

Research Article

Reflective Evaluation of Next-Generation Sequencing Data during Early Phase Detection of the Delta Variant

Upasana Ramphal^{1, 2, 3}, Derek Tshiabuila⁴, Yajna Ramphal⁴, Jennifer Giandhari^{1, *}, Carel van Heerden⁵, Cheryl Baxter^{2, 4}, Stephanie van Wyk⁴, Sureshnee Pillay¹, Oluwakemi Laguda-Akingba⁶, Eduan Wilkinson^{1, 4}, Richard Lessells¹, Tulio de Oliveira^{1, 2, 3, 4, 7}

1. KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa; E-Mails: upasanaramphal@gmail.com; GiandhariJ@ukzn.ac.za; pillaysureshnee@gmail.com; ewilkinson@sun.ac.za; LessellsR@ukzn.ac.za; tulio@sun.ac.za
2. Center for AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa; E-Mail: cbaxter@sun.ac.za
3. Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), Durban, South Africa
4. Centre for Epidemic Response and Innovation (CERI), School of Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa; E-Mails: derektsiabuil@gmail.com; yajna.ramphal@gmail.com; stephanie.vanwyk@gmail.com;
5. Central Analytical Facilities (CAF), Stellenbosch University, Stellenbosch, South Africa; E-Mail: cjvh@sun.ac.za
6. Faculty of Health Sciences, Walter Sisulu University and National Health Laboratory Service Gqeberha, Port Elizabeth, Eastern Cape, South Africa; E-Mail: Oluwakemi.LA@nhls.ac.za
7. Department of Global Health, University of Washington, Seattle, Washington, United States of America

* **Correspondence:** Jennifer Giandhari; E-Mail: GiandhariJ@ukzn.ac.za

Academic Editor: Andrés Moya

OBM Genetics

2024, volume 8, issue 2

doi:10.21926/obm.genet.2402239

Received: March 01, 2024

Accepted: May 16, 2024

Published: May 30, 2024

Abstract

During the SARS-CoV-2 pandemic, next-generation sequencing (NGS) technologies like the Ion Torrent S5 and Illumina MiSeq, alongside advanced software, improved genomic surveillance



© 2024 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

in South Africa. This study analysed anonymized samples from the Eastern Cape using Genome Detective and NextClade, showing Ion Torrent S5 and Illumina MiSeq success rates of 96% and 94%, respectively. The study focused on genomic coverage (above 80%) and mutation detection (below 100), with the Ion Torrent S5 achieving 99% coverage compared to Illumina MiSeq's 80%, likely due to different primers used in amplification. The Ion Torrent S5 was more effective in sequencing varied viral loads, whereas Illumina MiSeq had difficulties with lower loads. Both platforms were adept at identifying clades, successfully differentiating between Beta (<45%) and Delta variants (<30%), despite minor discrepancies in assignments due to Illumina MiSeq's lower coverage, leading to a failure rate of up to 6%. Manual library preparation showed similar sample processing and clade identification capabilities for both platforms. However, differences in sequencing duration (3.5 vs. 36 hours), automation level, genomic coverage (80% vs. 99%), and viral load compatibility were noted, highlighting each platform's unique advantages and challenges in SARS-CoV-2 genomic surveillance. In conclusion, the Illumina MiSeq and Ion Torrent S5 platforms are both efficacious in executing whole-genome sequencing (WGS) via amplicons, facilitating precise, accurate, and high-throughput examinations of SARS-CoV-2 viral genomes. However, it is important to note the existence of disparities in the quality of data produced by each platform. Each system offers unique benefits and limitations, rendering them viable choices for the genomic surveillance of SARS-CoV-2.

Keywords

Next-generation-sequencing (NGS); SARS-CoV-2; Illumina MiSeq; Ion Torrent S5; genomic surveillance; viral load; data analysis

1. Introduction

At the end of 2019, incidence rates of SARS-CoV-2 increased substantially within a short period, followed by a rapid global diffusion and evolution of the virus, resulting in five novel variants of concern (VOC) - each driving new infection waves [1, 2]. A VOC, as listed by the World Health Organisation and Centers for Disease Control and Prevention, is defined as variants that include attributes of increasing transmissibility, caused severe disease and increased hospitalizations, showing a reduction in therapeutics such as vaccine efficacy and diagnostics detection failures [3, 4]. Globally, more than 691 million confirmed cases of COVID-19 have been reported to date resulting in over 6, 9 million deaths [5].

Pathogen genome sequencing is a fundamental surveillance tool used to support the understanding of the molecular epidemiology of disease outbreaks [6]. Recent advances in sequencing technologies have shown their applicability for research use in outbreak situations, such as those observed with Ebola, Zika virus, SARS, MERS, and most recently, the novel SARS-CoV-2 [7-12]. Tracking signature genetic mutations of the virus allows researchers to estimate the influence of early outbreaks. It ensures more accurate detection and characterization of variants, possible drug resistance mutations, vaccine escape variants, virulence, and pathogenicity factors [13-17]. Therefore, genomic surveillance, complemented by real-time monitoring and data-sharing

networks, is valuable for understanding SARS-CoV-2 transmission and epidemic dynamics. Sequencing centers nationwide initiated genomic surveillance programs for the WGS of SARS-CoV-2 as part of the Network for Genomic Surveillance in South Africa (NGS-SA) [18]. Sequencing technologies used included Illumina, Ion Torrent, and Oxford Nanopore technologies. Together, the network kept abreast of the latest SARS-CoV-2 variants circulating within South Africa's infection waves and produced over 53,000 genomes, which have been made publicly available since May 2020 [19-22].

Although there are distinct differences between these two systems, the Ion Torrent S5 and Illumina MiSeq platforms are well-known for their mid- to high-throughput sequencing, variant calling, and overall good-quality short-read sequences [23-26]. Illumina, a benchmark in sequencing technology, uses a fluorescence-based paradigm for determining nucleotide sequence in which all of the enzymatic processes and imaging steps take place in a flow cell. Ion Torrent, an alternative sequencing technology, reads nucleotide sequences based on a measure of pH by proton release and makes use of a semiconductor sequencing chip and ion spheres bound to DNA [25, 27, 28]. Additionally, there are differences in the type of data generated by each platform. The sequence reads generated from Illumina data in a single run have the same length and are paired-end reads, whereas Ion Torrent reads vary in size and are single-ended [23, 29]. A comparative study involving SARS-CoV-2 further highlights the ease of use and operation with the automated Ion Torrent S5 workflow when coupled with the Ion Chef [30]. The Ion Chef was used to automate library preparation for small sample numbers and is an essential component in templating prepared libraries onto the Ion Torrent sequencing chip. Cost comparisons of these platforms are similar, provided an increase in multiplexing of samples is maintained on the Ion Torrent S5 [26]. The operation of the Illumina MiSeq at maximum capacity ensures an overall low cost due to the high efficiency of the platform [26, 30].

While SARS-CoV-2 has been broadly studied over the past two years, there are still concerns about emerging variants and mutations; therefore, continued surveillance of SARS-CoV-2 variants remains critical for identifying new emerging variants. For this reason, it is necessary to ensure that sequencing data generated by various WGS platforms is comparable in terms of performance and sequencing output. A recent study performed a benchmarking comparison on several different SARS-CoV-2 genome-sequencing protocols and reported performance variation across WGS technologies [31]. Another study assessed the WGS of SARS-CoV-2 using the Ion Torrent and Illumina technologies with their respective protocols in considerable detail [30]. Although the study's findings demonstrate that genomic coverage was high with faster turnaround times on the Ion Torrent platform, it is challenging to compare the data generated without using the same analysis pipeline to avoid discrepancies in assembly and quality control processes.

Therefore, in this retrospective comparative study, we reflect on the data generated using one set of remnant SARS-CoV-2 positive samples collected from the Eastern Cape to directly compare data generated by the Ion Torrent S5 and Illumina MiSeq platforms. Both platforms used the same analysis pipeline to assess if genomic coverage, quantification of mutations (deletions, insertions, and substitutions), and clade assignment were comparable for data generated across the two platforms.

2. Materials and Methods

2.1 Sample Population, Collection, and Processing

As part of the NGS-SA initiative, we used remnant routine Eastern Cape genomic surveillance sample swabs with collection dates ranging from 14 March 2021 to 9 June 2021 to assess the sequencing data generated from two NGS platforms [18]. The National Health Laboratory Service (NHLS) in Port Elizabeth, Eastern Cape collected nasopharyngeal and oropharyngeal swabs from inpatients and outpatients in clinics and hospitals. The NHLS team also determined SARS-CoV-2 positivity using qualitative polymerase chain reaction (qPCR) assays on the Cobas® SARS-CoV-2 (Roche Molecular, Pleasanton, CA, USA), Xpert Xpress SARS-CoV-2 (Cepheid, CA, USA) or Seegene Allplex™ 2019-nCoV Assay and the CFX96 DX™, Bio-Rad (Seegene, Inqaba Biotec, SA). Sample Ct values were provided as part of the metadata files accompanying all samples. Remnant nasopharyngeal and oropharyngeal swabs from 183 patients were used for this study (Table S1), irrespective of their Ct values. The RNA was extracted at the KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP) based in Durban, KwaZulu-Natal, South Africa. The extracted RNA was used for independent library preparation at the respective sequencing sites followed by sequencing on the Illumina MiSeq based at KRISP and on the Ion Torrent S5 based at the Central Analytical Facilities (CAF) in Stellenbosch, Western Cape. Samples were sequenced over two runs on each platform to directly compare the sequence data generated.

2.2 Nucleic Acid Extraction

All 183 samples were extracted as per the manufacturer's instructions using the CMG-1049 kit on the Chemagic 360 instrument (Perkin Elmer, Hamburg, Germany). Total nucleic acid (TNA) extraction was performed using 200 µL per sample added to 450 µL lysis buffer and 14 µL Poly A RNA/proteinase-K reaction mixture. The TNA was eluted in 100 µL elution buffer in which two aliquots were made and stored at -20 C until further use.

