Development of genomic methodologies for pathogen surveillance. Microfluidics-automated workflow: From sample to sequencing library.

Julio Cesar Serna Vazquez

Department of Human Genetics, Faculty of Medicine and Health Sciences,

McGill University, Montreal, Quebec, Canada

August 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Dedication

To my parents and grandparents.

Abstract

Sequencing is a fundamental tool for the monitoring of virus outbreaks, such as the recent COVID-19 pandemic, in which sequencing was implemented for the original identification of the SARS-CoV-2 virus, to establish variants of concern, and to track transmission routes throughout the different waves.

Sequencing platforms continue to become more widely available, given the development of portable, more economical sequencing devices. However, library preparation remains a challenge for the implementation of sequencing in remote locations, as the whole workflow from sample to sequencing library involves several steps that need to be carried out by qualified personnel in a dedicated laboratory.

To tackle this issue, our laboratory has collaborated with the National Research Council (NRC-CNRC) to develop an automated workflow for Nanopore library preparation from SARS-CoV-2 clinical samples using the NRC-developed automated microfluidics device, the PowerBlade. The combination of this automated, portable device with portable sequencing technologies, such as those developed by Nanopore, gives us a workflow that can be easily run almost anywhere without the need for a highly experienced technician due to the automation of the workflow. Furthermore, this adaptable technology could be modified and used for identification of other pathogens.

In this work, we present a novel Nanopore library preparation protocol suitable for on-chip implementation on the PowerBlade, which was tested against the standard ONT protocol, both on-chip and off-chip. Additionally, we present a performance comparison of on-chip and off-chip ARTIC SARS-CoV-2 PCR.

Résumé

Le séquençage est un outil fondamental pour la surveillance des épidémies de virus, comme la récente pandémie de COVID-19, dans laquelle le séquençage a été mis en oeuvre pour l'identification initiale du virus SARS-CoV-2, pour établir les variantes à suivre et pour suivre les voies de transmission tout au long des différentes vagues.

Les plateformes de séquençage continuent de se démocratiser grâce à la mise au point d'appareils de séquençage portables et plus économiques. Cependant, la préparation des librairies reste un défi pour la mise en oeuvre du séquençage dans des endroits éloignés, car l'ensemble du processus, de l'échantillon à la librairie de séquençage, implique un grand nombre d'étapes qui doivent être effectuées par un personnel qualifié dans un laboratoire spécialisé.

Pour faire face à ce problème, notre laboratoire a collaboré avec le Conseil National de Recherches Canada (CNRC-NRC) pour développer un flux de travail automatisé pour la préparation de librairies Nanopore à partir d'échantillons cliniques du SARS-CoV-2, en utilisant le dispositif microfluidique automatisé développé par le CNRC, le PowerBlade. La combinaison de ce dispositif automatisé et portable avec des technologies de séquençage portables, telles que celles développées par Nanopore, nous donne un flux de travail qui peut être facilement exécuté presque partout, sans avoir besoin d'un technicien hautement expérimenté en raison de l'automatisation du flux de travail. De plus, cette technologie adaptable pourrait être modifiée et utilisée pour l'identification d'autres agents pathogènes.

Dans ce travail, nous présentons un nouveau protocole de préparation de librairies Nanopore adapté à la mise en oeuvre sur puce sur le PowerBlade, qui a été testé par rapport au protocole ONT standard, à la fois sur puce et hors puce. En outre, nous présentons une comparaison des performances de la PCR ARTIC SARS-CoV-2 sur puce et hors puce.

Table of Contents

List of Abbreviations

COVID-19	Coronavirus disease 2019
DNA	Deoxyribonucleic acid
Mb	Megabase
NGS	Next generation sequencing
NRC-CNRC	National Research Council
ONT	Oxford Nanopore Technologies
PCR	Polymerase chain reaction
PacBio	Pacific Biosciences
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SARS	Severe acute respiratory syndrome
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SMRT	Single-molecule real-time
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
ZMW	Zero-mode waveline
bp	Base pair
cDNA	Complementary DNA
dsDNA	Double stranded DNA
kb	Kilobase
ssDNA	Single stranded DNA

List of Figures

List of Tables

Acknowledgments

Thank you to my supervisor Dr. Ioannis Ragoussis for giving me the opportunity to pursue a master's degree in his laboratory, providing me with all the support I needed to complete my studies, including funding, mentorship, and guidance. Thank you for giving me the opportunity to fulfill a dream and advance in my career.

Thank you to the Department of Human Genetics for their financial support in the form of a Differential Fee Waiver, allowing me to pursue graduate studies at McGill University.

Thank you to Mitacs for the opportunity to come and do research in Canada through the Globalink Research Internship and the Globalink Graduate Fellowship.

Thank you to the National Research Council for the collaboration.

Thank you to everyone in the laboratory.

Thank you to all my family and friends.

And finally, thank you to my parents. I cannot thank you enough for all your love and hard work, without which I would not be here.

Format of the Thesis

This thesis is presented in traditional format, in agreement with the guidelines for thesis preparation set out by the Graduate and Postdoctoral Studies Office (GPS) and the Department of Human Genetics of McGill University. A total of six chapters and an appendix make up this thesis. Chapter 1 is an introduction covering the history of DNA sequencing technologies, pathogen sequencing, and automated approaches for library preparation. Chapter 2 describes the materials and methods used in the present work. Chapter 3 shows the results obtained. Chapter 4 presents a discussion of the results. Chapter 5 includes conclusions and future directions. Chapter 6 shows a full reference list.

Contribution of Authors

This thesis was written by me and supervised by Prof. Ioannis Ragoussis. The experiment design was done by Prof. Ioannis Ragoussis, Dr. Sarah Reiling, Dr Spyros Oikonomopoulos, Dr. Jose Avila Cervantes, and me. The sequencing was performed by me at the Advanced Genomic Technologies laboratory, led by Prof. Ioannis Ragoussis. All analyses were performed by me with the help of Dr. Jose Avila Cervantes and Dr Spyros Oikonomopoulos.

Chapter 1: Introduction

1.1 History of DNA sequencing technologies

1.1.1 Background

One of the major landmarks in molecular biology and biology in general, has been the elucidation of the tridimensional structure of DNA by Watson and Crick with the critical support of the x-ray diffraction images produced by Rosalind Franklin and Maurice Wilkins (Elkin, 2003; Watson & Crick, 1953). DNA, which was discovered by Friedrich Miescher in 1869 (Dahm, 2005; Miescher, 1869) and proved to be the molecule responsible for the transmission of genetic material by the Avery-MacLeod-McCarty experiment in 1944 (Avery et al., 1944), had become one of the most intriguing concepts in science due to the evident next step in its research, the elucidation of its sequence.

DNA sequencing efforts started in the 1960s (Shendure et al., 2008), however, the first widely recognized sequencing technologies arrived in the 1970s, with the development of the Maxam-Gilbert and Sanger sequencing methodologies (Heather & Chain, 2016). These sequencing methodologies have been named first-generation sequencing technologies (Kchouk et al., 2017).

1.1.2 First-generation sequencing

Maxam-Gilbert sequencing was developed by Allan Maxam and Walter Gilbert and published in 1977 (Maxam & Gilbert, 1977). The process is based on radiolabeling of DNA fragments and cleaving them at specific bases. In this method, the DNA

fragments to be sequenced are radio-labelled in their 5’ end by replacing the terminal phosphate for the radioactive isotope ^{32}P. The radio-labelled DNA is then denatured and added into four different reactions, each one of them with specific reagents that will cleave the DNA strand at one or two specific nucleotides (A, C, G/A, C/T). Afterwards, the DNA fragments from each reaction are run separately in a polyacrylamide gel electrophoresis, and an autoradiography of the gel is generated. The DNA fragment's sequence can be obtained from the autoradiography as only the radio-labelled fragments will appear on it, displaying both the fragment size (position) and the nucleotide where it was cleaved (Dorado et al., 2019; Kchouk et al., 2017; Maxam & Gilbert, 1977).

After having created the plus and minus method for DNA sequencing in 1975 (Sanger & Coulson, 1975), in 1977, Frederick Sanger and his team developed a novel DNA sequencing method originally referred to as “DNA sequencing with chain-terminating inhibitors” that later gained popularity as “Sanger sequencing” (Sanger, Nicklen, et al., 1977). Originally, the method relied on the use of radio-labeled dideoxynucleotides (ddNTPs), deoxynucleotides (dNTPs) and DNA polymerase. Four different reactions are set up, each one with the DNA to be sequenced, DNA polymerase, a primer, the four dNTPs and one of the four ddNTPs in a very low concentration. In each reaction, the DNA polymerase starts adding the corresponding dNTPs until a ddNTP is added. As the ddNTPs lack the 3'-OH group needed to make a phosphodiester bond with another nucleotide, the polymerization reaction cannot continue, and the fragment remains of a certain length. Given that the concentration of ddNTPs is very low compared to that of the dNTPs, their addition does not always occur at the first available position resulting in

fragments of different lengths. Afterwards, each reaction is run through electrophoresis in a polyacrylamide gel, displaying the position of the radio-labelled ddNTP through autoradiography. When the four reactions are visualized, the sequence of the DNA fragment can be deduced from the size of the fragments from each of the radio-labelled ddNTPs (Men et al., 2008; Sanger, Nicklen, et al., 1977).

Maxam-Gilbert sequencing remained a commonly used technique for several years; however, Sanger sequencing positioned itself as the leader of the generation due to its simpler method and subsequent improvements to the original protocol (Giani et al., 2020). The Sanger sequencing method was improved in several ways over the years. The substitution of radio-labelled ddNTPs for fluorescent-labelled ddNTPs not only reduced radiation risk for the user but also allowed to have all four ddNTPs in a single reaction, which then could be analyzed in a single electrophoresis gel as each ddNTP has a different fluorescent marker. Furthermore, the implementation of capillary electrophoresis, laser-based fluorescence detection, and the use of computers to analyze the data allowed sequencing automation and the possibility of producing larger sequencing outputs (Dovichi & Zhang, 2000; Men et al., 2008; Slatko et al., 2018).

Sanger sequencing prevailed as the most used sequencing technology through the decades (Heather & Chain, 2016; Stranneheim & Lundeberg, 2012), being the technology used in both the Human Genome Project and the Celera Genomics efforts to decode the human genome (Lander et al., 2001; Venter et al., 2001), and it is still being used nowadays for clinical applications (Arteche-Lopez et al., 2021). However, with the arrival of second-generation sequencing technologies (also termed next-

generation and short-read), the paradigm changed due to their capacity to obtain much larger amounts of data in a shorter span of time and at a reduced cost (Kircher & Kelso, 2010).

1.1.3 Second-generation sequencing

Second-generation sequencing platforms include 454 pyrosequencing, Sequencing by Oligonucleotide Ligation and Detection (SOLiD), Illumina (Solexa) sequencing, and Ion Torrent semiconductor sequencing (Gupta & Verma, 2019; Hu et al., 2021). Throughput limitations previously encountered with Sanger sequencing were overcome by the development of these high-throughput sequencing technologies, which allowed massive parallel sequencing in a single device for the very first time (Hu et al., 2021).

454 pyrosequencing, now discontinued, was the first second-generation sequencing technology available. It consists of the detection of light from the release of a pyrophosphate during DNA polymerization. The detection relies on three enzymes: DNA polymerase, ATP sulfurylase, and luciferase. The process occurs in a well with a bead-bound DNA fragment, which is sequentially flooded with each one of the four different nucleotides. As the nucleotides are sequentially added and washed, the DNA polymerase produces pyrophosphate each time it incorporates a nucleotide. The pyrophosphate produced is converted to ATP by the ATP sulfurylase, and the luciferase uses this ATP and oxygen to transform luciferin into oxyluciferin, producing light. The intensity of the light is used to determine if the nucleotide was added more than one

time; however, this becomes an issue for homopolymers of more than 8 bp (Calabrese et al., 2020; Gupta & Verma, 2019; Slatko et al., 2018).

Illumina uses a sequencing-by-synthesis approach. Illumina sequencing (originally developed by Solexa) starts by ligating adapters to both ends of randomly fragmented dsDNA. The dsDNA is denatured into ssDNA, and as it is added to a flow cell, the ssDNA fragments attach to oligonucleotides fixed to the surface of the flow cell that are complementary to the adapter sequences. Complementary strands to the ssDNA fragments are generated with a DNA polymerase, and then, the original strand is denatured and removed. The unattached end of each fragment binds to another surface oligonucleotide; the complementary strand is generated and then denatured. Multiple cycles of this process result in clusters of multiple copies of each DNA fragment. Later, fluorescent-labelled nucleotides are added and incorporated one by one into the DNA strands. When a nucleotide is incorporated by a DNA polymerase, the fluorescent marker, which also serves as a terminator of synthesis, is excited and detected, allowing base identification. Then, the marker is cleaved off, allowing the next base to be incorporated and identified. (Bentley et al., 2008; Buermans & den Dunnen, 2014; Mardis, 2008; Slatko et al., 2018). Figure 1 shows the Illumina sequencing process.

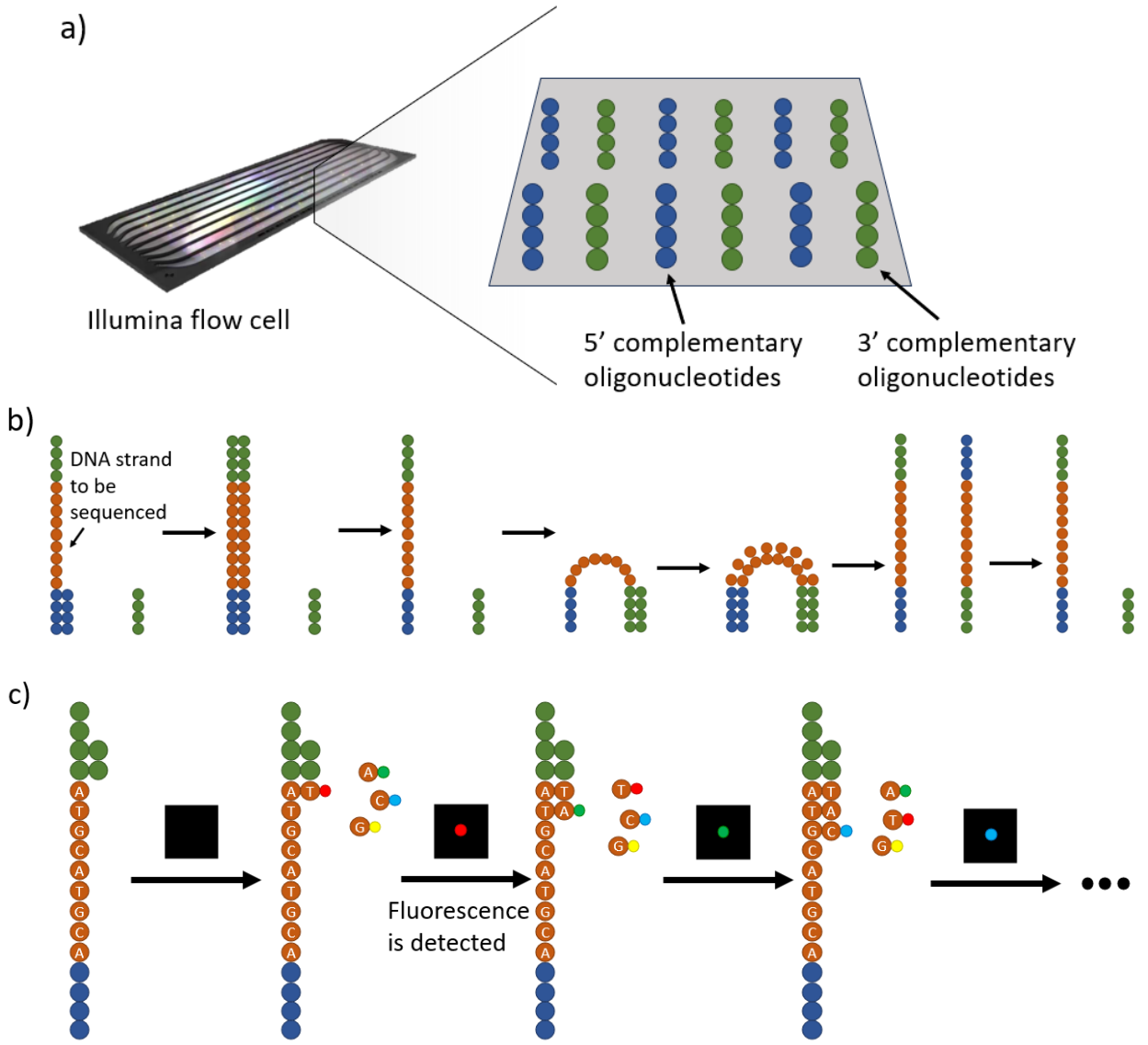


Figure 1. Illumina sequencing. a) Illumina flow cell. Illumina flow cells contain oligonucleotides that are complementary to the sequences at the ends of the DNA to be sequenced. b) Clonal amplification. The DNA strand to be sequenced is attached to the complementary nucleotide and clonally amplified on the flow cell. c) Sequencing by synthesis. Fluorescent-labeled nucleotides and a polymerase are added to the flow cell. Each nucleotide is incorporated onto the DNA strand and the fluorescence corresponding to that nucleotide is identified, allowing to determine the nucleotide in the original strand.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD) starts with the fragmentation of the DNA to be sequenced and an adaptor ligation, followed by attachment to beads and clonal amplification through emulsion PCR. The beads are attached to a glass slide, and a primer of length *n* is hybridized to the adapter. Then, fluorescent-labeled 8 bp oligonucleotides are added and if the first two nucleotides of the oligonucleotide are complementary to those of the sequence, it will be attached adjacent to the primer by a DNA ligase. Unbound oligonucleotides are washed off, and the fluorescent marker is excited and detected by a sensor. The fluorescent marker is cleaved off, and the process repeats until the entire sequence is covered. Then, the generated strand is washed off, and the process is repeated sequentially with primers of length n-1, n-2, n-3, and n-4, allowing to identify each individual base and sections that were not sequenced before. (Gupta & Verma, 2019; Mardis, 2008; Shokralla et al., 2012)

Ion Torrent semiconductor sequencing is based on the detection of hydrogen ion concentration change. First, DNA is fragmented, ligated to adapters, attached to beads, and amplified by emulsion PCR. The beads are then added to a chip with millions of wells, and a single bead will be deposited in each well. Solutions of one of the four nucleotides and DNA polymerase are added sequentially. Each well works as a pH detector, and when the correct nucleotide is added, the ion concentration change is detected, and the base identified. (Shokralla et al., 2012; Slatko et al., 2018)

It was not until the 2010s that the third generation of novel sequencing technologies was developed. These technologies expanded the limits of sequencing through the

sequencing of native DNA fragments in real time (Lee et al., 2016; McCarthy, 2010; Xiao & Zhou, 2020)

1.1.4 Third-generation sequencing

Long-read sequencing technologies, also known as third-generation sequencing, are defined by their capacity to produce much longer reads than those generated through previous sequencing methods, easily surpassing the 1,000 bp barrier (Bleidorn, 2015). These technologies were built on top of the high-throughput achievement of second-generation sequencing. They provided the possibility of kilo- and even mega-base read generation in real-time without imperative amplification or fragmentation steps, allowing easier de novo assembly, identification of complex regions and detection of structural variants (Hu et al., 2021; Pollard et al., 2018).

Nowadays, the two most prominent long-read sequencing technologies are those developed by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) (Adewale, 2020; Amarasinghe et al., 2020). Average read lengths are 15-20 kb for PacBio and 10-100 kb for ONT, with the possibility of generating ONT reads of up to a few megabases (Mb) (Pollard et al., 2018). The current record for an ONT read is ~ 2.3 Mb, the longest for any technology, given that the length of an ONT read is mainly limited by the size of the DNA fragment itself (Amarasinghe et al., 2020; Makalowski & Shabardina, 2020). At first, one of the main concerns of long-read sequencing technologies was their high error rates, but as with any technology, long-read sequencing technologies have continuously improved to advance towards more accurate reads (Amarasinghe et al., 2020; Hu et al., 2021). ONT currently offers a raw

read accuracy of up to 99.6%, while PacBio provides a >99% accuracy for short reads (Oxford Nanopore Technologies n.d.-a; PacBio 2023). Maximum read length, detection method, as well as single read accuracy for each one of the sequencing technologies described in this document, can be observed in Table 1.

PacBio's single-molecule real-time sequencing (SMRT) works by incorporation of fluorescent-labeled nucleotides onto the DNA strand to sequence (a dsDNA fragment previously circularized by ligation of hairpin adaptors) through a DNA polymerase fixed at the bottom of a structure called zero-mode waveline (ZMW). The ZMW is a nano-structure, which allows the observation of the fluorescence of only the nucleotide that is being incorporated at a given moment. The fluorescence is analyzed by a detector in order to identify the base that was incorporated. After incorporating all the nucleotides onto a strand, the polymerase can continue incorporating nucleotides onto the hairpin adapter, the other strand, back again on the previously polymerized strand, and so on, until the lifetime of the polymerase allows it. The number of sequencing "passes" in SMRT sequencing is limited by the lifetime of the DNA polymerase and the length of the strand; as the strand size increases, the number of passes that the polymerase can make around the circularized DNA will diminish. A circular consensus sequence is created from the passes generated. The accuracy of the sequence is determined by the number of passes (Rhoads & Au, 2015).

Table 1. Characteristics of the main sequencing technologies. First generation technologies are marked in orange, the second generation is marked in green, and third generation is marked in blue.

	Sanger	Maxam-Gilbert	Illumina	IonTorrent	454 Pyrosequencing	SOLiD	PacBio	ONT
Generation	First	First	Second	Second	Second	Second	Third	Third
Short/Long read	Short-read	Short-read	Short-read	Short-read	Short-read	Short-read	Long-read	Long-read
Maximum Read Length	~1000 bp	~1000 bp	300 bp	~600 bp	~1000 bp	85 bp	>25 kb	Only limited by DNA fragment size (>2 Mb)
Detection method	DNA autoradiograph / Fluorescence in electrophoresis	DNA autoradiograph	Fluorescence	Changes in pH	Light detection	Fluorescence	Fluorescence	Electrical conductivity
Single read accuracy	>99.9%	99%	99.9%	>99%	>99.5%	99.99%	~90% long reads >99% short reads	>95%
High-throughput	No	No	Yes	Yes	Yes	Yes	Yes	Yes
References	[1-4]	[1, 5]	[6-8]	[9, 10]	[11-13]	[14]	[15-17]	[18-21]

1.1.4.1 Oxford Nanopore Technologies

ONT's nanopore sequencing relies on the measurement of changes in ionic current as a DNA strand passes through a protein nanopore embedded in an electro-resistant membrane subjected to electric current. The changes in ionic current are dependent on the sequence of the DNA strand; that is, each nucleotide produces a distinct signal when passing through the nanopore. A sensor detects and stores the signal, which is later basecalled by the computer and presented as reads to the user (Deamer et al., 2016; Wang et al., 2014). The nanopore sequencing process is shown in Figure 2.

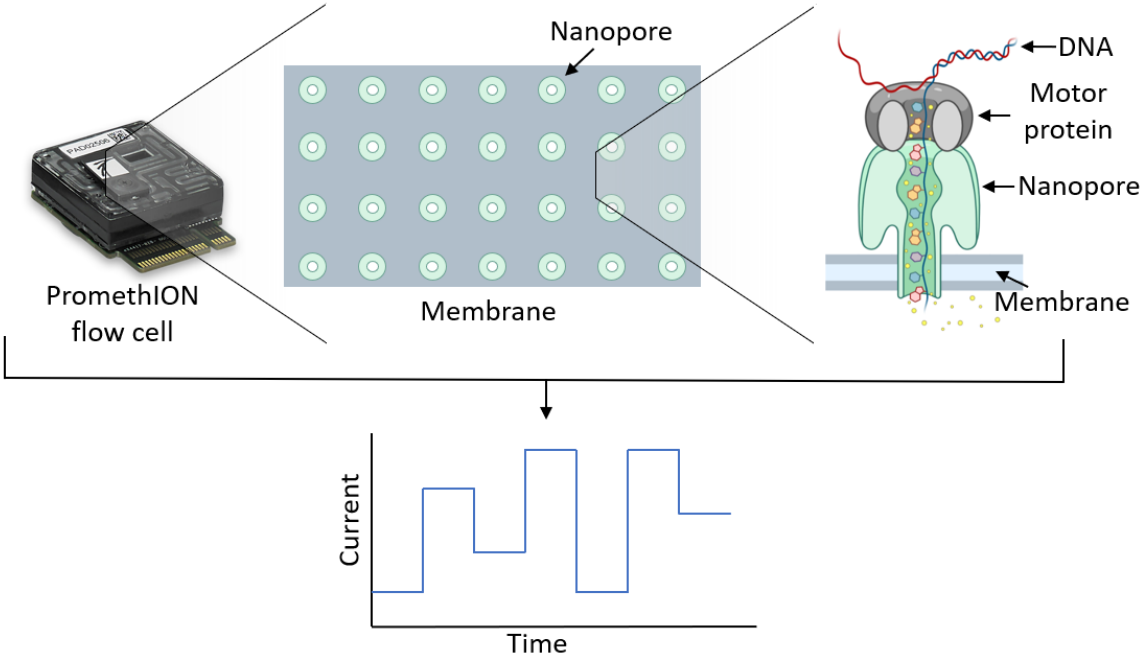


Figure 2. Nanopore Sequencing Process. A PromethION or any other nanopore flow cell contains a membrane with embedded nanopores. When sequencing begins, a motor protein attaches to the top of the nanopore, separating the DNA strands and controlling the speed at which the DNA passes through. The electric current flowing through the membrane is modified based on the nucleotide that is passing through the nanopore, allowing the identification of each nucleotide. PromethION flow cell image from ONT (Oxford Nanopore Technologies 2023a).

ONT offers high portability with devices suitable for RNA and DNA sequencing, with the capacity of real-time sequencing and basecalling; displaying the number and quality of reads in real-time, allowing fast pathogen identification, and even permitting selective sequencing, in which undesired DNA strands can be forced out of a pore to improve desired throughput (Bao et al., 2021; Bruno et al., 2021; Bull et al., 2020; Sun et al., 2022; Zhang et al., 2021) The company is consistently doing research to improve the

nanopores, continually releasing new versions of flow cells with novel pores (Oxford Nanopore Technologies n.d.-i).

