	- 5.7	- 7.74

¹ Numbers in italic lies above the critical threshold for hairpin formation as suggested by IDT.

A search using BLAST showed that all P/P matched only with HIV-1 genome except for the left primer in the pol2000 P/P set which, except for HIV-1, also matched with the species *Neostethus bicornis* (fish), *Struthio camelus australis* (southern ostrich), *Culicoides sonorensis* (blood midges) and *Myripristis murdjan* (fish).

The primer-BLAST search showed that both pol4000 and nef9000FP1 primers and probes could potentially bind to human DNA.

The left primer that targeted the *nef* gene was difficult to manufacture due to the fact it contained more than three G repeats and therefore a new forward primer had to be designed. The new primer is presented in Table 11.

Table 11. Potential primer to replace the forward primer in the *nef* gene. The primer is created with primer3. Conservation rate (%) is calculated in Excel by comparing the selected primer or probe sequence with the HXB2 sequence in the same region, data from the HIV Sequence Compendium 2013 (Abfalterer, et al., 2013). The hairpin temperature is the calculated value of the T_m of the most stable hairpin structure that the primer or probe can form.

P/P set	Type	T _m (°C)	Conservation rate (%)	Hairpin (°C)
nef9000FP2	Forward primer	57.65	80	39.3

A further analysis with OligoAnalyzer was made to investigate the primers ability to form dimers and hairpins, see Table 12. All values were higher than the lowest result from previous primers. Based on this the primer was a good substitute. A BLAST search was performed and showed only hits with HIV-1 sequences.

Table 12. Results from OligoAnalyzer. A low value (delta G) indicate a risk for hairpin, self-dimer and hetero-dimer to occur.

P/P set	Type	Hairpin (kcal/mole)	Self-dimer (kcal/mole)	Hetero-dimer (kcal/mole)
nef9000FP2	Forward primer	- 0.45	- 5.36	- 10.88 ¹

¹ Numbers in italic lies below the critical threshold for hairpin formation as suggested by IDT.

5.2 Primer test

5.2.1 First evaluation of the P/P sets

To see if the P/P sets could amplify HIV-1 a first evaluation was conducted with all four P/P sets. At a primer and probe concentration of 300 nM pol4000 did not amplify HIV-1 with Pos B as eluate. When increasing the primer concentration to 600 nM positive controls were detected. Nef9000FP2 also had the ability to amplify HIV-1 RNA at a primer concentration of 600nM. Pol2000 and nef9000FP1 showed no clear amplification and were therefore omitted from further analysis. The results are shown in Table 13. Negative control was used to see if the P/P set was amplifying human DNA instead of HIV. There was no amplification of the negative control in any of the tests.

Table 13. Results of the first evaluation.

P/P set	C _{primers} (nM)	C _{probe} (nM)	F _{pos} average ¹ ±SD ²	F _{max} ³	Detected positive %	Cp average ±SD ²
pol4000	300	300	-	1.22	0	-
pol2000	300	300	-	0.97	0	-
nef9000FP1	600	300	-	1.59	0	-
nef9000FP2	600	300	4.55±1.03	5.71	37.5	31.68±3.55
pol4000	600	300	2.76±0.63	4.02	75	23.35±0.44
pol2000	600	300	-	1.59	0	-

¹ The average fluorescence signal calculated from the detected positive controls

² Standard deviation

³ The maximum fluorescence signal

Changing the eluate to the Pos A (3.4 IU/μL) gave stronger fluorescence signals as compared to using eluate from the Pos B (1.6 IU/μL), see Table 14. The average fluorescence signal calculated from the positive detected controls were 10.71±7.72 for the P/P set nef9000FL2 and 4.53±0.91 for the P/P set pol4000 with eluate from Pos A. This can be compared to the results in Table 13 from the previous experiment when nef9000FL2 achieved an average fluorescence of 4.55±1.03 and pol4000 an average fluorescence of 2.76±0.63.

Table 14. Results of Pos A as eluate using nef900FL2 and pol4000 P/P sets.

P/P set	C _{primers} (nM)	C _{probe} (nM)	F _{pos} average ¹ ±SD ²	F _{max} ³	Detected positive %	Cp average ±SD ²
nef9000FP2	600	300	10.71±7.72	21.38	50	34.70±0.88
pol4000	600	300	4.53±0.91	6.60	66.7	23.77±0.17

¹ The average fluorescence signal calculated from the detected positive controls

² Standard deviation

³ The maximum fluorescence signal

When increasing the primer concentration to 900nM stronger fluorescence signals were obtained for both pol4000 and nef9000FP2, see Table 15. Increasing the probe concentration to 450nM resulted in an even stronger signal at 14.21±0.45 for nef9000FL2 and 5.76±0.50 for pol4000.

Table 15. Results of increased concentration of primers and probe.

P/P set	C _{primers} (nM)	C _{probe} (nM)	F _{pos} average ¹ ±SD ²	F _{max} ³	Detected positive %	Cp average ±SD ²
pol4000	900	300	5.76±0.50	6.51	67	23.77±0.065
pol4000	900	450	8.39±2.29	10.90	100	24.16±0.60
nef9000FP2	900	300	14.21±0.45	14.58	50	35.94±2.68
nef9000FP2	900	450	17.53±12.07	34.78	67	35.62±2.62

¹ The average fluorescence signal calculated from the detected positive controls

² Standard deviation

³ The maximum fluorescence signal

The cp-values for pol4000 in all experiments in the first evaluation were all between 23 and 25 cycles. Nef9000FP2 displayed higher cp-values, between 34 and 37 cycles (Figure 9). Due to the weak signal of the nef9000FP2 P/P it was excluded from further analysis and in the following experiments only pol4000 P/P was used.