2.3 Complementary DNA (cDNA) Synthesis

cDNA synthesis of samples sequenced on the Illumina MiSeq was performed with the SuperScript IV reverse transcriptase using random hexamers (Life Technologies), while cDNA synthesis on samples sequenced on the Ion Torrent S5 was performed with the SuperScript Vilo cDNA synthesis kit (Life Technologies).

2.4 Library Preparation and Next-Generation Sequencing Strategies

Sequence libraries for Illumina MiSeq and Ion Torrent S5 sequencing were manually prepared using the Nextera DNA Flex Library Prep kit and the Ion AmpliSeq Library Kit Plus, respectively. Templating of the prepared libraries onto the sequencing chip for the Ion Torrent S5 was automated using the Ion Chef the Ion AmpliSeq SARS-CoV-2 Research Panel, and Ion AmpliSeq kit for Chef DL8.

2.5 Multiplex Tiling PCR, Illumina MiSeq Library Preparation, and Sequencing

As previously published, samples sequenced using Illumina MiSeq were amplified using a multiplex PCR [21]. ARTIC primers were designed on Primal Scheme (<http://primal.zibraproject.org/>) to generate 400 base pair (bp) amplicons with 70 bp overlaps [32]. The primers (v3 as of June 2021) were used to amplify the SARS-CoV-2 whole genome (30 kb). Amplicons were purified using Ampure XP purification beads (Beckman Coulter, High Wycombe, UK), using a 1:1 ratio. All purified amplicons were quantified on the Qubit 4.0 instrument using the Qubit double-stranded DNA (dsDNA) High Sensitivity assay kit (Life Technologies). Purified amplicons were stored at 4°C prior to further use. Indexed paired-end libraries were prepared using the Nextera DNA Flex Library Prep kit and the Nextera DNA CD indexes (Illumina, San Diego, USA) per the manufacturer's instructions. Libraries were purified and normalized to 4 nM prior to the pooling. The pooled library was denatured using 0.2 N sodium hydroxide followed by dilution to obtain a final concentration of 8 pM. At least two controls were included in each sequencing run, and 96 samples were processed in total. The library was spiked with 1% PhiX Control v3 (adapter-ligated library used as a control) and was sequenced using a 500-cycle v2 MiSeq Reagent Kit on the Illumina MiSeq instrument (Illumina, San Diego, CA, USA) [21].

2.6 Ion Torrent S5 Library Preparation and Sequencing

Manual library preparation workflow was performed using the overlapping amplicon strategy and a 2-pool primer panel on the Ion AmpliSeq Library Kit Plus. Primer pool 1 consisted of 125 primer pairs and primer pool 2 consisted of 122 primer pairs generating amplicons within a range of 125 bp to 275 bp in length. It is imperative to note that the panel design allows for the tiling of ~237 amplicons across the SARS-CoV-2 genome (~30kb), resulting in a sequencing coverage of 99.0%, covering positions 43 to position 29,842 (positions relative to the SARS-CoV-2 reference, GenBank accession number NC_045512). An additional five primer pairs targeting human expression were used as controls within the panel. SARS-COV-2 targets were set up using a 16-cycle target amplification for samples with a broad range of viral load. Amplified targets for the two primer pools were combined and ligated using the Ioncode Barcode Adapters 1-96 kit. Automated templating of 70 pM libraries was loaded onto two high sequencing data output Ion 540 Chips per sequencing run using the Ion Chef followed by sequencing on the Ion Torrent S5 as per manufacturer's protocol (Ion 540™ Kit - Chef User Guide Pub. No. MAN0010851). A minimum of two controls were included in each run, and 96 samples were processed per sequence run using two 540 chips. All runs were pre-planned and set up using the Ion Torrent suite software (v5.16.0). All information on the Ion AmpliSeq SARS-CoV-2 Research Panel is available at <https://ampliseq.com>.

2.7 Sequence Data Analysis

The data analysis was processed by one bioinformatician at KRISP and the sequences generated by the Illumina MiSeq were initially analyzed before those generated on the Ion Torrent S5 due to routine sequencing schedules. The raw paired-end reads generated from Illumina MiSeq and the raw single-end reads from Ion Torrent sequencing (FASTQ files) were assembled using the web-based application Genome Detective, version 1.126 (<https://www.genomeDetective.com/>) [33]. Genome Detective is a web-based assembly tool incorporating de novo and reference-based

mapping algorithms to assemble whole viral genomes. The initial assemblies obtained from Genome Detective were refined by aligning mapped reads to the reference and generating consensus sequences for each comparison run on both sequencing platforms. Consensus sequences for both platforms were assessed using NextClade (<https://clades.nextstrain.org/>, version 1.7.4) for sequence clade assignment, identification, quantification of mutations, and sequence quality analyses. NextClade is a classification tool that utilizes Nextstrain nomenclature to distinguish differences between a given sequence and a reference sequence to identify various clades and VOCs [34]. Additional data regarding S-gene coverage was obtained by uploading consensus sequences to Genome Detective. Sequences (FASTA files) that passed quality control with greater than 80% genomic coverage and less than 100 mutations were deposited onto Global Initiative on Sharing Avian Influenza Data (GISAID) (<https://www.gisaid.org/>) [19]. Internal quality control of 100 mutations was obtained from previous sequencing data using the mutational rate of the virus and the time-lapse into the SARS-CoV-2 pandemic. All the raw sequence data generated for this research are available publicly in the NCBI Sequence Read Archive (project no. PRJNA636748). Epidemic data is freely available on the open-source GISAID database and accessed using the GISAID Identifier: EPI_SET_230203ym; doi: <https://doi.org/10.55876/gis8.230203ym>. The GISAID accession identifiers are also included as part of File S1 and Table S2. Most sequences uploaded onto GISAID were obtained from data generated on the Illumina MiSeq platform, as these were initially sequenced and analysed for uploading purposes only. Outstanding sequences that passed quality control were obtained from sequencing data generated on the Ion Torrent S5 and uploaded onto GISAID after that.

2.8 Statistical Evaluation and Considerations

Data visualization and statistical analysis were performed using ggplot2 v3.3.6 package and R v.4.2. A spearman's ranked sum correlation test was performed to determine the relationship between viral load and coverage (genomic and S-gene) obtained from sequences generated on each platform. The Wilcoxon test was used to establish the difference in the range of genomic coverage obtained between platforms and to assess the difference in the quantification of mutations (total mutations, insertions, deletions, and substitutions) detected between the Ion Torrent S5 and the Illumina MiSeq platforms.

This study falls under the approval of the Biomedical Research Ethics Committee (BREC) of the University of KwaZulu-Natal, South Africa, with protocol reference number BREC/00004745/2022. As no human subjects were involved, no informed consent was required.

3. Results

3.1 Process Flow

The Illumina MiSeq and Ion Torrent S5 sequencing platforms followed specific workflows for WGS of SARS-CoV-2 samples. Figure 1 provides a detailed overview of the processes involved, from sample receipt to data analysis. Post nucleic acid extraction. The Illumina MiSeq and Ion Torrent S5 workflows take on a separate direction in sample processing.