The concept of nanopore sequencing began in the 1980s when the idea of sequencing strands of DNA by passing them through a nanometric pore was devised (Deamer et al., 2016). In 1996, the first successful experiments were carried out using the α-hemolysin heptameric pore from *Staphylococcus aureus* (Kasianowicz et al., 1996). With an inner diameter of just 1 nm, the pore is ideal for allowing the passage of a single strand of DNA at a time while blocking the passage of dsDNA fragments since the diameters of DNA are 0.9–0.1 nm for a single strand and 2 nm for dsDNA. It is stable and keeps open when exposed to an electric current. (Deamer et al., 2016; Kasianowicz et al., 1996; Oxford Nanopore Technologies n.d.-i; Wang et al., 2001).

However, wild-type α-hemolysin pores are not ideal to use as a DNA sequencing sensor since the DNA strands pass through them too quickly to be recognized, and they have a 5 nm stem in which, even at the right speed, more than 12 nucleotides would need to be identified collectively, making it difficult to identify individual nucleotides. Further research was necessary after it was determined that a different pore would be required and that the speed at which the DNA strand passed through the pore needed to be controlled (Deamer et al., 2016; Pennisi, 2012). It was not until a modified porin A from *Mycobacterium smegmatis* was used as the pore and a ϕ29 DNA polymerase was used to control the strand speed that nanopore sequencing became a reality (Deamer et al., 2016; Manrao et al., 2012; Pennisi, 2012).

Oxford Nanopore Technologies was founded in 2005 initially as “Oxford Nanolabs”, whose aim was the development and commercialization of nanopores that could act as biosensors for the detection of agents such as metals and proteins. The company began research on sequencing in 2006, when they applied for and received a grant for sequencing technology development. In 2008, now focused on sequencing, the company was renamed Oxford Nanopore Technologies, continuing with the development of a sequencing technology that could hit the market (Pennisi, 2012).

ONT released the first nanopore sequencing results in 2012. Their first sequencing device, the MinION, was made available to early-access users in 2014 and became commercially available in 2015 (Deamer et al., 2016; Oxford Nanopore Technologies n.d.-b).

The company's initial product, the MinION, is a compact, portable sequencer that uses flow cells with up to 2048 pores operated in sets of 512, thus having a maximum of 512 pores sequencing DNA strands simultaneously (Jain et al., 2016; Oxford Nanopore Technologies n.d.-d). Later, in 2017, ONT released the GridION, a device that allows sequencing up to five MinION flow cells at the same time (Oxford Nanopore Technologies n.d.-d). Continuing with the development of bigger devices that can run several flow cells in parallel, ONT introduced the PromethION in 2018. Along with the device, new flow cells were developed. PromethION flow cells have up to 10,700 pores and 2,675 sensors. The PromethION comes in two different versions, PromethION 24 and PromethION 48, which allow simultaneous sequencing of 24 and 48 PromethION flow cells, respectively (Oxford Nanopore Technologies n.d.-h).

An adapter for the MinION and the GridION was launched in 2019. This adapter, called the Flongle, allows sequencing with a new type of flow cell. Flongle flow cells have 126 pores, each of them controlled individually, and are the least expensive of all the flow cells available. The Flongle is a low-cost option that can generate as much as 1.8 Gb of data. (Oxford Nanopore Technologies n.d.-e). The MinION Mk1C, released in 2020, is a modified version of the original MinION that includes a screen and a computer all in a single portable device, removing the need for an additional computer to start a MinION run (Oxford Nanopore Technologies n.d.-b, n.d.-g). Figure 3 shows a timeline with the major releases and events from ONT.

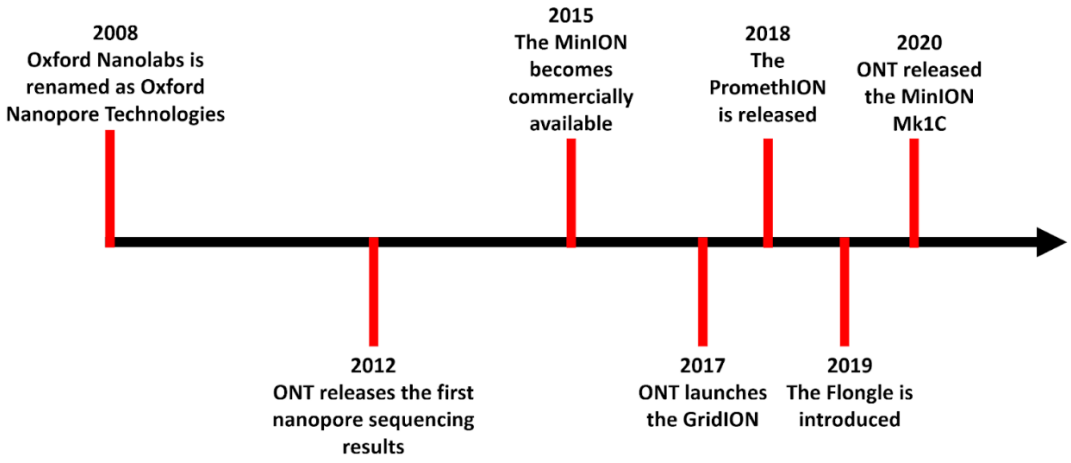


Figure 3. Timeline of major events and product releases from ONT.

1.2 Pathogen Sequencing

1.2.1 Pathogen sequencing approaches

Nowadays, sequencing is a vital diagnostic tool for both genetic and infectious disease diagnosis (Adams & Eng, 2018; Duan et al., 2021). Regarding the latter, sequencing during pathogen outbreaks is crucial as its use allows a better

understanding of the pathogen's identity, its phylogenetics, and how it is being transmitted (Wohl et al., 2016).

When an unknown pathogen is believed to be the causative agent of a disease, metagenomic sequencing can be used as a powerful tool to determine the virus, parasite, fungi, or bacteria causing the infection. Metagenomic sequencing refers to the sequencing of all the genetic material in a sample, and it was made available with the development of NGS due to their high throughput. The use of metagenomics sequencing in clinical settings is crucial as it overcomes the burden of applying tests that only detect some specific pathogens, providing conclusive results in a single test. Furthermore, sequencing provides much more information than classic molecular tests, such as strain identification and a transcriptomic landscape of the pathogen and the host response to it (Chiu & Miller, 2019; Gu et al., 2021; Quince et al., 2017).

Another approach is targeted sequencing, which can be used when there is only a limited number of pathogens suspected as the causative agent of the disease. Among the advantages that targeted sequencing provides is a higher number of reads coming specifically from the pathogen of interest, which leads to a higher detection sensitivity, and depending on the targeted sequencing approach used, it can also provide whole genomes. Two of the techniques used for targeted sequencing are capture methods, which work by designing probes with high affinity to the genetic material of interest; and PCR (Chiu & Miller, 2019; Gaudin & Desnues, 2018).

Furthermore, through the integration of epidemiological and sequencing data, it is possible to identify the source of infection and track transmission networks, providing

valuable information for governments in order to timely generate policies with respect to an outbreak (Behrmann & Spiegel, 2020; Brejova et al., 2021; Bull et al., 2020). In this context, short-read sequencing technologies are generally used for conventional pathogen detection, while long-read platforms are typically used for de novo sequencing due to their capability of producing overlapping reads big enough to sort out complex repetitive regions (Gwinn et al., 2019).

1.2.2 Initial developments and recent SARS-CoV-2 pandemic

The first pathogen to be fully sequenced, and the first whole genome sequenced, was the bacteriophage ΦX174, sequenced by Sanger and his team in 1977 with the plus and minus sequencing method (Sanger, Air, et al., 1977). In 2003, during the SARS outbreak, sequencing was first applied to detect a pathogen during an outbreak. The severe acute respiratory syndrome coronavirus was identified through electron microscopy, and its identity as a novel coronavirus was confirmed by several assays, including sequencing. The obtention of the whole genome of SARS in the early stages of the pandemic allowed the development of molecular tests in a more accurate and timely manner due to a better understanding of the virus and tackle the outbreak in a better way (Goldsmith et al., 2004; Marra et al., 2003; Rota et al., 2003; Stadler et al., 2003).

The current COVID-19 pandemic represents the biggest challenge for public health since modern sequencing technologies were developed. Since early on in the pandemic, sequencing has played a critical role. Metagenomic sequencing was used for the identification of the virus, and genomic surveillance through targeted sequencing is

carried out across the globe (Brito et al., 2022; Wu et al., 2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19, which has led to a global pandemic responsible for more than 6.3 million deaths as of June 26, 2022 (WHO, 2022). SARS-CoV-2 infection in humans was first identified in December 2019 in Wuhan, China (Wu et al., 2020; Zhou et al., 2020). A patient reporting chest tightness, dry cough, pain, weakness and presenting fever and dyspnea was hospitalized on December 26, 2021. Commercial pathogen antigen-detection kits and quantitative PCR excluded the symptoms from being caused by the presence of common respiratory pathogens such as influenza viruses, human adenoviruses, *Chlamydia pneumoniae* and *Mycoplasma pneumoniae*. Metagenomic RNA sequencing of bronchoalveolar lavage fluid from the patient was performed to identify the etiological agent. The longest contig built was found to have a nucleotide identity of 89.1% with the bat SARS-like coronavirus isolate bat-SL-CoVZC45, later identifying it as a novel coronavirus strain that was later named SARS-CoV-2 by the International Committee on Taxonomy of Viruses (Coronaviridae Study Group of the International Committee on Taxonomy of, 2020; Wu et al., 2020).

SARS-CoV-2 is a virus from the Coronaviridae family with a positive-sense single-stranded RNA genome of approximately 30 kb. It shares a 79.6% sequence identity with SARS-CoV, and 96% with the bat coronavirus RaTG13 (Wu et al., 2020; Zhou et al., 2020). It has a genome of ~29.9 kb, and as other coronaviruses, it comprises four different structural proteins: nucleocapsid (N), envelope (E), membrane (M), and spike (S). It also contains sixteen non-structural proteins and eleven accessory proteins

(Redondo et al., 2021; M. Y. Wang et al., 2020; Wu et al., 2020). A linear map of the SARS-CoV-2 genome can be observed in Figure 4. The spike protein of SARS-CoV-2, through its receptor-binding domain, allows entry into host cells for infection by forming a complex with the angiotensin-converting enzyme 2 (ACE2) in humans (Shang et al., 2020; Walls et al., 2020; Zhou et al., 2020).

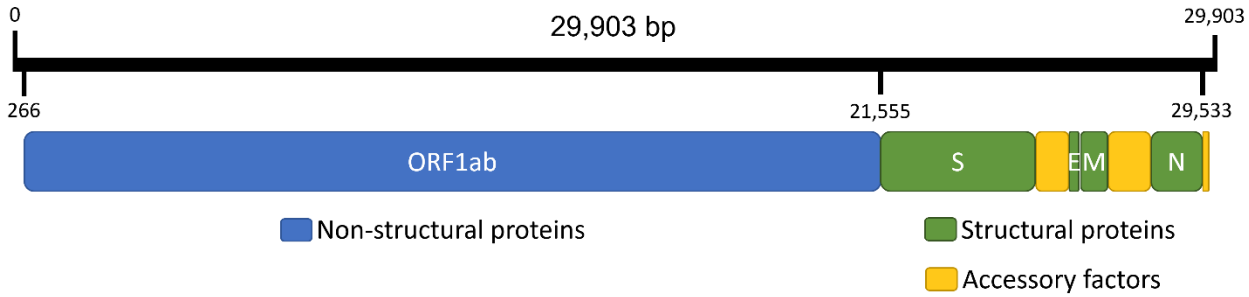


Figure 4. SARS-CoV-2 linear genome map. The ORF1ab which encodes for non-structural proteins is colored blue. In green are displayed its four structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N). And in yellow the regions that encode for accessory factors. Adapted from (Gordon et al., 2020) and the NIH graphics tool for the severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 complete genome (National Library of Medicine (US) National Center for Biotechnology Information, 2020).

As the common signs and symptoms of COVID-19 (e.g., cough, fever, dyspnoea) are present in other upper respiratory tract infections, and given its high transmissibility, molecular tests need to be performed to diagnose COVID-19 and prevent virus propagation (Chu et al., 2020; Harrison et al., 2020; Kevadiya et al., 2021). For SARS-CoV-2 identification and diagnosis, rapid antigen tests have been widely used due to

their cost, rapidness, and simplicity, allowing to obtain a result within minutes and without the need of expensive laboratory equipment or a highly expert individual. However, rapid antigen tests generally present a lower sensitivity than reverse transcription polymerase chain reaction (RT-PCR), and thus, its use for diagnosis highly depends on the clinical context of the patient (Dinnes et al., 2021; Kruttgen et al., 2021; Mak et al., 2020). Nucleic acid amplification tests, and specifically, RT-PCR, are better for clinical detection of SARS-CoV-2 due to their specificity and sensitivity (Kevadiya et al., 2021; Kruttgen et al., 2021). Nevertheless, when it comes to a more detailed surveillance, genomic sequencing is required (Brejova et al., 2021).

1.2.3 Sequencing approaches during the SARS-CoV-2 pandemic

Research groups have developed primer sets that allow the generation of complete SARS-CoV-2 genomes through PCR. The ARTIC Network developed a set of primers for tiled, multiplex PCR amplification that produces ~400 bp amplicons. Multiplexing refers to the use of more than 1 primer pair in a single PCR reaction, and tiling refers to having amplicons that overlap each other. The overlapping of amplicons aids in generating sequences with higher coverages and building more reliable consensus sequences. The first version was released in January 2020, the ARTIC SARS-CoV-2 primers set are continuously updated to improve genome coverage by avoiding amplicon dropouts as new variants with mutations in primer-binding sites appear (ARTIC Network; Davis James et al.; Tyson et al., 2020).

The Midnight primers are another multiplex, tiled primer set created for whole genome amplification of the SARS-CoV-2 genome. This set consists of 29 primers and

generates amplicons of ~1200 bp (Freed et al., 2020). Another primer set developed by a research group is the Entebbe panel. This set generates amplicons of ~1500 bp (Cotten et al., 2021). Detailed characteristics of these primer sets can be observed in Table 2.

Table 2. ARTIC, Midnight, and Entebbe SARS-CoV-2 primer sets information.

	ARTIC	Midnight	Entebbe
Amplicon length	~400 bp	~1200 bp	~1500 bp
Number of amplicons	99	29	20
Number of pools	2	2	2
Coordinates covered	50-29,827	30–28,985	31-29,866
Coverage	99.7%	97%	99.7%
Reference	(ARTIC Network, n.d.; Integrated DNA Technologies 2023a)	(Freed et al., 2020; Integrated DNA Technologies 2023b)	(Cotten et al., 2021)

Furthermore, companies have also produced primer panels for SARS-CoV-2, such as the AmpliSeq for Illumina SARS-CoV-2 Research Panel from Illumina, the xGen™ SARS-CoV-2 Amplicon Panel from IDT, and the CleanPlex SARS-CoV-2 Kit from Paragon Genomics. Each one of these primer sets is directed toward specific sequencing applications (Illumina 2023a; Integrated DNA Technologies 2023; Paragon Genomics n.d.).

1.2.4 Previous approaches for library preparation and sequencing in the field

Due to the portability of the MinION and given that it does not need any specialized setup or calibration, ONT sequencing has been employed for quick deployment in outbreak situations when sequencing facilities are not accessible, such as during the Ebola pandemic in West Africa, in which MinION devices were set up for the first time for on-site sample sequencing, out of a dedicated laboratory. All equipment necessary for sequencing, including the MinION sequencing device and reagents, were taken to Liberia and Guinea as checked commercial luggage, and promptly installed, allowing to promptly obtain Ebola virus sequences from positive clinical samples and analyze them (Hoenen et al., 2016; Quick et al., 2016).

Another early approach was the development of a mobile sequencing laboratory using MinION devices by the Zika in Brazil Real-time Analysis (ZiBRA) project during the 2016 Zika virus outbreak in Brazil and travelled throughout the areas with the highest number of Zika virus cases. This was a relevant strategy because, according to the ZiBRA project, the typical diagnostic process typically involves sample transportation to the state capital for analysis, which can delay results by several months, highlighting the importance of on-site portable sequencing during outbreaks (Faria et al., 2016).

On-site outbreak surveillance with ONT's MinION has been implemented in other outbreaks including the Lassa virus (Kafetzopoulou et al., 2019), yellow fever (Giovanetti et al., 2019), swine flu (Jia et al., 2020), avian influenza virus (de Vries et al., 2022), pathogens causing bacterial meningitis (Pallerla et al., 2022), and the ongoing SARS-CoV-2 pandemic, among others (Freed et al., 2020; Li et al., 2020; Oxford Nanopore Technologies n.d.-c; M. Wang et al., 2020; Wang et al., 2021).

On-site sequencing of wildlife has also been done using portable Nanopore sequencing. Examples include the implementation of sample collection, library preparation, and sequencing for the identification of reptiles and amphibians in Ecuadorian and Tanzanian rainforests (Menegon et al., 2017; Pomerantz et al., 2018).

Additional directions have been proposed, and advance has been made for portable Nanopore sequencing in a number of other fields, like crop and livestock monitoring (Boykin et al., 2018; Lamb et al., 2020), and human DNA re-identification for forensics and other applications (Zaaijer et al., 2017). Moreover, Nanopore sequencing has even been performed in space, when libraries were prepared and sequenced on the International Space Station (Castro-Wallace et al., 2017).

1.3 Automated approaches for sequencing library preparation

Genetic material must be subjected to several preparation steps before loading into a sequencing device. This process, called library preparation, is typically a lengthy methodology that involves several steps, including DNA repair, amplification, addition of sample identifiers for sample multiplexing, and magnetic bead clean-ups, among other steps crucial for a successful sequencing run. Library preparation requires several hours of work from a trained technician and the use of plastic consumables such as pipette tips and tubes. These factors have led to the development and implementation of a number of automated library preparation devices (Hess et al., 2020; Illumina n.d.; Oxford Nanopore Technologies n.d.-f; Shi et al., 2023). Automated library preparation devices allow users to standardize procedures, reduce contamination, increase precision, reduce hands-on user time, and obtain ready-to-sequence libraries in a rapid

way (Farias-Hesson et al., 2010). Currently, the two main approaches for automation are pipetting workstations and microfluidic devices (Hess et al., 2020).

Pipetting workstations essentially substitute the technician's pipetting by doing it themselves using robotic arms and pipetting heads. Magnetic plates and thermal zones can be coupled with the devices to provide the user with the possibility of performing a fully automated library preparation. Due to their ability to manage multiple plates at once, these workstations are ideal when a huge number of libraries need to be prepared simultaneously. Some of the companies producing pipetting workstations are Beckman Coulter, Eppendorf, PerkinElmer, and Hamilton (Hess et al., 2020; Illumina 2023c).

1.3.1 Microfluidics for automated library preparation

The use of droplet-based microfluidic devices for sequencing library preparation is a novel approach, even if the fundamental concepts behind these developments were established more than a century ago. Gabriel Lippmann's work on electrocapillarity laid the groundwork for the study of electrowetting, which is the driving force behind the latest generation of automated microfluidic devices (Mugele & Baret, 2005). Electrowetting is the manipulation of a droplet's interfacial tension through the application of an electric field. This effect allows the movement of droplets in electrode boards without having them in contact with any other material apart from the reagents themselves and the matrix in which they move (Hess et al., 2020; Illumina 2023b; Mugele & Baret, 2005; Pollack et al., 2000; Shen et al., 2023). In contrast, centrifugal microfluidic systems can produce sequencing libraries using centrifugal force for molecule separation, pneumatic pumps for fluid movement, and heating modules can

be added as well. (Hess et al., 2021; Park et al., 2020). Library preparation through automated microfluidic devices eliminates contamination risk during the process as it is carried out in enclosed cartridges, reduces the amount of reagents and materials used, and avoids extensive shearing of the genetic material through the elimination of the majority of the pipetting steps; however, single-use cartridges need to be used for every experiment (Hess et al., 2020; Illumina 2023b).

One of the companies working in the development of droplet-based microfluidic library preparation devices is Miroculus. Their device, the Miro Canvas, allows the movement, mixing, merging, and splitting of droplets and has thermal and magnetic zones. The Miro Canvas has an electrode board at the bottom of a slot in which a single-use cartridge is inserted. The cartridge contains one hydrophobic layer on top and one on the bottom, which create an environment for the droplets to move as the electrode board on which the cartridge sits applies a series of electric potentials, moving droplets around. The reagents are inserted into the cartridge, the protocol is selected, and the device performs the library preparation (Miroculus 2023).

Another company working on the development of automated microfluidic library preparation platforms is Volta Labs, which is working on the development of a device using electrowetting as the moving mechanism, with magnetic and thermal zones as well (Volta Labs n.d.). A commercially available device is still in development.

Moreover, ONT has developed its own microfluidics automatic library preparation platform, the VolTRAX, which is a portable device that contains a heater, a peltier, an optical fluorescent detector for nucleic acid quantification, and magnets for magnetic

bead cleanups. Samples are loaded into a disposable cartridge, and in a similar fashion to Miroculus and Volta, the droplets are moved around by applying an electric charge via electrowetting, with protocols programmed and controlled by software on a computer (Oxford Nanopore Technologies 2023b).

On the other hand, the National Research Council (NRC) has developed an automated microfluidics device for protocol automation that differs in the motion forces used for droplet movement, as it uses centrifugation and pneumatic pressure to perform protocols such as library preparation. The device, called the PowerBlade, uses cheap, disposable thermoplastic microfluidic chips that contain microchannels and reaction chambers in which samples and reagents are loaded. The chips are then placed in the PowerBlade, and the protocol is carried out automatically (National Research Council, 2022). The PowerBlade is shown in Figure 5.

Figure 5. PowerBlade. NRC-developed microfluidic device for protocol automation. PowerBlade image from the NRC (National Research Council, 2022).

PowerBlade's use of thermoplastic chips offers flexibility in design, as new chips can be designed and integrated for automating different protocols while keeping each reaction at a low cost. Additionally, when compared to other microfluidic automation platforms

like ONT's VolTRAX, Miroculus Canvas, and Volta Labs device, the PowerBlade is currently the only one that offers the entire sample-to-workflow process and uses cheaper cartridges (Miroculus 2023; National Research Council, 2022; Oxford Nanopore Technologies 2023b; Volta Labs n.d.).

1.4 Rationale

Sequencing is still limited by library preparation. Library preparation can take several hours and needs to be performed by a trained technician in an appropriate laboratory or by a pipetting workstation, which is not possible in remote areas, resource-limited settings, or places where a molecular biology technician is not always available, such as hospitals and clinics. In the context of outbreaks such as the current SARS-CoV-2 pandemic, the capability of generating sequencing data in a timely manner is crucial for etiological agent identification, classification, and a rapid response for the benefit of individual patients or for entire communities in the sense of public health policies.

To tackle this issue, our laboratory has collaborated with the National Research Council to develop an automated workflow for Nanopore library preparation from SARS-CoV-2 clinical samples using the NRC-developed automated microfluidics device, the PowerBlade. The combination of this automated, portable device with portable sequencing technologies, such as those developed by Nanopore, gives us a workflow that can be easily run almost anywhere, not limited to hospitals or laboratories but also in remote locations, without the need for a highly experienced technician due to the automation of the workflow. Furthermore, this is an adaptable technology that can be modified and used for the identification of other pathogens.

The entire process, from sample RNA extraction to final library, takes place on a set of two microfluidic chips that are run in the PowerBlade. Chip #1 was designed to perform RNA extraction, reverse transcription, multiplex SARS-CoV-2 PCR with 98 primers in two separate pools for whole genome SARS-CoV-2 amplification (PCR ramp rate of 35°C/minute), and DNA purification. Chip #2, previously designed by the NRC, was used for library preparation, which includes end repair/dA-tailing (End-prep), barcode ligation, sequencing adapter ligation, and DNA purification. A diagram of the PowerBlade workflow and the chips can be observed in Figure 6.

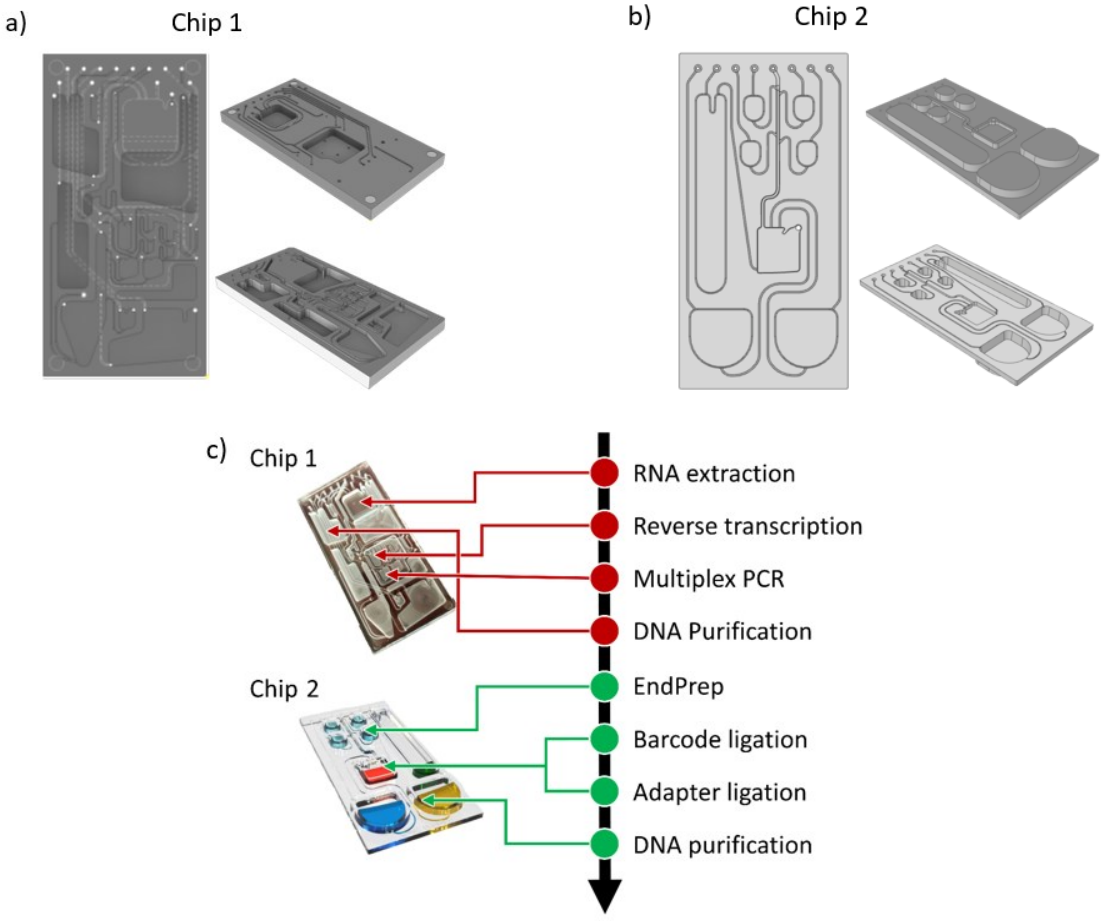


Figure 6. a) Chip 1. b) Chip 2 c) On-chip workflow. RNA extraction, reverse transcription, multiplex PCR, and DNA purification are carried out on chip 1. Library preparation is done on chip 2.