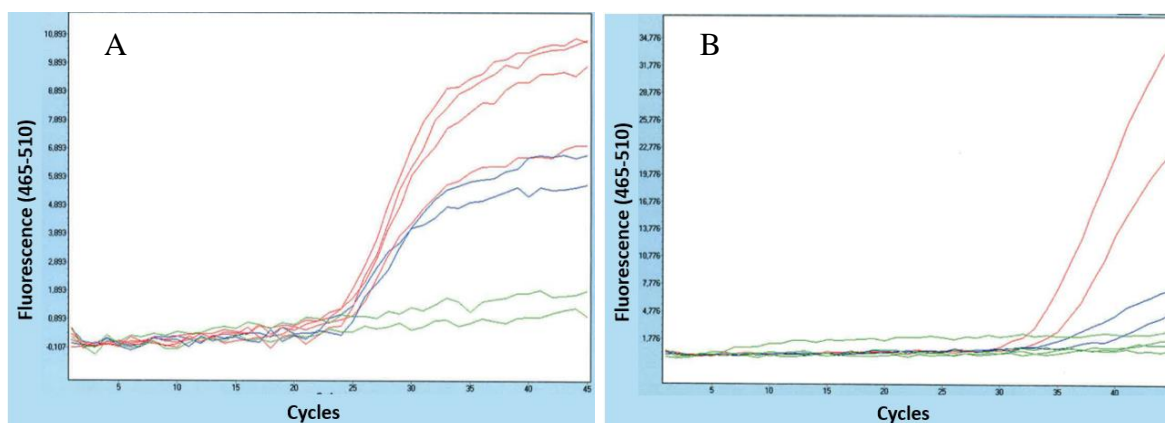


Figure 9. A) P/P set pol4000 with 900 nM primers and 450 nM probe. B) P/P set nef9000FP2 with 900nM primers and 450nM probe.

5.2.2 Internal control

To evaluate if the P/P set, pol4000, could amplify HIV-1 when IC P/P also was present in the master mix, IC P/P was added to the master mix in one of the two otherwise identical mixes. The maximum fluorescence from amplification of HIV-1 without IC P/P was 7.6 and with IC P/P 11.5. Eighty-three percentages of the HIV-1 positive control replicates were detected without IC P/P present in the master mix and 100% were detected when P/P targeting IC was present in the master mix. Without IC P/P the average fluorescence from detected positive controls was 6.51 ± 0.84 and the average cp-value from positive detected controls was 23.89 ± 0.38 . With IC P/P the average fluorescence was 6.01 ± 2.26 and the average cp-value from positive detected controls was 24.23 ± 0.58 (Figure 10).

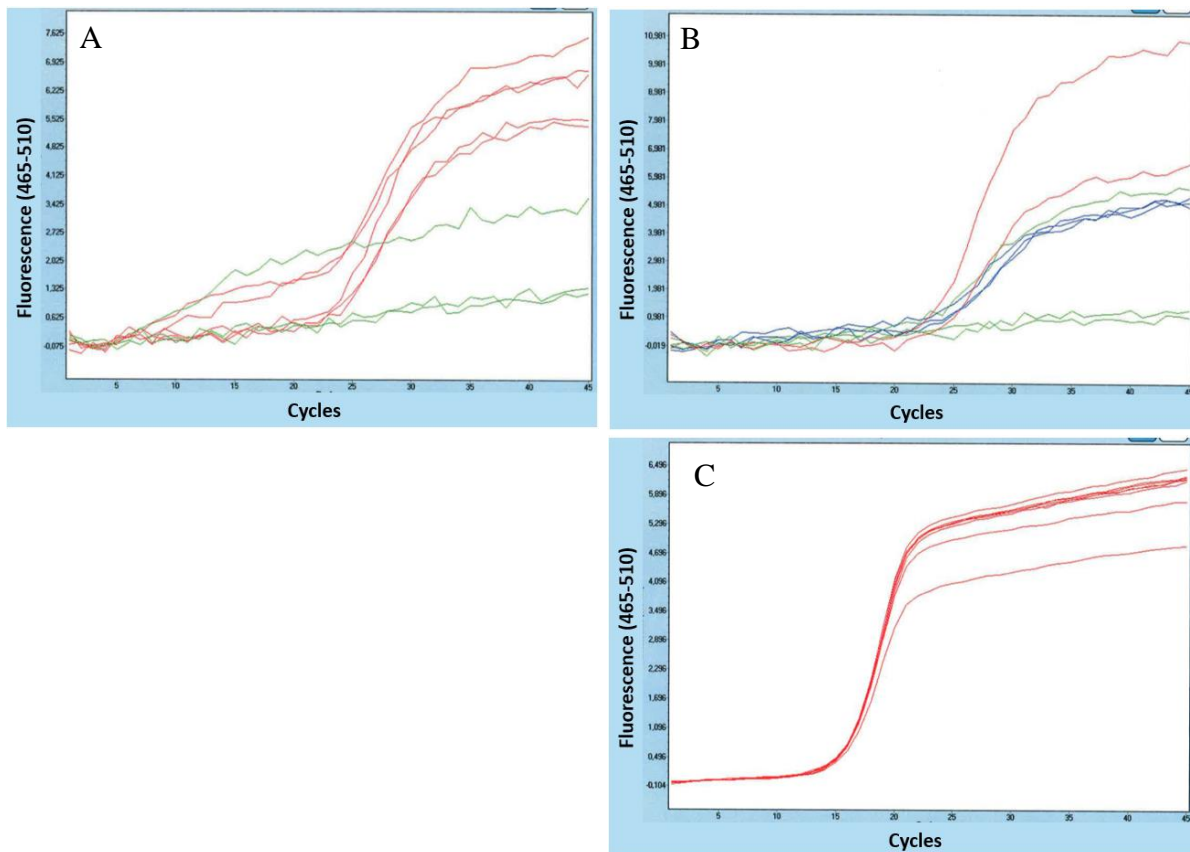


Figure 10. Both runs analyse the amplification of HIV-1 with pol4000 as P/P set. The runs has six positive and two negative samples and 900 nM primers and 450 nM probe. A) There was no IC P/P set in the master mix. Maximum fluorescence was 7.6. B) IC P/P set from HEV routine analysis was added into the mater mix. The IC results can be found in panel C. Maximum fluorescence was 11.5. C) Results from analysing the emitted fluorescence from the amplification of IC. All samples got a cp-value ranging between 15.47 and 15.67.

The cp-values for the IC amplification were between 15.47 and 15.67 (Figure 10 C). The IC results were approved according to acceptance criteria of the HEV routine analyse method (see section 4.3.2 Internal control (IC)).

5.2.3 Probe concentration optimization

To optimize the P/P set, pol4000, four different probe concentrations were tested, 150, 450, 600, and 900 nM. All probe concentrations were tested together with a primer concentration of 900 nM. The average florescence that the detected positive samples emitted are shown in Table 16. As expected, the experiment with 900 nM probe had the highest average florescence value. Whereas no signal from the positive controls was received for the experiment where 150 nM probe was used, all the positive controls were detected with probe concentration 450 nM. The graph from the assay with 450 nM of probe looks similar visually to the assay where 900nM probe was used but the statistical analysis showed that there was a difference (Figure 11).

Table 16. The results from the probe concentration optimization. Four concentrations of probe was investigated; 150, 450, 600, and 900 nM. The primer concentration was kept constant at 900 nM.

