International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Retrospective Sequence Data Analysis of SARS - CoV-2 in Goa using Oxford Nanopore Technologies: A District Hospital Experience

Dr. Varsha Munj[1], Navin Shirodkar[2], Dr. Patricia Fernandes[3], Zyrell Fernandes[4]

[1]Senior Pathologist, Department of Pathology
Email: *varshadr[at]rediffmail.com*
Contact Number: 8830601924

[2]Internship Student, Department of Microbiology
Email: *navinshirodkar007[at]gmail.com*
Contact Number: 9881251970

[3]Microbiologist, Department of Microbiology,
Email: *patricia15_fernandes[at]rediffmail.com*
Contact Number: 9823458772

[4]Microbiologist, Department of Microbiology
Email: *zy1595[at]gmail.com*
Contact Number: 9960189740

Abstract: ***Background: Over the past decade, with the development and refinement of third - generation sequencing technologies such as Nanopore Sequencing, Whole genome Sequencing (WGS) has become accessible to smaller institutes worldwide. The recent COVID - 19 Pandemic caused by the SARS - CoV - 2 virus has demonstrated the importance of Genome sequencing in observing the acquisition of new viral mutations, proving vital in monitoring the evolution of new variants of concerns and their spread within the population. Method: The study retrospectively analyzed the sequenced data of 508 samples obtained after nanopore sequencing from January 2023 to July 2023 of samples with cycle threshold <30. Results: It had shown that there was a surge of COVID - 19 cases observed during March, April and May, which correlated with recombinant variants of SARS - CoV 2 in positive samples for the state of Goa belonging to the XBB 1.16 lineage. Conclusion: This study highlighted the challenges of performing such an analysis in a resource - limited environment and artefacts associated with using Midnight Primers.***

Keywords: Nanopore Sequencing, COVID - 19, viral mutations, XBB 1.16, North District Hospital Goa

1. Introduction

The Coronavirus Disease 2019 (COVID - 19) is a respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS - CoV - 2), a member of the Coronavirus group of viruses. The highly contagious virus was first identified in Wuhan in 2019 and swiftly triggered a global outbreak, prompting the World Health Organization (WHO) to declare a Pandemic ([1]). This significantly strained and pressured public healthcare systems and profoundly impacted the global economy ([2]). Due to the prioritisation of healthcare towards the fight against COVID - 19, there was a significant disruption in the treatment of other conditions, leading to a substantial backlog of patients in healthcare systems around the world ([3; 4]).

The strategies involved in combating the COVID - 19 Pandemic demonstrated the importance of genome sequencing in tracking viral mutations, proving vital in monitoring the evolution of new variants of concerns and their spread within the population ([5; 6; 7; 8]). The development of new variants, more infectious than the original L strain, led to a drastic surge in case numbers. This was seen in April 2022 when there was a surge in Covid case numbers and mortality rates driven by the spread of Lineage B.1.617, of which the Delta subvariant was responsible for a large number of cases in India leading to the Second wave of Covid infections ([9]). This led to a considerable strain on the healthcare system, including a shortage of medical oxygen, hospital beds, and other items essential to treating COVID - 19 patients ([10]). Similarly, B.1.1.529 omicron variants ([11]) also increased cases worldwide and in India and soon became the dominant strain, replacing the Delta variant ([12]). Sequencing data regarding the flow of these variants of concern is vital to track the spread of the virus and variants and to instruct public policy to direct healthcare ([13]).

In tandem with the COVID - 19 pandemic, sequencing technologies have become much more widely adopted in the global fight against COVID - 19, making it more accessible to smaller institutions worldwide. Third - generation sequencers, including those developed by Oxford Nanopore Technology (ONT), were widely adopted for quick, high - throughput, real - time COVID diagnosis and as a platform for variant analysis ([5]). Although Second or Next generation Short - read sequencing technologies, such as Illumina MiSeq, represent the current standard for accurate sequence detection in pathogen genomics with a higher accuracy rate at the single nucleotide level; however, third - generation sequencers like the ONT MinIon have been widely adopted in the clinical setting due to their portability, ease of Setup, and relatively faster results which can be analyzed in real -

time allowing for quick, high - throughput sample diagnosis and variant monitoring in the population ([14; 15]).