Considering the limited on-chip space and the amount of time that the ligation reagents would have to endure at room or higher temperatures before starting the reaction, the standard ONT barcoding ligation protocol (Figure 7a) was not appropriate for on-chip library preparation. We developed a novel Nanopore library preparation protocol in which both barcode and sequencing adapter are ligated in the same step (Figure 7b). This one-step ligation protocol would be suitable for on-chip implementation since it requires fewer steps and less space to function as a result of the merging of the two ligation steps. The protocol, previously developed in our laboratory, will be tested against the standard ONT protocol to prove its efficacy.

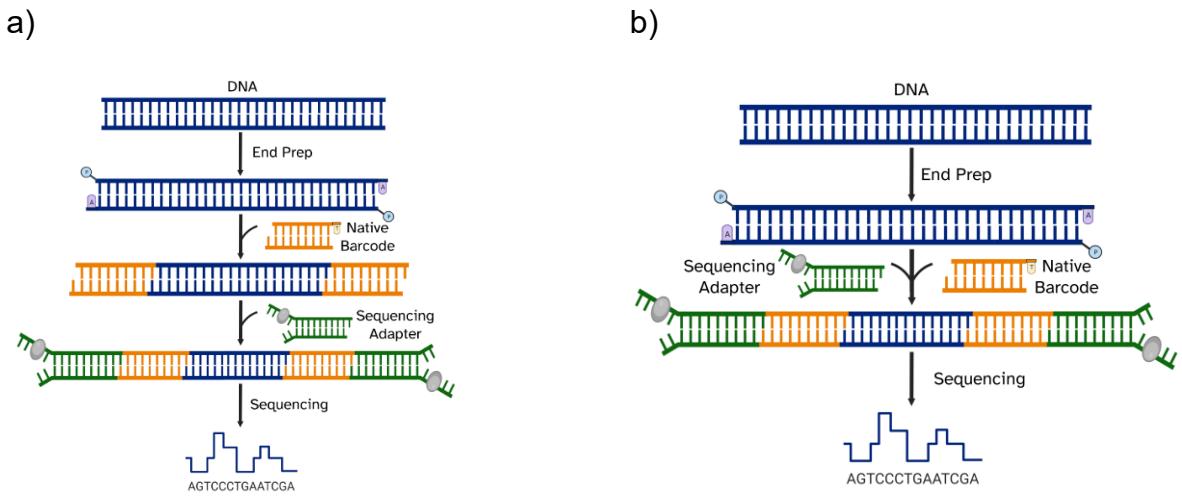


Figure 7. Library preparation protocols. a) ONT standard library preparation protocol (Two-step ligation). b) One-step ligation library preparation protocol developed in our laboratory.

1.5 Hypothesis and objectives

The aim of the present work is to test three key aspects of the PowerBlade on-chip workflow: the novel one-step ligation library preparation protocol, on-chip PCR, and on-chip library preparation using the one-step ligation protocol. The novel one-step ligation protocol was recently developed in our laboratory for on-chip implementation on the PowerBlade. Its performance will be compared against the standard ONT library preparation protocol in this thesis. The SARS-CoV-2 ARTIC PCR and primer set was developed by the ARTIC consortium (ARTIC Network, n.d.). The PowerBlade and Chip #1 were developed by the NRC, and chip #2 was developed by the NRC in collaboration with our laboratory. We hypothesized that the correct functioning of these elements will contribute towards the implementation of the PowerBlade automated workflow for Nanopore library preparation from SARS-CoV-2 clinical samples.

The objectives of this thesis are the following:

I. Evaluate our novel one-step ligation library preparation protocol in comparison to the standard (two-step ligation) library preparation protocol.
II. Evaluate on-chip PCR performance in comparison against standard off-chip PCR carried out in a thermocycler.
III. Evaluate on-chip library preparation performance in comparison with standard off-chip library preparation. Both on-chip and off-chip library preparations will be performed using the one-step ligation library preparation protocol.

Chapter 2: Materials and Methods

2.1 Novel library preparation protocol evaluation

Two sets of experiments were carried out to test our novel library preparation protocol. The genetic material used for both sets of experiments was SARS-CoV-2 ARTIC amplicons generated from cDNA coming from RNA extracted from clinical nasopharyngeal samples of COVID-19 positive patients. The first set of experiments was meant to evaluate the performance of our novel library preparation protocol against the standard protocol by preparing libraries on the bench with different amounts of ARTIC amplicons as starting material. The amounts used were 100, 75, and 50 ng. Libraries were prepared with three samples for each amplicon amount, this information can be observed in Table 3. The libraries produced were sequenced on a PromethION flow cell without normalizing them.

Table 3. DNA (ARTIC amplicon) amount and number of samples used for the second set of library preparation experiments.

Amplicon amount used (ng)	Number of samples
100	3
75	3
50	3

The second set of experiments involved library preparation with amplicons coming from clinical samples of different Ct values, with both protocols. The Ct value is defined as the number of amplification cycles that a sample needs to go through in real-time PCR to

cross a predefined fluorescence threshold that corresponds to the amount of DNA generated during the PCR reaction. Ct values provide an insight into the original sample in an inverse correlation manner: low Ct values indicate a higher nucleic acid concentration in the sample, while higher Ct values indicate a lower nucleic acid concentration in the sample (Ontario Agency for Health Protection and Promotion (Public Health Ontario), 2020). The Ct values used were 15, 20, 25, and 30. Three different samples were used for each Ct value, which is displayed in Table 4. The libraries generated were adjusted to the same concentration and sequenced on a single PromethION flow cell.

Table 4. Ct values and number of samples used for the first set of library preparation evaluation experiments.

Ct value	Number of samples
15	3
20	3
25	3
30	3

2.2 On-chip PCR evaluation

PCR performed on-chip on the PowerBlade was evaluated against off-chip controls. Figure 8a shows the workflow carried out. cDNA produced from Twist Synthetic SARS-CoV-2 RNA controls was used to perform PCR both on-chip on the PowerBlade and on a thermocycler on the bench. Then, library preparation was carried out on the bench and the libraries of each experiment set were sequenced together on

PromethION flow cells after concentration normalization. There were 2 stages of tests for on-chip PCR evaluation, the number of on-chip and off-chip runs per round can be observed in Figure 8b. This was done given that after the first on-chip PCR experimental stage, a calibration of the temperature in the PCR chambers was performed to improve the results obtained. During the first experiment round, two on-chip PCR reactions were carried out along with one off-chip control, and on the second stage, after calibration of the temperature inside the PCR chambers of the chip, two on-chip PCRs were performed with two off-chip controls.

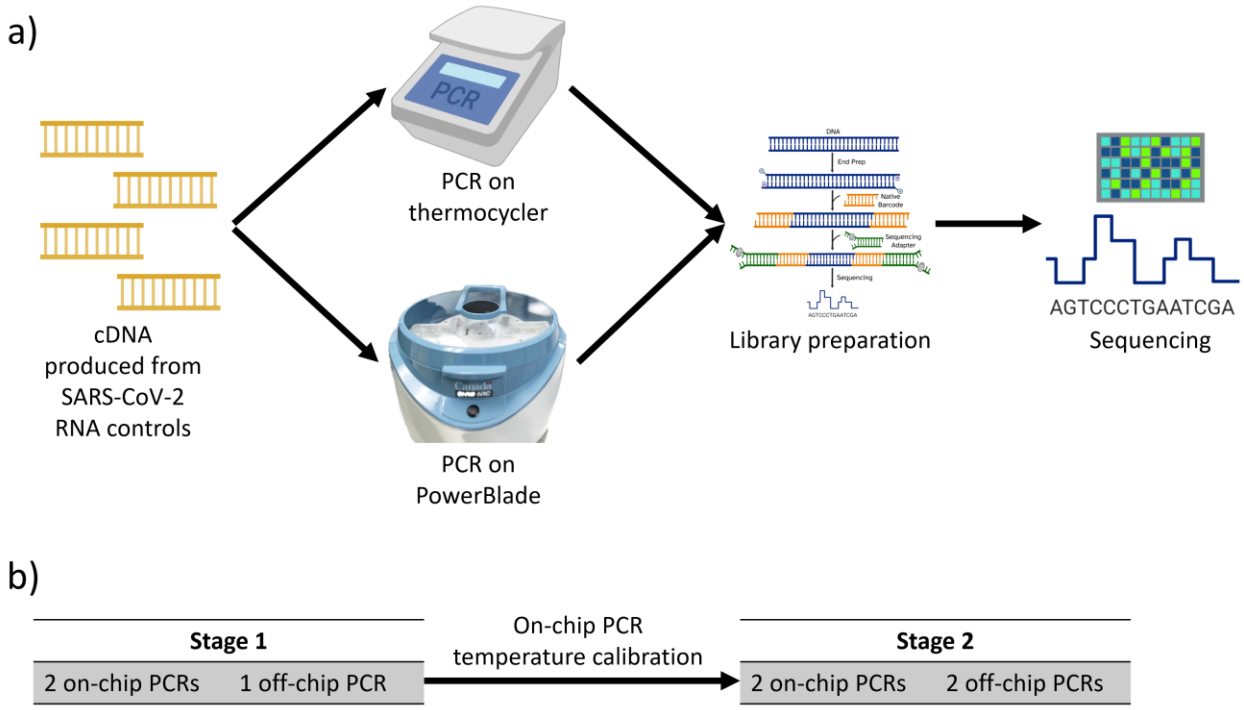
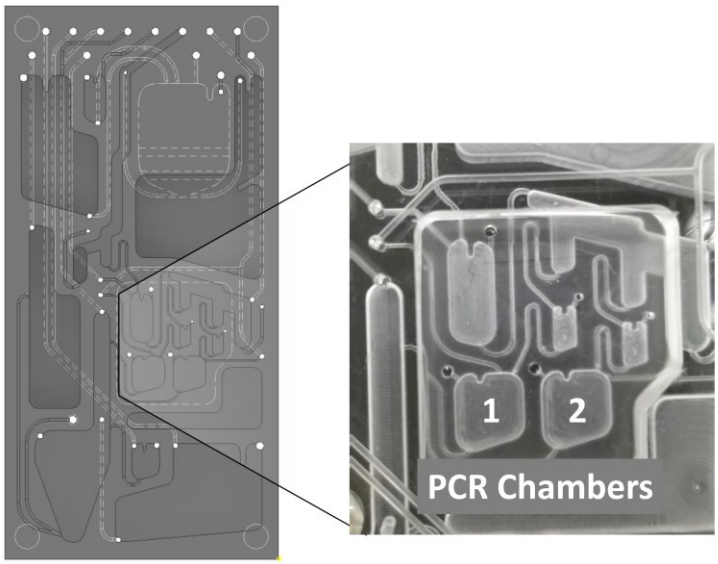


Figure 8. On-chip PCR evaluation workflow. a) Experiment workflow carried out in both stages of experimentation b) Stages of on-chip PCR testing. During stage 1, two on-chip PCR reactions were carried out, with an off-chip PCR control. For stage 2, after temperature calibration, two on-chip PCR reactions were performed, with two off-chip PCR controls.

As mentioned before, for each stage, two on-chip PCRs were carried out, however, these are not replicates. Given that the PCR is carried out in two pools in two different chambers of the chip, named chamber A & H (Figure 9a), the first PCR reaction from both stages was performed with pool 1 being amplified in chamber 1, and pool 2 being amplified on chamber 2. The second PCR reaction from both stages was performed vice versa, with pool 1 in chamber 2, and pool 2 in chamber 1, this experiment design can be observed in Figure 9b.

a)



b)

	Pool 1	Pool 2
PCR 1	Chamber **1**	Chamber **2**
PCR 2	Chamber **2**	Chamber **1**

Figure 9. PCR chambers and primer pools amplified in chambers. a) PCR chambers located on chip 1. b) PCR pool amplification on the two different chambers for both stages. For one of the PCR reactions of each stage, pool 1 was amplified on chamber 1 and pool 2 on chamber 2. For the other PCR reaction, pool 1 was amplified on chamber 2 and pool 2 on chamber 1.

2.3 On-chip library preparation evaluation

Nanopore sequencing library preparation was evaluated against off-chip controls. The workflow carried out can be observed in Figure 10. Sequencing libraries were prepared both on-chip on the PowerBlade, and off-chip on the bench from DNA SARS-CoV-2 ARTIC amplicons produced through standard on-bench PCR from cDNA from Twist Synthetic SARS-CoV-2 RNA controls. The libraries were sequenced independently without normalization, all in PromethION flow cells.

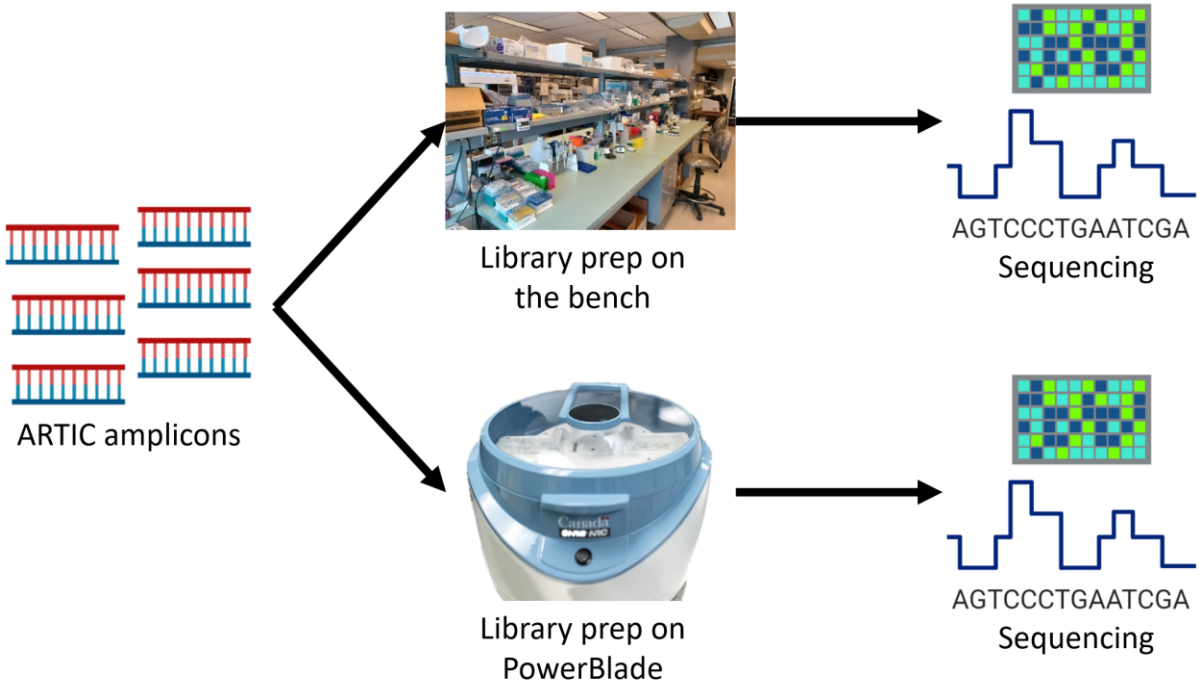


Figure 10. On-chip library preparation evaluation workflow.

2.4 Protocols

RNA extracted from nasopharyngeal swabs of SARS-CoV-2 positive patients was reverse transcribed and PCR amplified using the ARTIC SARS-CoV-2 primers. As positive controls, Twist Synthetic SARS-CoV-2 RNA controls were reverse transcribed, and PCR amplified following the same methods as the samples. The protocols are described in detail in the following subsections.

2.4.1 Reverse Transcription

Reverse transcription was carried out using a modified LunaScript protocol developed at our lab (Chen et al., 2020), which is based on the Protocol for LunaScript RTSuperMix Kit (E3010) from NEB.

For each sample, 11 µL of RNA were mixed with 4 µL of LunaScript RT SuperMix (5x) (NEB), and 5 µL of nuclease-free water. The components were then gently mixed by pipetting and pulse-centrifuged. The reaction was incubated at 25°C for 2 minutes, 55°C for 20 minutes, and 95°C for 1 minute. The reaction can then be placed on ice for 1 minute or stored at -20°C.

2.4.2 PCR for SARS-CoV-2 genome amplification

SARS-CoV-2 genome amplification was carried out through PCR using the ARTIC SARS-CoV-2 primers (ARTIC Network) with the protocol that is currently being used at our lab (J Reiling et al., 2022).

For each sample, two reactions were set, one with ARTIC primer pool 1 and the other with ARTIC primer pool 2 (Integrated DNA Technologies). Each reaction was set up

with 3.7 µL of primer pool, 3.8 µL of nuclease-free water, 12.5 µL of Q5 Hot Start High-Fidelity 2X Master Mix (NEB), and 5 µL of reverse transcription product obtained in the Reverse Transcription step. The reactions were then gently mixed by pipetting and pulse-centrifuged. The reaction was incubated at 98°C for 30 seconds. This was followed by 36 cycles of 98°C for 15 minutes and 63°C for 5 minutes. The reaction can then be placed on ice for 1 minute or stored at -20°C. A post-PCR magnetic bead cleanup was performed using SPRIselect beads (Beckman Coulter). The entire contents of the reaction products of pool 1 and pool 2 were combined for each sample. An equal volume of SPRI beads was added to the combined products, mixed by pipetting, and incubated at RT for 5 minutes. Then, they were incubated on a magnetic rack for 5 minutes and the supernatant was removed. While still on the magnetic rack, 200 µL of fresh 80% ethanol were added and discarded after 30 seconds, this step was done twice for a total of two 80% ethanol washes. The beads were left to dry for 3 minutes, then removed from the magnetic rack, and 30 µL of Buffer EB (QIAGEN) were added. The beads were resuspended and incubated at RT for 3 minutes. Then, they were incubated on the magnetic rack for 5 minutes and the eluate was recovered.

2.4.3 ONT Library preparation

The standard ONT barcoded library preparation process comprises 3 steps: EndPrep, barcode ligation, and sequencing adapter ligation. The novel one-step ligation library preparation protocol developed in our laboratory combines the two ligation steps into a single one, thus reducing the amount of time needed for library preparation, and in the case of implementation on a microfluidic device, reducing the number of chambers/inlets needed to use.

The standard library preparation protocol was carried out based on the nCoV-2019 McGill Nanopore LibPrep Protocol used at our lab for library preparation of ARTIC amplicons from SARS-CoV-2 clinical samples (Reiling et al., 2020). First, for the EndPrep reaction, 5 µL of DNA amplicons generated in the PCR step are mixed with 7.5 µL of nuclease-free water, 1.75 µL of Ultra II End Prep Reaction Buffer (NEB), and 0.75 µL of Ultra II End Prep Enzyme Mix (NEB). The reaction was incubated at RT for 10 minutes, 65°C for 5 minutes, and placed on ice for 1 minute. Then, for barcode ligation, 2.5 µL of a nanopore native barcode was added to each sample, making sure that each sample had a different barcode. 10 µL of Ultra II Ligation Master Mix (NEB), 0.3 µL of Ligation Enhancer (NEB) and 4.2 µL of nuclease-free water were added too. Then, the reaction was incubated at RT for 15 minutes, 70°C for 10 minutes, and placed on ice for 1 minute. The reaction was then cleaned with SPRIselect beads (Beckman Coulter) by adding 32 µL of beads and gently mixing by pipetting. The reaction was incubated at RT for 5 minutes, then incubated on a magnetic rack for 5 minutes and the supernatant was removed. 200 µL of fresh 80% ethanol were added and discarded after 30 seconds, this step was done twice for a total of two 80% ethanol washes. The beads were left to dry for 3 minutes, then removed from the magnetic rack, and 30 µL of Buffer EB (QIAGEN) were added. The beads were resuspended and incubated at RT for 3 minutes. Then, they were incubated on the magnetic rack for 5 minutes and the eluate was recovered. For adapter ligation, the following reagents were added to the 30 µL of eluate: 10 µL of NEBNext Quick Ligation Reaction Buffer (5X) (NEB), 5 µL of AMII adapter mix (ONT), and 5 µL of Quick T4 DNA Ligase. The reaction was incubated at RT for 20 minutes and placed on ice for 1 minute. After this, 50 µL of SPRIselect beads were added, the

reaction was mixed and incubated at RT for 5 minutes. Then, the sample was placed on a magnetic rack for 5 minutes and the supernatant was removed. The sample was resuspended in 200 µL of SFB, incubated for 5 minutes at RT, and then for 5 minutes on a magnetic rack. The supernatant was discarded and the SFB wash step was repeated for a total of 2 SFB washes. The sample was removed from the magnetic rack and 15 µL of Buffer EB were added to the bead pellet, mixed, incubated at RT for 5 minutes, and then placed on the magnetic rack for 5 minutes. The eluate containing the ONT library was then recovered.

Our novel one-step ligation library preparation protocol in which both barcode and adapter ligation occur in a single step was performed as follows: For the EndPrep, 200 ng in 6.67 µL, 10 µL of nuclease-free water, 2.33 µL of Ultra II End Prep Reaction Buffer (NEB), and 1 µL of Ultra II End Prep Enzyme Mix (NEB) were mixed. The reaction was incubated at RT for 10 minutes, 65°C for 5 minutes, and placed on ice for 1 minute. For the barcode and adapter ligation in a single step the following reagents were added to the product of the EndPrep reaction: 5 µL of nanopore native barcode (ONT) (making sure that each sample had a different barcode), 20 µL of NEBNext Quick Ligation Reaction Buffer (NEB), 0.6 µL of Ligation Enhancer (NEB), 20 µL of Ultra II Ligation Master Mix (NEB), 5 µL of AMII adapter mix (ONT), 10 µL of Quick T4 DNA Ligase (NEB), and 19.4 µL of nuclease-free water. The reaction was incubated at RT for 15 minutes and placed on ice for 1 minute. 100 µL of SPRIselect beads were added to the reaction, mixed by pippeting and incubated at RT for 5 minutes. Then, the sample was placed on a magnetic rack for 5 minutes and the supernatant was removed. The sample was resuspended in 200 µL of SFB, incubated for 5 minutes at RT, and then for 5

minutes on a magnetic rack. The supernatant was discarded and the SFB wash step was repeated for a total of 2 SFB washes. The sample was removed from the magnetic rack and 15 µL of Buffer EB were added to the bead pellet, mixed, incubated at RT for 5 minutes, and then placed on the magnetic rack for 5 minutes. The eluate containing the ONT library was then recovered.

2.5 Sequencing and data analysis

All libraries were sequenced using the PromethION and basecalled using MinKNOW in high accuracy mode. The fastq.gz files produced by MinKNOW were concatenated and unzipped using Linux commands (cat, and gunzip, respectively). Sequencing adapters and native barcodes were removed using porechop (v0.2.4) (Wick et al., 2017). The generated fastq files with a >20 quality were mapped against the complete genome of the severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1 (National Library of Medicine (US) National Center for Biotechnology Information, 2020), using minimap2 (v2.24) (Li, 2018). The generated SAM files were converted to BAM, sorted, indexed using samtools (v1.16.1) (Danecek et al., 2021). Additionally, samtools was used to generate depth, coverage, and read number metrics. Fragment size metrics and plots, as well as coverage plots, and bigWig files were generated using deeptools (v3.5.1) (Ramirez et al., 2016). PCR pool contribution was calculated using the SARS-CoV-2 ARTIC primer sequences, labeling them as either pool 1 or pool 2, and aligning them to the generated reads. Primer pool contribution was calculated by dividing the number of reads that aligned to primers in pool 1 or 2, divided by the total number of reads. Lineage calling was performed using the online tool of Pangolin: Pangolin COVID-19 Lineage Assigner (v4.3) (O'Toole et al., 2021). Phylogeny

graphics were produced using Pathogenwatch (v20.2.8) (Argimon et al., 2021). A diagram displaying the data analysis performed can be observed in Figure 11.