C_{probe} (nM)	Average fluorescence positive \pm SD ¹	Detected positive %
150	Not detected	0
450	6.23 \pm 2.15	100
600	5.74 \pm 1.61	50
900	10.16 \pm 1.44	92

¹ Standard deviation

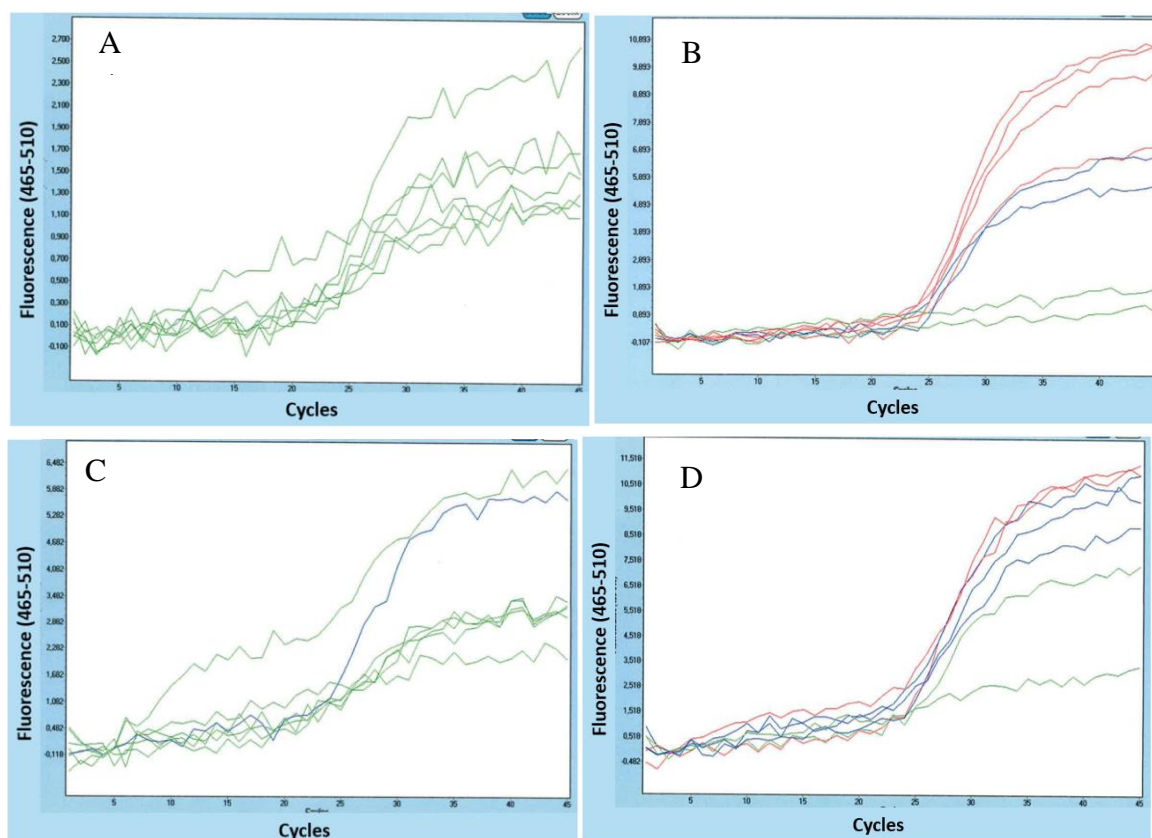


Figure 11. Six positive samples were tested using A) Assay with 900 nM primers and 150 nM probe and one negative control. B) Assay with 900 nM primers and 450 nM probe and one negative control. C) Assay with 900 nM primers and 600 nM probe and two negative controls. D) Assay with 900 nM primers and 900 nM probe and one negative control.

In Figure 12 the average fluorescents for the four different probe concentrations from Table 16 are plotted as bars. To investigate if there was a difference between the different probe concentrations an ANOVA test was conducted. The p value was less than 0.05 and showed that there was a significant difference between at least two concentration. To investigate which concentration of probe that gave the best result a t-test was conducted. The result of the t-test concluded that 900nM had significant higher fluorescence than 600 nM ($p < 0.001$) and significant higher fluorescence than 450 nM ($p < 0.01$). The t-test also showed that there was no significant difference between 450 nM and 600 nM when comparing the average positive

fluorescence. No statistical analysis was made with the 150nM concentration of probe because no positive control was detected with that concentration.

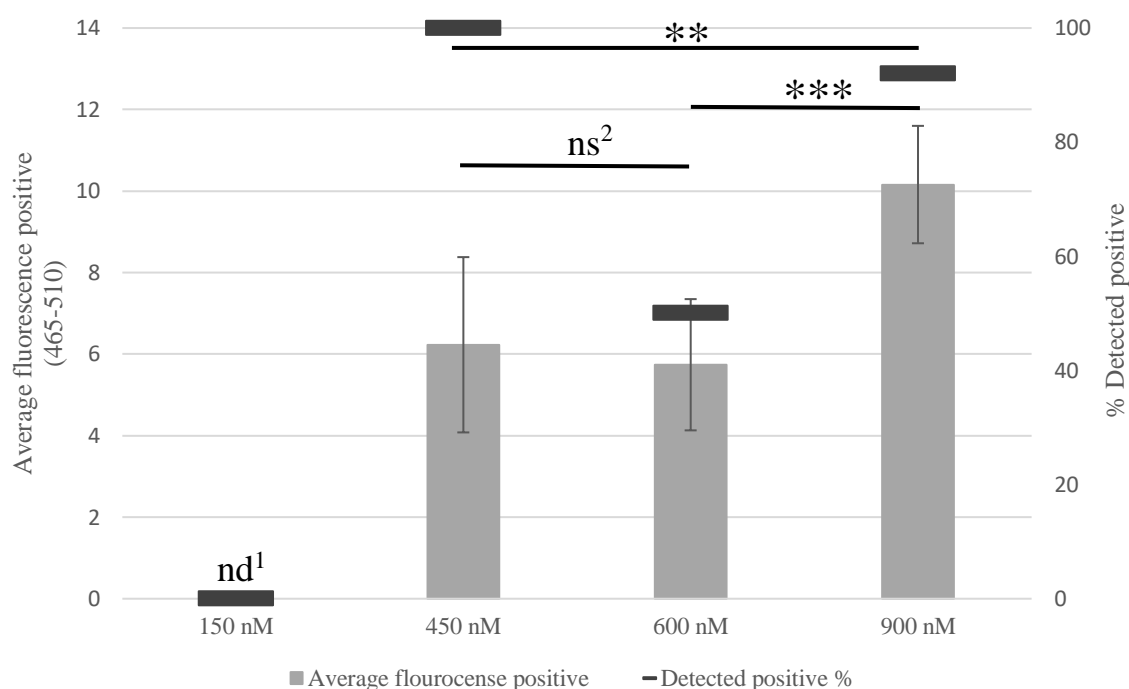


Figure 12. Average fluorescence at the detected positive controls indicated as bars. The rate of detected positive controls is shown as bold black lines. No positive controls were found with the probe concentration 150 nM. The concentration 900 nM had statistically significant stronger fluorescence than the other probe concentrations.