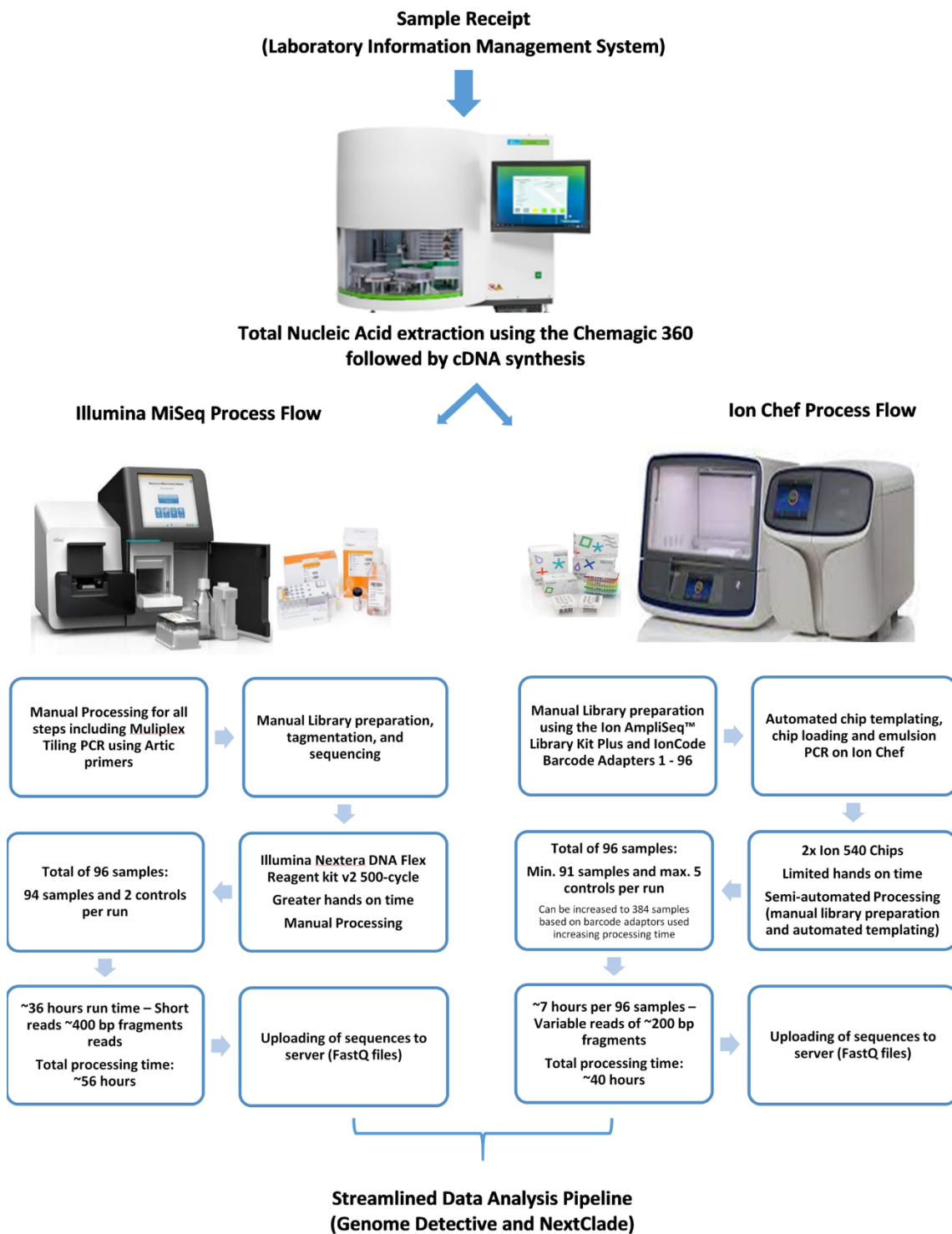


Figure 1 Overview of the Next-Generation Sequencing process flow for the WGS of SARS-CoV-2 using the Illumina MiSeq and Ion Torrent S5 sequencing platforms. The Illumina MiSeq and Ion Torrent S5 workflows accommodate manual libraries of 96 samples in total. Illumina MiSeq accommodates 94 samples and 2 controls per sequence run; Ion Torrent S5 accommodates a minimum of 91 samples and a maximum of 5 controls per sequence run using 2 Ion 540 chips.

3.2 Sequencer-Specific Attributes

Sequencing data for the independent runs (Table S3 and Table S4) were subjected to WGS on the Illumina MiSeq and Ion Torrent S5, respectively. This sequencing data for each platform was tabulated separately upon run completion due to platform-specific parameters.

Each platform was independently assessed based on kit and platform-specific outputs as listed in Table 1. Each platform displays similarities in the use of 2-pool protocol-specific primer sets, processing capacity of a broad range of sample quantities, semi-automation of process, high sequence success rates, and genomic coverage. The platforms differ across the platform-specific kit requirements and primer sets used, as well as the performance output in read length, fragment size, and varied sequencing duration observed per sequence run.

Table 1 Comparison of (a) Illumina MiSeq and (b) Ion Torrent S5 sequencing methodologies used for WGS of SARS-CoV-2.

Attributes	Illumina MiSeq	Ion Torrent S5 and Chef
Primer details	Artic Primers (version 3) - 2 pools	AmpliSeq Primers - 2 pools
Sequencer performance and kit details	Nextera Flex V2 500 cycle kit 24-30 million paired-ended reads 12-15 million single reads Throughput: 8.5 Gb Sequencing Run time: ~36 hours	AmpliSeq Research Panel using the Ion 540 Chip 60-80 million reads Throughput: 10-15 Gb Sequencing Run time: ~3.5 hours per chip
Sample quantity	Processing low to high sample numbers	Processing of low to high sample numbers
Sample processing (automated vs manual)	Can be automated using the addition of an automated liquid handler to reduce hands on time and error rates Automation allows for handling of large sample numbers High throughput with manual library preparation and indexing of 96 libraries at a time Increased hands on times and increase error rates	Automation of library preparation on Ion Chef allows for small sample number processing (8 samples per 7.5 hours) Automation limits hands on time and limits error rates Higher throughput is achievable by preparing libraries manually with the use greater IonCode Barcode Adaptors Manual library preparation increases hands on time and increases error rates
Read lengths	Assembled Paired-end reads 400 bp fragments	Single-end reads 125-275 bp fragments

3.3 Sequencing Platform Performance Comparison

Run A and Run B, constituting 183 libraries in total, were sequenced on both platforms, generating 86 consensus sequences for each Illumina MiSeq run, followed by 92 and 83 consensus sequences for the Ion Torrent S5 sequencing runs (Table 2).

Table 2 Summary of Consensus sequence data from Illumina MiSeq and Ion Torrent S5 platforms.

<i>Properties</i>	<i>Sequencing Platform</i>	
	<i>Illumina MiSeq</i>	<i>Ion Torrent S5 and Chef</i>
<i>Total number of samples processed per platform</i>	183	
Sequencing Run Time per run (hrs)	36	7 (3.5 hrs per 540 chip)
Sequence Success Rate (n/N (%))	172/183 (93.9%)	175/183 (95.6%)
Sequence Failure Rates (n/N (%))	11/183 (6.0%)	8/183 (4.4%)
Greater than 80% genomic coverage (n/N (%))	138/172 (80.2%)	174/175 (99.4%)
Sequences with less than 100 mutations (%)	100%	100%
Mean Genomic Coverage (%)	83.4%	98.9%
Mean S-gene coverage (%)	78.7%	99.0%
Total mutations per platform	7036	8116
Total Substitutions per platform	4671	5039
Total Insertions per platform	17	46
Total Deletions per platform	2348	3031
Paired consensus genomes for both platforms (n/N (%))	164/183 (89.6%)	
Matched clade assignments between platforms (n/N (%))	147/183 (80.3%)	
Mismatched clade assignments between platforms (n/N (%))	17/183 (9.3%)	

The sequence success rate was 93.9% on the Illumina MiSeq and 95.6% on the Ion Torrent S5. The sequencing runtime per run on the Illumina MiSeq was 36 hours, whereas the Ion Torrent S5 platform had a total sequencing runtime of seven hours per 96 samples. Of the 183 sequenced samples, 164 (89.6%) had paired consensus sequences from both platforms. Table 2 combines the samples sequenced in Run A and Run B on both the Illumina MiSeq and the Ion Torrent S5 for a minimum of 91 samples and a maximum of five controls per run. A total of 183 libraries were sequenced on each platform, and 172 and 175 successful consensus sequences were produced from the Illumina MiSeq and Ion Torrent S5, respectively. The total hands-on time for processing and sequencing is illustrated in Figure 1. Consensus genomes from the Illumina MiSeq had a mean coverage of 83.4%, with 80.2% having a coverage of 80% and above. Consensus genomes from the

Ion Torrent S5 had a mean coverage of 98.9% with 99.4% of sequences having coverage of 80% and above.

3.4 Genome Sequence Quality Metrics

The consensus sequences generated on Genome detective were evaluated to determine genome coverage and the quantity of assigned mutations. In total, 99.4% of the genomes that were generated on the Ion Torrent S5 passed the genomic coverage quality metrics for GISAID submissions based on KRISP's internal specification compared to 80.2% on the Illumina MiSeq (Table 2). All consensus sequences generated from both platforms had less than 100 mutations quantified for each sequence. Table S6 lists the total mutations for each sample sequenced on the Ion Torrent S5 and the Illumina MiSeq, respectively.

3.5 Genomic and S-Gene Coverage of SARS-CoV-2

Whole-genome assemblies generated on the Ion Torrent S5 showed a generally higher mean genomic coverage than those generated on the Illumina MiSeq (Figure 2). Mean genomic coverages of 99.0% and 83.7% was observed on the Ion Torrent S5 and Illumina MiSeq, respectively. A highly statistically significant difference (Wilcoxon, $p < 0.0001$) was observed in genomic coverage obtained between the two platforms. Genomes generated on the Ion Torrent S5 ranged from 63, 0% to 100.0%, whereas genomes generated on the Illumina MiSeq ranged from 1.9% to 99.5%. Furthermore, the S-gene coverage ranged from 64.2% to 100.0% on the Ion Torrent S5 and 0.9% to 100.0% on the Illumina MiSeq, with average S-gene coverage of 99.0% and 78.7%, respectively. A highly significant difference (Wilcoxon, $p < 0.0001$) in the S-gene coverage was observed across both platforms (Figure 3). Overall, the genomic and the S-gene coverages were consistently higher on the Ion Torrent S5 platform than on the Illumina MiSeq.

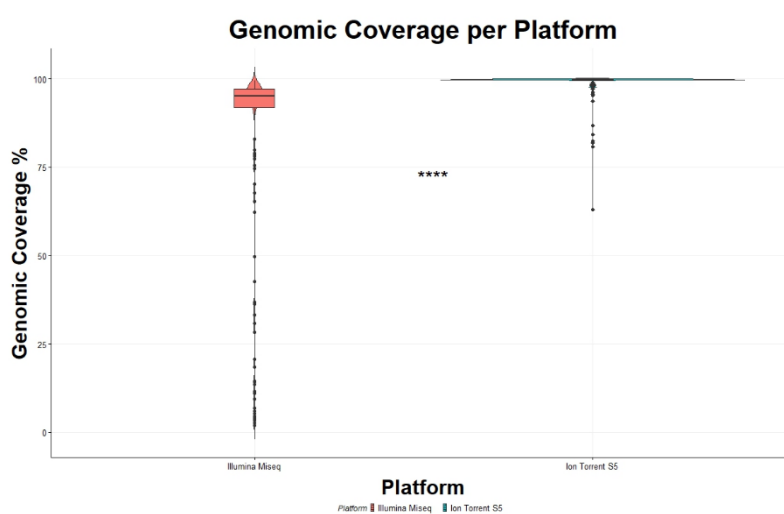


Figure 2 Comparison of the genomic sequence coverage for sequences generated on the Illumina MiSeq and Ion Torrent S5 platforms. Statistical comparison was performed using a Wilcoxon rank sum test. Statistical significance (Wilcoxon rank sum tests) was represented by “*” (****: $p < 0.0001$).

