CHAPTER I

INTRODUCTION

1.1. Background

Coronavirus Disease 2019 (COVID-19) global pandemic has affected approximately 226 million people globally and has contributed to more than 4 million deaths as of September 2021 (World Health Organization, n.d.). The causing agent of this infectious disease is a single-stranded positive-sense RNA virus, called the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), that attacks the respiratory system (Kim et al., 2020). SARS-CoV-2 is classified into the *Coronaviridae* family, and its genome consists of around 30,000 nucleotides, which is why SARS-CoV-2 is considered one of the largest RNA viruses (Esakandari et al., 2020). SARS-CoV-2 genome is further cleaved into at least six open reading frames (ORFs) and translated into 16 nonstructural proteins and four structural proteins, including spike (S), envelope (E), membrane (M), and nucleocapsid (N) (Dearlove et al., 2020). The S protein has two subunits and has a crucial role in recognizing and binding to the host cell membrane (Huang et al., 2020). In addition, the E and M proteins facilitate viral assembly and budding, which are essential for viral pathogenesis (Schoeman & Fielding, 2019). Meanwhile, the N protein forms the viral capsid, protecting the viral genome from the environment (Surjit & Lal, 2009).

The same as other viruses, SARS-CoV-2 is actively adapting to the environment via genome mutations. For the past two years, several new variants of SARS-CoV-2 have emerged and been divided into variants of concerns (VOC) and variants of interest (VOI) (Konings et al., 2021). According to the World Health Organization (n.d.), Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529) variants of SARS-CoV-2 are considered as VOC because they are found to have an increase in transmissibility and virulence. Meanwhile, Lambda (C.37) and Mu (B.1.621) variants are included in the list of VOI because they are predicted to have genetic changes that could

alter the viral characteristics.

At the same time, various vaccines have been developed to fight the COVID-19 global pandemic (Dong et al., 2020). The currently available vaccines for COVID-19 in Indonesia, including Sinovac, Sinopharm, Moderna, BioNTech/Pfizer, and Oxford/AstraZeneca utilized the ancestral strain SARS-CoV-2 to trigger the adaptive immunity, which includes antibody and T-cell responses (Ophinni et al., 2020). Antibody produced by B-cells neutralizes the virus by binding to the virus, disrupting the virus's ability to infect the host cells (Tarke et al., 2021). Unfortunately, the high rate of mutations of SARS-CoV-2 that further resulted in the emergence of new VOC and VOI have raised concerns regarding a decrease in the antibodies' neutralization ability triggered by the vaccines (Noh, Jeong, & Shin, 2021). However, another component of the adaptive immunity called T-cells, which include CD4+ and CD8+ T-cells, is reported to also offer protection against SARS-CoV-2 by recognizing epitopes from the antigen presented by human leukocyte antigen (HLA) molecules (Niesen et al., 2021). The CD4+ T-cell produces cytokines that will initiate the activation of other immune cells, while the CD8+ T-cell is able to kill virus-infected cells (Tarke et al., 2021). Nevertheless, the question regarding the effect of mutation on T-cell epitopes on vaccine effectiveness and efficacy remains unresolved (Williams & Burgers, 2021).

Accordingly, next-generation sequencing was done to see the currently circulating variants of SARS-CoV-2 in Indonesia. Furthermore, an immunoinformatics study was done to predict T-cell epitopes from structural proteins of circulating SARS-CoV-2 variants in Indonesia. Finally, conservancy analysis was also performed to further study the consequences of mutations in the epitopes of the circulating variants.

1.2. Objectives

The objectives of this study include conducting next-generation sequencing on SARS-CoV-2 circulating variants in Indonesia for a genomic surveillance study. In addition, a

conservancy analysis study was done by comparing structural proteins T-cell epitopes from ancestral SARS-CoV-2 (hCoV-19/Wuhan/WIV04) and SARS-CoV-2 variants circulating from October 2021 to January 2022.

1.3. Scope of Work

The scope of work for this thesis project includes next-generation sequencing of circulating variants of SARS-CoV-2 samples, genome annotation and translation, identification of HLA alleles prevalent in the Indonesian population, prediction of the epitopes for S, M, E, and N proteins of SARS-CoV-2, conservancy analysis of T-cell epitopes from ancestral SARS-CoV-2 with circulating isolates of SARS-CoV-2 in Indonesia, and data analysis and interpretation.