¹ Not detected

² Not significant

*p < 0.05, **p < 0.01, ***p < 0.001.

An ANOVA test showed that there were no significant differences in average cp-values between the different probe concentrations tested (p > 0.05) (Table 17). The probe concentration of 150 nM was not included in the ANOVA test because of no detected positive control.

Table 17. Different probe concentrations average cp-value. An ANOVA test showed that there was no significant difference between the different probe concentrations average cp-value.

C _{probe}	Average cp-value ±SD ¹
150 nM	Not detected
450 nM	24.27±0.52 ²
600 nM	23.73±0.67 ²
900 nM	24.46±0.56 ²

¹ Standard deviation

² Not significant

5.2.4 Further primer concentration optimization

The result of testing the primer concentration of 1350 nM with the probe concentration of 900 nM gave an 83% detection rate and an average positive fluorescence at 7.88 ± 0.72 with a maximum fluorescence at 8.75. Comparing the results to the previous results with 900 nM primer and 900 nM probe (Table 16) shows that there is no further improvement when increasing the primer concentration above 900 nM with a probe concentration of 900 nM.

5.2.5 Betaine

To assess if betaine could enhance the pol4000 P/P set performance with its ability to lower the risk for primer or probe dimerization (see section 4.2.3 Betaine) betaine was added to the master mix in one experiment. The result of the betaine test showed no amplification of any positive control. The maximum fluorescence was at 1.59.

5.2.6 Double eluate

Eluate from the positive control, Pos B, had a low concentration of HIV-1. To evaluate if adding more eluate to the PCR wells would give stronger fluorescence signals, 10 μL instead of 5 μL of eluate from Pos B was added to each PCR well. Double added volume of eluate showed lower fluorescence signals than with 5 μL eluate. The maximum fluorescence was at 3.20 (Figure 13).

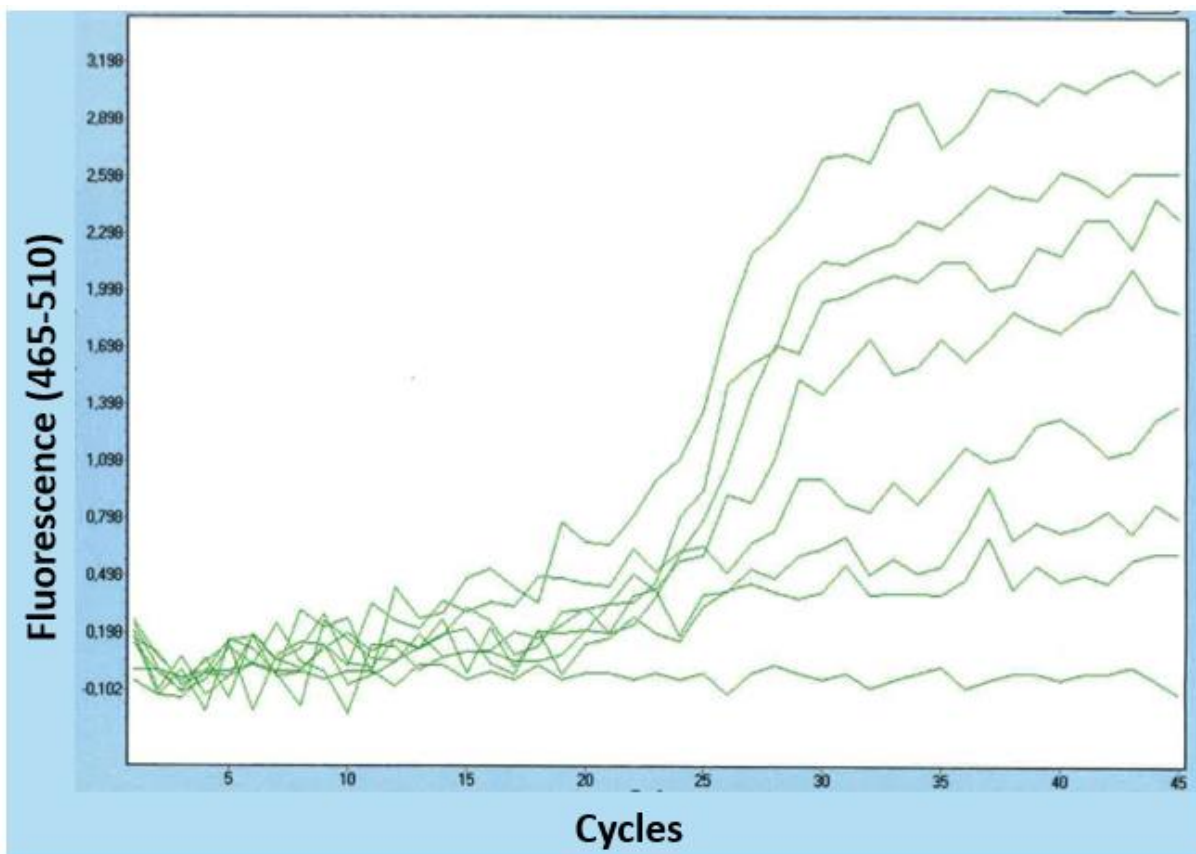


Figure 13. In this test the P/P set pol4000 with 600 nM primers and 300 nM probe was used. There was 10 μL of eluate from Pos B in each PCR well.

5.3 T_m analysis in OligoAnalyzer

As described in the method under section 5.2 primer and probe design, OligoAnalyzer was used to simulate how oligo, Na⁺, and Mg⁺⁺ concentration can effect T_m for the P/P set pol4000. Oligo concentration is the concentration of P/P. In Figure 14 the different oligo concentrations from 0.1 μM to 1.5 μM are plotted against T_m. The Na⁺, Mg⁺⁺ and nucleotide concentrations were kept constant at 50 mM, 3 mM and 0.8 mM respectively. A higher oligo concentration gave a higher T_m in the tested interval. At 900 nM, which was the P/P concentration with the best results in previous experiments, the probe had a T_m of 70.6°C, the forward primer had a T_m of 66.2°C and reverse primer had a T_m of 65.8°C.