Goa, a popular tourist destination on the Konkan coast of India, faces a heightened risk of infectious diseases, such as COVID - 19, due to a constant influx of international and domestic tourists ([16]). From March to May 2023, a surge in COVID - 19 - positive cases was observed in India and has been widely attributed to the evolution and spread of the XBB 1.16 subvariant of SARS - CoV - 2 in India ([17]). In May 2023, XBB 1.16 was deemed a variant of interest by the WHO ([18]).

In Goa, there was a sudden surge of covid positive patients. However, a thorough analysis of the different variants in the Goan population is lacking, causing a gap in information regarding the regional COVID - 19 landscape.

Aims and Objectives

1) Retrospective analysis of sequenced data to establish a pipeline to generate reads with high quality scores and genomic coverage in - order to monitor the variants.
2) To detect the presence of new recombinant variants and check the correlation with increase in cases.
3) Identify the limitations of long read sequencing data for COVID - 19 analysis.

2. Materials and Methods

This is a retrospective cohort study carried outat the North District Hospital Molecular and Microbiology Laboratory, Mapusa Goa from the months of January 2023 to July 2023 after the permission of institutional ethics committee.508 SARS - CoV - 2 positive samples with a qPCR CT value less than 30 and 10x genomic coverage more than 85% on sequencing were considered for the study using nanopore sequencing with Oxford Nanopore Technologies (ONT) MinION mk1B. Sequences with a Phred quality score below 8 and sequences with a 10x genomic coverage less than 85% on sequencing were excluded from the study.

Sample Collection and Processing

Nasopharyngeal and oropharyngeal swabs were collected from patients suspected to be infected with the SARS - CoV - 2 virus in Goa. Samples were collected and stored in viral transport medium and delivered to the Microbiology and Molecular biology laboratory at North District Hospital Mapusa Goa. These included samples directly collected at the hospital or received for sequencing from other ICMR - approved testing facilities all over Goa. Samples were checked to ensure they met all the acceptance criteria and labeled. Viral RNA was extracted using a Roche Magnapure 24 fully automated extraction machine using the manufacturer's pathogen 3.2.1 protocol (Roche, 2022).

qRT - PCR

qRT - PCR runs were performed using Applied Biosystems Quantstudio 5 (Thermo Fisher, Waltham, MA, USA) instrument using ICMR - NIV SARS - CoV - 2 detection kit using positive and negative controls, targeting E (envelope) gene, RdRp (RNA dependent RNA polymerase) gene, and Orf1a gene. The steps followed were the same as the kit protocol. Samples with a Cycle threshold (Ct) up to 30 were selected for sequencing.

Library Preparation and Sequencing with MinlON Mk1B

(A) Arctic Protocol

A shorter, adapted version of the SARS - CoV - 2 sequencing protocol (nCoV 2019 sequencing protocol v3 (LoCost) V.3) developed and adapted by the ARTIC Network (htps: //articnetwork /) using the reagent from New England BioLabs (NEB, Ipswich, MA, USA) and Oxford Nanopore Technology (ONT, Oxford, UK) was utilized.

Each sequencing was performed using the Ligation Sequencing Kit 109 (SQK - LSK109, ONT, Oxford, UK). For cDNA synthesis, 2 uL of LunaScript RT SuperMix (M3010, NEB, Ipswich, MA., USA) was mixed with 8 uL of viral RNA, incubated 2 min at 25 °C, 10 min at 55 °C, and 1 min at 95 °C for enzyme inactivation, then held at 4 °C.

The overlapping amplicons were generated (- 400 bp) by mixing 12.5 uL of Hot Start High Fidelity Master Mix (M0494, NEB, Ipswich, MA, USA), 3.7 uL of primer pool V3 (ARTIC nCoV - 2019 V3 Panel and ARTIC nCoV - 2019 V4 Panel, IDT, Coralville, IA, USA), 3.8 uL of nuclease - free water (NFW, NEB, Ipswich, MA, USA), and 5 uL of cDNA. Two separate reactions were performed for each sample using the two primer pools. The cycling program was the following: initial denaturation: 30 s at 98 °C, followed by annealing: 15 s at 98 °C, Denaturing at 5 min at 65 °C for 28 cycles and a final holding step at 4 C.