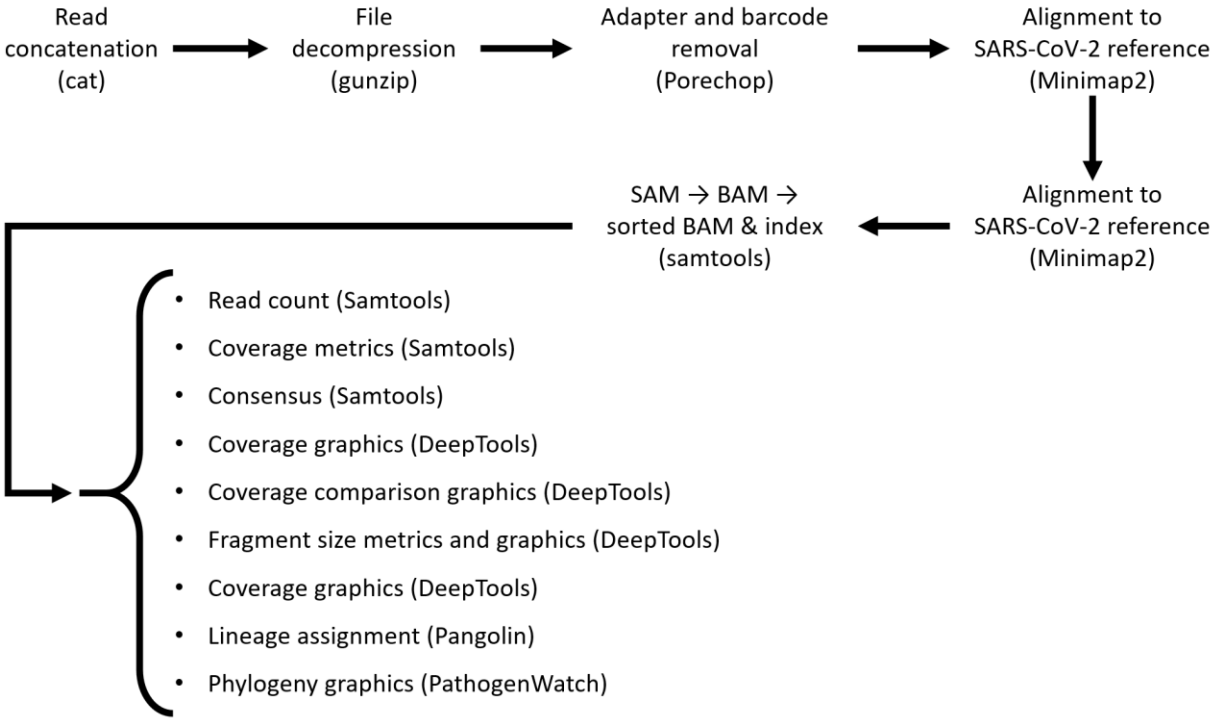


Figure 11. Diagram showing the data analysis performed.

Chapter 3: Results

3.1 Novel library preparation protocol evaluation

3.1.1 Library preparation protocol testing with different amounts of starting material

Nanopore sequencing libraries were prepared with SARS-CoV-2 ARTIC amplicons generated from nasopharyngeal clinical samples, using both the standard two-step ligation Nanopore protocol and our novel one-step ligation protocol. The libraries were prepared with different amounts of starting material: 100, 75, and 50 ng of amplicons. The libraries were prepared with different input amounts of DNA to determine how the DNA amount affects the performance of the protocols. We used these DNA amounts, given that these are common DNA inputs in our process for SARS-CoV-2 sequencing in our laboratory. Each sample was prepared with both the one-step ligation protocol and the standard ONT library preparation protocol, and the DNA came from the same amplification batch; hence, no other variable played a part in the performance observed other than the protocol itself, allowing us to compare how the protocols perform against each other.

Figure 12 shows the sequencing metrics plotted, as well as DNA recovery after library preparation. While the total read and mapped read number varies for each library, both the one-step and two-step protocols present genome coverages of more than 98% for all the libraries. All the libraries prepared presented the expected fragment length for ARTIC SARS-CoV-2 amplicons (~420 base pairs). Figure 13 shows the fragment size, comparing libraries prepared with the one-step and the two-step ligation protocols.

Figure 14 shows that even when all libraries present a high and similar coverage, these do not present the same coverage depth. Figures 15-18 show differences in coverage depth along the SARS-CoV-2 genome.

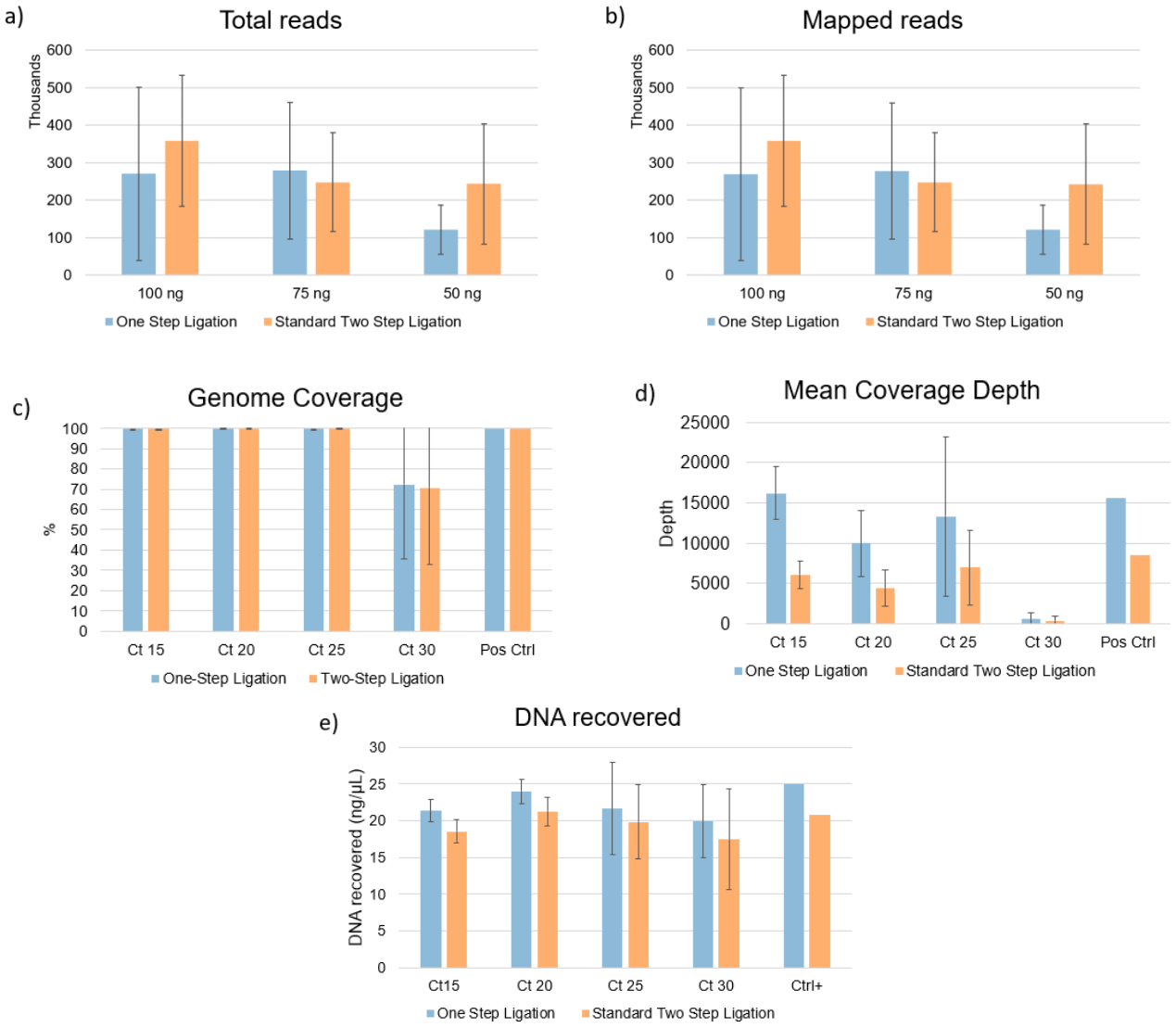


Figure 12. Sequencing metrics and DNA recovery of libraries prepared with different amounts of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. a) Total reads. b) Mapped reads. c) Genome Coverage. d) Mean Coverage Depth. e) DNA recovered after library preparation.

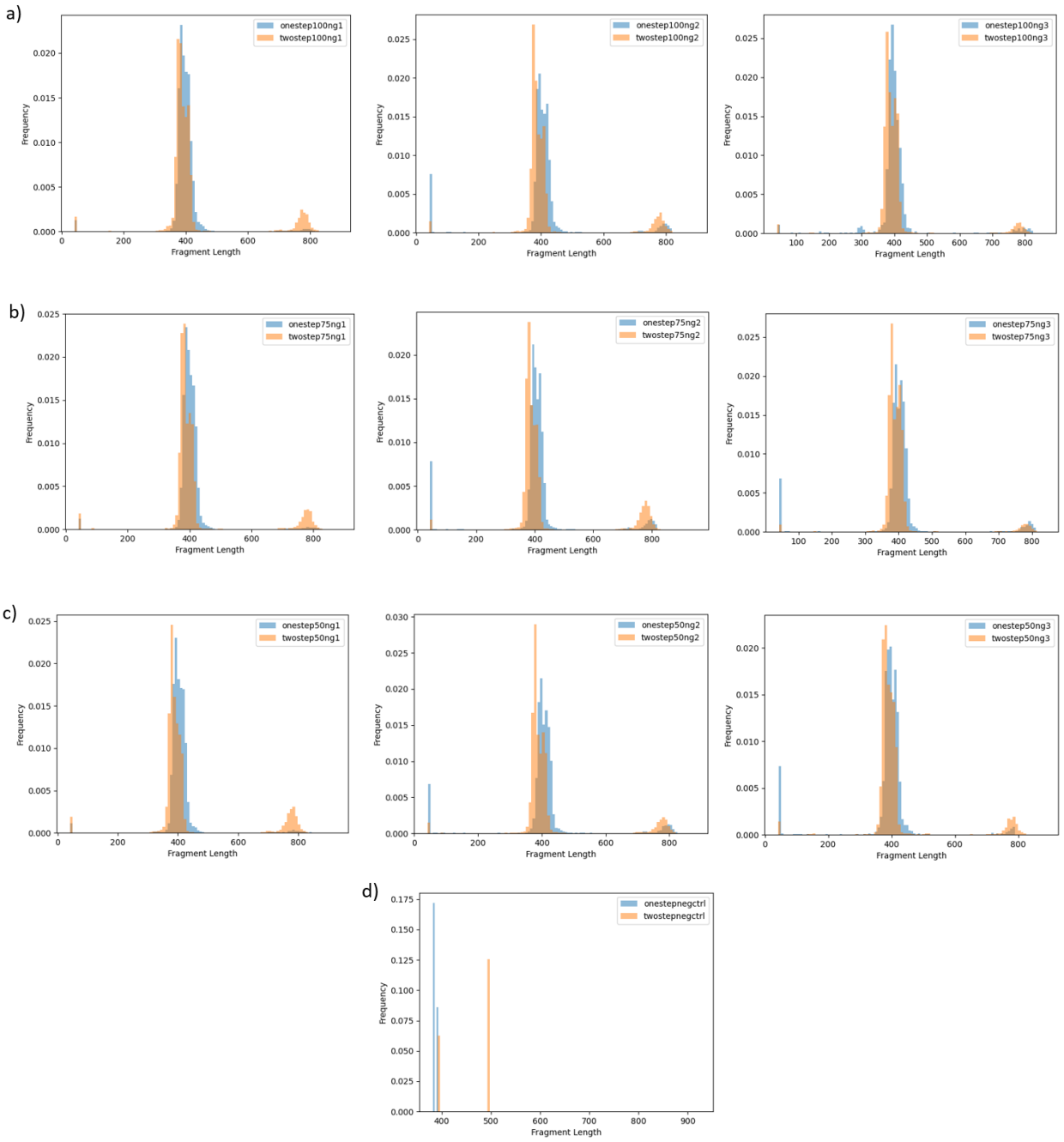


Figure 13. Distribution of sequenced fragment length of libraries prepared with different amounts of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. a) 100 ng samples. b) 75 ng samples. c) 50 ng samples. d) Negative control.

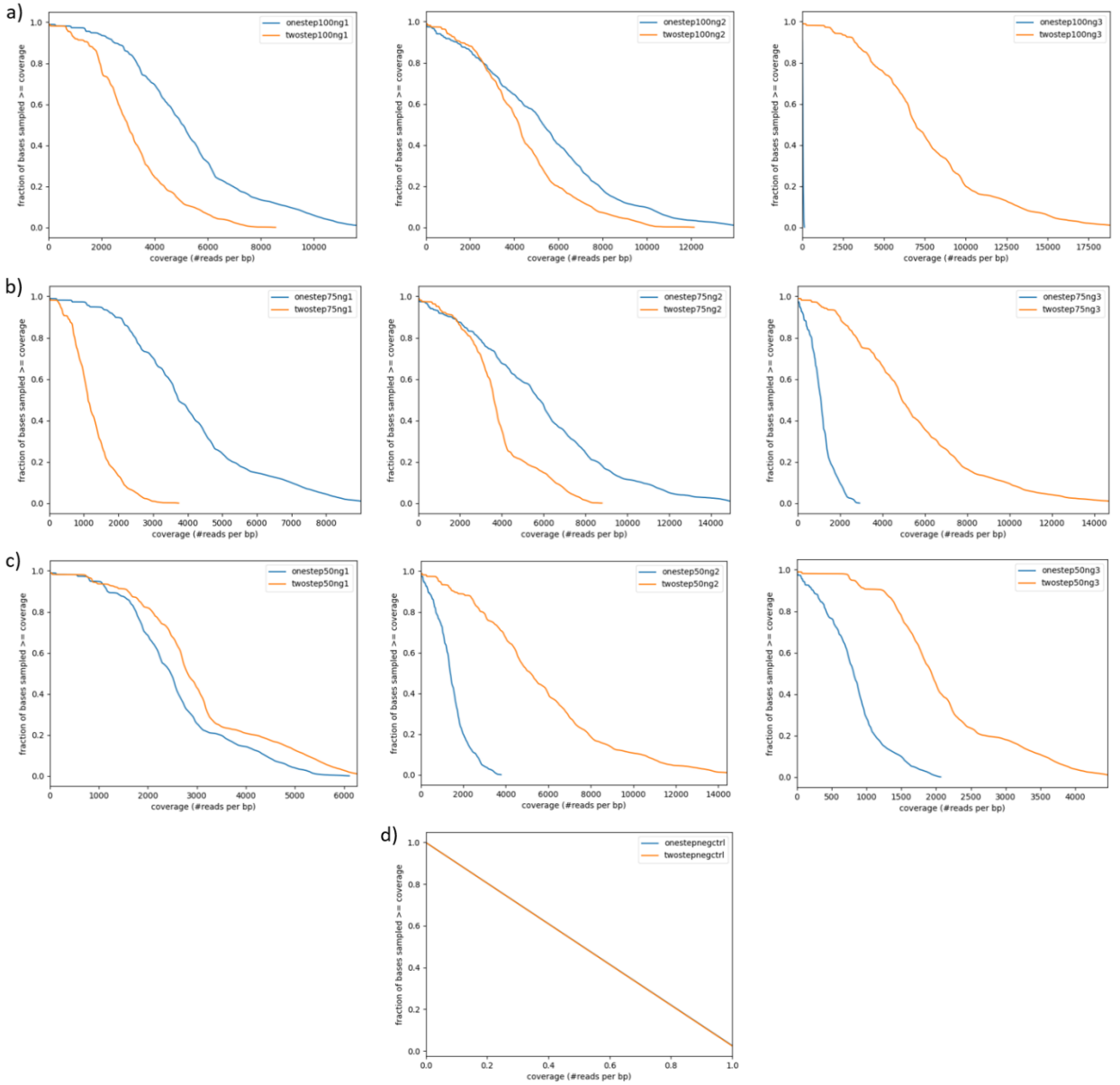


Figure 14. Genome coverage comparison of libraries prepared with different amounts of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Coverage as number of reads per base pair is shown on the x-axis, and the y-axis shows the fraction or percentage of bases sampled with a coverage such as that of the x-axis or higher. a) Ct 15 samples. b) Ct 20 samples. c) Ct 25 samples. d) Ct 30 samples.

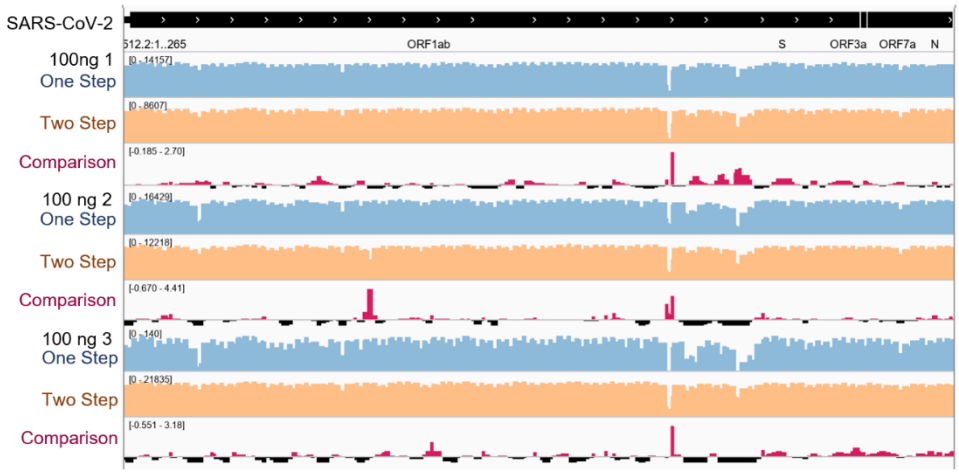


Figure 15. IGV Genome coverage of libraries prepared with 100 ng of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

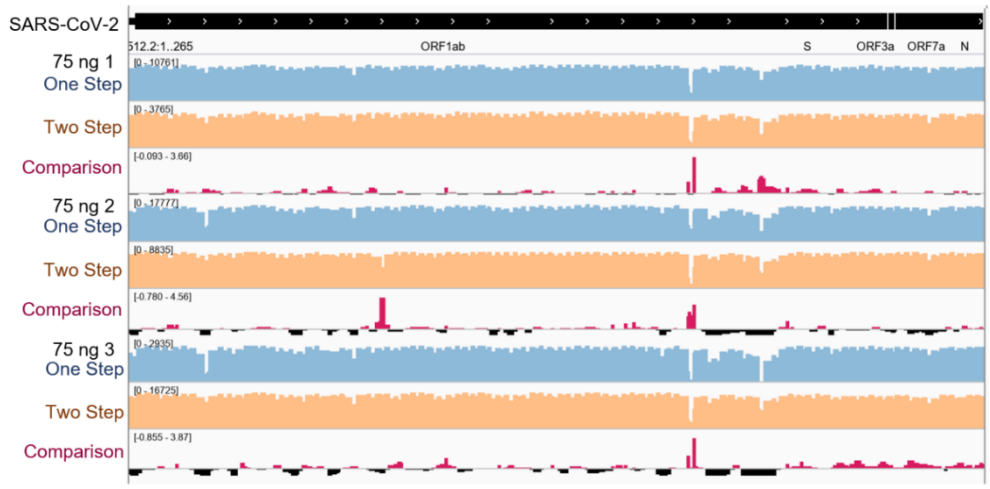


Figure 16. IGV Genome coverage of libraries prepared with 75 ng of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

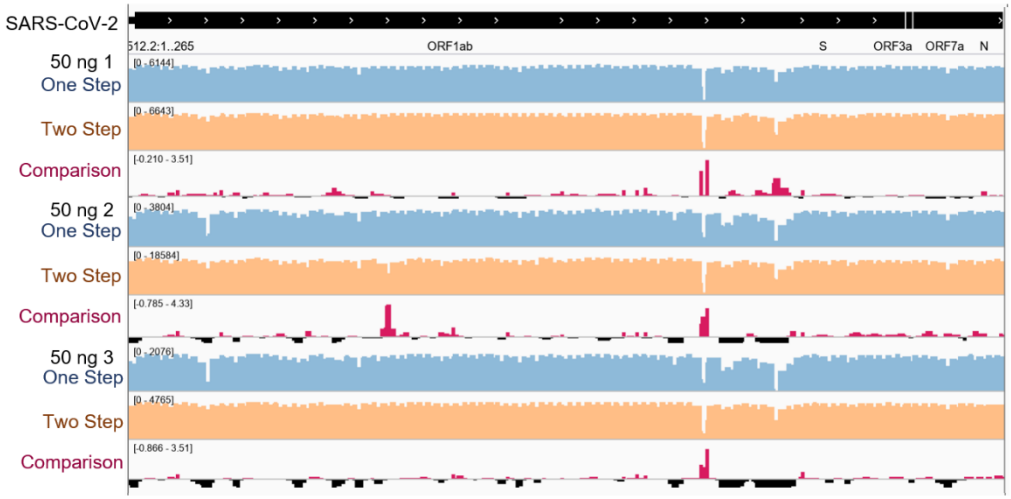


Figure 17. IGV Genome coverage of libraries prepared with 50 ng of DNA as input. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

Figure 18. IGV Genome coverage of libraries prepared as a negative control. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

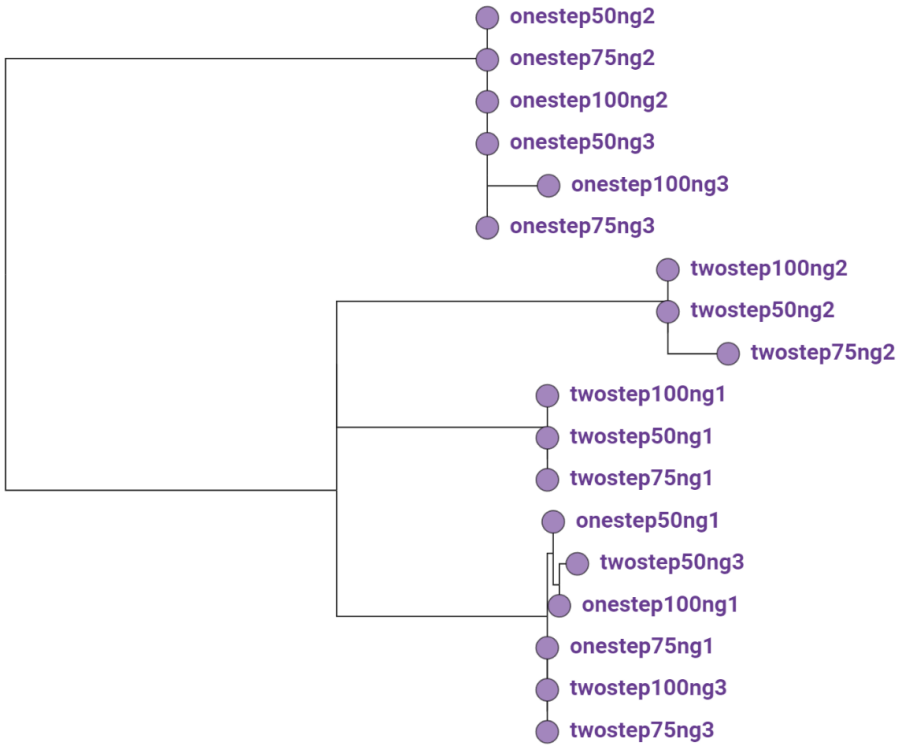


Figure 19. Phylogenetic tree of libraries prepared with different amounts of DNA as input.

3.1.2 Library preparation protocol testing with different Ct values

Nanopore sequencing libraries were prepared with SARS-CoV-2 ARTIC amplicons generated from nasopharyngeal clinical samples, using both the standard Nanopore protocol and our novel one-step ligation protocol. Material of four different Ct values (15, 20, 25, and 30) was used to prepare the libraries. These Ct values were selected as they are common in the samples received from clinical sites. This set of clinical samples is different from the one used in the previous section. However, as each sample was prepared with both the one-step ligation library preparation protocol and the standard ONT library preparation protocol, and the DNA used was produced in the same amplification batch, there is no other variability added other than the library preparation process itself, allowing us to determine how the one-step ligation protocol compares to the standard one.

The sequencing metrics observed in Figure 20 show that more reads were sequenced for the one-step ligation-prepared libraries and that both methods present a similar and complete genome coverage. Nevertheless, we observe a lower coverage, number of mapped reads, and mean depth for the Ct 30 samples. Figure 21 shows the fragment length of the amplicons generated, and most of them lie within the expected size. However, two Ct 30 samples show a larger proportion of fragments differing from the expected size. Genome coverage depth appears to be slightly higher for one-step prepared libraries, as observed in Figure 22. This effect is related to the number of mapped reads sequenced for each library. Generally, as the number of mapped reads increases, the coverage depth increases too, given that all the primers are amplified. We observe the same effect but now on a genome-wide scale in Figures 23-27, in

which the one-step prepared libraries have a higher coverage of most regions of the SARS-CoV-2 genome.

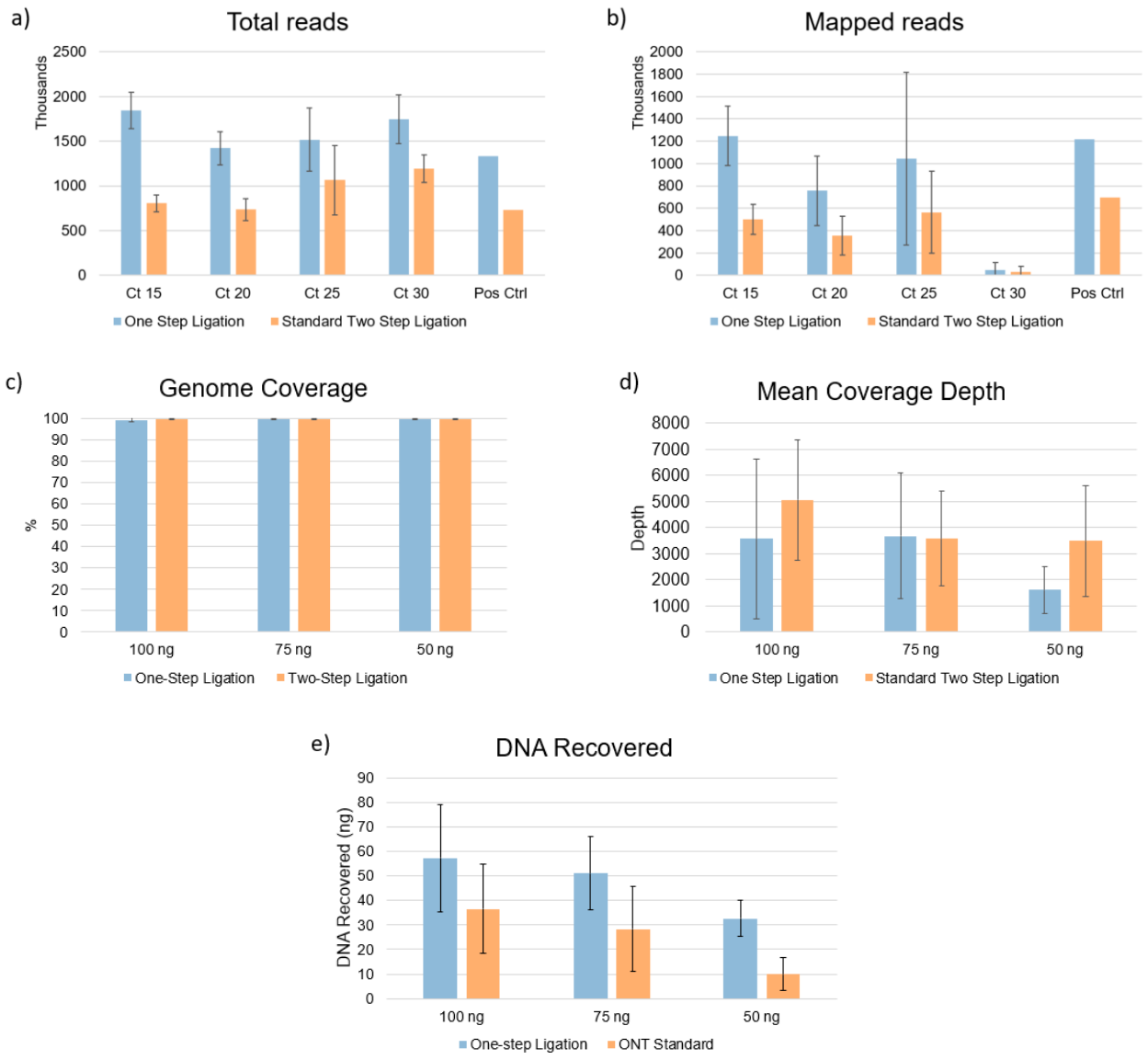


Figure 20. Sequencing metrics and DNA recovery of libraries prepared from samples with different Ct values. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. a) Total reads. b) Mapped reads. c) Genome Coverage. d) Mean Coverage Depth. e) DNA recovered after library preparation.