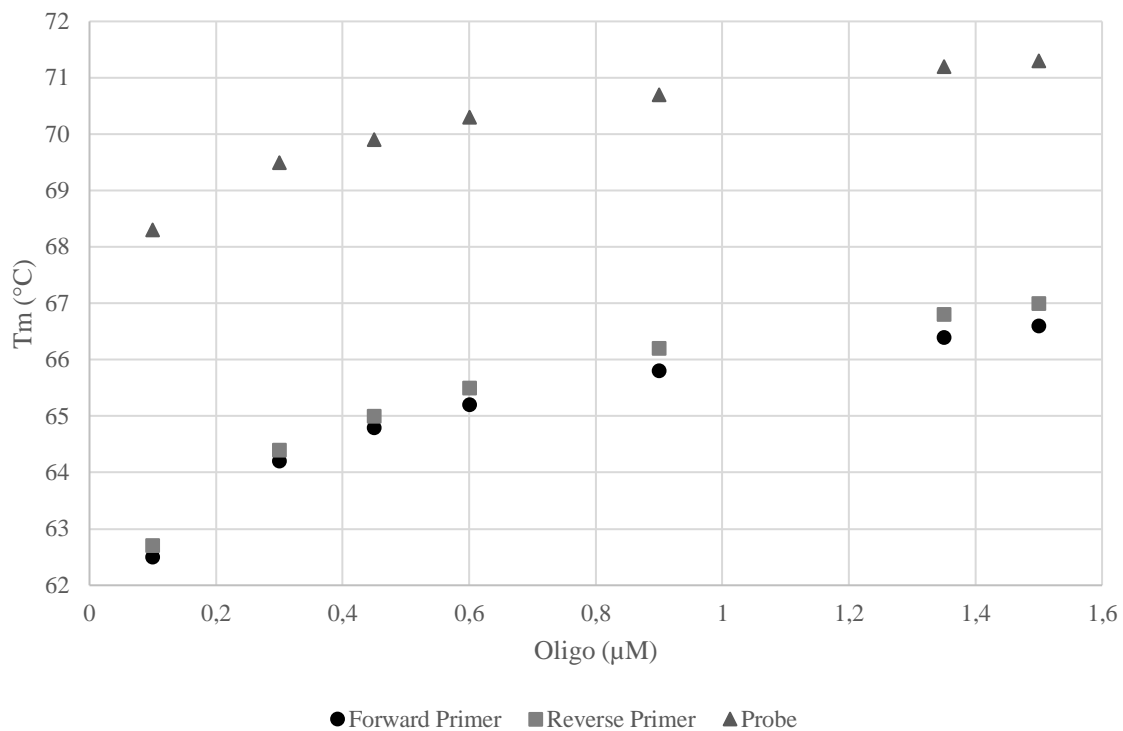


Figure 14. T_m at different oligo concentrations for forward primer, reverse primer and probe. The Na, Mg and nucleotide concentrations were kept constant at 50 mM, 3 mM and 0.8 mM.

An analysis of how Na⁺ concentration affected T_m for the P/P is presented in Figure 15 A. The oligo, Mg⁺⁺ and nucleotide concentrations were kept constant at 0.9 μM, 3 mM and 0.8 mM respectively. There is no difference in the T_m between 10 nM and 25 mM of Na⁺ but between 25 nM and 75 Mm there is a temperature increase. In Figure 15 B different concentrations of Mg⁺⁺ are plotted against T_m. The oligo, Na⁺ and nucleotide concentrations were kept constant at 0.9 μM, 50 mM and 0.8 mM. The T_m is increasing with a higher amount of Mg⁺⁺.

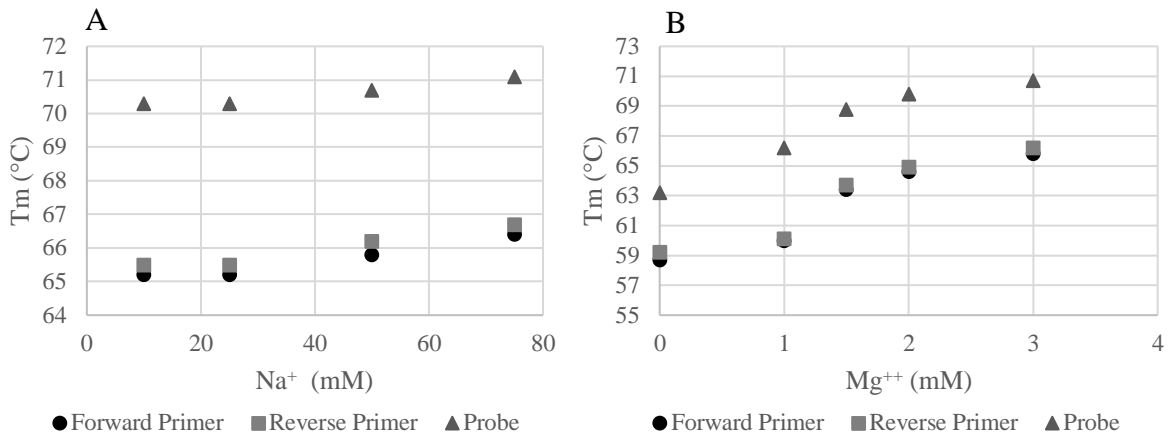


Figure 15. A) T_m for P/P set pol4000 at A) different Na^+ concentrations between 10 nM to 75 nM or B) different Mg^{++} concentrations between 0 nM to 3 nM.

5.4 Genotype test

To evaluate the ability of the pol4000 P/P set to detect a broad range of different genotypes of HIV-1 a genotype test was performed. There were three groups of HIV-1 tested, M, N and O and eight subgroups from group M, A, B, C, D, AE, F, G, and H. Two experiments tested all groups and subgroups. In the first experiment, there were four replicates of each genotype and positive control. In the second experiment only two replicates of every genotype and positive control was tested, see Figure 16 for the two experiments positive controls result. The maximum fluorescence was 4.5 in the first experiment and 4.7 in the second experiment. In the first experiment only 2 out of 4 positive HIV-1 controls were amplified but in the second experiment 2 out of 2 positive HIV-1 controls were amplified.

