I. INTRODUCTION

1.1 Background

Acquired immunodeficiency syndrome (AIDS) is a disease caused by Human Immunodeficiency Virus (HIV) and has already become a global burden disease as in 2019, it was reported that there were around 3.68 million cases of HIV/AIDS where it is estimated that the prevalence rate is 476.23 per 100,000 people (Tian et al., 2023). AIDS is the advanced development stage of an HIV infection in which the patient would experience major damage to the immune system particularly the CD4+ T cells making them susceptible to opportunistic infections and cancer (Schwetz & Fauci, 2018). During an infection, HIV can turn into a group of mutated viruses called a quasispecies in which they have similar mutations between each other. HIV utilizes quasispecies to escape the immune system and it has also been reported to be responsible in creating drug resistant strains of HIV hence disrupting the control of HIV infection (Ode et al., 2015). This occurrence is due to the insertion and deletion of certain nucleotides in a genome causing genetic diversity within the quasispecies as well as mismatches due to error-prone reverse transcriptase enzymes (Alisoltani et al., 2022; Lloyd et al., 2014). Hence it is necessary to analyze the HIV genome in order to gain information regarding HIV genomic diversity in order to administer a more personalized and effective antiretroviral therapy (ART) to HIV patients.

In analyzing HIV genetic sequences, Sanger sequencing is still considered the gold standard as until today it is demonstrated to be able to detect HIV drug resistance mutations while also being able to determine the presence of insertions or deletions (indels) in HIV (Manyana et al., 2021) with high accuracy. Other sequencing technologies such as Illumina MiSeq and PacBio have also been reported to be able to detect HIV drug resistant mutation at the quasispecies level specifically and sensitively (Huang et al., 2016; Zhang & Ma, 2021). In the last decade, long-read sequencers such as PacBio have

been available with advantages compared to short-read sequencers such as Sanger sequencing and Illumina in terms of how they are better in analyzing repeat regions (homopolymer) as well as high diversity regions in a genome (Nakano et al., 2017). Nevertheless, that advantage is traded with a relatively low accuracy. PacBio, has improved their sequencing chemistry, claiming to have raw reads with 99.9% accuracy (PacBio, 2023). If the claim is true, it would revolutionize quasispecies analysis as bioinformatics processing would not be needed to analyze HIV-1 genome diversity in an individual; each raw read will be equivalent to a HIV-1 molecule. However, it was observed in a preliminary experiment that the raw sequencing reads contain irregularly high levels of indels and stop codons in the middle of an important gene for HIV-1 survival, i.e., the protease-reverse-transcriptase (PRRT) gene. Therefore, a confirmation of present indels should be done with the current HIV gold standard sequencing platform, which is Sanger Sequencing.

1.2 Research Aim

The aim of this research is to confirm the validity of indels found in PacBio sequencing of HIV PRRT gene using Sanger sequencing as a reference method for indel detection.

1.3 Scope of Research

The scope of this research includes:

- Amplification of HIV genome by PCR
- Ligation of selected PCR product into plasmid vector
- Transfection of plasmids to competent cells for propagation of the specific HIV genetic sequence
- Incubation of LB plate for competent cells propagation
- Cloning Pick-up by picking the present colonies on the agar and analyzing them on PCR to confirm successful cloning

- Cloning Purification of the present colonies by subculturing the colonies which would then be followed by MiniPrep to isolate the DNA
- Sanger sequencing
- Data Analysis of counting the indels found from the Sanger sequencing results by either bioinformatics or manual methods

1.4 Research Question/Hypothesis

The indels would be confirmed in Sanger sequencing due to the high accuracy of PacBio in analyzing the HIV sequence allowing it to be a tool for indel detection in the world of HIV sequencing.