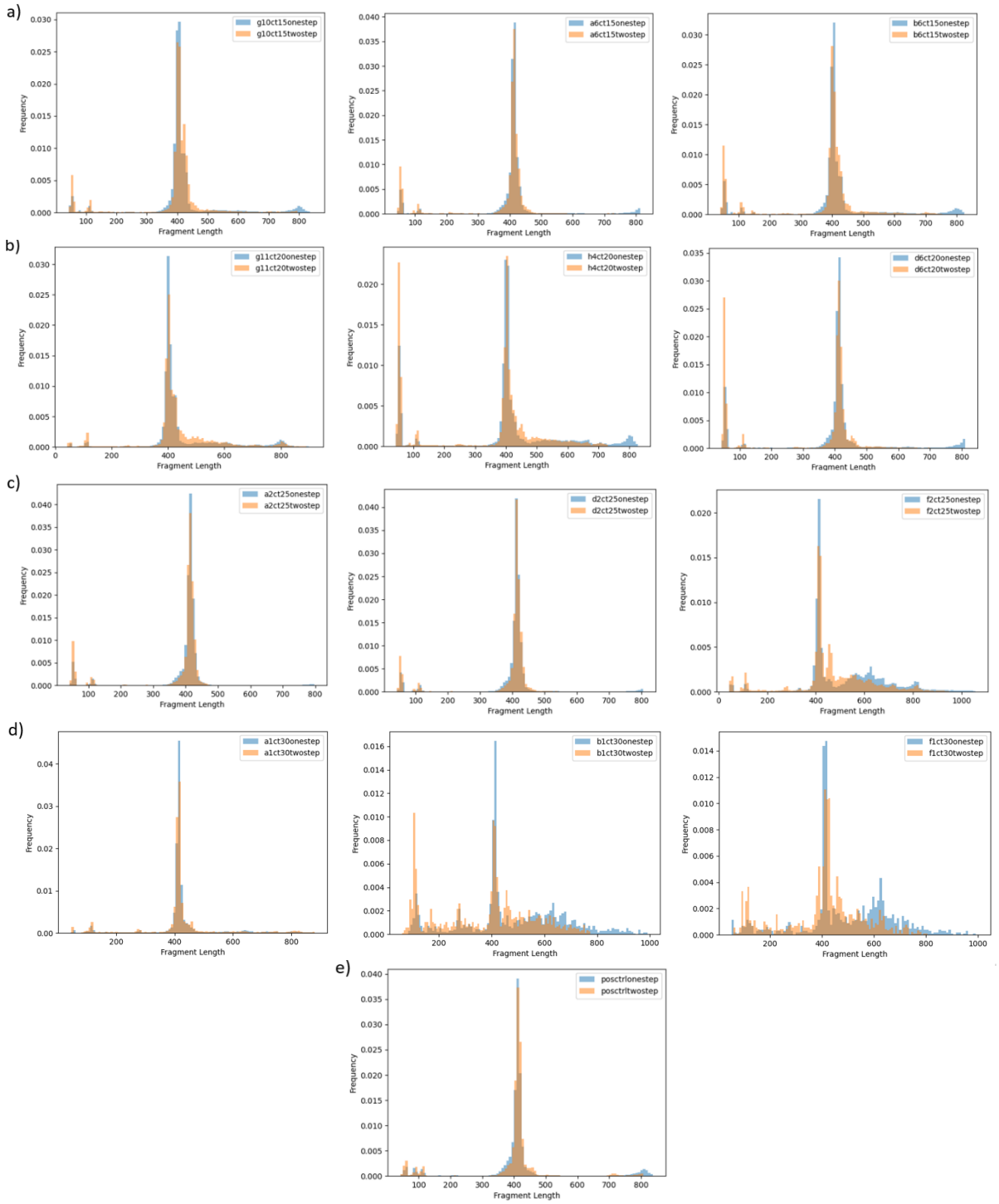


Figure 21. Distribution of sequenced fragment length of libraries prepared from samples with different Ct values. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard protocol are shown in orange. a) Ct 15 samples. b) Ct 20 samples. c) Ct 25 samples. d) Ct 30 samples. e) Positive control

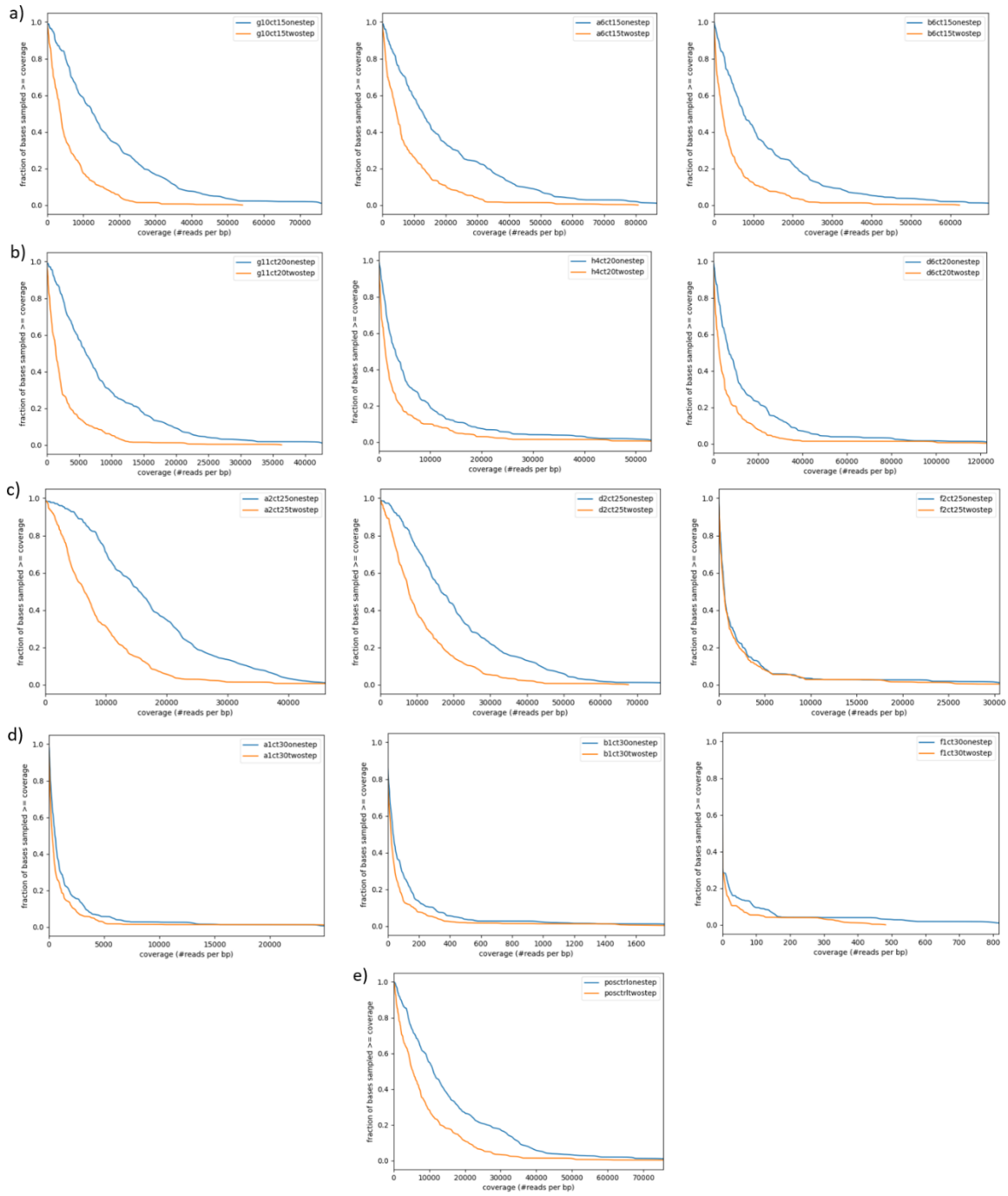


Figure 22. Genome coverage comparison of libraries prepared from samples with different Ct values. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. a) Ct 15 samples. b) Ct 20 samples. c) Ct 25 samples. d) Ct 30 samples. e) Positive control

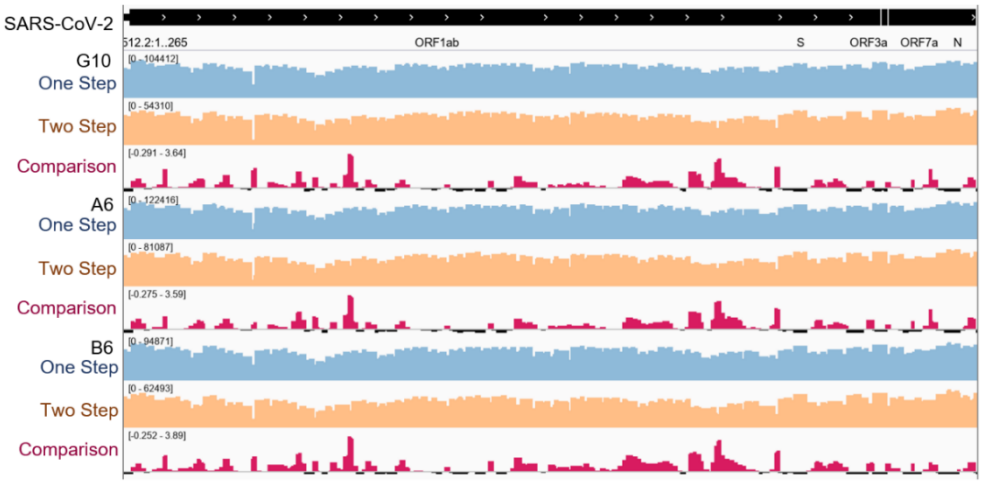


Figure 23. IGV Genome coverage of libraries prepared with Ct 15 samples. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

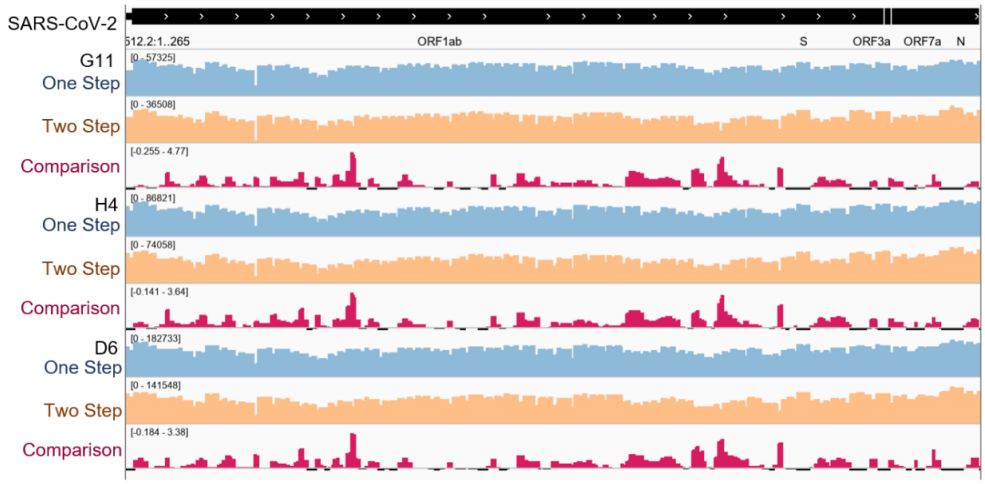


Figure 24. IGV Genome coverage of libraries prepared with Ct 20 samples. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

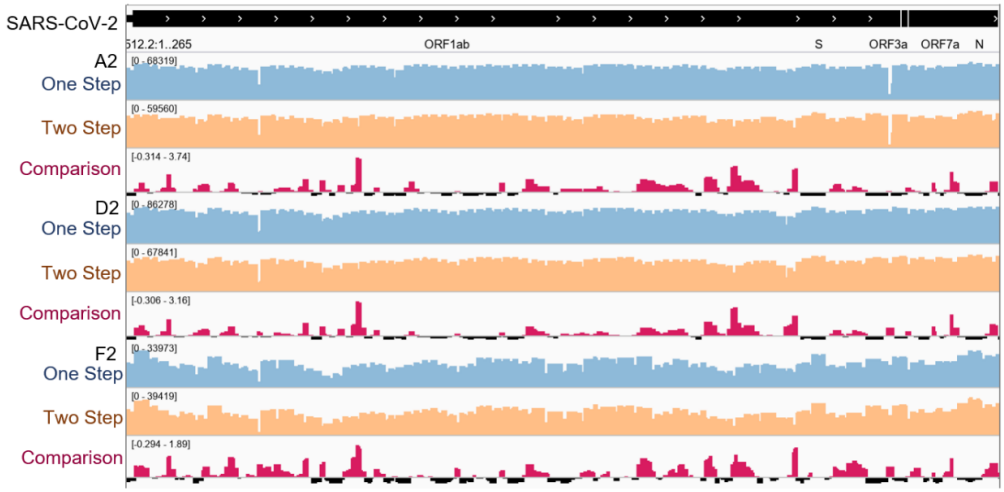


Figure 25. IGV Genome coverage of libraries prepared with Ct 25 samples. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

Figure 26. IGV Genome coverage of libraries prepared with Ct 30 samples. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

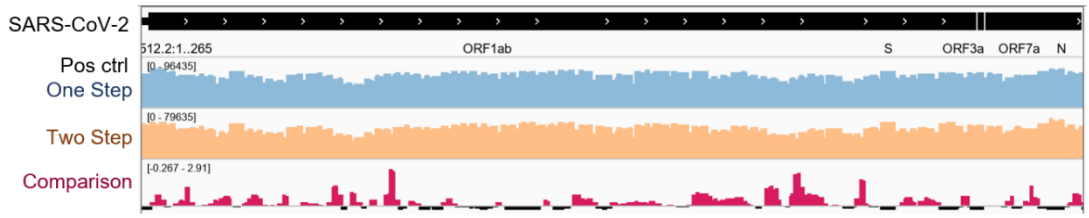


Figure 27. IGV Genome coverage of libraries prepared with as positive controls. Libraries prepared with the one-step ligation protocol are shown in blue while libraries prepared with the standard two-step protocol are shown in orange. Positive (pink) values show higher coverage for one-step prepared libraries while negative (black) values show higher coverage for two-step prepared libraries.

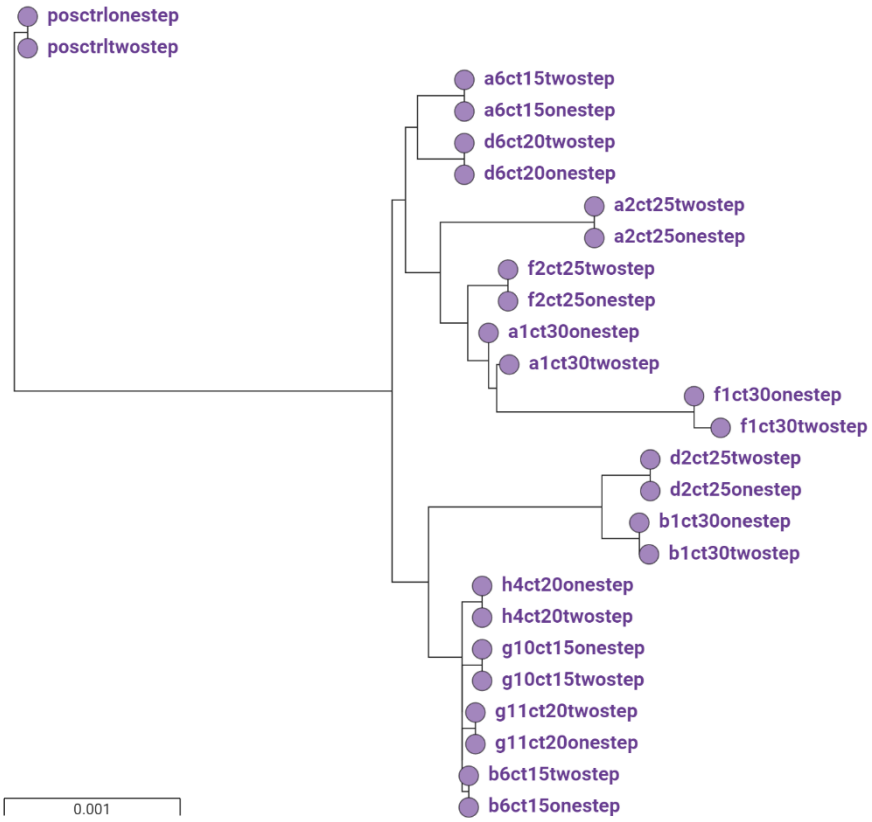


Figure 28. Phylogenetic tree of all samples sequenced. Produced with Pathogenwatch.

3.2 On-chip library preparation evaluation

Libraries prepared both on-chip on the PowerBlade and off-chip on the bench were sequenced on different flow cells without normalization, hence the differences in number of total and mapped reads observed in Figure 29. Nevertheless, even with these differences, the covered bases and coverage graphs show a high and similar coverage for both the on-chip and off-chip prepared libraries. Figure 30 shows a high similitude in the fragment size profile between the libraries prepared on and off the PowerBlade.

Figure 31 displays a big difference in coverage depth, which is explained by the discrepancy in number of reads between on and off-chip libraries, however as seen in Figure 32, the off-chip libraries present a full genome coverage too, showing that the libraries generated on-chip on the PowerBlade are completely functional and resemble those prepared off-chip in coverage and fragment size.

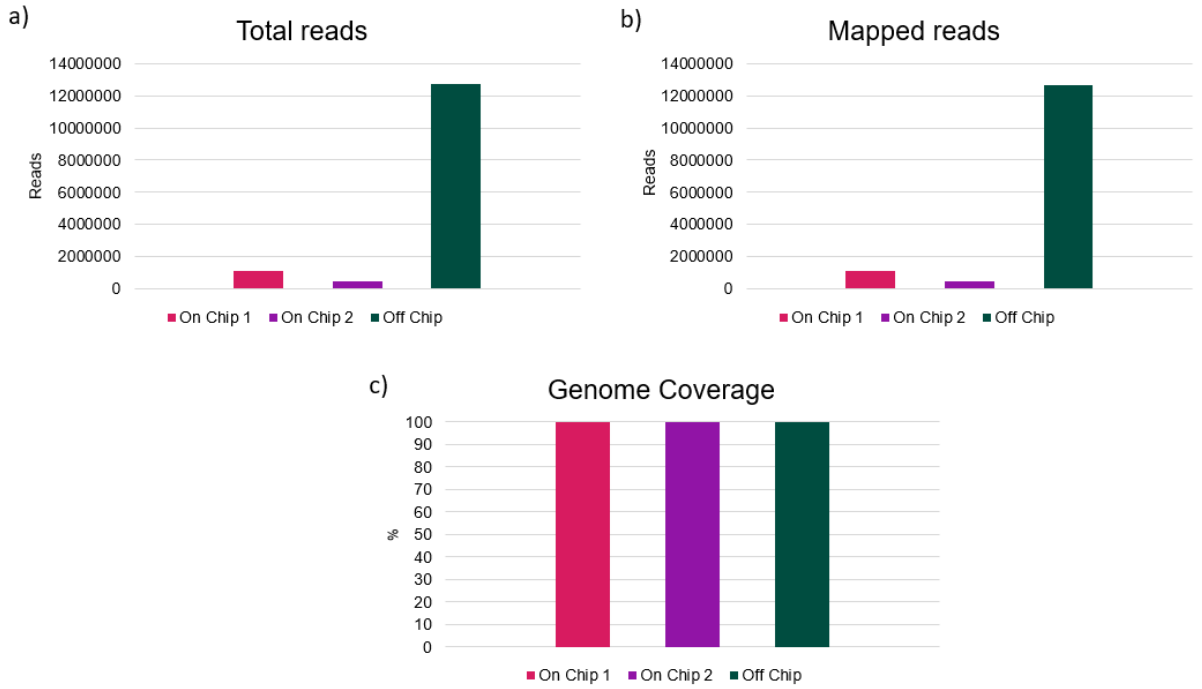


Figure 29. Sequencing metrics of libraries prepared on-chip. Libraries prepared on chip are shown in pink and purple while those produced off-chip are shown in green. a) Total reads. b) Mapped reads. c) Covered bases. d) Coverage.

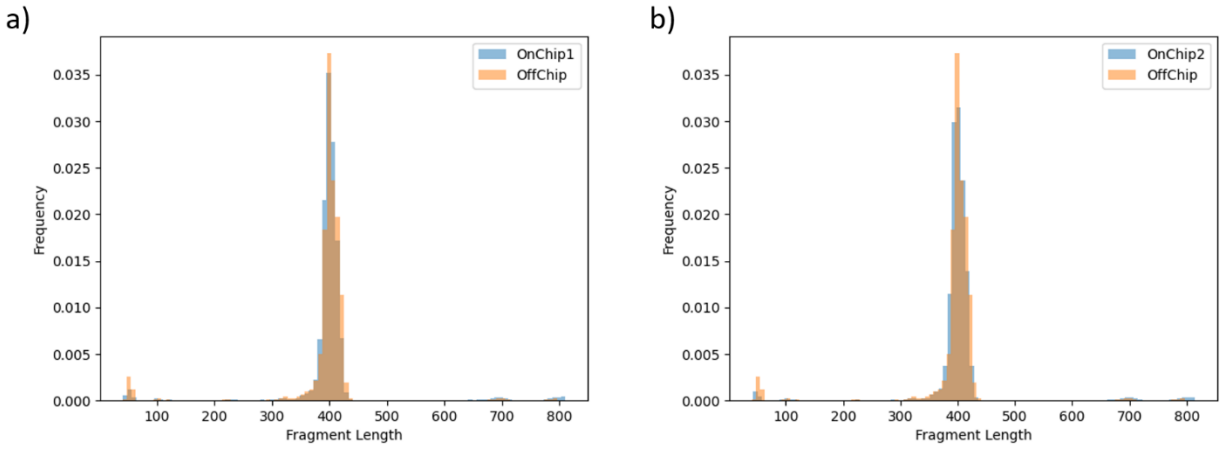


Figure 30. Distribution of sequenced fragment length of libraries prepared on-chip. Libraries prepared on chip are shown in blue while those prepared off-chip are shown in orange. a) On-chip 1 vs off-chip. b) On-chip 2 vs off-chip.

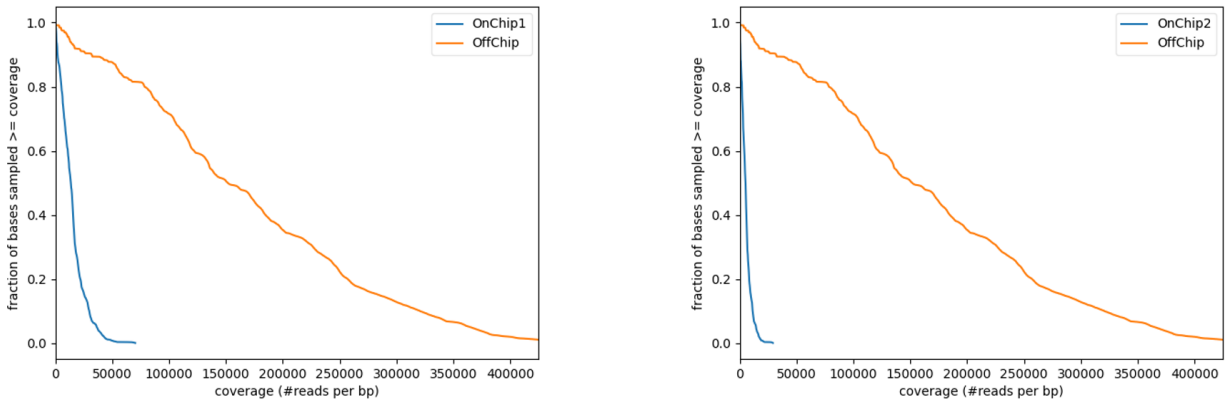


Figure 31. Genome coverage comparison of libraries produced on-chip. Libraries prepared on chip are shown in blue while those prepared off-chip are shown in orange. Coverage as number of reads per base pair is shown on the x-axis, and the y-axis shows the fraction or percentage of bases sampled with a coverage such as that of the x-axis or higher. a) On-chip 1 vs off-chip. b) On-chip 2 vs off-chip.

