CHAPTER 1

INTRODUCTION

1.1. Research Background

The rapid emergence of new Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) variants within the past two years has led scientists to continuously monitor the genomic changes of the virus to compensate for the advanced mutation rate. Real-time genomic surveillance has become one of the essential elements of epidemiology monitoring. However, since SARS-CoV-2 has a relatively large genome size, reconstructing the entire genome effectively and inexpensively was a challenge with the typical Next-generation Sequencing (NGS) platform (Freed et al., 2020; Khailany et al., 2020).

ARTIC Network, a project to develop a sequencing system to support real-time molecular epidemiology from viral outbreaks, designed 98 tiled primer pairs targeting 400 bp amplicons named ARTIC V3. In collaboration with Oxford Nanopore Technologies (ONT) sequencing, a brand-new third-generation sequencing, the primer is used with SQK-LSK109, a ligation-based library preparation kit as a part of the ARTIC V3 Classic protocol. At the same time, Freed and colleagues (2020) also designed the same tiled primer targeting 1200 bp amplicons named Midnight primer. This primer is also a part of the ONT protocol, the Midnight protocol, which uses a transposase-based library preparation kit named SQK-RBK110.96. Since 2020, these protocols have been claimed to be the easy and high-throughput workflow for a turnaround SARS-CoV-2 sequencing.

Despite the rapid, simple, and high-throughput sequencing using ARTIC V3 Classic and Midnight protocol, the preliminary data during sequencing at GSI Lab demonstrated that the result from both methods had shown considerable amplicon drops, affecting the recovery of the viral sequences. Although the drops are more prominent in ARTIC V3 rather than Midnight, this problem is quite concerning as a complete SARS-CoV-2 sequence is important to support genomic monitoring, variant detection, and vaccine/drug development. In a report by Itokawa et al. (2020), the ARTIC V3 primer set was tested at eight different annealing temperatures ranging from 63.1 to 68.6°C and

reported the optimal temperature of 65°C, which is opposite to the preliminary data at GSI Lab as the amplicon drops were observed during sequencing with annealing temperature of 65°C. For the Midnight primer set, currently, there are no publications that reported amplicons dropout, such as found in sequencing runs at GSI Lab. Therefore, this study will determine the optimal annealing temperature for both ARTIC V3 and Midnight primer sets.

Preliminary data also shows that sequencing using SQK-LSK109 yields a higher quality in sequencing depth, resulting in higher confidence in variant detection and eliminating ambiguous bases. In contrast, a higher genome coverage is achieved with SQK-RBK110.96. A study by González-Recio and colleagues (2021) reported similar findings using SQK-LSK109 and SQK-RBK004 with ARTIC V2 primer. However, the study is conducted using two different flow cells (R9.4 and R10). Hence, this study will evaluate both kits with the same flow cells and the newer version of the primer. Furthermore, the performance of the two methods, ARTIC V3 Classic and Midnight protocol, will also be evaluated to see which combination of primer and library preparation kit works best to generate a high-quality sequencing with minimal amplicon drops.

1.2. Research Objective

The objectives of this study are:

- To find out the optimized PCR tiling condition for ARTIC V3 and Midnight primer set to use in SARS-CoV-2 Whole Genome Sequencing using ONT Sequencing Protocol.
- To evaluate the performance of different ONT SARS-CoV-2 sequencing chemistry; Ligation-based Sequencing (SQK-LSK109), and Transposase-based Sequencing (SQK-RBK110.96) in terms of quality.
- To evaluate the performance of the ARTIC V3 Classic protocol (SQK-LSK109 with ARTIC V3 primer) and the Midnight protocol (SQK-RBK110.96 with Midnight primer) in terms of quality.

1.3. Scope of Work

The scope of work of this study is limited to wet-lab works only and does not include the bioinformatics processing of the sequencing.

- Gradient PCR to determine the optimum annealing temperature (T_a) of ARTIC V3 and Midnight primer set.
- Agarose Gel Electrophoresis as a confirmation step of the gradient PCR tiling.
- Sequencing the gradient PCR tiling sample of both ARTIC V3 and Midnight primer set with each SQK-LSK109 and SQK-RBK110.96.