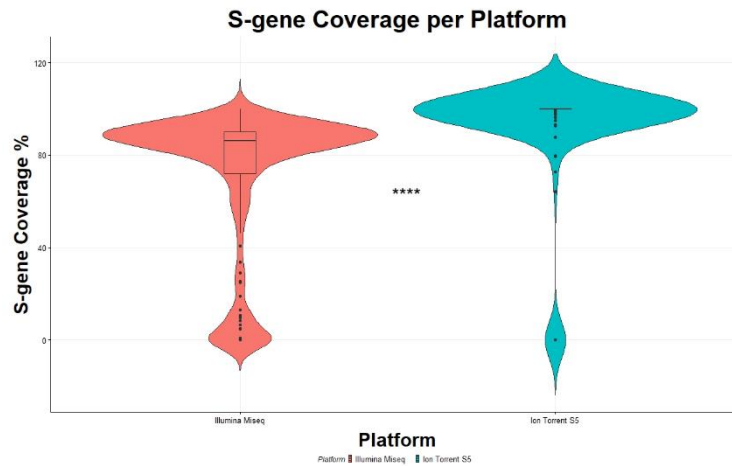


Figure 3 Comparison of the S-gene coverage for sequences generated on the Illumina MiSeq and Ion Torrent S5 platforms. Statistical comparison was performed using a Wilcoxon rank sum test. Statistical significance (Wilcoxon rank sum tests) was represented by “*” (****: $p < 0.0001$).

3.6 Effect of Viral Load on Genomic and S-Gene Coverage of SARS-CoV-2

A Spearman’s ranked sum correlation test was performed to determine the effect of increasing viral load on genomic and S-gene coverage on sequences generated from each sequencing platform (Figure 4 and Figure 5).

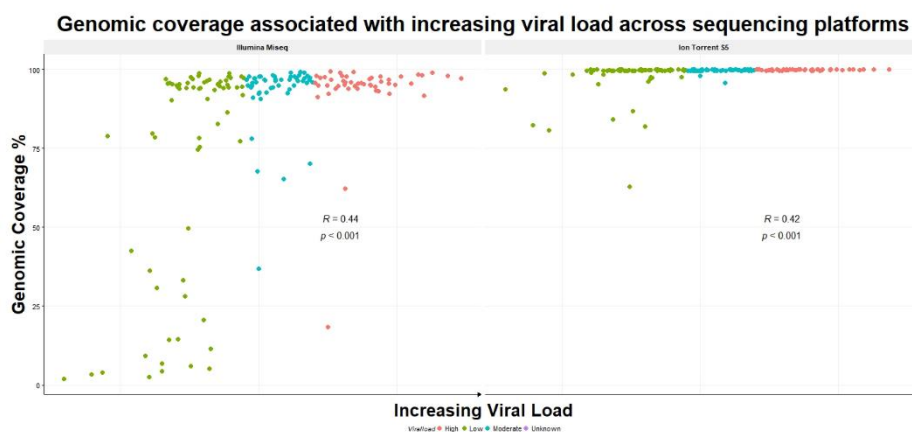


Figure 4 Comparison of the influence on viral load and resulting genome coverage produced on the Illumina MiSeq and Ion Torrent S5 platforms. Illustration of the correlation between SARS-CoV-2 genomic coverage obtained and increasing viral load concentration for samples sequenced on the Illumina MiSeq and Ion Torrent S5. Statistical significance was evaluated using the Spearman’s rank sum correlation test to determine the effect of increasing viral load on the genomic coverage obtained for sequences generated on the Illumina MiSeq and Ion Torrent S5 ($p < 0.001$). Genomic coverage (in percentage) was plotted on the y-axis, and viral load estimates were plotted on the X-axis, highlighting increasing viral load concentration as unknown, low, moderate, and high for each sample sequenced.

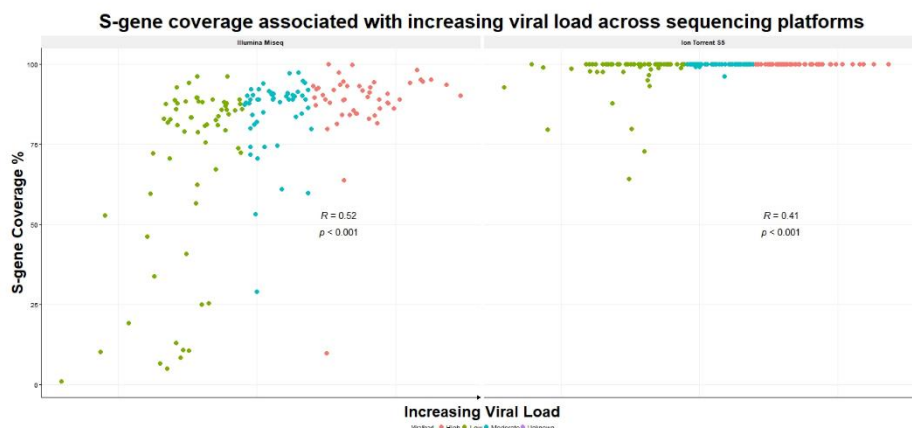


Figure 5 Comparison of the influence on viral load and resulting spike gene coverage produced on the Illumina MiSeq and Ion Torrent S5 platforms. Illustration of the correlation between SARS-CoV-2 spike coverage (S-gene) obtained and increasing viral load concentration for samples sequenced on the Illumina MiSeq and Ion Torrent S5. Statistical significance was evaluated using the Spearman’s rank sum correlation test to determine the effect of increasing viral load on the coverage of the S-gene obtained for sequences generated on the Illumina MiSeq and Ion Torrent S5 ($p < 0.001$). Coverage of the S-gene (in percentage) was plotted on the y-axis, and viral load estimates were plotted on the X-axis highlighting increasing viral load concentration as unknown, low, moderate, and high for each sample sequenced.

Of the 183 samples sequenced on each platform, 180 sequences had corresponding Ct scores available. Estimated viral loads were qualitatively based on mean Ct values provided with available metadata and grouped as per Table 3. Sequences generated from samples with mean Ct values ≤ 25 (high viral loads) accounted for 28.4% of the total consensus genomes, followed by 29.5% with mean Ct > 25 and ≤ 30 (moderate viral loads) and 40.4% with mean Ct values > 30 (low viral load). High statistical significance obtained for sequences generated on the Illumina MiSeq ($p < 0.001$) and Ion Torrent S5 ($p < 0.001$) showed that estimated viral load directly influences genomic coverage. Mean genomic coverages were also tabulated in Table 3 for each of the estimated viral loads assigned. Although consensus sequences generated on the Ion Torrent S5 had higher overall mean genomic coverages compared to the Illumina MiSeq, genomic coverage is known to gradually decline with decreasing viral load. A similar trend was also for S-gene coverage in association with viral load. In addition, several samples sequenced on the Illumina MiSeq resulted in the absence of coverage in the S-gene region (Table S5). These sequences occurred in samples with very low viral load and low template material and, therefore, obtained low genomic coverages.

Table 3 Mean Ct score range for viral load estimation.

Ct score (mean)	Sample No. (n (%))	Estimated Viral Load (Qualitative)	Mean Genomic Coverage (Ion Torrent S5/Illumina MiSeq)
Ct ≤ 25	52 (28.4%)	High	99.9/93.7
25 < Ct ≤ 30	54 (29.5%)	Moderate	99.7/93.0
Ct >30	74 (40.4%)	Low	97.7/71.4

No Ct score is available	3 (1.6%)	Unknown	99.3/39.0
--------------------------	----------	---------	-----------

Table 3 summarises the range of Ct values with estimated viral loads used in this study. There were No Ct values for three samples, so the viral load was not estimated and was listed as unknown. The mean genomic coverage obtained for each viral load estimate is also provided to illustrate the decrease in coverage obtained with decreasing viral load (template material).

3.7 SARS-CoV-2 Clade Assignments

The consensus sequences were uploaded onto NextClade, and the clade assignments were determined and compared between sequences generated on the two platforms. Clade assignments obtained from sequences run on the Ion Torrent S5 and Illumina MiSeq were comparable across the majority of the clades identified. As illustrated in Table 4, Beta and Delta variants were identified in 41.5% and 30.0% of samples sequenced on the Ion Torrent S5 followed by 44.8% and 29.5% on the Illumina MiSeq, respectively. The number of samples that were unsuccessfully sequenced (eight and 11 on the Ion Torrent S5 and Illumina MiSeq, respectively) were not assigned a clade and are detailed in Table 5.

Table 4 Clade assignment summary between sequencing platforms.

Clade	Ion Torrent S5 (n (%))	Illumina MiSeq (n (%))
19A	14 (7.7%)	7 (3.8%)
19B	-	2 (1.1%)
20A	8 (4.4%)	4 (2.2%)
20B	13 (7.1%)	13 (7.1%)
20C	7 (3.8%)	8 (4.4%)
20H (Beta, V2)	76 (41.5%)	82 (44.8%)
20I (Alpha, V1)	2 (1.1%)	2 (1.1%)
21A (Delta)	1 (0.5%)	1 (0.5%)
21J (Delta)	54 (29.5%)	53 (29.0%)
Blanks	8 (4.4%)	11 (6.0%)

Table 5 Samples unsuccessfully sequenced on the Ion Torrent S5 and Illumina MiSeq platforms.