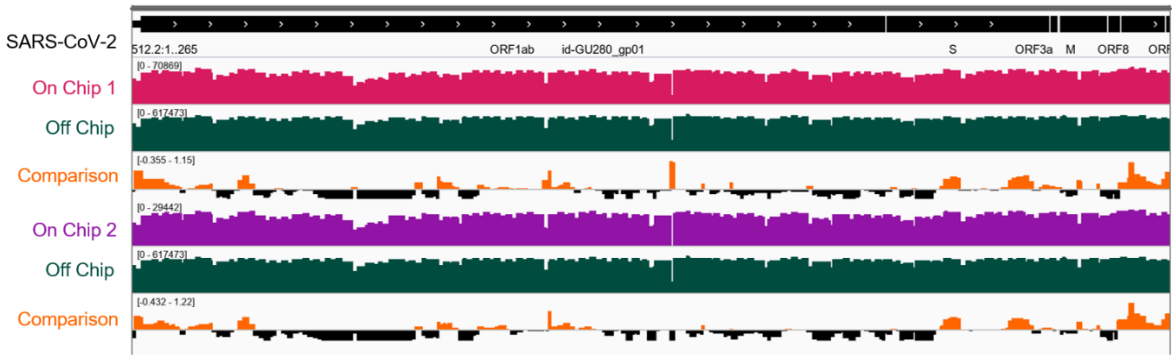


Figure 32. IGV Genome coverage comparison of libraires prepared on chip. Libraries produced on chip are shown in pink and purple while those produced off-chip are shown in green. Positive (orange) values show higher coverage for libraries prepared on-chip while negative (black) values show a higher coverage for libraries prepared off-chip.

3.3 On-chip PCR evaluation

3.3.1 On-chip PCR evaluation - 1st set of experiments.

On-chip PCR was carried out, and libraries prepared with on-chip and off-chip amplicons were sequenced with normalization. Sequencing metrics are presented in Figure 33. One of the libraries produced with amplicons produced on-chip presents a lower amount of total and mapped reads; however, overall, the covered bases and coverage percentage is similar between both the libraries prepared with on-chip and the off-chip amplicons. Nevertheless, Figure 34 shows that the size of the amplicons generated on-chip does not fully correspond to the expected value obtained off-chip. Moreover, a detailed look at the genome coverage in Figures 35 and 36 show that the coverage of the off-chip amplicons is higher than that of the on-chip amplicons.

Nonetheless, pool contribution (Figure 37), a key aspect of a 2-pool PCR, is highly dissimilar between the on-chip and off-chip amplicons. The amplicons produced on-chip on the PowerBlade have an average contribution from pool 2 of about 90%. In contrast, the amplicons produced off-chip had a more equitable distribution, with only 60% of the total amplicons originating from pool 2. This represents a ~30% difference between them. These results derived in a calibration of the temperature inside the PCR chamber of chip 1, to ensure that the PCR is conducted in a proper way and amplicons are

generated properly.

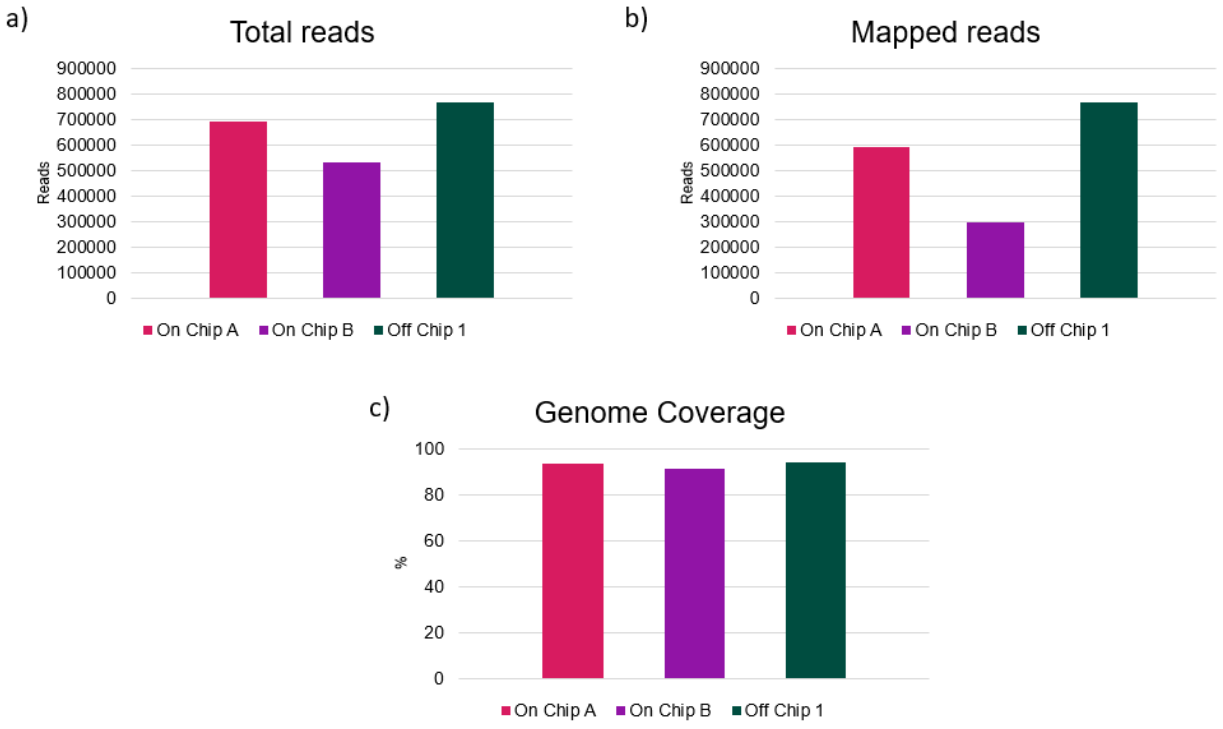


Figure 33. Sequencing metrics of libraries prepared from amplicons produced on-chip. 1st experiment. Libraries prepared with amplicons produced on chip are shown in pink and purple while those prepared with amplicons produced off-chip are shown in green. a) Total reads. b) Mapped reads. c) Covered bases. d) Coverage.

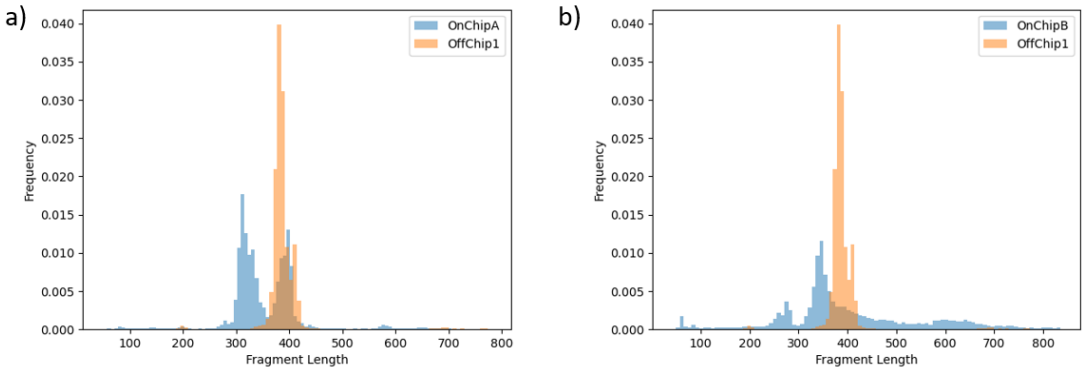


Figure 34. Distribution of sequenced fragment length of libraries prepared from amplicons produced on-chip. 1st experiment. Libraries prepared with amplicons produced on chip are shown in blue while those prepared with amplicons produced off-chip are shown in orange. a) On-chip A vs off-chip 1. b) On-chip B vs off-chip 1.

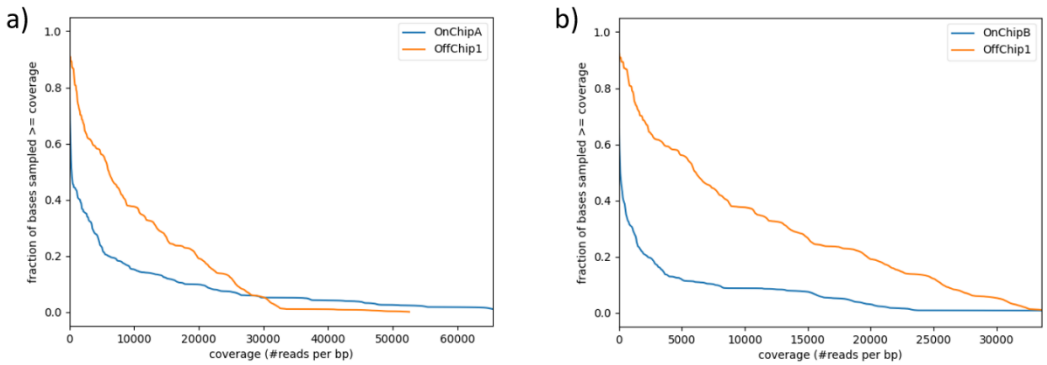


Figure 35. Genome coverage comparison of libraries prepared from amplicons produced on-chip. 1st experiment. Libraries prepared with amplicons produced on chip are shown in blue while those prepared with amplicons produced off-chip are shown in orange. Coverage as number of reads per base pair is shown on the x-axis, and the y-axis shows the fraction percentage of bases sampled with a coverage such as that of the x-axis or higher. a) On-chip A vs off-chip 1. b) On-chip B vs off chip 1

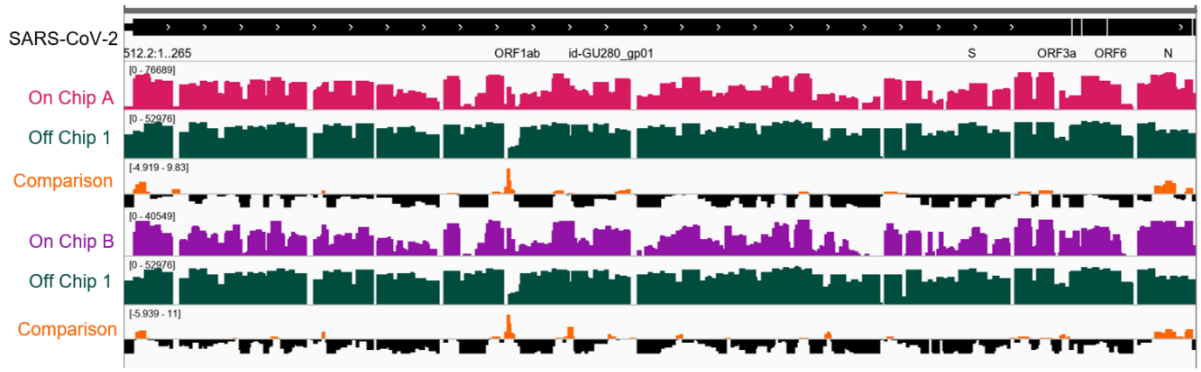


Figure 36. IGV Genome coverage comparison of libraries prepared from amplicons produced on-chip. 1st experiment. Libraries prepared with amplicons produced on chip are shown in pink and purple while those produced with amplicons produced off-chip are shown in green. Positive (orange) values show higher coverage for libraries prepared with on-chip produced amplicons while negative (black) values show a higher coverage for libraries prepared with off-chip produced amplicons.

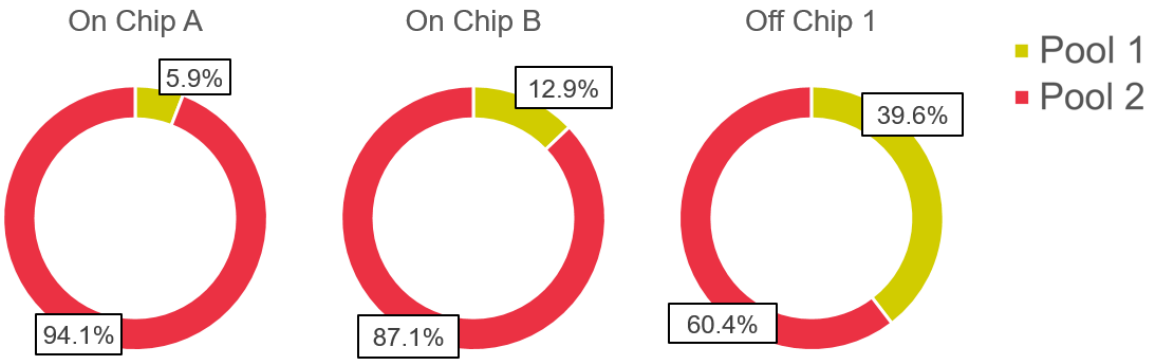


Figure 37. Primer pool contribution of libraries prepared from amplicons produced on-chip. 1st experiment.

3.3.2 On-chip PCR evaluation - 2nd set of experiments.

Calibration of the temperature on the chip 1 PCR chambers was carried out and subsequently, another round of on-chip PCR experiments was performed. The libraries produced this time are called on-chip C & D for those prepared with amplicons produced on the PowerBlade (to avoid mixing with the on-chip A & B libraries produced in the previous round of experiments), and the libraries made with off-chip generated amplicons were named off-chip 2 & 3.

In this second round, after temperature calibration of the PCR chambers, we observe that the covered bases and the coverage are more similar between on-chip and off-chip amplicons than in the original experiment, and the number of reads is balanced. Figure 39 shows that the on-chip and off-chip amplicons have more similar amplicon sizes, peaking all libraries at the expected size. Figures 40 and 41 display a similar coverage profile and depth between on-chip and off-chip amplicons.

In particular, the pool contribution of the on-chip PCR resembles more closely that of the off-chip produced amplicons, reducing the difference between on-chip and off-chip from ~30% in the first round of experiments to less than 10% in this second round.

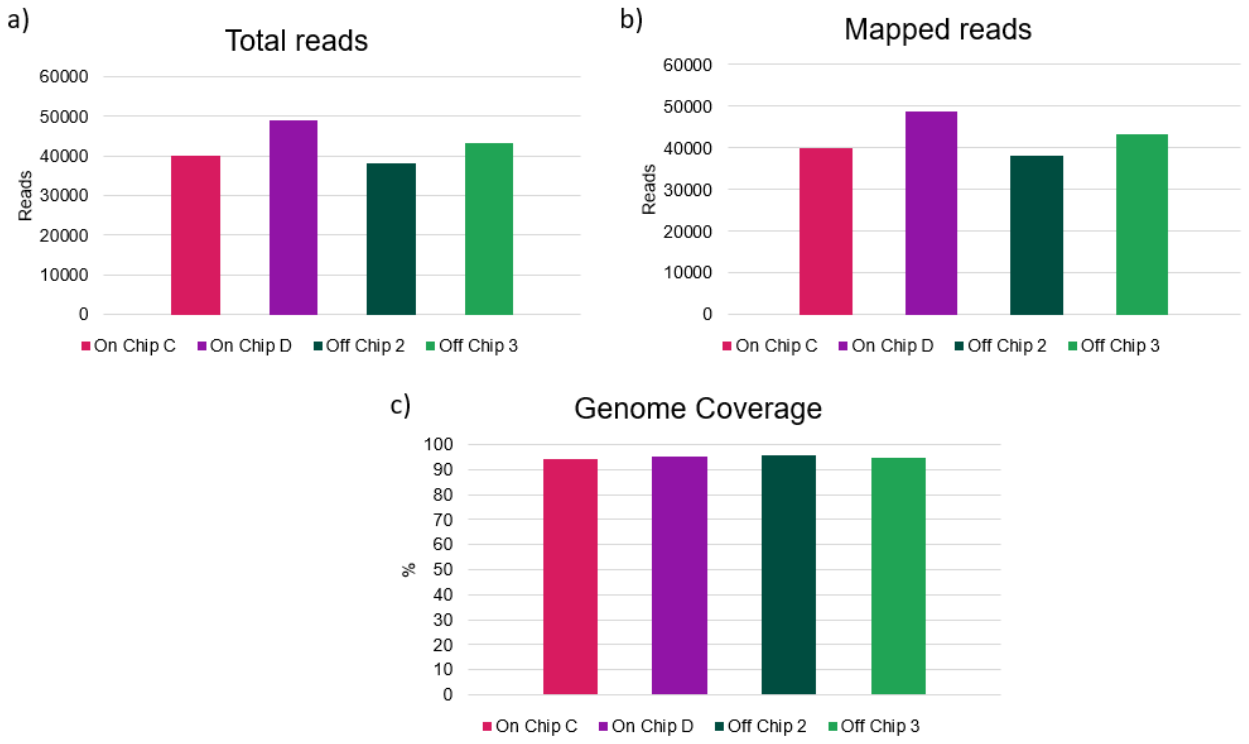


Figure 38. Sequencing metrics of libraries prepared from amplicons produced on-chip. 2nd experiment. Libraries prepared with amplicons produced on chip are shown in pink and purple while those prepared with amplicons produced off-chip are shown in dark and light green. a) Total reads. b) Mapped reads. c) Covered bases. d) Coverage.

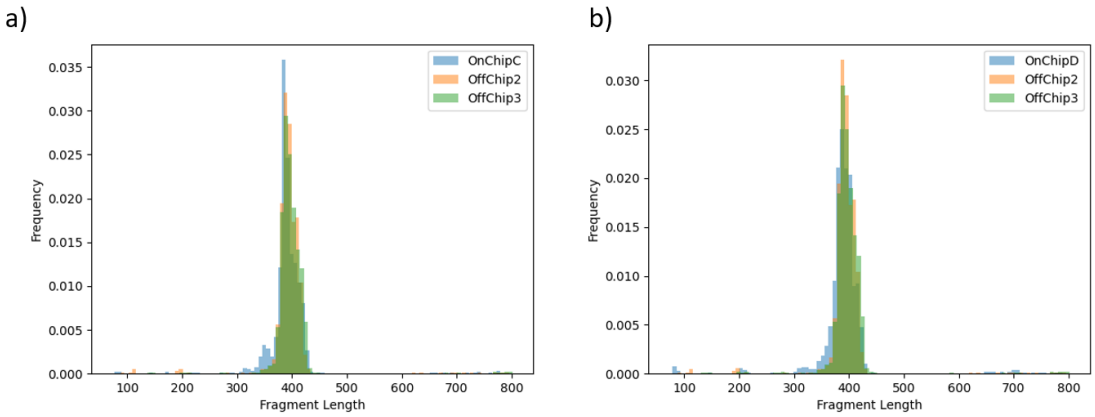


Figure 39. Distribution of sequenced fragment length of libraries prepared from amplicons produced on-chip. 2nd experiment. Libraries prepared with amplicons produced on chip are shown in blue while those prepared with amplicons produced off-chip are shown in orange and green. a) On-chip C vs off-chip 2 and 3. b) On-chip D vs off-chip 2 and 3.

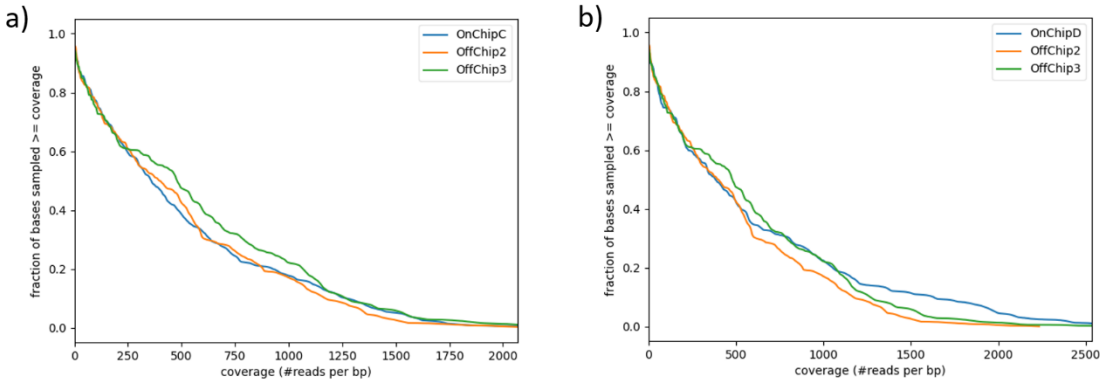


Figure 40. Genome coverage comparison of libraries prepared from amplicons produced on-chip. 2nd experiment. Libraries prepared with amplicons produced on chip are shown in blue while those prepared with amplicons produced off-chip are shown in orange and green. Coverage as number of reads per base pair is shown on the x-axis, and the y-axis shows the fraction or percentage of bases sampled with a coverage such as that of the x-axis or higher. a) On-chip C vs off-chip 2 and 3. b) On-chip D vs off-chip 2 and 3.

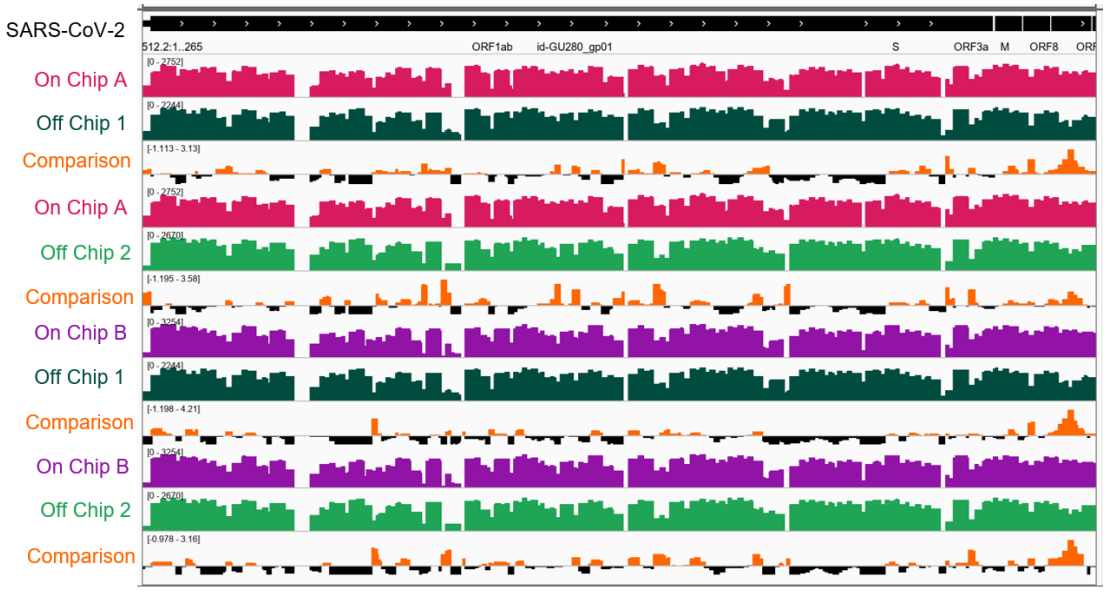


Figure 41. IGV Genome coverage comparison of libraries prepared from amplicons produced on-chip. 2nd experiment. Libraries prepared with amplicons produced on chip are shown in pink and purple while those prepared with amplicons produced off-chip are shown in dark and light green. Positive (orange) values show higher coverage for libraries prepared with on-chip produced amplicons while negative (black) values show a higher coverage for libraries prepared with off-chip produced amplicons.

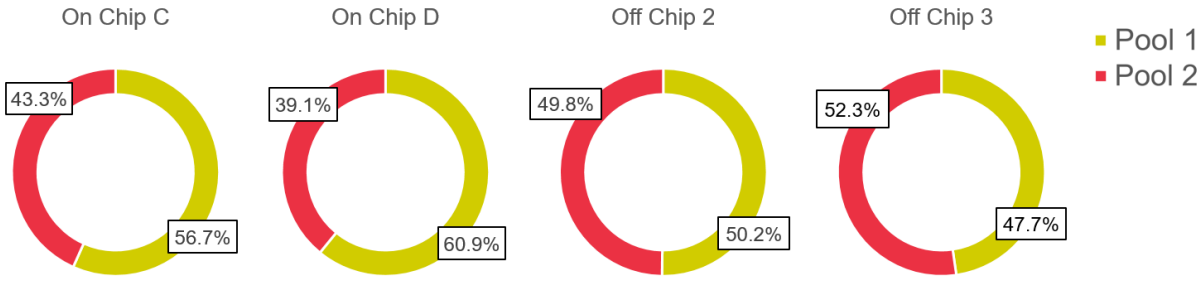


Figure 42. Primer pool contribution of libraries prepared from amplicons produced on-chip. 2nd experiment.

Chapter 4: Discussion

4.1 Novel library preparation protocol evaluation

Our first objective was to assess the correct performance of our novel one-step ligation library preparation protocol. The results displayed show in the first place, that this protocol produces libraries that are able to be sequenced with Nanopore instruments. Furthermore, the libraries produced present similar or higher numbers of mapped reads, and a similar or higher DNA recovery than the standard library preparation protocol. This can be explained given that there is one bead cleanup less in the one-step ligation protocol, thus losing less DNA in the process. A key aspect observed was that the sequenced fragments obtained with both protocols are located in the same size range.

The genome coverage of the SARS-CoV-2 genome obtained with both protocols shows no major differences, apart from coverage depth in some cases, which is explained by variations in the number of fragments sequenced. Overall, the first section of the results chapter shows a good performance of our novel one-step ligation protocol compared to the standard protocol, rendering it available for on-chip testing on the PowerBlade and, of course, for its use on the bench for other applications. The relevance of this protocol for its use both on and off-chip relies on the fact that it requires fewer steps, thus taking less time without affecting the efficiency of the library preparation and the sequencing outcome.

4.2 On-chip library preparation evaluation

After evaluating our novel one-step ligation library preparation protocol on the bench against the standard two-step ligation library preparation protocol, and showing that the protocol works, we tested it on-chip on the PowerBlade. Two libraries were prepared on the PowerBlade, while one was prepared as a control on the bench. The sequencing of these libraries was carried out in separate flow cells without normalization, hence showing a large difference in the number of reads and coverage depth. Still, when we observe the fragment size, percentage of the genome covered, and the visual IGV comparisons of the SARS-CoV-2 genome coverage, we notice almost no differences between the libraries prepared on and off-chip. With these results, we demonstrate that the PowerBlade does generate libraries that can be successfully sequenced and that these libraries generated are of good quality.

4.3 On-chip PCR evaluation

In the second section of results, we tested the functioning of the on-chip PCR on the PowerBlade. Initially, we observed in the first round of experiments that even if the libraries generated did present reads and coverage of the SARS-CoV-2 genome, meaning that ARTIC amplicons were indeed produced, they did not have an optimum quality. The sequenced fragment size of the off-chip amplicons did not peak at 420 bp as that of the library prepared with amplicons produced off-chip, the genome coverage was not as high as the off-chip control, and most significantly, the primer pool contribution was highly uneven.