Samples were then pooled together and washed by first adding equal amounts of Ampure XP beads and incubating at room temperature for 10 minutes on a hula mixer, the beads were pelleted using a magnetic stand until the eluate was clear and colorless, supernatant was pipetted and discarded. The pellet was washed with 200 ul of freshly prepared 80% ethanol. After pipetting the ethanol, the pellet was left to dry for 30 seconds. The pellet was resuspended in 15ul of nuclease - free water and kept at room temperature for 2 minutes. The tube was kept on the magnetic stand until the eluate was colorless.15 ul of the eluate containing the DNA library was collected and stored.

The concentration of DNA was quantified using a Qubit dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA) on a Qubit 2.0 instrument (Invitrogen, Waltham, MA, USA). For the end preparation, we used NebNext Ultra lI End Repair/dA - Tailing module NEB, Ipswich, MA, USA), as follows: 1.75 uL of Ultra II End Prep Reaction Buffer (NEB, Ipswich, MA, USA), 0.75 uL Ultra II End Prep Enzyme Mix (NEB, Ipswich, MA, USA), and 50 ng of sample from the previous step diluted to 12.5 ul. The reaction mixture was incubated for 5 min at 25 C.5 min at 65 °C, and Cooled at 4 °C.

Samples were barcoded using 2.5 uL of EXP - NBD104 (barcodes 1 - 12, ONT, Oxford, UK) and EXP - NBD114 (barcodes 13 - 24. ONT Oxford. UK), 10 ul of Blunt TA Ligase Master Mix (MO367, NEB, Ipswich, MA, USA).6 ul. of NFW (NEB. Ipswich, MA, USA), and 1.5 uLof reaction

mixture from the previous step, then incubated as follows: 20 min at 20 C, 10 min at 65 °C, and cooling for 1 min at 4 C. Next, 10 uL of all barcoded samples was Pooled together and purified using 6 uL of AMPure XP Magnetic Beads (Beckman Coulter, Brea, CA, USA). Samples were quantified using a Qubit 2.0 spectrophotometer (Invitrogen, Waltham, MA, USA) and dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA).

For adaptor ligation, about 30 ng of the barcoded samples were mixed with 10 uLNEBNext Quick Ligation Reaction Buffer (NebNext Quick Ligation Module, E6056, NEB, Ipswich, MA, USA), 5 uL of adaptor MIX (AMII, ONT, Oxford, UK), and 5 uL of Quick T4 DNA Ligase (NebNext Quick Ligation Module, E6056, NEB, Ipswich, MA, USA). The mixture was incubated at room temperature for 10 min, followed by purification with AMPure XP Magnetic Beads (Beckman Coulter, Brea, CA, USA) 1: 1 and another Qubit quantification. About 15 ng of the library was loaded in a final volume of 75 uL on a primed R9.4.1 flow cell (ONT, Oxford, UK) fitted in a MinlON Mk1B (ONT, Oxford, UK) instrument.

(B) Midnight Protocol

Library Preparation and Sequencing with MinlON Mk1B We used an adapted version of the SARS - CoV - 2 Midnight sequencing protocol developed by Oxford Nanopore Technology (ONT, Oxford, UK) using the reagent from New England BioLabs (NEB, Ipswich, MA, USA) and Oxford Nanopore Technology (ONT, Oxford, UK).

For cDNA synthesis, 2 uL of LunaScript RT SuperMix (M3010, NEB, Ipswich, MA., USA) was mixed with 8 uL of viral RNA, incubated 2 min at 25 °C, 10 min at 55 °C, and 1 min at 95 °C for enzyme inactivation, then kept at 4 °C until the next step. the overlapping amplicons were generated (- 1200 bp) by mixing 6.25 uL of Hot Start High Fidelity Master Mix (M0494, NEB, Ipswich, MA, USA), 0, 05 uL of priers (midnight nCoV - 2019 V3 Panel and ARTIC nCoV - 2019 V4 Panel, IDT, Corralville, IA, USA), 3.7 uL of nuclease - free water (NFW, NEB, Ipswich, MA, USA), and 2.5 uL of cDNA.

Two separate reactions were performed for each sample using the two primer pools. The cycling program was the following: initial step 30 s at 98 °C, followed by 15 s at 98 °C, 2 min at 61 °C and 3 min at 65 C, for 35 cycles and cooling at 4 C. The contents of pool A were transferred to the corresponding well of pool B and mixed thoroughly using a pipette.5ul of pool B (now containing pooled PCR products) was transferred to the corresponding well of the barcode attachment plate along with 2.5 uL of NFW and 2.5 uL of rapid barcodes (SQK - RBK110.96, Nanopore technologies) from the rapid barcoding plate the plate was sealed and incubated in a thermal cycler at 30 C for 2 min and 80 C for two min.10 uL of all barcoded samples were Pooled and purified using equal volumes of AMPure XP Magnetic Beads (Beckman Coulter, Brea, CA, USA).