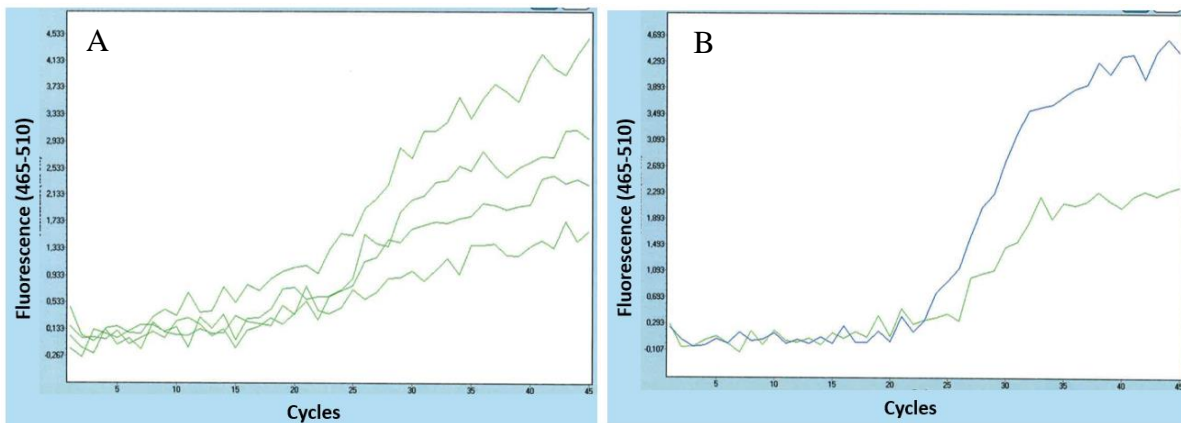


Figure 16. A) The positive control, Pos A at the first run. The maximum fluorescence was 4.5. B) The positive control, Pos A at the second run. The maximum fluorescence was 4.7.

All subtypes from group M had at least one sample that the PCR reaction had amplified HIV-1. No amplification was observed on the N and O group (Table 18).

Table 18. The rate of amplified replicates from the genotype experiment. The value on detected replicates is an average from the three independent assessor's results. There are 3 groups, M, N and O, and 8 subgroups from group M that have been tested, A, B, C, D, AE, F, G, and H.

Group	Subgroups	% Detected replicates
M	A	22.0
	B	61.3
	C	66.7
	D	39.0
	AE	50.0
	F	33.3
	G	39.0
	H	44.7
O	-	0.0
N	-	0.0

To investigate if the rate of positive detected replicates correlated with the total number of mismatches for the P/P set a comparison between the P/P in pol4000 and each genotype were made. The result of the count of the nucleotide mismatch between P/P set and genotype reference target is shown in table 19.

Table 19. Average number and nucleotide mismatches in P/P set compared to genotype reference.

		Average nucleotide mismatches in P/P set compared to genotype reference			
Group	Subgroups	Forward primer	Probe	Reverse primer	Total
M	A	0.5	2.7	0.3	3.5
	B	0.5	1.2	0.2	1.9
	C	0.3	5.0	0.0	5.3
	D	0.3	2.3	0.0	2.6
	AE	0.3	3.0	0.1	3.4
	F	0.5	2.2	0.1	2.8
	G	0.6	3.0	0.0	3.6
	H	0.3	2.5	0.0	2.8
O	-	3.0	4.6	2.9	10.4
N	-	1.7	5.0	1.0	7.7

Three independent assessors were deciding how many detected positive replicates there were on each genotype. The detected replicates in percent from each assessor were plotted against the average nucleotide mismatches for each genotype from Table 19 (Figure 17). The regression analysis showed that 27.5% of the detection result could be explained by nucleotide mismatches with a p-value below 0.01.

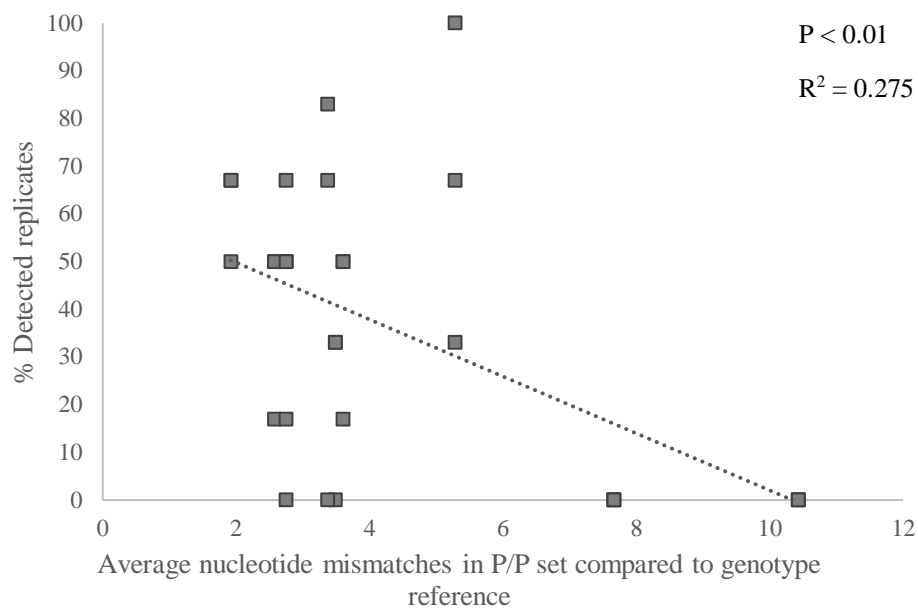


Figure 17. There are 3 groups, M, N and O and 8 subgroups from group M that has been tested, A, B, C, D, AE, F, G, and H. Average nucleotide mismatches in P/P set compared to genotype reference has been calculated for each genotype (Table 19). The detected replicate rate for each genotype has been analysed by three independent assessors and plotted against the number of mismatches. This generated 30 points on the graph. A regression analysis was conducted and showed that 27.5% of the detection result could be explained by nucleotide mismatches with a p-value below 0.01.

6 Discussion

The aim of this study was to develop a P/P system for a novel quantitative PCR analysis method for detection of HIV-1. First, four target regions were identified using the results from a literature review and a review of the HIV compendium (see section 6.1 Primer design). Of the four initial P/P sets, three were chosen for further tests in wet lab. The P/P set targeting the *gag* gene, gag800 was excluded since no probe could be designed within the PCR fragment to target different subtypes, and furthermore the forward primer had a low rate of conserved nucleotides, 50%, among the various subtypes assessed (see section 5.2 Primer and probe design).

The P/P set that targeted the *pol* gene at around 2000 bp (pol2000) was problematic and gave no amplification of HIV-1. Possibly because the designed probe could not obtain a T_m within the desired range, around 65-68°C, unlike the probes in the other P/P sets. This due to the low GC content in the targeted region of this P/P set. A T_m for the probe above the T_m for the primers is important because it makes the probe bind into the targeted sequence before the primers. This may be the reason behind the failed performance of the pol2000 P/P in the wet lab tests.