Primer pool contribution is a critical factor in the evaluation of the PCR, given that an uneven amplification would result in the irregular and incomplete coverage of the target organism's genome, in this case, SARS-CoV-2. Due to the observation of the highly uneven amplification from the two primer pools and the underperforming amplicon size and coverage, we decided that this was not the ideal performance that would be required for the actual implementation of the protocol on the microfluidics device for sample preparation.

Thus, a temperature calibration was performed by inserting a thermocouple inside the PCR chambers shown in Figure 7 and running the PCR program, observing the actual temperature inside the chambers. After calibrating the temperature in the controlling software, a new set of tests was conducted to ensure the temperature within the chambers was appropriate. This time, two on-chip PCR reactions and two off-chip PCR controls were carried out.

The results obtained in the second round of on-chip PCR experiments, after temperature calibration on the PCR chambers, show a clear improvement in the quality of the amplicons generated. The covered bases and coverage presented from the libraries produced with amplicons generated on-chip were highly similar to that of the off-chip amplicons. Moreover, the sequenced fragment length shows a high similarity between the on and off-chip amplicons, both peaking at the expected size of 420 bp. Given that the number of fragments sequenced from all libraries prepared was similar, the genome coverage comparison presented in Figure 34 shows a tight similarity between all libraries sequenced. When observing the pool contribution, we noticed that the pool contribution was much more balanced this time, with an average contribution of

~59% from pool 2; in the first round of experiments, this was ~91%. The difference in pool 2 contribution between on and off-chip amplicons went from ~30% in the first set of experiments to less than 10% in this second set.

Given the results obtained, showing a clear enhancement in the quality of the amplicons generated on chip 1 on the PowerBlade, we can note that the performance of the on-chip PCR resembles a high similarity to that of the off-chip standard PCR on a conventional thermocycler, thus making feasible its use for our workflow. Additionally, this process can be modified to perform any other PCR for any other organism, given that the temperature in the chambers is adequate.

4.4 Overall work

The results presented in this thesis show that the evaluated sections of the sample-to-library workflow perform adequately, with a similar performance on the PowerBlade to that on the bench. The one-step ligation library preparation workflow performed as well as the standard ONT protocol, producing libraries viable for sequencing and with similar sequencing metrics between the two protocols. Then, evaluating the performance of the protocol on-chip on the PowerBlade demonstrated to generate viable libraries with sequencing metrics comparable to those of libraries prepared off-chip. Finally, the on-chip PCR on the PowerBlade was evaluated against the standard PCR on a thermocycler, showing a good amplification corresponding to that on the bench.

Testing the entire workflow is the next step in the process to validate its efficacy for future applications outside the laboratory. Thorough testing with clinical samples of different Ct values will be needed to determine its range of application in a real-life setting. This workflow will enable sequencing to be performed on suspected or confirmed COVID-19 patients in remote geographical locations or under limited economic situations where sequencing was not previously feasible, given that this is an automated instrument that does not require a complex installation or manipulation, available to be performed by individuals that are not necessarily molecular biology technicians. Additionally, the one-step ligation protocol presented in this thesis is meant to be used in a standard laboratory setting as well, as it reduces library preparation time without compromising the library output or sequencing result.

The PowerBlade, developed by the NRC, is suitable for the automation of a wide range of protocols that are typically performed in a series of different instruments in a laboratory; this, along with its portability, makes it an excellent tool for the implementation of the workflow. Nevertheless, this workflow can be modified for optimization regarding reagent volumes or the protocol themselves, aiming for better yields or higher quality of the sequencing libraries generated. This workflow could also be transferred to other protocol automation platforms using technologies previously described, such as the electrowetting-based droplet microfluidics instruments, if the application requires so. Furthermore, this workflow is not meant to be solely used during the SARS-CoV-2 pandemic, as the RNA extraction and amplification steps could be modified for the identification of other pathogens and even for treatment of other kinds of samples, aiding in future pandemics or routine testing of samples for other means.

Chapter 5: Conclusions and Future Directions

In this thesis, we evaluate three critical aspects of the operation of a sample to sequencing library preparation workflow on a set of cartridges on the PowerBlade, a microfluidics protocol automation device developed by the National Research Council. These three elements evaluated are a novel library preparation protocol developed in our lab, on-chip PCR performance, and on-chip library preparation protocol efficiency.

The development of the novel one-step library preparation protocol was crucial, given the on-chip limitations that we encountered. We noticed that a more straightforward and faster protocol needed to be created to carry out nanopore library preparation on the PowerBlade. Our protocol was evaluated against the standard library preparation protocol, which includes one ligation step for the addition of the barcode onto the DNA fragment to be sequenced and another ligation step for the attachment of the sequencing adapter. In contrast, our protocol combines the ligation of both the barcode and sequencing adapter in a single step.

We demonstrated that our novel library preparation protocol works as efficiently as the standard protocol, generating proper nanopore libraries with sequencing metrics comparable to those prepared with the standard protocol. This is important not only as we have developed a faster protocol that requires fewer steps to generate a nanopore library, but also because we could then test it on the PowerBlade.

A correct amplification of the whole SARS-CoV-2 genome is a fundamental aspect of the entire workflow. On-chip PCR evaluation showed that the amplicons generated on-chip on the PowerBlade are of the same quality as those generated off-chip.

Furthermore, the on-chip PCR can be adapted to amplify other genomes by changing the temperature program and the primers.

The last step, testing the on-chip library preparation with our novel protocol, showed that the PowerBlade can generate libraries satisfactorily, comparable to those generated on the bench. Overall, we have demonstrated that the novel one-step library preparation protocol works both on the bench and on-chip, and that the SARS-CoV-2 ARTIC PCR produces adequate amplicons on-chip. These aspects are crucial for the successful execution of the entire workflow.

As a future direction, testing the whole workflow with a clinical sample should be done on the PowerBlade and compared to an off-PowerBlade control workflow prepared on the bench. Moreover, due to the significance of the recent pandemic and the availability of samples, SARS-CoV-2 was used for this workflow; nevertheless, this workflow is intended to be modified to identify any other organism. Implementing this automated sample-to-sequencing library workflow will enable sample sequencing in remote locations and different settings where it is not currently available but is required, allowing them to keep an accurate track of pathogens and take precise and valuable measures on time.

Chapter 6: References

Adams, D. R., & Eng, C. M. (2018). Next-Generation Sequencing to Diagnose Suspected Genetic Disorders. *N Engl J Med*, *379*(14), 1353-1362. https://doi.org/10.1056/NEJMra1711801

Adewale, B. A. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med*, *9*(1), 1340. https://doi.org/10.4102/ajlm.v9i1.1340

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*, *21*(1), 30. https://doi.org/10.1186/s13059-020-1935-5

Argimon, S., David, S., Underwood, A., Abrudan, M., Wheeler, N. E., Kekre, M., Abudahab, K., Yeats, C. A., Goater, R., Taylor, B., Harste, H., Muddyman, D., Feil, E. J., Brisse, S., Holt, K., Donado-Godoy, P., Ravikumar, K. L., Okeke, I. N., Carlos, C., . . . Resistance, N. G. H. R. U. o. G. S. o. A. (2021). Rapid Genomic Characterization and Global Surveillance of Klebsiella Using Pathogenwatch. *Clin Infect Dis*, *73*(Suppl_4), S325-S335. https://doi.org/10.1093/cid/ciab784

Arteche-Lopez, A., Avila-Fernandez, A., Romero, R., Riveiro-Alvarez, R., Lopez-Martinez, M. A., Gimenez-Pardo, A., Velez-Monsalve, C., Gallego-Merlo, J., Garcia-Vara, I., Almoguera, B., Bustamante-Aragones, A., Blanco-Kelly, F., Tahsin-Swafiri, S., Rodriguez-Pinilla, E., Minguez, P., Lorda, I., Trujillo-Tiebas, M. J., & Ayuso, C. (2021). Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Sci Rep*, *11*(1), 5697. https://doi.org/10.1038/s41598-021-85182-w

ARTIC Network. *SARS-CoV-2*. https://artic.network/ncov-2019

ARTIC Network. (n.d.). *SARS-CoV-2*. https://artic.network/ncov-2019

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med*, *79*(2), 137-158. https://doi.org/10.1084/jem.79.2.137

Bao, Y., Wadden, J., Erb-Downward, J. R., Ranjan, P., Zhou, W., McDonald, T. L., Mills, R. E., Boyle, A. P., Dickson, R. P., Blaauw, D., & Welch, J. D. (2021). SquiggleNet: real-time, direct classification of nanopore signals. *Genome Biol*, *22*(1), 298. https://doi.org/10.1186/s13059-021-02511-y

Behrmann, O., & Spiegel, M. (2020). COVID-19: from rapid genome sequencing to fast decisions. *The Lancet Infectious Diseases*, *20*(11), 1218. https://doi.org/10.1016/s1473-3099(20)30580-6

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., . . . Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53-59. https://doi.org/10.1038/nature07517

Bleidorn, C. (2015). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, *14*(1), 1-8. https://doi.org/10.1080/14772000.2015.1099575

Boykin, L. M., Ammar, G., Bruno Rossitto De, M., Anders, S., James, M. W., Tonny, K., Stephen, L., Myriam, R., Monica, A. K., Joseph, N., Fred, T., Peter, S., Charles, K., Deogratius, M., Joel, E., Hilda, B., Titus, A., Geoffrey, O.-O., Phillip, A., . . . Samuel, K. (2018). Real time portable genome sequencing for global food security. *bioRxiv*, 314526. https://doi.org/10.1101/314526

Brejova, B., Borsova, K., Hodorova, V., Cabanova, V., Gafurov, A., Fricova, D., Nebohacova, M., Vinar, T., Klempa, B., & Nosek, J. (2021). Nanopore sequencing of SARS-CoV-2: Comparison of short and long PCR-tiling amplicon protocols. *PLoS One*, *16*(10), e0259277. https://doi.org/10.1371/journal.pone.0259277

Brito, A. F., Semenova, E., Dudas, G., Hassler, G. W., Kalinich, C. C., Kraemer, M. U. G., Ho, J., Tegally, H., Githinji, G., Agoti, C. N., Matkin, L. E., Whittaker, C., Bulgarian, S.-C.-s. g., Communicable Diseases Genomics, N., Project, C.-I., Danish Covid-19 Genome, C., Fiocruz, C.-G. S. N., team, G. c. c., Network for Genomic Surveillance in South, A., . . . Faria, N. R. (2022). Global disparities in SARS-CoV-2 genomic surveillance. *Nat Commun*, *13*(1), 7003. https://doi.org/10.1038/s41467-022-33713-y

Bruno, A., Aury, J. M., & Engelen, S. (2021). BoardION: real-time monitoring of Oxford Nanopore sequencing instruments. *BMC Bioinformatics*, *22*(1), 245. https://doi.org/10.1186/s12859-021-04161-0

Buermans, H. P., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta*, *1842*(10), 1932-1941. https://doi.org/10.1016/j.bbadis.2014.06.015

Bull, R. A., Adikari, T. N., Ferguson, J. M., Hammond, J. M., Stevanovski, I., Beukers, A. G., Naing, Z., Yeang, M., Verich, A., Gamaarachchi, H., Kim, K. W., Luciani, F., Stelzer-Braid, S., Eden, J. S., Rawlinson, W. D., van Hal, S. J., & Deveson, I. W. (2020). Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat Commun*, *11*(1), 6272. https://doi.org/10.1038/s41467-020-20075-6

Calabrese, C., Gomez-Duran, A., Reyes, A., & Attimonelli, M. (2020). Methods for the identification of mitochondrial DNA variants. In *The Human Mitochondrial Genome* (pp. 243-275). https://doi.org/10.1016/b978-0-12-819656-4.00011-5

Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., Dworkin, J. P., Lupisella, M. L., Smith, D. J., Botkin, D. J., Stephenson, T. A., Juul, S., Turner, D. J., Izquierdo, F., Federman, S., Stryke, D., Somasekar, S., Alexander, N., Yu, G., . . . Burton, A. S. (2017). Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Scientific Reports*, *7*(1), 18022. https://doi.org/10.1038/s41598-017-18364-0

Chen, S.-H., J Reiling, S., Quick, J., & Ragoussis, I. (2020). nCoV-2019 McGill RT Protocol, Lunascript. *protocols.io*. https://doi.org/10.17504/protocols.io.bjgekjte

Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature Reviews Genetics*, *20*(6), 341-355. https://doi.org/10.1038/s41576-019-0113-7

Chu, D. K. W., Pan, Y., Cheng, S. M. S., Hui, K. P. Y., Krishnan, P., Liu, Y., Ng, D. Y. M., Wan, C. K. C., Yang, P., Wang, Q., Peiris, M., & Poon, L. L. M. (2020). Molecular

Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clin Chem*, *66*(4), 549-555. https://doi.org/10.1093/clinchem/hvaa029

Coronaviridae Study Group of the International Committee on Taxonomy of, V. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*, *5*(4), 536-544. https://doi.org/10.1038/s41564-020-0695-z

Cotten, M., Lule Bugembe, D., Kaleebu, P., & V.T. Phan, M. (2021). Alternate primers for whole-genome SARS-CoV-2 sequencing. *Virus Evolution*, *7*(1). https://doi.org/10.1093/ve/veab006

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental Biology*, *278*(2), 274-288. https://doi.org/https://doi.org/10.1016/j.ydbio.2004.11.028

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

Davis James, J., Long, S. W., Christensen Paul, A., Olsen Randall, J., Olson, R., Shukla, M., Subedi, S., Stevens, R., Musser James, M., & Pride David, T. Analysis of the ARTIC Version 3 and Version 4 SARS-CoV-2 Primers and Their Impact on the Detection of the G142D Amino Acid Substitution in the Spike Protein. *Microbiology Spectrum*, *9*(3), e01803-01821. https://doi.org/10.1128/Spectrum.01803-21

de Vries, E. M., Cogan, N. O. I., Gubala, A. J., Mee, P. T., O'Riley, K. J., Rodoni, B. C., & Lynch, S. E. (2022). Rapid, in-field deployable, avian influenza virus haemagglutinin characterisation tool using MinION technology. *Scientific Reports*, *12*(1), 11886. https://doi.org/10.1038/s41598-022-16048-y

Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nat Biotechnol*, *34*(5), 518-524. https://doi.org/10.1038/nbt.3423

Dinnes, J., Deeks, J. J., Berhane, S., Taylor, M., Adriano, A., Davenport, C., Dittrich, S., Emperador, D., Takwoingi, Y., Cunningham, J., & et al. (2021). Rapid, point‐of‐care antigen tests for diagnosis of SARS‐CoV‐2 infection. *Cochrane Database of Systematic Reviews*(3). https://doi.org/10.1002/14651858.CD013705.pub2

Dorado, G., Gálvez, S., Budak, H., Unver, T., & Hernández, P. (2019). Nucleic-Acid Sequencing. In *Encyclopedia of Biomedical Engineering* (pp. 443-460). https://doi.org/10.1016/b978-0-12-801238-3.08998-4

Dovichi, N. J., & Zhang, J. (2000). How Capillary Electrophoresis Sequenced the Human Genome This Essay is based on a lecture given at the Analytica 2000 conference in Munich (Germany) on the occasion of the Heinrich-Emanuel-Merck Prize presentation. *Angew Chem Int Ed Engl*, *39*(24), 4463-4468. https://doi.org/10.1002/1521-3773(20001215)39:24<4463::aid-anie4463>3.0.co;2-8

Duan, H., Li, X., Mei, A., Li, P., Liu, Y., Li, X., Li, W., Wang, C., & Xie, S. (2021). The diagnostic value of metagenomic next-generation sequencing in infectious diseases. *BMC Infect Dis*, *21*(1), 62. https://doi.org/10.1186/s12879-020-05746-5

Elkin, L. O. (2003). Rosalind Franklin and the Double Helix. *Physics Today*, *56*(3), 42-48. https://doi.org/10.1063/1.1570771

Faria, N. R., Sabino, E. C., Nunes, M. R., Alcantara, L. C., Loman, N. J., & Pybus, O. G. (2016). Mobile real-time surveillance of Zika virus in Brazil. *Genome Med*, *8*(1), 97. https://doi.org/10.1186/s13073-016-0356-2

Farias-Hesson, E., Erikson, J., Atkins, A., Shen, P., Davis, R. W., Scharfe, C., & Pourmand, N. (2010). Semi-automated library preparation for high-throughput DNA sequencing platforms. *J Biomed Biotechnol*, *2010*, 617469. https://doi.org/10.1155/2010/617469

Freed, N. E., Vlkova, M., Faisal, M. B., & Silander, O. K. (2020). Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol Methods Protoc*, *5*(1), bpaa014. https://doi.org/10.1093/biomethods/bpaa014

Gaudin, M., & Desnues, C. (2018). Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. *Front Microbiol*, *9*, 2924. https://doi.org/10.3389/fmicb.2018.02924

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9-19. https://doi.org/10.1016/j.csbj.2019.11.002

Giovanetti, M., de Mendonça, M. C. L., Fonseca, V., Mares-Guia, M. A., Fabri, A., Xavier, J., de Jesus, J. G., Gräf, T., dos Santos Rodrigues, C. D., & Dos Santos, C. C. (2019). Yellow fever virus reemergence and spread in Southeast Brazil, 2016–2019. *Journal of Virology*, *94*(1), e01623-01619.

Goldsmith, C. S., Tatti, K. M., Ksiazek, T. G., Rollin, P. E., Comer, J. A., Lee, W. W., Rota, P. A., Bankamp, B., Bellini, W. J., & Zaki, S. R. (2004). Ultrastructural characterization of SARS coronavirus. *Emerg Infect Dis*, *10*(2), 320-326. https://doi.org/10.3201/eid1002.030913

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., . . . Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), 459-468. https://doi.org/10.1038/s41586-020-2286-9

Gu, W., Deng, X., Lee, M., Sucu, Y. D., Arevalo, S., Stryke, D., Federman, S., Gopez, A., Reyes, K., Zorn, K., Sample, H., Yu, G., Ishpuniani, G., Briggs, B., Chow, E. D., Berger, A., Wilson, M. R., Wang, C., Hsu, E., . . . Chiu, C. Y. (2021). Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nature Medicine*, *27*(1), 115-124. https://doi.org/10.1038/s41591-020-1105-z

Gupta, N., & Verma, V. K. (2019). Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. In *Microbial Technology for the Welfare of Society* (pp. 313-341). https://doi.org/10.1007/978-981-13-8844-6_15

Gwinn, M., MacCannell, D., & Armstrong, G. L. (2019). Next-Generation Sequencing of Infectious Pathogens. *JAMA*, *321*(9), 893-894. https://doi.org/10.1001/jama.2018.21669

Harrison, A. G., Lin, T., & Wang, P. (2020). Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. *Trends Immunol*, *41*(12), 1100-1115. https://doi.org/10.1016/j.it.2020.10.004

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1-8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hess, J. F., Hess, M. E., Zengerle, R., Paust, N., Boerries, M., & Hutzenlaub, T. (2021). Automated library preparation for whole genome sequencing by centrifugal microfluidics. *Anal Chim Acta*, *1182*, 338954. https://doi.org/10.1016/j.aca.2021.338954

Hess, J. F., Kohl, T. A., Kotrova, M., Ronsch, K., Paprotka, T., Mohr, V., Hutzenlaub, T., Bruggemann, M., Zengerle, R., Niemann, S., & Paust, N. (2020). Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv*, *41*, 107537. https://doi.org/10.1016/j.biotechadv.2020.107537

Hoenen, T., Groseth, A., Rosenke, K., Fischer, R. J., Hoenen, A., Judson, S. D., Martellaro, C., Falzarano, D., Marzi, A., Squires, R. B., Wollenberg, K. R., de Wit, E., Prescott, J., Safronetz, D., van Doremalen, N., Bushmaker, T., Feldmann, F., McNally, K., Bolay, F. K., . . . Feldmann, H. (2016). Nanopore Sequencing as a Rapidly Deployable Ebola Outbreak Tool. *Emerg Infect Dis*, *22*(2), 331-334. https://doi.org/10.3201/eid2202.151796

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Hum Immunol*, *82*(11), 801-811. https://doi.org/10.1016/j.humimm.2021.02.012

Illumina (2023a). *AmpliSeq for Illumina SARS-CoV-2 Research Panel*. https://www.illumina.com/products/by-brand/ampliseq/community-panels/sars-cov-2.html

Illumina (2023b). *Automating lab protocols*. https://www.illumina.com/science/technology/digital-microfluidics.html

Illumina (2023c). *Automation Methods that Span Our Library Prep Portfolio*. https://www.illumina.com/techniques/sequencing/ngs-library-prep/automation.html

Illumina (n.d.). *NeoPrep*. https://www.illumina.com/content/dam/illumina-marketing/images/neoprep/index.html

Integrated DNA Technologies (2023). *xGen™ SARS-CoV-2 Amplicon Panel*. https://www.idtdna.com/pages/products/next-generation-sequencing/workflow/xgen-ngs-amplicon-sequencing/predesigned-amplicon-panels/sars-cov-2-amp-panel#product-details

J Reiling, S., Loranger, K., Roy, A. M., Chen, S.-H., & Ragoussis, I. (2022). nCoV-2019 McGill Artic PCR Protocol, V4.1 at 63C V.2. *protocols.io*. https://doi.org/10.17504/protocols.io.ewov18e4ygr2/v2

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, *17*(1), 239. https://doi.org/10.1186/s13059-016-1103-0

Jia, L., Jiang, M., Wu, K., Hu, J., Wang, Y., Quan, W., Hao, M., Liu, H., Wei, H., Fan, W., Liu, W., Hu, R., Wang, D., Li, J., Chen, J., & Liu, D. (2020). Nanopore sequencing

of African swine fever virus. *Science China Life Sciences*, *63*(1), 160-164. https://doi.org/10.1007/s11427-019-9828-1

Kafetzopoulou, L. E., Pullan, S. T., Lemey, P., Suchard, M. A., Ehichioya, D. U., Pahlmann, M., Thielebein, A., Hinzmann, J., Oestereich, L., Wozniak, D. M., Efthymiadis, K., Schachten, D., Koenig, F., Matjeschk, J., Lorenzen, S., Lumley, S., Ighodalo, Y., Adomeh, D. I., Olokor, T., . . . Duraffour, S. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science*, *363*(6422), 74-77. https://doi.org/10.1126/science.aau9343

Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A*, *93*(24), 13770-13773. https://doi.org/10.1073/pnas.93.24.13770

Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, *09*(03). https://doi.org/10.4172/0974-8369.1000395

Kevadiya, B. D., Machhi, J., Herskovitz, J., Oleynikov, M. D., Blomberg, W. R., Bajwa, N., Soni, D., Das, S., Hasan, M., Patel, M., Senan, A. M., Gorantla, S., McMillan, J., Edagwa, B., Eisenberg, R., Gurumurthy, C. B., Reid, S. P. M., Punyadeera, C., Chang, L., & Gendelman, H. E. (2021). Diagnostics for SARS-CoV-2 infections. *Nat Mater*, *20*(5), 593-605. https://doi.org/10.1038/s41563-020-00906-z

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. *Bioessays*, *32*(6), 524-536. https://doi.org/10.1002/bies.200900181

Kruttgen, A., Cornelissen, C. G., Dreher, M., Hornef, M. W., Imohl, M., & Kleines, M. (2021). Comparison of the SARS-CoV-2 Rapid antigen test to the real star Sars-CoV-2 RT PCR kit. *J Virol Methods*, *288*, 114024. https://doi.org/10.1016/j.jviromet.2020.114024

Lamb, H. J., Hayes, B. J., Nguyen, L. T., & Ross, E. M. (2020). The Future of Livestock Management: A Review of Real-Time Portable Sequencing Applied to Livestock. *Genes (Basel)*, *11*(12). https://doi.org/10.3390/genes11121478

Lander, E. S., Consortium, I. H. G. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., . . . Conso, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860-921. https://doi.org/10.1038/35057062

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., Richard McCombie, W., & Schatz, M. C. (2016). Third-generation sequencing and the future of genomics. *bioRxiv*, 048603. https://doi.org/10.1101/048603

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100. https://doi.org/10.1093/bioinformatics/bty191

Li, J., Wang, H., Mao, L., Yu, H., Yu, X., Sun, Z., Qian, X., Cheng, S., Chen, S., Chen, J., Pan, J., Shi, J., & Wang, X. (2020). Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. *Sci Rep*, *10*(1), 17492. https://doi.org/10.1038/s41598-020-74656-y

Mak, G. C. K., Cheng, P. K. C., Lau, S. S. Y., Wong, K. K. Y., Lau, C. S., Lam, E. T. K., Chan, R. C. W., & Tsang, D. N. C. (2020). Evaluation of rapid antigen test for detection of SARS-CoV-2 virus. *Journal of Clinical Virology*, *129*, 104500. https://doi.org/https://doi.org/10.1016/j.jcv.2020.104500

Makalowski, W., & Shabardina, V. (2020). Bioinformatics of nanopore sequencing. *J Hum Genet*, *65*(1), 61-67. https://doi.org/10.1038/s10038-019-0659-4

Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., Pavlenok, M., Niederweis, M., & Gundlach, J. H. (2012). Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol*, *30*(4), 349-353. https://doi.org/10.1038/nbt.2171

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, *9*, 387-402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., . . . Roper, R. L. (2003). The Genome sequence of the SARS-associated coronavirus. *Science*, *300*(5624), 1399-1404. https://doi.org/10.1126/science.1085953

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proc Natl Acad Sci U S A*, *74*(2), 560-564. https://doi.org/10.1073/pnas.74.2.560

McCarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. *Chem Biol*, *17*(7), 675-676. https://doi.org/10.1016/j.chembiol.2010.07.004

Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA sequencing. *Next Generation Genome Sequencing: Towards Personalized Medicine*, 1-11.

Menegon, M., Cantaloni, C., Rodriguez-Prieto, A., Centomo, C., Abdelfattah, A., Rossato, M., Bernardi, M., Xumerle, L., Loader, S., & Delledonne, M. (2017). On site DNA barcoding by nanopore sequencing. *PLoS One*, *12*(10), e0184741. https://doi.org/10.1371/journal.pone.0184741

Miescher, F. (1869). Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher–Aus dem Wissenschaftlichen Briefwechsel von F. Miescher. *FCW Vogel (Leibzig)*, *1*, 33-38.