Accession Identifiers	Mean Ct (Viral Load)	Ion Torrent S5		Illumina MiSeq	
		Genomic Coverage (%)	Clade Assignment	Genomic Coverage (%)	Clade Assignment
EPI_ISL_3275349	26.8 (Moderate)	99.8	20H (Beta, V2)	-	-
EPI_ISL_3275351	22.4 (High)	99.8	20B	-	-
EPI_ISL_3275352	24.4 (High)	99.8	21J (Delta)	-	-
EPI_ISL_3275353	33.0 (Low)	99.8	21J (Delta)	-	-

EPI_ISL_3275354	28.2 (Moderate)	99.7	21J (Delta)	-	-
-	33.1 (Low)	99.7	21A (Delta)	-	-
EPI_ISL_3275374	35.7 (Low)	100.0	20H (Beta, V2)	-	-
EPI_ISL_3275376	34.1 (Low)	99.9	20C	-	-
EPI_ISL_3275378	19.2 (High)	100.0	20C	-	-
EPI_ISL_2727188	18.1 (High)	-	-	91.7	20A
EPI_ISL_3275379	33.1 (Low)	100.0	20H (Beta, V2)	-	-
EPI_ISL_2727191	33.7 (Low)	97.1	19A	-	-
EPI_ISL_2727197	22.6 (High)	-	-	95.8	20H (Beta, V2)
EPI_ISL_2727205	35.8 (Low)	-	-	95.4	20H (Beta, V2)
EPI_ISL_2727214	32.3 (Low)	-	-	86.4	20H (Beta, V2)
EPI_ISL_2727215	25.7 (High)	-	-	94.9	20H (Beta, V2)
EPI_ISL_2727222	30.8 (Low)	-	-	95.0	20H (Beta, V2)
-	36.5 (Low)	-	-	14.2	20H (Beta, V2)
EPI_ISL_2727233	15.3 (High)	-	-	97.3	21J (Delta)
Unsuccessful sequences		8		11	

Of the 183 consensus genomes compared, 17 sequences were classified as different clades between the NGS platforms (Table 6). We highlighted the mismatched clades assigned to each sequence generated per NGS platform and the genomic and S-gene coverages obtained for each, followed by the respective amino acid mutations identified. It is evident from the data obtained that genomic coverage influenced the clade classification of the variants sequenced by each platform.

Table 6 Clade calling discrepancy between the Illumina MiSeq and Ion Torrent S5 platforms with their corresponding mutations identified (mutations in bold signify key S-gene mutations specific to the VOC identified).

Sample ID	Coverage (%) per Platform				Clade Calling per Platform		Key Mutations	
	Illumina MiSeq		Ion Torrent S5		Illumina MiSeq	Ion Torrent S5	Illumina MiSeq	Ion Torrent S5
	GC %	S-gene Coverage %	GC %	S-gene Coverage %				
K016691	18.5	9.7	99.8	100.0	19A	20H (Beta, V2)	ORF7b:T40I	E:P71L, N:P13S, N:T205I, ORF1a:T265I, ORF1a:K1655N, RF1a:K3353R, ORF1b:P314L, ORF1b:A1057S, ORF3a:Q57H, ORF3a:S171L, S:L18F, S:D80A, S:D138Y, S:D215G, S:R246G, S: K417N , S: E484K , S: N501Y , S:D614G, S:A701V
K016709	78.1	71.7	99.7	100.0	20H (Beta, V2)	20A	E:P71L, N:T205I, ORF1a:T265I, ORF1a:G507R, ORF1a:P2046L, ORF1a:N2596S, ORF1a:M2796C, ORF1a:K3353R,	<i>ORF1b:P314L</i> , S:D80A, S: N501Y , S:D614G

								<i>ORF1b:P314L,</i> <i>ORF1b:I1074V,</i> <i>ORF3a:S171L,</i> <i>S:L18F, S:D80A,</i> <i>S:D215G, S:K417N,</i> <i>S:E484K, S:N501Y,</i> <i>S:D614G, S:A701V,</i> <i>S:S940F, S:A1087S</i>	
K016715	11.0	4.8	99.7	100.0	20C	20A	<i>ORF1a:T265I,</i> <i>ORF1a:K3353R,</i> <i>ORF1a:D4335Y,</i> <i>ORF1b:P314L</i>	S:N501Y, S:D614G	
K016754	36.3	46.2	99.0	97.7	21J (Delta)	19A	<i>M:I82T, N:D63G,</i> <i>ORF1b:E513*,</i> <i>ORF7b:T40I,</i> <i>ORF9b:T60A,</i> <i>S:T19R, S:L452R,</i> <i>S:T478K, S:D614G,</i> <i>S:D950N</i>	<i>M:I82T, N:D63G,</i> <i>N:D377Y,</i> <i>ORF1a:T3255I,</i> <i>ORF1b:G662S,</i> <i>ORF1b:A1918V,</i> <i>ORF7a:V82A,</i> <i>ORF7b:T40I,</i> <i>ORF9b:T60A,</i> S:L452R, S:D614G, S:P681R	
K016760	28.2	10.8	99.7	100.0	19A	20A	ORF1a:K3353R	<i>M:I82T, N:D63G,</i> <i>N:G215C,</i> <i>ORF1a:A1306S,</i> <i>ORF1a:P2046L,</i> <i>ORF1a:P2287S,</i> <i>ORF1a:T3255I,</i>	

								ORF1b:P314L, ORF1b:G662S, ORF1b:P1000L, ORF1b:A1918V, ORF3a:S26L, ORF7a:V82A, ORF9b:T60A, S:T19R, S:T478K, S:N501Y, S:D614G
K016870	11.5	25.4	97.5	98.3	20H (Beta, V2)	19A	ORF1a:K1197N, ORF1a:T1638I, ORF1a:P1640S, ORF1a:K1655N, S:K417N	ORF3a:S26L, ORF3a:Q57H, ORF3a:S171L, S:D614G, S:A701V
<i>K016899</i>	13.5	0.0	98.1	98.6	20C	19A	ORF1b:R1315C	E:P71L, S:A701V
K016902	42.7	19.1	98.4	98.6	20H (Beta, V2)	19A	N:G60F, N:K61P, N:E62S, N:D63S, N:K65P, N:R68L, N:G69L, N:I74L, N:D81Y, N:D82Y, <i>ORF1a:T265I</i> , ORF1a:T2154I, ORF1a:N2767H, ORF1a:K3353R, ORF1b:V22I, ORF1b:C44S, ORF1b:L271I, ORF9b:E65S,	<i>ORF1a:T265I</i> ; S:D614G

								ORF9b:D66Y, ORF9b:Q70H, ORF9b:Q77H, ORF9b:M78I, S:A701V	
K016908	30.8	33.7	95.4	97.6	20B	20A	N:A152S, N:R203K, N:G204R, N:N213I, ORF1a:T395I, ORF9b:R32L, S:D614G	ORF1b:S1779I, S:N450K, S:D614G, S:P681R	
K016917	77.3	73.7	97.7	98.8	20H (Beta, V2)	19A	E:P71L, N:D128Y, N:T205I, N:Y268N, ORF1a:K1655N, ORF1a:K3353R, ORF1b:S1182L, ORF3a:W131L, ORF3a:S171L, ORF8:D63N, S:L18F, S:D215G, S:K417N, S:E484K, S:N501Y, S:D614G	N:D128Y, ORF1a:K3353R, ORF3a:Q57H, S:L18F, S:K417N, S:D614G, S:A701V	
K016920	33.1	8.3	99.7	97.6	20C	19A	ORF1a:T265I, ORF1a:T2087S, ORF1a:K3353R, S:D614G	N:T205I, S:D614G	
K016923	5.9	10.6	86.8	79.7	20A	20H (Beta, V2)	S:K417N	E:P71L, M:I82T, N:T135I, N:T205I, ORF1a:T265I,	

									ORF1a:I547F, ORF1a:K1655N, ORF1b:P314L, ORF1b:L1698F, ORF3a:Q57H, ORF3a:G100S, ORF3a:S171L, S:T19A, S:L24F, S:P25T, S:K182N, S:D215G
K016926	3.9	10.2	98.9	99.0	19B	20A	<i>S:D614G</i>		M:I82T, ORF1a:P2046L, ORF1a:S2048F, ORF1a:T3255I, ORF1b:G662S, ORF1b:P1000L, ORF1b:A1918V, <i>S:D614G</i>
<i>K016931</i>	9.3	0.0	99.7	100.0	21A (Delta)	20H (Beta, V2)	M:I82T, N:D63G, ORF9b:T60A		E:P71L; N:T205I, ORF1a:T265I, ORF1a:K1655N, ORF1a:K3353R, ORF3a:Q57H, ORF3a:S171L, S:K417N, S:E484K, S:N501Y , S:D614G, S:A701V
K016940	4.4	6.6	98.5	97.6	20H (Beta, V2)	19A			ORF3a:Q57H,

								RF3a:S171L, S:D614G, S:A701V
K016943	5.2	0.0	97.7	93.1	19B	20A	ORF1a:A540T, ORF1a:K1655N	S:D614G
K016945	79.9	59.6	99.9	100.0	20H (Beta, V2)	19A	E:P71L, N:T205I, N:T271I, ORF1a:K1655N, ORF1a:K3353R, ORF1b:P314L, ORF1b:A941S, ORF1b:G1129V, ORF3a:G100S, ORF3a:S171L, S:D215G, S:K417N , S:D614G, S:A701V	S:D614G
Clade Summary per Platform (total = 17)							20H Beta V2 (n = 6), 20A (n = 1), 20B (n = 1), 20C (n = 3), 19A (n = 2), 19B (n = 2) 21J Delta (n = 2)	20H Beta V2 (n = 3); 20A (n = 6); 19A (n = 8)

Table 6 highlights the 17 samples sequenced on the Illumina MiSeq and Ion Torrent S5 and were classified as different clades by Nexclade online tool. Clades identified by the Illumina MiSeq include 20H Beta V2 (n = 6), 20A (n = 1), 20B (n = 1), 20C (n = 3), 19A (n = 2), 19B (n = 2), and 21J Delta (n = 2). Clades identified by the Ion Torrent S5 include 20H Beta V2 (n = 3); 20A (n = 6); and 19A (n = 8)). An overall higher genomic coverage was observed for sequences generated on the Ion Torrent S5 than on the Illumina MiSeq. Amino acid mutations identified by both platforms for a given sample are highlighted in italics, and S-gene specific key mutations for the respective clade assignment are listed in bold for each sample sequenced per platform. Furthermore, it was observed that a few sequences generated on the Illumina MiSeq obtained no coverage for the S-gene region of the SARS-CoV-2 genome.