Samples were quantified using a Qubit 2.0 spectrophotometer (Invitrogen, Waltham, MA, USA) and Qubit dsDNA HS Assay Kit (Thermo Fisher, Waltham, MA, USA). For adaptor ligation, about 600 - 800 ng of the barcoded samples were mixed with 1 uL of rapid adapter (RAP F) and incubated at room temperature for 5 min. to this, we added 37.5 uL of sequencing buffer II (SBII) and 25.5 ul of loading beads II (LBII) to make 75ul of the final library, which is loaded on a primed R9.4.1 flow cell (ONT, Oxford, UK) fitted in a MinlON Mk1B (ONT, Oxford, UK) instrument.

(C) Base - calling and demultiplexing

For both the above procedures, base - calling and demultiplexing were performed with the MinKNOW 20.10 (ONT, Oxford, UK) software, which is integrated into MinION Mk1B (ONT, Oxford, UK); further data analysis was carried out on the Commander NGS (Version 2022.04.26 - 13521) developed by Genotypic Ltd. (Bengaluru, India), which contains pipelines for COVID analysis for both Arctic and Midnight primers.

The software investigates the depth of coverage for each barcoded sample. It runs quality checks to determine phred quality scores using Nanostat ([19]) and information regarding contig lengths. Pangolin (version v4.3, pangolin data v1.21) and Nexclade (version 2.14.1, commit: 85e00e8, branch: release; reference strain - Wuhan - Hu - 1/2019 (MN908947), last updated: 2023 - 08 - 09 12: 00 (UTC)) were used to predict the most likely variant via their web applications ([20; 21]).

Visualisation Tools

R (version 4.1.2; R Core Team 2021) andRStudio (2022.07.3 Build 586; RStudio Team, 2022) were used to generate all plots (except coverage plots which used Commander NGS) used in this study. Notably, the Tidyverse ([22]) package library was used, with ggplot2 ([22]) and ggstatsplot [23] used for plot generation. The Scripts, data and session information used in this study are included in the supplementary section.

Primer Study

The Primer comparison study was done using 8 samples (Samples 1 - 8), which underwent RNA extraction followed by library preparation and sequencing using either Arctic v3 primers with native barcoding expansion kits (EXP - NBD104Oxford Nanopore Technologies), Arctic v3 primers with Rapid barcoding kits (SQK - RBK110.96, Oxford Nanopore Technologies), Midnight (Genotypic) with rapid barcoding kits (SQK - RBK110.96, Oxford Nanopore Technologies) and relatively newer Midnight (Oxford Nanopore Technologies) with the same rapid barcoding kits. Of the 8 samples, samples 3 and 4 were excluded from the analysis as they were of poor quality with RT - PCR Cycle Threshold values of 33 and 31, respectively, which was above the base cutoff of 30 used for samples and including these would not be representative of actual primer - protocol performance.

3. Results

1) There was a Surge in COVID - 19 Cases Across Goa.