Miroculus (2023). *Technology*. https://miroculus.com/technology/

Mugele, F., & Baret, J.-C. (2005). Electrowetting: from basics to applications. *Journal of Physics: Condensed Matter*, *17*(28), R705-R774. https://doi.org/10.1088/0953-8984/17/28/r01

National Library of Medicine (US) National Center for Biotechnology Information. (2020). *Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome [Internet]. Bethesda (MD)*. https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2?report=fasta&to=29903

National Research Council. (2022). *NRC PowerBlade*. https://nrc.canada.ca/en/research-development/research-collaboration/nrc-powerblade

O'Toole, A., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S. W., Aanensen, D. M., Holmes, E. C., Pybus, O. G., & Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol*, *7*(2), veab064. https://doi.org/10.1093/ve/veab064

Ontario Agency for Health Protection and Promotion (Public Health Ontario). (2020). *Focus on: an overview of cycle threshold values and their role in SARS-Cov-2 real-time PCR test interpretation*. Toronto, ON: Queen's Printer for Ontario

Oxford Nanopore Technologies (2023a). *Flow cells* https://nanoporetech.com/how-it-works/flow-cells-and-nanopores

Oxford Nanopore Technologies (2023b). *VolTRAX*. Retrieved June 7 from https://nanoporetech.com/products/voltrax

Oxford Nanopore Technologies (n.d.-a). *Accuracy*. Retrieved February 5, 2023 from https://nanoporetech.com/accuracy

Oxford Nanopore Technologies (n.d.-b). *Company history* Retrieved February 5, 2023 from https://nanoporetech.com/about-us/history

Oxford Nanopore Technologies (n.d.-c). *COVID-19*. Retrieved February 6, 2023 from https://nanoporetech.com/covid-19

Oxford Nanopore Technologies (n.d.-d). *Devices*. Retrieved February 5, 2023 from https://store.nanoporetech.com/devices.html

Oxford Nanopore Technologies (n.d.-e). *Flongle Starter pack*. Retrieved February 5, 2023 from https://store.nanoporetech.com/flongle-intro-pack.html

Oxford Nanopore Technologies (n.d.-f). *LIGATION SEQUENCING GDNA - NATIVE BARCODING KIT 24 V14 - Overview of the Protocol*. https://community.nanoporetech.com/docs/prepare/library_prep_protocols/ligation-sequencing-gdna-native-barcoding-v14-sqk-nbd114-24/v/nbe_9169_v114_reve_15sep2022/overview-of-the-protocol?devices=promethion

Oxford Nanopore Technologies (n.d.-g). *MinION Mk1C*. Retrieved February 5, 2023 from https://store.nanoporetech.com/minion-mk1c.html

Oxford Nanopore Technologies (n.d.-h). *Product specifications*. Retrieved February 5, 2023 from https://nanoporetech.com/products/specifications

Oxford Nanopore Technologies (n.d.-i). *Types of nanopores*. Retrieved February 4, 2023 from https://nanoporetech.com/how-it-works/types-of-nanopores

PacBio (2023). *HiFi Sequencing* Retrieved February 5, 2023 from https://www.pacb.com/technology/hifi-sequencing/

Pallerla, S. R., Van Dong, D., Linh, L. T. K., Van Son, T., Quyen, D. T., Hoan, P. Q., Trung, N. T., The, N. T., Ruter, J., Boutin, S., Nurjadi, D., Sy, B. T., Kremsner, P. G., Meyer, C. G., Song, L. H., & Velavan, T. P. (2022). Diagnosis of pathogens causing bacterial meningitis using Nanopore sequencing in a resource-limited setting. *Ann Clin Microbiol Antimicrob*, *21*(1), 39. https://doi.org/10.1186/s12941-022-00530-6

Paragon Genomics (n.d.). *CleanPlex SARS-CoV-2 Kit*. https://www.paragongenomics.com/product/cleanplex-sars-cov-2-panel/

Park, J., Han, D. H., & Park, J. K. (2020). Towards practical sample preparation in point-of-care testing: user-friendly microfluidic devices. *Lab Chip*, *20*(7), 1191-1203. https://doi.org/10.1039/d0lc00047g

Pennisi, E. (2012). Search for pore-fection. *Science*, *336*(6081), 534-537. https://doi.org/10.1126/science.336.6081.534

Pollack, M. G., Fair, R. B., & Shenderov, A. D. (2000). Electrowetting-based actuation of liquid droplets for microfluidic applications. *Applied Physics Letters*, *77*(11), 1725-1726. https://doi.org/10.1063/1.1308534

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: their purpose and place. *Hum Mol Genet*, *27*(R2), R234-R241. https://doi.org/10.1093/hmg/ddy177

Pomerantz, A., Penafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., Barrio-Amoros, C. L., Salazar-Valenzuela, D., & Prost, S. (2018). Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience*, *7*(4). https://doi.org/10.1093/gigascience/giy033

Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouedraogo, N., Afrough, B., Bah, A., Baum, J. H., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sanchez, A., Carter, L. L., . . . Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, *530*(7589), 228-232. https://doi.org/10.1038/nature16996

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833-844. https://doi.org/10.1038/nbt.3935

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dundar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, *44*(W1), W160-165. https://doi.org/10.1093/nar/gkw257

Redondo, N., Zaldivar-Lopez, S., Garrido, J. J., & Montoya, M. (2021). SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Front Immunol*, *12*, 708264. https://doi.org/10.3389/fimmu.2021.708264

Reiling, S. J., Roy, A. M., Chen, S.-H., Quick, J., & Ragoussis, I. (2020). nCoV-2019 McGill Nanopore LibPrep Protocol, 5 ng NB. https://doi.org/10.17504/protocols.io.bjkvkkw6

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, *13*(5), 278-289. https://doi.org/10.1016/j.gpb.2015.08.002

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., . . . Bellini, W. J. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, *300*(5624), 1394-1399. https://doi.org/10.1126/science.1085952

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M., & Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, *265*(5596), 687-695. https://doi.org/10.1038/265687a0

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, *94*(3), 441-448. https://doi.org/https://doi.org/10.1016/0022-2836(75)90213-2

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, *74*(12), 5463-5467. https://doi.org/10.1073/pnas.74.12.5463

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, *581*(7807), 221-224. https://doi.org/10.1038/s41586-020-2179-y

Shen, R., Lv, A. m., Yi, S., Wang, P., Mak, P.-I., Martins, R. P., & Jia, Y. (2023). Nucleic acid analysis on electrowetting-based digital microfluidics. *TrAC Trends in Analytical Chemistry*, *158*. https://doi.org/10.1016/j.trac.2022.116826

Shendure, J. A., Porreca, G. J., & Church, G. M. (2008). Overview of DNA sequencing strategies. *Curr Protoc Mol Biol*, *Chapter 7*, Unit 7 1. https://doi.org/10.1002/0471142727.mb0701s81

Shi, C., Feng, Y., Sun, R., Chen, J., Zhao, Y., Wang, Z., Xie, S., Zhou, J., Yang, L., Cao, X., Feng, J., Zhang, C., Wei, B., Wang, X., Chang, Y., Zhao, J., Wang, Z., Zheng, J., Liu, J., . . . Ma, J. (2023). Development and clinical applications of an enclosed automated targeted NGS library preparation system. *Clin Chim Acta*, *540*, 117224. https://doi.org/10.1016/j.cca.2023.117224

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Mol Ecol*, *21*(8), 1794-1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*, *122*(1), e59. https://doi.org/10.1002/cpmb.59

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H. D., & Rappuoli, R. (2003). SARS--beginning to understand a new virus. *Nat Rev Microbiol*, *1*(3), 209-218. https://doi.org/10.1038/nrmicro775

Stranneheim, H., & Lundeberg, J. (2012). Stepping stones in DNA sequencing. *Biotechnol J*, *7*(9), 1063-1073. https://doi.org/10.1002/biot.201200153

Sun, Y., Li, X., Yang, Q., Zhao, B., Wu, Z., & Xia, Y. (2022). Genome Enrichment of Rare, Unknown Species from Complicated Microbiome by Nanopore Selective Sequencing. *bioRxiv*, 2022.2002.2013.480078. https://doi.org/10.1101/2022.02.13.480078

Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J. H., Lapointe, H., Kamelian, K., Smith, A. D., Prystajecky, N., Goodfellow, I., Wilson, S. J., Harrigan, R., Snutch, T. P., Loman, N. J., & Quick, J. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*, 2020.2009.2004.283077. https://doi.org/10.1101/2020.09.04.283077

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., & Holt, R. A. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351.

Volta Labs (n.d.). *Technology*. Retrieved June 7 from https://www.voltalabs.com/technology

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281-292 e286. https://doi.org/10.1016/j.cell.2020.02.058

Wang, H., Tang, Z., Li, Z., & Wang, E. (2001). Self-assembled monolayer of ssDNA on Au(111) substrate. *Surface Science*, *480*(1), L389-L394. https://doi.org/https://doi.org/10.1016/S0039-6028(01)01007-X

Wang, M., Fu, A., Hu, B., Tong, Y., Liu, R., Liu, Z., Gu, J., Xiang, B., Liu, J., Jiang, W., Shen, G., Zhao, W., Men, D., Deng, Z., Yu, L., Wei, W., Li, Y., & Liu, T. (2020). Nanopore Targeted Sequencing for the Accurate and Comprehensive Detection of SARS-CoV-2 and Other Respiratory Viruses. *Small*, *16*(32), e2002169. https://doi.org/10.1002/smll.202002169

Wang, M. Y., Zhao, R., Gao, L. J., Gao, X. F., Wang, D. P., & Cao, J. M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front Cell Infect Microbiol*, *10*, 587269. https://doi.org/10.3389/fcimb.2020.587269

Wang, Y., Yang, Q., & Wang, Z. (2014). The evolution of nanopore sequencing. *Front Genet*, *5*, 449. https://doi.org/10.3389/fgene.2014.00449

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*, *39*(11), 1348-1365. https://doi.org/10.1038/s41587-021-01108-x

Watson, J. D., & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737-738. https://doi.org/10.1038/171737a0

WHO. (2022). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved June 29, 2022 from https://covid19.who.int/

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom*, *3*(10), e000132. https://doi.org/10.1099/mgen.0.000132

Wohl, S., Schaffner, S. F., & Sabeti, P. C. (2016). Genomic Analysis of Viral Outbreaks. *Annu Rev Virol*, *3*(1), 173-195. https://doi.org/10.1146/annurev-virology-110615-035747

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., Yuan, M. L., Zhang, Y. L., Dai, F. H., Liu, Y., Wang, Q. M., Zheng, J. J., Xu, L., Holmes, E. C., & Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265-269. https://doi.org/10.1038/s41586-020-2008-3

Xiao, T., & Zhou, W. (2020). The third generation sequencing: the advanced approach to genetic diseases. *Translational Pediatrics*, *9*(2), 163-173. https://doi.org/10.21037/tp.2020.03.06

Zaaijer, S., Gordon, A., Speyer, D., Piccone, R., Groen, S. C., & Erlich, Y. (2017). Rapid re-identification of human samples using portable DNA sequencing. *eLife*, *6*, e27798. https://doi.org/10.7554/eLife.27798

Zhang, H., Li, H., Jain, C., Cheng, H., Au, K. F., Li, H., & Aluru, S. (2021). Real-time mapping of nanopore raw signals. *Bioinformatics*, *37*(Suppl_1), i477-i483. https://doi.org/10.1093/bioinformatics/btab264

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., . . . Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, *579*(7798), 270-273. https://doi.org/10.1038/s41586-020-2012-7

Appendix I: Supplemental Figures and Tables

Supplemental table 1. Library preparation protocol testing with different amounts of starting material. Sequencing metrics. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

DNA amount used for library prep	Sample	Total reads	Mapped reads	Covered bases	Coverage	Mean depth	Mean base q	Mean map q
100	onestep100ng1	400827	400372	29868	99.88	5275	21.1	60
	twostep100ng1	220539	220266	29888	99.95	3184	21.2	60
	onestep100ng2	408197	405460	29876	99.91	5400	21.1	59.9
	twostep100ng2	299634	299197	29863	99.87	4357	21	60
	onestep100ng3	3853	3816	29374	98.23	51	21.2	59.9
	twostep100ng3	555252	554264	29901	99.99	7614	21	60
75	onestep75ng1	305744	305250	29899	99.99	4038	21.2	60
	twostep75ng1	201953	201599	29868	99.88	3043	21.2	60
	onestep75ng2	446495	443182	29903	100.00	5886	21.1	59.9
	twostep75ng2	397065	396288	29867	99.88	5592	21.1	59.9
	onestep75ng3	84552	84081	29663	99.20	1114	21	59.9
	twostep75ng3	145897	145517	29869	99.89	2097	21	60
50	onestep50ng1	193564	193309	29900	99.99	2582	20.9	60
	twostep50ng1	84801	84662	29850	99.82	1244	21	59.9
	onestep50ng2	108616	107829	29873	99.90	1434	21	59.9
	twostep50ng2	241064	240793	29869	99.89	3784	21	60
	onestep50ng3	64316	63769	29696	99.31	822	21.1	59.9
	twostep50ng3	405319	404524	29889	99.95	5434	21	60
	onestepnegctrl	67	2	756	2.53	0	21	60
	twostepnegctrl	631	2	709	2.37	0	16.2	60

Supplemental table 2. Library preparation protocol testing with different amounts of starting material. Sequenced fragment length. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

DNA amount used for library prep	Sample	Read Len. Min.	Read Len. 1st. Qu.	Read Len. Mean	Read Len. Median	Read Len. 3rd Qu.	Read Len. Max	Read Len. Std.
100	onestep100ng1	42	384	416	396	411	2041	119
	twostep100ng1	41	378	444	390	409	2710	182
	onestep100ng2	43	391	409	403	420	2446	142
	twostep100ng2	41	377	445	389	409	2302	183
	onestep100ng3	43	389	416	398	411	1623	113
	twostep100ng3	41	378	427	390	406	3009	154
75	onestep75ng1	41	389	421	400	414	1972	123
	twostep75ng1	41	378	444	388	408	2359	189
	onestep75ng2	42	393	409	404	420	2001	142
	twostep75ng2	41	378	474	390	412	2737	223
	onestep75ng3	42	389	406	402	416	1532	133
	twostep75ng3	41	378	419	390	405	2439	137
50	onestep50ng1	41	392	427	404	418	2029	128
	twostep50ng1	41	378	461	391	412	2380	201
	onestep50ng2	43	393	412	406	421	2091	139
	twostep50ng2	41	377	439	388	409	2731	171
	onestep50ng3	43	386	394	399	413	1318	117
	twostep50ng3	41	379	442	391	408	2351	180
	onestepnegctrl	382	382	385	382	386.5	391	4
	twostepnegctrl	392	444.5	462	497	497	497	49

Supplemental table 3. Library preparation protocol testing with different amounts of starting material. Pango lineage calling. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

DNA amount used for library prep	Sample	Lineage	Conflict	Scorpio call	Scorpio support	Scorpio conflict
100	onestep100ng1	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep100ng1	AY.119.2	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep100ng2	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep100ng2	AY.103	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep100ng3	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep100ng3	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08
75	onestep75ng1	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep75ng1	AY.119.2	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep75ng2	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep75ng2	AY.103	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep75ng3	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep75ng3	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08
50	onestep50ng1	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep50ng1	AY.119.2	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep50ng2	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep50ng2	AY.103	0	Delta (B.1.617.2-like)	0.77	0.08
	onestep50ng3	AY.74	0	Delta (B.1.617.2-like)	0.77	0.08
	twostep50ng3	AY.44	0	Delta (B.1.617.2-like)	0.77	0.08

Supplemental table 4. Library preparation protocol testing with different Ct values. Sequencing metrics. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

Ct value	Sample	Total reads	Mapped reads	Covered bases	Coverage	Meand depth	Mean base q	Mean map q
15	g10ct15onestep	2007766	1278398	29859	99.85	16885	23.5	59.5
	g10ct15twostep	871710	481834	29859	99.85	5874	23.8	59.2
	a6ct15onestep	1905650	1497363	29849	99.82	19060	23.8	59.5
	a6ct15twostep	849582	640770	29844	99.80	7880	23.5	59.3
	b6ct15onestep	1616645	968048	29852	99.83	12661	23.6	59.4
	b6ct15twostep	696352	372832	29839	99.79	4535	23.5	59.2
20	g11ct20onestep	1397277	643595	29848	99.82	8558	23.5	59.4
	g11ct20twostep	591631	219373	29854	99.84	2645	23.3	58.8
	h4ct20onestep	1251467	518961	29864	99.87	6816	23.5	59.3
	h4ct20twostep	802404	297904	29866	99.88	3614	23.4	58.9
	d6ct20onestep	1617195	1105540	29864	99.87	14531	23.8	59.4
	d6ct20twostep	811290	552102	29868	99.88	6956	23.4	59.4
25	a2ct25onestep	1430809	1353254	29860	99.86	17208	23.8	59.5
	a2ct25twostep	694116	660163	29894	99.97	8233	23.6	59.4
	d2ct25onestep	1904568	1613264	29839	99.79	20660	23.8	59.5
	d2ct25twostep	1031300	873188	29818	99.72	10865	23.9	59.4
	f2ct25onestep	1213359	164291	29845	99.81	2095	23.3	59.3
	f2ct25twostep	1467046	158770	29838	99.78	1836	23.4	58.5
30	a1ct30onestep	1947930	125947	29832	99.76	1521	23.7	59.3
	a1ct30twostep	1363938	87201	29828	99.75	1040	23.2	59.2
	b1ct30onestep	1433668	10204	25862	86.49	115	23.3	58.7
	b1ct30twostep	1156330	7124	25121	84.01	71	23.2	57.6
	f1ct30onestep	1850802	3554	9168	30.66	40	23.2	58.9
	f1ct30twostep	1060225	1956	8473	28.33	21	23.3	58.4
	posctrlonestep	1331244	1215700	29847	99.81	15660	23.6	59.3
	posctrltwostep	765996	695312	29872	99.90	8477	23.4	59.1

Supplemental table 5. Library preparation protocol testing with different Ct values. Sequenced fragment length. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

Ct value	Sample	Read Len. Min.	Read Len. 1st. Qu.	Read Len. Mean	Read Len. Median	Read Len. 3rd Qu.	Read Len. Max	Read Len. Std.
15	g10ct15onestep	43	398	422	405	420	2224	137
	g10ct15twostep	43	398	397	406	423	5789	137
	a6ct15onestep	42	405	407	413	419	1940	140
	a6ct15twostep	42	403	376	413	420	1987	148
	b6ct15onestep	43	395	412	404	420	2185	160
	b6ct15twostep	41	393	377	405	422	1852	169
20	g11ct20onestep	42	399	453	407	432	2568	147
	g11ct20twostep	43	400	446	412	467	3093	150
	h4ct20onestep	42	392	414	404	445	2743	211
	h4ct20twostep	43	117	365	403	445	2137	215
	d6ct20onestep	42	400	406	412	420	2443	188
	d6ct20twostep	42	112	350	411	420	2197	204
25	a2ct25onestep	41	405	399	413	420	1631	121
	a2ct25twostep	42	404	376	413	420	1643	134
	d2ct25onestep	42	405	403	413	421	2956	127
	d2ct25twostep	41	405	382	414	422	1921	133
	f2ct25onestep	44	412	529	456	631	2193	191
	f2ct25twostep	42	411	475	438	555	2418	180
30	a1ct30onestep	42	409	440	414	422	2602	153
	a1ct30twostep	45	408	441	414	426	2708	187
	b1ct30onestep	83	405	496	431	629	2159	220
	b1ct30twostep	60	219	387	410	510	1320	199
	f1ct30onestep	51	411	501	461	623	1329	186
	f1ct30twostep	58	379	419	423	485.25	1785	177
	posctrlonestep	42	403	419	412	419	2156	148
	posctrltwostep	42	405	403	414	420	2724	137

Supplemental table 6. Library preparation protocol testing with different Ct values. Pango lineage calling. Libraries prepared with the one-step ligation protocol are shown in gray and libraries prepared with the standard two-step protocol are shown in white.

Ct value	Sample	Lineage	Conflict	Scorpio call	Scorpio support	Scorpio conflict
15	g10ct15onestep	BA.4.6	0	Omicron (BA.4-like)	0.94	0
	g10ct15twostep	BA.4.6	0	Omicron (BA.4-like)	0.95	0
	a6ct15onestep	BA.5.2	0	Omicron (BA.5-like)	0.92	0.02
	a6ct15twostep	BA.5.2	0	Omicron (BA.5-like)	0.93	0.02
	b6ct15onestep	BA.4.6	0	Omicron (BA.4-like)	0.94	0
	b6ct15twostep	BA.4.6	0	Omicron (BA.4-like)	0.92	0
20	g11ct20onestep	BA.4.6	0	Omicron (BA.4-like)	0.94	0
	g11ct20twostep	BA.4.6	0	Omicron (Unassigned)	0.88	0
	h4ct20onestep	BA.4.6.5	0.5	Omicron (BA.4-like)	0.94	0
	h4ct20twotep	BA.4.6.5	0.5	Omicron (BA.4-like)	0.92	0
	d6ct20onestep	BF.10	0	Omicron (BA.5-like)	0.9	0.02
	d6ct20twostep	BF.10	0	Omicron (BA.5-like)	0.9	0.02
25	a2ct25onestep	XBR	0	Omicron (Unassigned)	0.91	0.03
	a2ct25twostep	XBR	0	Omicron (Unassigned)	0.91	0.03
	d2ct25onestep	XBB.1.5	0.333333	Omicron (BA.2-like)	0.89	0.02
	d2ct25twostep	XBB.1.5	0.333333	Omicron (BA.2-like)	0.91	0.02
	f2ct25onestep	BQ.1.1.40	0	Omicron (Unassigned)	0.85	0.03
	f2ct25twostep	BQ.1.1.40	0	Omicron (Unassigned)	0.88	0.03
30	a1ct30onestep	BQ.1.1	0	Omicron (Unassigned)	0.9	0.02
	a1ct30twostep	BQ.1.1	0	Omicron (BA.5-like)	0.9	0.02
	b1ct30onestep	XBB.1.5	0.5	Omicron (Unassigned)	0.88	0
	b1ct30twostep	XBB.1	0.5	Omicron (Unassigned)	0.79	0.03
	posctrlonestep	B.1	0			
	posctrltwostep	B.1	0			

Supplemental table 7. On chip library preparation testing. Sequencing metrics.

Sample	Total reads	Mapped reads	Covered bases	Coverage	Mean depth	Mean base q	Mean map q
On Chip 1	1104873	1103323	29845	99.81	14756	19.2	59.9
On Chip 2	421402	420984	29837	99.78	5612	17.5	60
Off Chip	12749782	12695103	29873	99.90	168873	22.1	59.9

Supplemental table 8. On chip library preparation testing. Sequenced fragment length.

Sample	Read Len. Min.	Read Len. 1st. Qu.	Read Len. Mean	Read Len. Median	Read Len. 3rd Qu.	Read Len. Max	Read Len. Std.
OnChip1	41	394	405	401	409	2494	89
OnChip2	43	394	408	401	410	1490	86
OffChip	40	395	398	402	412	27941	96

Supplemental table 9. On chip library preparation testing. Pango lineage calling.

Sample	Lineage	Conflict	Scorpio call	Scorpio support	Scorpio conflict
On Chip 1	B.1	0			
On Chip 2	B.1	0			
Off Chip	B.1	0			

Supplemental table 10. On chip PCR testing. 1st experiment. Sequencing metrics.

Sample	Total reads	Mapped reads	Covered bases	Coverage	Mean depth	Mean base q	Mean map q
Off Chip 1	766627	765825	28037	93.76	9781	21.9	59.9
On Chip A	692104	589146	27908	93.33	5807	21.4	59.3
On Chip B	530807	297670	27251	91.13	2301	21.5	56.9

Supplemental table 11. On chip PCR testing. 1st experiment. Sequenced fragment length

Sample	Read Len. Min.	Read Len. 1st. Qu.	Read Len. Mean	Read Len. Median	Read Len. 3rd Qu.	Read Len. Max	Read Len. Std.
OnChipA	51	314	356	338	392	1987	79
OnChipB	41	335	418	364	488	2245	174
OffChip1	48	378	390	384	393	1937	49

Supplemental table 12. On chip PCR testing. 1st experiment. Pango lineage calling.

Sample	Lineage	Conflict	Scorpio call	Scorpio support	Scorpio conflict
On Chip A	B	0			
On Chip B	B	0.5			
Off Chip 1	B	0			

Supplemental table 13. On chip PCR testing. 2nd experiment. Sequencing metrics. Experiments carried out after temperature calibration of the PCR chambers.

Sample	Total reads	Mapped reads	Covered bases	Coverage	Mean depth	Mean base q	Mean map q
On Chip C	40231	40014	28149	94.13	502	20.3	59.7
On Chip D	49020	48636	28539	95.44	598	20.2	59.7
Off Chip 2	38131	38063	28662	95.85	495	20.2	60
Off Chip 3	43394	43323	28389	94.94	564	20.2	60

Supplemental table 14. On chip PCR testing. 2nd experiment. Sequenced fragment length. Experiments carried out after temperature calibration of the PCR chambers.

Sample	Read Len. Min.	Read Len. 1st. Qu.	Read Len. Mean	Read Len. Median	Read Len. 3rd Qu.	Read Len. Max	Read Len. Std.
OnChipC	76	383	393	389	403	1200	57
OnChipD	79	380	394	390	402	1426	68
OffChip2	88	386	397	393	405	1196	57
OffChip3	71	387	401	395	408	1304	56

Supplemental table 15. On chip PCR testing. 2nd experiment. Pango lineage calling Experiments carried out after temperature calibration of the PCR chambers.

Sample	Lineage	Conflict	Scorpio call	Scorpio support	Scorpio conflict
On Chip C	B	0			
On Chip D	B	0			
Off Chip 2	B	0			
Off Chip 3	B	0			