A more successful gene to target was the *nef* gene. The nef9000 P/P set had two possible forward primers. The two P/P sets were named nef9000FP1 and nef9000FP2. The first contained more than 4 G repeats and was therefore difficult to manufacture due to the formation of tetrad secondary structure. The nef9000FP1 showed no amplification of HIV-1. Likely because of the G repeats forming a secondary structure. Therefore, a new forward primer was designed that did not contain the 4 G repeats and indeed exhibited a better performance. The P/P set with nef9000FP2 gave the highest maximum fluorescence of the designed P/P sets but only detected 33-67% of the positive controls. This may in part be explained by the reverse primer targeting the *nef* gene that had a self-dimer value that was lower than the IDT threshold of -9 kcal/mole (Integrated DNA Technologies, Inc, 2020), (see table 10). However, for comparison the forward primer for the method for detecting Hepatitis B used in the routine analyses at Octapharma AB has an even lower value for self-dimerization, as determined using OligoAnalyzer, than the nef9000 reverse primer and works fine. Therefore, the delta G value may not give the full explain. The Octapharma routine Hepatitis E assay is using betaine in the master mix, which is preventing dimerization. The assay with nef9000FP2 did not contain betaine and hence it could be that problems with dimerization are lowering the detection rate. The average cp-value for the P/P set, nef9000FP2, was approximately 10 cycles higher than that for pol4000. Because of limits in project time, the high PCR failing rate and high cp-values no further experiments were made with nef9000FP2.

The best performing P/P set was the pol4000, targeting the *pol* gene. The P/P set had a risk of forming hetero dimers because the dimerization value was lower than the threshold from IDT (Table 10). However, when comparing to the P/P in the Octapharma routine HEV and HBV method the risk for dimers is similar and works fine, but both contain betaine in their master

mixes. Therefore, betaine was included in the pol4000 PCR recipe to see if better fluorescence signals could be obtained. However, adding betaine to the master mix gave no amplification of HIV-1. One reason could be that the P/P T_m got lower than the PCR program annealing temperature. According to the T_m analysis, the T_m was 0.5 degrees higher at 900 nM than at 600 nM of P/P. Further analysis testing a range of different amounts of betaine to the master mix and with 900 nM P/P would be interesting to investigate if a better performance of the pol4000 P/P system could be achieved. It would also be interesting to add some Mg^{++} to the master mix because according to the T_m analysis in Table 16 B the T_m of the P/P set gets higher when increasing the Mg^{++} concentration

A low concentration of HIV-1 gave low signals (see section 6.2.1, First evaluation of the P/P sets). A test to elevate the concentration of HIV-1 in the PCR reaction was performed to see if this would improve the sensitivity by adding 10 μ L eluate instead of 5 μ L into the PCR wells. To maintain the total PCR volume at 20 μ L the water had to be excluded from the PCR mix. However, this attempt did not result in higher fluorescence signals, instead the signals got lower. The reason for this could be that when adding more eluate to the PCR wells the levels of salt in the PCR wells probably became higher and effected the T_m of the P/P set. This because the eluation buffer likely contained more salt than the normally added water.

To get stronger signals from the pol4000 assay an optimization of the primer and probe concentration were performed. The best result was achieved when using a concentration of 900nM primer and a concentration of 900nM probe. According to the enzyme manufacture, the optimum concentrations should be 300-900nM for primers and 50-250nM for the probe (Thermofisher, 2020). Hence, in our experiments the optimum probe concentration found was more than three times as high as the recommended concentration. This shows that it is important to test the system outside of the recommendations. A further optimization could be to separate the concentrations on the forward and reverse primer and test individual primer concentrations together with probe concentrations in the recommended interval, and also at 900 nM.

After the optimization of pol4000 a genotype test was conducted to assess if the P/P set could detect different groups and subtypes of HIV-1. The genotype tests positive controls had much lower fluorescence signals than in the optimizing experiments. One reason behind this could be that the components in the master mix did not mix properly due to the much larger amount of master mix stock prepared, compared to the previous experiments. Another explanation could be that the extraction was not optimal. A lower RNA concentration gives a lower maximum fluorescence as seen comparing using eluate from Pos B (36 IU/mL) (Table 13) instead of Pos A (201 IU/mL) (Table 14) suggesting that the eluates from the genotype extraction had a lower concentration of HIV-1 than previous extractions. No amplification of group N and O was observed but all tested subgroups of group M had at least one amplification (Table 18). Even though the guidelines in the current pharmacopeias does not obligate detection of group N and O, to cover for stricter guidelines in the future an aim for this novel HIV-1 assay is to enable detection of these groups. Furthermore, 27.5% of the HIV-1 amplification success rate could be explained by nucleotide mismatches between the P/P set and the targeted sequence at the different genotypes. Redesigning the pol4000 P/P set with

some of the nucleotide switched to fit better in with the N and O groups and thereby lower the number of mismatches would be interesting to test in further studies.

7 Conclusion

The results reveal that there is a potential in the P/P set targeting the *pol* gene region around nucleotide position 4000 in the reference sequence HXB2. There is a need for further investigation in order to determine whether the pol4000 can be used in a new routine analysis by Octapharma AB. I propose three possible improvements:

- Gain a stronger fluorescence signal by optimizing the forward and reverse primers' concentrations separately.
- Redesign the P/P set by changing some of the nucleotides lower the number of mismatches with the N and O group genome to be able to detect all tested groups
- Add betaine to the master mix and increase the Mg^{++} concentration lower the risk for dimerization.

8 Acknowledgements

This project would not have been possible without the support and help received from everyone at PCR R&D and support at Octapharma AB. It has been a pleasure to work with you. A special thanks to Rasmus Gustafsson, my supervisor at Octapharma AB. Without his guidance and encouragement, my work would not have been as successful. I also want to thank the PCR analysis group for all the help in the lab.

Thank you!