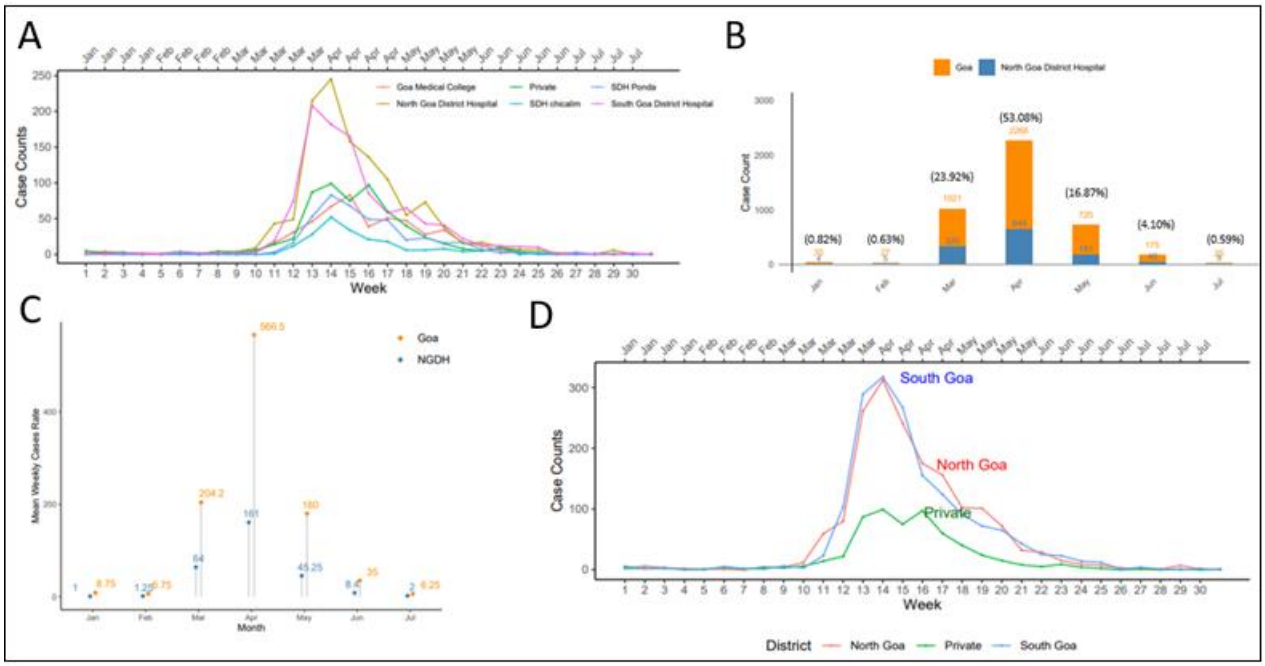

Fig 1. **Surge in COVID - 19 Rates Across Goa**. (A) Positive case counts from five major public Government testing facilities in Goa and from private testers over 30 weeks. The x - axis represents the week and month of testing, while the Y - axis represents case counts (SDH = Sub District Hospital). (B) Stacked bar graph showing the total number of positive cases per month. Percentage values reflect the percentage of total cases recorded monthly during 30 weeks. (C) Mean weekly cases reported each month (NGDH = North Goa District Hospital. Goa counts reflect all the facilities, including NGDH). (D) From the 6 facilities above, counts were arranged at the district level.

A surge in COVID - 19 cases was reported among the 9 ICMR - approved testing laboratories in Goa between March and June (**Fig 1. A).** From the 2nd of January (beginning of Week 1) till the end of July (end of Week 30), 93.86% of all cases recorded in Goa occurred during the Months of March, April and May, with each having 23.91%, 53.37%, and 16.96% of the positive COVID 19 cases respectively (**Fig 1B; S. Fig 1A).**

The Month of April saw the most significant number of positive cases, with a mean of 566.5 cases per week observed in Goa (**Fig 1C**). Case data uploaded by the ICMR - approved testing facilities for COVID - 19 revealed that the District of South Goa had higher cases than the District of North Goa (**Fig 1A, D**). With regards to individual testing facilities, The North Goa District Hospital recorded the most significant number of positive samples (1205) overall, followed by the South Goa District Hospital (1031), private labs (593), Goa Medical College (496), Sub District Hospital Ponda (413) and lastly Sub District Hospital Chicalim (211) (**Fig1 B**).

The positivity (%) rates across Goa similarly increased, with April having the highest positivity, peaking at 14.91% during Week 14 (April) of the study (**S. Fig 1 C, D, E**). This trend of SARS - CoV - 2 positivity correlated with the trend of positivity seen across the country.

2) The Majority of Sequenced Cases are Cases Driven by Recombinant Variants.