The high coverage obtained for sequences on the Ion Torrent S5 contributed to a reliable clade assignment of variants to Illumina MiSeq data. Specific key mutations for each mismatched clade assignment were also examined with a focus on specific critical mutations in the spike region. Not all specific key mutations were identified for each assigned clade. We also observed inconsistency in mutations identified in sequences between the platforms. That is, each sequence identified had a different set of mutations called. This discrepancy may be attributed to several sequencing factors, such as the primers used, the number of reads obtained, and platform-specific processing.

3.8 Quantification of Mutations (Insertions, Deletions, and Substitutions) for Sequences Generated on the Illumina MiSeq and Ion Torrent S5

The number of mutations detected for each sample was individually compared in order to establish if there was a significant difference in the assignment of mutations. These analyses included the total substitutions, insertions, and deletions detected by the Ion Torrent S5 and Illumina MiSeq platforms (Figure 6). In total, we analysed 347 consensus sequences. There was a highly significant difference in total mutations observed across the Ion Torrent S5 and Illumina MiSeq (Wilcoxon, $p < 0.0001$), with a greater number of mutations detected by the Ion Torrent S5 (6-94 mutations, total: 8116) than the Illumina MiSeq (1-92 mutations, total: 7036). A marked significant difference was also noted for the number of substitutions (Wilcoxon, $p < 0.05$) and deletions (Wilcoxon, $p < 0.0001$) identified by both platforms; however, no significant difference was observed for insertions (Wilcoxon, $p = 0.25$ ns) across both platforms. The variation in the number of mutations detected across the sequencing platforms are listed in Table 2.



Figure 6 Mutational analysis of all sequences from runs A and B generated on the Illumina MiSeq and Ion Torrent S5 platforms. Consensus sequences were produced using Genome Detective and uploaded to NextClade for quantification of mutations. A comparison in the number and type of mutations detected by each platform was performed using a Wilcoxon rank sum test and statistical significance was represented by “*” (ns: non-significant, *: $p < 0.05$, ****: $p < 0.0001$). The total number of mutations (A), substitutions (B), insertions (C), and deletions (D) on the Illumina MiSeq and Ion Torrent S5 platforms for the combined two runs as illustrated above.

4. Discussion

The Ion Torrent S5 and Illumina MiSeq provide alternative methods for researchers to study SARS-CoV-2 at a genomic level [30]. This study compared the performance and data generated by the two WGS platforms. We hypothesized the sequencing methodologies that were used for the genomic surveillance of SARS-CoV-2 in a high throughput laboratory setting, and we generated sequencing data that differed when analyzed with the same analysis pipeline. In a brief overview of the data generated for genomic coverage, clade assignments, and quantification of mutations, we concluded that the platforms were similar in sequencing capabilities but differed in sequencing data outcomes. Our findings indicate that the Ion Torrent S5 produced sequences with higher genomic coverage over a broader range of viral loads in a shorter time than the Illumina MiSeq. These findings agreed with previous comparison studies [25, 26, 30, 35].

Regarding the sequencing process, the Ion Torrent S5 and Illumina MiSeq followed a streamlined process flow, allowing the platforms to display their adaptability in the WGS of SARS-CoV-2. The sequencing runtime for a sample set 96 on the Illumina MiSeq was 36 hours, whereas, on the Ion Torrent S5, it was seven hours using two Ion 540 sequencing chips. The difference in sequencing time allows more samples to be sequenced on the Ion Torrent S5 in 36 hours than on the Illumina

MiSeq. Although the sequencing duration is much shorter on the Ion Torrent S5, it is important to note that the remaining processes, such as amplification and library preparation, consume a shorter duration on the Illumina MiSeq. The automated process for templating the manually prepared libraries onto the Ion Torrent sequencing chip is approximately 15.5 hours whereas the amplification and tagmentation step utilize less than 10 hours. These findings confirmed previous observations that found similarities in processing times with each respective workflow [30]. With limited hands-on time, full automation with the Ion chef allows for faster turnaround times but limits sample numbers that can be processed at once (eight libraries per seven and a half hours on the Ion Chef) on the Ion Torrent S5. Thus making the Ion Torrents selling point of full automation its major downfall in a high-throughput laboratory setting. However, considering the manual route for library preparation, the Ion Torrent S5 is similar to the Illumina MiSeq in handling larger sample numbers provided that manual library preparation is followed. The Illumina MiSeq workflow can also be automated by including an external liquid handler to reduce hands-on time and overall processing duration. In addition, reagents used in the upstream preparation processes on the Illumina MiSeq can be further optimised and validated to accommodate greater sample numbers with reduced reagent volumes by miniaturisation of process workflow [36]. In contrast, the use of full automation on the Ion Torrent S5 coupled with the Ion chef allows for limited handling of small sample numbers. It increases the overall turnaround time in a day-shift laboratory site. It is therefore, feasible to incorporate a manual library preparation for such platforms to minimize turnaround times and increase sample numbers processed in a high throughput laboratory setting as illustrated in this study.

The same remnant sample set of 183 was used to limit variability between samples and to compare the data generated during analysis from the two platforms. The sequence process directly impacted the sequencing outcomes with genomic coverage and sequence quality metrics. Sequence quality was based on in-house quality control specifications established at KRISP based on previous data for GISAID submissions. These included sequences with more than 80% genomic coverage and less than 100 mutations. Although the Ion Torrent S5 and Illumina MiSeq are both capable of producing complete SARS-CoV-2 genomes, sequences generated on the Ion Torrent S5 maintained an overall higher mean genomic coverage in comparison to sequences generated on the Illumina MiSeq. Various factors contribute to the genomic coverage obtained from both platforms. Ct values are semi-quantitative numbers that generally categorise the concentration of viral RNA in a given sample following qPCR testing. An inverse correlation was observed between viral load and Ct values. Low Ct scores are associated with high viral loads and influence sample quality and overall sequence quality [37, 38]. Echoing previous findings, we observed an association of viral load on genomic coverage for all sequences generated. Moderate to low viral load samples sequenced on the Ion Torrent S5 resulted in an overall good mean genomic coverage (>60%), higher success rates, and increased test eligibility. These findings imply that the Ion Torrent S5 sequencing capabilities are less likely to be affected by sample Ct values and, therefore, can be employed in sequencing samples during the early stages of infection when viral load is lower. However, further investigation may be required to assess this using a larger sample cohort across various laboratories. In contrast, the Illumina MiSeq relied on samples with higher viral load for better coverage of genomes as observed in other studies [20-22].

Additionally, the increase in genomic coverage obtained from the Ion Torrent S5 may be attributed to the greater number of reads obtained per sample using two Ion 540 chips in a

sequencing run of 96 samples [26]. As highlighted in Table 1, the AmpliSeq Research Panel on Ion Torrent S5 produces over twice the number of reads of the Illumina MiSeq, and half the size of fragments sequenced compared to the Illumina MiSeq (200 bp versus 400 bp) [39]. In essence, the number of reads achievable for the samples sequenced in this dataset would be at least double on the Ion Torrent S5 to obtain a coverage that is higher or equal to the Illumina MiSeq. It is possible that the greater number of reads obtained per sample could also have contributed to the greater coverage and reliability in clade assignments observed in sequences generated on the Ion Torrent S5.

Furthermore, the Ion Torrent S5 detected a significantly greater number of total mutations (insertions, substitutions, and deletions) than the Illumina MiSeq. Previous studies have reported that unlike the Illumina MiSeq, semiconductor sequencing platforms like the Ion Torrent S5 are known to produce a predominated homopolymer-associated base-call error utilizing INDELS [26, 40, 41]. Interestingly, these were often deletions instead of insertions, similar to the findings of this study, which may have contributed to the larger number of total mutations from Ion Torrent S5 sequences. According to Marine *et al.*, 2020, while such INDELS may be adjusted and corrected for well characterised viruses, this may not be the case when characterizing novel viruses [26]. This further validates the need for in-depth quality control parameters during the analysis of such sequences.

We eliminate the variability between other assembly methods by using Genome Detective as the prime assembly method for generating consensus genomes for both platforms. Additionally, the advantage of using NextClade is that it allows the user to determine the difference in the quality of the consensus sequences generated, classify clades accordingly, and establish similarity in identifying evolutionary changes between sequences from each platform [34]. In contrast to our findings above, highlighting higher genomic coverage obtained for sequences generated on the Ion Torrent S5, we find most of these sequences to be grouped as lower quality on NextClade compared to those generated on the Illumina MiSeq (data not shown). This, however, can be attributed to sequencing errors or miscalled bases generated on the Ion Torrent systems as previously observed [8, 26, 41, 42]. Furthermore, other findings indicate that the Ion Torrent S5 and Illumina MiSeq sequences can easily differentiate between the Beta and Delta VOCs based on mutation calling and respective clade assignment. NextClade assigned the same clades for 147/183 (80.3%) samples during the early delta-replacing-beta phase observed in the Eastern Cape, South Africa. A mismatch in clade assignment was observed in 17/183 (9.3%) samples successfully sequenced on both platforms followed by a low failure rate of 4.4% and 6.0% on the Ion Torrent S5 and Illumina MiSeq, respectively. It would be interesting to expand the clade classification and sequencing across a larger cohort of known VOCs using both platforms to assess the reliability and accuracy of clade assignment of NextClade. The Pangolin lineage assignment tool is an alternative software for lineage classification; however, it was not included in this study [43]. Furthermore, we did not perform additional testing on samples that failed sequencing, nor did we performed concordance studies to evaluate the agreement of our findings with outcomes obtained from alternative PCR-based testing methods for this particular collection of samples.