9 Bibliography

- Abfalterer, W. et al., 2013. *HIV Sequence Compendium 2013*. 1st ed. Los Alamos: Theoretical Biology and Biophysics.
- Ampliqon A/S, 2020. *Ampliqon*. [Online]
Available at: <https://ampliqon.com>
[Accessed 08 March 2020].
- Berg, J. M., Tymoczko, J. L., Gatto, G. J. & Stryer, L., 2015. *Biochemistry*. 8th ed. New York: Macmillan Learning .
- Boivin, S., Cusack, S. & Ruigrok, R. W. H., 2010. Influenza A Virus Polymerase: Structural Insights into Replication and Host Adaptation Mechanisms. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 285(37), pp. 28411-28417.
- Bouvier, N. & Palese, P., 2008. The biology of influenza viruses. *Vaccine*, 26(4), pp. 49-53.
- Brown, T. A., 2016. *Gene Cloning and DNA Analysis*. 7th ed. New Jersey: Wiley-Blackwell.
- Bukh, J., 2016. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. *Journal of Hepatology* , Volume 65, pp. 2-21.
- Désiré, N., Dehée, A., Veriniq, S. & Jacomet, C., 2001. Quantification of Human Immunodeficiency Virus Type 1 Proviral Load by a TaqMan Real-Time PCR Assay. *JOURNAL OF CLINICAL MICROBIOLOGY*, 39(4), pp. 1303-1310.
- Division for Sustainable Development Goals, 2020. *Sustainable Development Goals*. [Online]
Available at: <https://sustainabledevelopment.un.org/sdg3>
[Accessed 04 April 2020].
- Eberle, J. & Gurtler, L., 2012. HIV Types, Groups, Subtypes and Recombinant Forms: Errors in Replication, Selection Pressure and Quasispecies. *Intervirology*, Volume 55, pp. 79-83.
- Enzo Life Sciences, Inc, 2020. *What are the differences between PCR, RT-PCR, qPCR, and RT-qPCR?*. [Online]
Available at: <https://www.enzolifesciences.com>
[Accessed 30 January 2020].
- Fraiman, D. & Fraiman, R., 2018. An ANOVA approach for statistical comparisons of brain networks. *SCIENTIFIC REPORTS* , Volume 8, pp. 1-14.
- Freed, E. O., 2015. HIV-1 assembly, release and maturation. *Nature*, Volume 13, pp. 484-496.
- Guan, W., Wong, S., Zhi, N. & Qiu, J., 2009. The Genome of Human Parvovirus B19 Can Replicate in Nonpermissive Cells with the Help of Adenovirus Genes and Produces Infectious Virus. *Journal of Virology* , 83(18), pp. 9541-9553.
- Heid, C., Stevens, J., Livak, K. & Williams, P., 1996. Quantitative Real Time PCR. *Genome Res*, 6(10), pp. 986-994.

- Integrated DNA Technologies, Inc, 2020. *idtdna*. [Online]
Available at: <https://eu.idtdna.com>
[Accessed 11 February 2020].
- Ishino, S. & Ishino, Y., 2014. DNA polymerases as useful reagents for biotechnology – the history of developmental research in the field. *Front Microbiol.* 2014; 5: 465., Volume 5, pp. 1-8.
- Jensen, M., Fukushima, M. & RW, D., 2010. DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. *PLoS One*, 5(6), pp. 1-5.
- Khodakaram-Tafti, A. & Farjanikish, G., 2017. Persistent bovine viral diarrhoea virus (BVDV) infection in cattle herds. *Iran J Vet Res.* , 18(3), pp. 154-163.
- Kim, T. K., 2015. T test as a parametric statistic. *Korean J Anesthesiol*, 68(6), pp. 540-546.
- Lodish, H. et al., 2000. *Molecular Cell Biology*. 4th ed. New York: W. H. Freeman.
- Los Alamos national laboratory, 2014. *HIV sequence database*. [Online]
Available at: <https://www.hiv.lanl.gov/content/sequence>
[Accessed 02 February 2020].
- Mahdi, M., Szojka, Z., Mótyán, J. & Zsér, T., 2018. Inhibitory Effects of HIV-2 Vpx on Replication of HIV-1. *Journal of Virology*, 92(14), pp. 1-16.
- Mossoro-Kpinde, C. D. et al., 2016. Analytical Performances of Human Immunodeficiency Virus Type 1 RNA-Based Amplix Real-Time PCR Platform for HIV-1 RNA Quantification. *AIDS Research and Treatment*, pp. 1-12.
- Mouquet, H. et al., 2011. Memory B Cell Antibodies to HIV-1 gp140 Cloned from Individuals Infected with Clade A and B Viruses. *PLoS One*, 6(9).
- Murphy, K. & Weaver, C., 2017. *Immunobiology*. 9th ed. New York: Garland Science.
- NCBI Taxonomy Browser, 2020. *NCBI Taxonomy Browser*. [Online]
Available at: <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser>
[Accessed 28 January 2020].
- Ngandu, N. K. et al., 2008. Extensive purifying selection acting on synonymous sites in HIV-1 Group M sequences. *Virology Journal* , Volume 5, p. 160.
- Nyamweya, S. et al., 2013. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Rev Med Virol.* , 23(4), pp. 221-240.
- Roche, 2001. *gene-quantification*. [Online]
Available at: <https://www.gene-quantification.de>
[Accessed 25 January 2020].
- Sabin, C. A., 2013. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy?. *BMC Med.* 2013 Nov 27;11:251. doi: 10.1186/1741-7015-11-251., 11(251).
- Santoro, M. M. & Perno, C. F., 2013. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol.* , Volume 481314.

- Schneider, A., Hommel, G. & Blettner, M., 2010. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.*, 107(44), p. 776-782.
- Seitz, R., 2016. Human Immunodeficiency Virus (HIV). *Transfus Med Hemother*, Volume 43, pp. 203-222.
- Skittrall, J., Ingemarsdotter, C., JR, G. & AML, L., 2019. A scale-free analysis of the HIV-1 genome demonstrates multiple conserved regions of structural and functional importance. *PLoS Comput Biol* , 15(9).
- Swanstrom, R. & Coffin, J., 2012. HIV-1 Pathogenesis: The Virus. *Cold Spring Harb Perspect Med.*, 2(12), pp. 1-18.
- Tamamis, P. & Floudas, C. A., 2014. Molecular Recognition of CCR5 by an HIV-1 gp120 V3 Loop. *PLoS One*, 9(4), p. e95767.
- ThermoFisher, 2020. *ThermoFisher*. [Online]
Available at: <https://www.thermoFisher.com/order/catalog/product/4444432#/>
[Accessed 07 February 2020].
- Tough, R. H. & McLaren, P. J., 2019. Interaction of the Host and Viral genome and their influence on HIV disease. *Frontiers in Genetics*, 9(720).
- UNAIDS, 2020. *UNAIDS*. [Online]
Available at: <https://www.unaids.org>
[Accessed 11 March 2020].
- Yuryev, A., 2007. *PCR Primer Design, Anton Yuryev (Ed), 1 edition, Humana Press, (Aug 2007)*,. 1st ed. New Jersey: Humana Press.
- Zhu, J. & Paul, W. E., 2008. CD4T cells: fates, functions, and faults. 112(5), pp. 1557-1568.