A major limitation and consideration for genomic surveillance laboratories are the use of updated primer sets. It is important to note that unlike the Illumina MiSeq Artic V3 primers, which were found to be problematic with novel variants such as Delta, the Ion Torrent primers (AmpliSeq Primers) covered 99% of the SARS-CoV-2 genome, including all serotypes, therefore attributing to

higher genomic and S-gene coverage [44-46]. Due to the consistent evolution of SARS-CoV-2, difficulty in mutational regions arose, resulting in poor coverage of some ARTIC V3 primers that were located in areas with key Delta mutations [45]. Since the initial primers were designed and based on the reference SARS-CoV-2 genome sequence, it was expected that there would be difficulty in identifying large structural variants. As a result, systematic limitations were observed in the presence of high levels of genomic variation. Subsequently, several S-gene target failures (SGTF) were observed during diagnostic qPCR testing for CoVID-19 [47-49]. A decreased coverage of specific regions within the SARS-CoV-2 genome was also observed, and novel variants emerged since the beginning of the pandemic [48]. Low coverage sequences generated on the Illumina MiSeq may have contributed to the discrepancy in clade assignment during the initial surge of the Delta variant. It was previously reported that the G142D amino acid substitution was substantially underrepresented among early Delta variant genomes identified [45]. Furthermore, Kuchinski *et al.*, 2022, reported a disruption in genomic sequencing of SARS-CoV-2 as a result of emerging mutations identified in novel variants [46]. Since the ARTIC primer set is one of the most widely used SARS-CoV-2 sequencing primers, the V3 primers were updated to address the amplicon drop-off observed among the Delta VOC, resulting in version 4 being released in June 2021. Unfortunately, V4 primers were not used during the execution of the study, as they were not procured at the time. Lambisia *et al.*, 2022, subsequently conducted a study to assess the impact of the updated V4 ARTIC primers on genome recovery using the ONT and concluded a great improvement in the recovery of the Delta variant amongst others [50]. The Ion Torrent sequencing panel was also updated to accommodate the amplicon drop-off in novel variants. The updated panel, Ion AmpliSeq SARS-CoV-2 Insight Research Assay, was designed to improve the coverage and uniformity of the previous Ion AmpliSeq SARS-CoV-2 Research Panel used in this study. Therefore, continuous improvement of current primers irrespective of kit specifications is an essential requirement of an effective genomic surveillance regime.

Plitnick *et al.*, 2021, directly compared the performance of the SARS-CoV-2 AmpliSeq Research Panel to the results obtained with the Illumina MiSeq-based ARTIC Nextflow analysis pipeline [30]. Post-bioinformatic analysis of data from such studies showed that both methods produced similar levels of coverage (>98%) across a broad range of viral loads (Ct values of 15.56 to 32.54 [median, 22.18]) and that both approaches sequenced SARS-CoV-2 effectively [30, 35]. Although the bioinformatic analysis pipelines used in this study differ, the findings of our study are similar to those documented in the above research by Plitnick *et al.*, 2021. Standardisation of analysis regimes accommodates the comparison of data from different NGS technologies without bias from independent assembly and analysis tools. Assembly software affects the overall genomic coverage of sequences obtained from various platforms. There is an additional need for quality control processes to improve the overall quality of such sequences made publicly available as recommended by Jacot *et al.*, 2021 for diagnostic purposes [51]. Such achievements have included the removal of frame shifts and unknown stop codons in some instances. In this study, we observed sequences generated on the Illumina MiSeq to be simpler to process, with quality control easily implemented across such sequences yielding sequences of better quality as per NextClade analysis. Although sequencing capabilities are similar on both platforms, higher genomic coverage of sequences was generated on the Ion Torrent S5. However, most of these sequences were of lower quality as per analysis on NextClade. It is, therefore, important to note that standardising assembly and analysis software allows for improved comparison of the data generated by the different

platforms and data analysed by different software. Nonetheless, this study complements previous research investigating the efficacy of Ion Torrent and Illumina platforms for sequencing viral pathogens [23, 24, 52].

In light of the preceding discoveries, it is of utmost significance to address that although samples were collected from mid-March to early June 2021, sequencing commenced only in mid-June 2021 upon receiving these samples. This observation underscores a significant challenge in the real-time genomic surveillance monitoring, given that the Delta wave in South Africa was officially reported from 17 May 2021 to 14 November 2021 [53]. Nonetheless, the surveillance samples collected in this study on 29 March 2021 revealed an earlier presence of the Delta variant in the Eastern Cape region. It is, therefore, necessary to acknowledge that these circumstances concerning the delayed initiation of sequencing for novel emerging variants can exert a direct influence on the transmission dynamics within communities and provinces. The implications of such challenges may hinder the timely reporting and monitoring of emergent variants, potentially leading to delays in implementing appropriate public health measures to mitigate their spread and impact [18, 54].

This study provides insight into the requirements and challenges of employing different methods for genomic surveillance in a high-throughput research laboratory, including the obstacles faced within such a program. Users should exercise caution when utilizing available sequences publicly, considering the technologies, assembly, analysis processes, and the quality of sequences from various samples, particularly when comparing data across different platforms. It is critically important for users to regularly update and manage the primers tailored to each technology, especially in response to the emergence of new variants. Such practices are vital for accurately tracking current variants and quickly identifying new ones, thereby enabling precise and timely genomic surveillance.

5. Conclusions

Genomic monitoring plays a crucial role in mitigating the spread of SARS-CoV-2, significantly aiding in the rapid detection of new mutations within the Delta VOC [16, 17]. Both the Ion Torrent S5 and Illumina MiSeq platforms could accurately distinguish between the Beta and Delta VOCs. Notably, the Ion Torrent S5 demonstrated superior performance in processing samples with lower viral concentrations (indicated by higher Ct values) compared to the Illumina MiSeq, though further analysis is warranted. Regarding the capabilities of these sequencers, including genomic coverage, the production of high-quality sequences, and the overall data output, both the Illumina MiSeq and Ion Torrent S5 were deemed suitable for whole-genome sequencing (WGS) of SARS-CoV-2. Nonetheless, distinctions in data quality exist between the two. Consequently, this study's findings highlight that, despite their strengths and weaknesses, the Ion Torrent S5 and Illumina MiSeq, in conjunction with analytical tools like Genome Detective and NextClade, stood as dependable options for SARS-CoV-2 genomic surveillance.

Acknowledgments

The authors would like to acknowledge the Next Generation Sequencing Network of South Africa (NGS-SA) which provided the WGS information used in this study as well as the non-financial support received from the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE).

Author Contributions

Conceptualization: Upasana Ramphal, Eduan Wilkinson, Jennifer Giandhari; Methodology: Upasana Ramphal, Carel van Heerden, Sureshnee Pillay, Jennifer Giandhari, Oluwakemi Laguda-Akingba; Formal Analysis: Upasana Ramphal, Derek Tshiabuila, Stephanie van Wyk; Resources: Upasana Ramphal, Jennifer Giandhari, Sureshnee Pillay; Data Curation: Upasana Ramphal, Yajna Ramphal, Derek Tshiabuila; Writing - original draft: Upasana Ramphal; Writing - review & editing: Upasana Ramphal, Yajna Ramphal, Jennifer Giandhari, Richard Lessells, Stephanie van Wyk, Carel v Heerden, Cheryl Baxter, Oluwakemi Laguda-Akingba; Visualization: Upasana Ramphal, Yajna Ramphal, Derek Tshiabuila; Supervision: Jennifer Giandhari, Richard Lessells, Cheryl Baxter, Tulio de Oliveira; Project administration: Tulio de Oliveira; Funding acquisition: Eduan Willkinson, Tulio de Oliveira; Final approval of manuscript: All authors.

Funding

The South African Department of Science and Innovation (DSI), and Department of Technology and Innovation as part of the Network for Genomic Surveillance in South Africa (NGS-SA) funded the study. The Strategic Health Innovation Partnerships Unit of the South African Medical Research Council supported this research reported in this publication. The Genomics Surveillance in South Africa was supported in part through National Institutes of Health USA grant U01 AI151698 for the United World Antiviral Research Network (UWARN) and by the Rockefeller Foundation. KRISP have received donations from Chan Soon-Shiong Family Foundation (CSSFF) and Illumina. SANTHE, is a DELTAS Africa Initiative [grant# DEL-15-006], and is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant # 107752/Z/15/Z] and the government of the United Kingdom (UK). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. The contents are purely the responsibilities of the authors and did not represent and reflect the view of the funder or those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Competing Interests

TdO received fees from Illumina as a member of the Infectious Diseases Testing Advisory Board. The remaining authors have declared that no competing interests exist.

Data Availability Statement

All the raw sequence data generated for this research are available publicly in the NCBI Sequence Read Archive (project no. PRJNA636748). Epidemics data is freely available on the open-source Global Initiative on Sharing All Influenza Data (GISAID) database and accessed using the GISAID Identifier: EPI_SET_230203ym; doi: <https://doi.org/10.55876/gis8.230203ym>.

Additional Materials

The following additional materials are uploaded at the page of this paper.

1. File S1: gisaid_supplemental_table_epi_set_230203ym.
2. File S2: Data Summary.
3. Table S1: Summary of sample set used for comparison of NGS platforms.
4. Table S2: GISAID Accession identifiers (EPI_SET ID: EPI_SET_230203ym, doi: 10.55876/gis8.230203ym).
5. Table S3: Sequencing data for runs performed on the Illumina MiSeq.
6. Table S4: Sequencing data for runs performed on the Ion Torrent S5.
7. Table S5: Absence of S-gene coverage on the Illumina MiSeq due to low viral load.
8. Table S6: Total mutations obtained between samples sequenced on the Ion torrent S5 and Illumina MiSeq platforms.

References

1. Konings F, Perkins MD, Kuhn JH, Pallen MJ, Alm EJ, Archer BN, et al. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. *Nat Microbiol.* 2021; 6: 821-823.
2. Tay JH, Porter AF, Wirth W, Duchene S. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. *Mol Biol Evol.* 2022; 39: msac013.
3. World Health Organisation. Historical working definitions and primary actions for SARS-CoV-2 variants [Internet]. Geneva, Switzerland: World Health Organisation; 2023. Available from: <https://www.who.int/publications/m/item/historical-working-definitions-and-primary-actions-for-sars-cov-2-variants>.
4. Centers for Disease Control and Prevention (CDC). Classifications & Definitions [Internet]. Atlanta, GA: Centers for Disease Control and Prevention (CDC); 2023. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html#print>.
5. Worldometer. COVID-19 coronavirus pandemic [Internet]. Dover, DE: Worldometer; 2024. Available from: <https://www.worldometers.info/coronavirus/>.
6. Babb de Villiers C, Blackburn L, Cook S, Janus J, Johnson E, Kroese M. Next generation sequencing for SARS-CoV-2 [Internet]. Berlin, Germany: ResearchGate; 2021. Available from: https://www.researchgate.net/publication/351688917_Next_generation_sequencing_for_SARS-CoV-2/link/60a4d25092851ccc66b85e18/download?tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIn19.
7. Faria NR, Azevedo RD, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. *Science.* 2016; 352: 345-349.
8. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature.* 2016; 530: 228-232.
9. Butera Y, Mukantwari E, Artesi M, Umuringa JD, O'Toole AN, Hill V, et al. Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming travelers on lineage diversity. *Nat Commun.* 2021; 12: 5705.
10. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat Med.* 2021; 27: 440-446.

11. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020. doi: 10.1101/2020.12.21.20248640.
12. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*. 2018; 23: 130-137.
13. Giandhari J, Pillay S, Wilkinson E, Tegally H, Sinayskiy I, Schuld M, et al. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. *Int J Infect Dis*. 2021; 103: 234-241.
14. Khan K, Karim F, Cele S, Reedoy K, San JE, Lustig G, et al. Omicron infection enhances delta antibody immunity in vaccinated persons. *Nature*. 2022; 607: 356-359.
15. Engelbrecht S, Delaney K, Kleinhans B, Wilkinson E, Tegally H, Stander T, et al. Multiple early introductions of sars-cov-2 to cape town, South Africa. *Viruses*. 2021; 13: 526.
16. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021; 592: 438-443.
17. Wilkinson E, Giovanetti M, Tegally H, San JE, Lessells R, Cuadros D, et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science*. 2021; 374: 423-431.
18. Msomi N, Mlisana K, de Oliveira T, Willianson C, Bhiman JN, Goedhals D, et al. A genomics network established to respond rapidly to public health threats in South Africa. *Lancet Microbe*. 2020; 1: e229-e230.
19. GISAID. In Focus [Internet]. Munich, Germany: GISAID; 2024. Available from: <https://www.gisaid.org/>.
20. Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evol*. 2020; 6: veaa075.
21. Pillay S, Giandhari J, Tegally H, Wilkinson E, Chimukangara B, Lessells R, et al. Whole genome sequencing of SARS-CoV-2: Adapting Illumina protocols for quick and accurate outbreak investigation during a pandemic. *Genes*. 2020; 11: 949.
22. Tshiabuila D, Giandhari J, Pillay S, Ramphal U, Ramphal Y, Maharaj A, et al. Comparison of SARS-CoV-2 sequencing using the ONT GridION and the Illumina MiSeq. *BMC Genomics*. 2022; 23: 319.
23. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: Comparison of ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*. 2012; 13: 341.
24. Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al. Performance comparison of illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. *Appl Environ Microbiol*. 2014; 80: 7583-7591.
25. Lahens NF, Ricciotti E, Smirnova O, Toorens E, Kim EJ, Baruzzo G, et al. A comparison of illumina and ion torrent sequencing platforms in the context of differential gene expression. *BMC Genomics*. 2017; 18: 602.
26. Marine RL, Magaña LC, Castro CJ, Zhao K, Montmayeur AM, Schmidt A, et al. Comparison of Illumina MiSeq and the ion torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses. *J Virol Methods*. 2020; 280: 113865.

27. Merriman B, Ion Torrent R&D Team, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*. 2012; 33: 3397-3417.
28. Thermo Fisher Scientific. Ion GeneStudio™ S5 system [Internet]. Shanghai, China: Thermo Fisher Scientific; 2024. Available from: <https://www.thermofisher.com/order/catalog/product/A38194>.
29. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012; 2012: 251364.
30. Plitnick J, Griesemer S, Lasek-Nesselquist E, Singh N, Lamson DM, George KS. Whole-genome sequencing of sars-cov-2: Assessment of the ion torrent ampliseq panel and comparison with the illumina miseq artic protocol. *J Clin Microbiol*. 2021; 59. doi: 10.1128/jcm.00649-21.
31. Liu J, Chen X, Liu Y, Lin J, Shen J, Zhang H, et al. Characterization of SARS-CoV-2 worldwide transmission based on evolutionary dynamics and specific viral mutations in the spike protein. *Infect Dis Poverty*. 2021; 10: 112.
32. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017; 12: 1261-1276.
33. Genome Detective. Genome detective virus tool [Internet]. Herent, Belgium: Genome Detective; 2024. Available from: <https://www.genomedetective.com/app/typingtool/virus/>.
34. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: Clade assignment, mutation calling and quality control for viral genomes. *Zenodo*. 2021. doi: 10.5281/zenodo.5607694.
35. Rachiglio AM, De Sabato L, Roma C, Cennamo M, Fiorenza M, Terracciano D, et al. SARS-CoV-2 complete genome sequencing from the Italian Campania region using a highly automated next generation sequencing system. *J Transl Med*. 2021; 19: 246.
36. Pillay S, San JE, Tshiabuila D, Naidoo Y, Pillay Y, Maharaj A, et al. Evaluation of miniaturized Illumina DNA preparation protocols for SARS-CoV-2 whole genome sequencing. *PLoS One*. 2023; 18: e0283219.
37. Rabaan AA, Tirupathi R, Sule AA, Aldali J, Mutair AA, Alhumaid S, et al. Viral dynamics and real-time RT-PCR Ct values correlation with disease severity in COVID-19. *Diagnostics*. 2021; 11: 1091.
38. Zuckerman NS, Bucris E, Erster O, Mandelboim M, Adler A, Burstein S, et al. Prolonged detection of complete viral genomes demonstrated by SARS-CoV-2 sequencing of serial respiratory specimens. *PLoS One*. 2021; 16: e0255691.
39. Thermo Fisher Scientific. Target selection for next-generation sequencing workflows [Internet]. Shanghai, China: Thermo Fisher Scientific; 2024. Available from: <https://www.thermofisher.cn/cn/zh/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets.html>.
40. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 2012; 30: 434-439.
41. Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief Bioinform*. 2016; 17: 154-179.
42. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. Shining a light on dark sequencing: Characterising errors in ion torrent PGM data. *PLoS Comput Biol*. 2013; 9: e1003031.

43. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 2020; 5: 1403-1407.
44. Alessandrini F, Caucci S, Onofri V, Melchionda F, Tagliabracci A, Bagnarelli P, et al. Evaluation of the ion AmpliSeq SARS-CoV-2 research panel by massive parallel sequencing. *Genes.* 2020; 11: 929.
45. Davis JJ, Long SW, Christensen PA, Olsen RJ, Olson R, Shukla M, et al. 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. *Microbiol Spectr.* 2021; 9: e01803-e01821.
46. Kuchinski KS, Nguyen J, Lee TD, Hickman R, Jassem AN, Hoang LM, et al. Mutations in emerging variant of concern lineages disrupt genomic sequencing of SARS-CoV-2 clinical specimens. *Int J Infect Dis.* 2022; 114: 51-54.
47. Wolter N, Jassat W, Walaza S, Welch R, Moultrie H, Groome M, et al. Early assessment of the clinical severity of the SARS-CoV-2 omicron variant in South Africa: A data linkage study. *Lancet.* 2022; 399: 437-446.
48. Vogels CB, Breban MI, Ott IM, Alpert T, Petrone ME, Watkins AE, et al. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol.* 2021; 19: e3001236.
49. Challen R, Dyson L, Overton CE, Guzman-Rincon LM, Hill EM, Stage HB, et al. Early epidemiological signatures of novel SARS-CoV-2 variants: Establishment of B.1.617.2 in England. *medRxiv.* 2021. doi: 10.1101/2021.06.05.21258365.
50. Lambisia AW, Mohammed KS, Makori TO, Ndwiga L, Mburu MW, Morobe JM, et al. Optimization of the SARS-CoV-2 ARTIC network V4 primers and whole genome sequencing protocol. *Front Med.* 2022; 9: 836728.
51. Jacot D, Pillonel T, Greub G, Bertelli C. Assessment of SARS-CoV-2 genome sequencing: Quality criteria and low-frequency variants. *J Clin Microbiol.* 2021; 59. doi: 10.1128/jcm.00944-21.
52. Szargut M, Cytacka S, Serwin K, Urbańska A, Gastineau R, Parczewski M, et al. SARS-CoV-2 whole-genome sequencing by Ion S5 technology-challenges, protocol optimization and success rates for different strains. *Viruses.* 2022; 14: 1230.
53. Tegally H, Wilkinson E, Althaus CL, Giovanetti M, San JE, Giandhari J, et al. Rapid replacement of the beta variant by the delta variant in South Africa. *medRxiv.* 2021. doi: 10.1101/2021.09.23.21264018.
54. Boehm E, Kronig I, Neher RA, Eckerle I, Vetter P, Kaiser L. Novel SARS-CoV-2 variants: The pandemics within the pandemic. *Clin Microbiol Infect.* 2021; 27: 1109-1117.