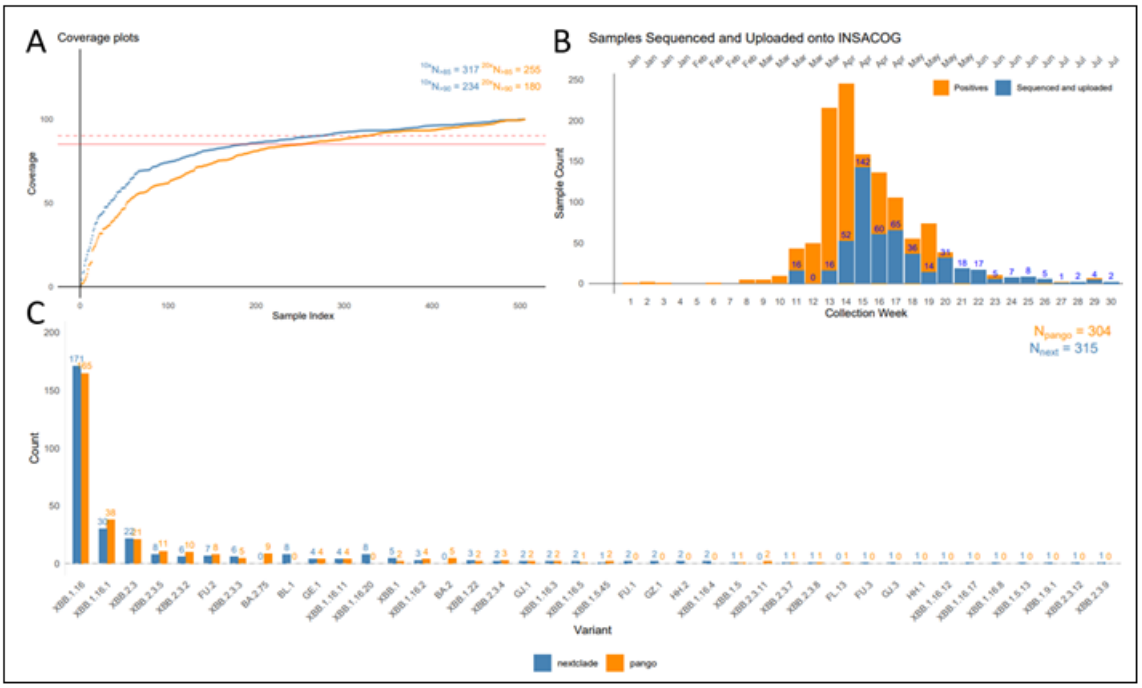

Figure 2: XBB 1.16 Lineage is Driving COVID - 19 Cases in Goa. (A) The total Coverage covered by the consensus sequenced is shown. The Blue line represents the 10X Coverage while the Orange line represents the 20X Coverage. The Y axis represents Coverage, while the X axis represents the sample number. (B) Shows when the samples which were sequenced and uploaded onto INSACOG were collected. The X - axis shows the collection week, while the Y axis shows the number of samples (sample count). (C) Shows the number of variants called after the samples underwent successful variant calling via either Pangolin or Next clade software. The Y - axis represents the number, while the X - axis represents the variant lineage.

508 positive samples, with a cycle threshold value of < 30, underwent sequencing. A coverage analysis revealed that only 317 out of the 508 samples sequenced had a 10x coverage above 85%, 255 samples had a 10x coverage above 90, and 234 samples had a 20x coverage above 85% (**Fig 2A).** The data from 501 of these samples were uploaded onto INSACOG (**Fig 2B**). The sequencing data uploaded onto INSACOG was from March to July 2023. The largest samples were sequenced in April, corresponding with a prominent case spike. Positive samples collected in January and February either had a CT value above 30 or did not receive sufficient data post - sequencing and, hence, were excluded from sequencing analysis.

Samples having a 10x coverage above 85% underwent variant analysis using Pangolin's and Nexclade's website tools. Default settings were used with both methods, and the Pango - learn model was used with pangolin. Lineage analysis revealed that the XBB 1.6, a recombinant variant, was the primary driver of cases observed in Goa, with 54% of cases being attributed to the subvariant and 12.50% being caused by XBB 1.16.1, a descendent subvariant (**Fig 3C**). The next three strains were descendants of the XBB 2.3 subvariant with XBB 2.3, XBB 2.3.5 and XBB 2.3.2 detected in 6.91%, 3.62% and 3.30% of samples. Most strains in circulation were recombinant variants of the Omicron variant. There were conflicts between Nextclade and Pangolin, with Nexclade being more lenient to low coverage values while assigning lineages. Although lineage calling was successful, all NextClade's QC checker samples were scored as Bad, including samples with 20x coverage values above 95%. In the case of Pangolin, it was noted that it was much stricter in assigning lineages with samples having lower coverage value being unassigned. Of the 507 sequenced samples, 304 and 315 samples had their variants called by Pangolin and Nextclade, respectively. Hence, due to poor consensus coverage, almost 200 (38 - 40%) of the sequenced samples could not have their variants called.

3) Loss of DNA observed following the Arctic Protocol drastically decreases sequencing Quality.

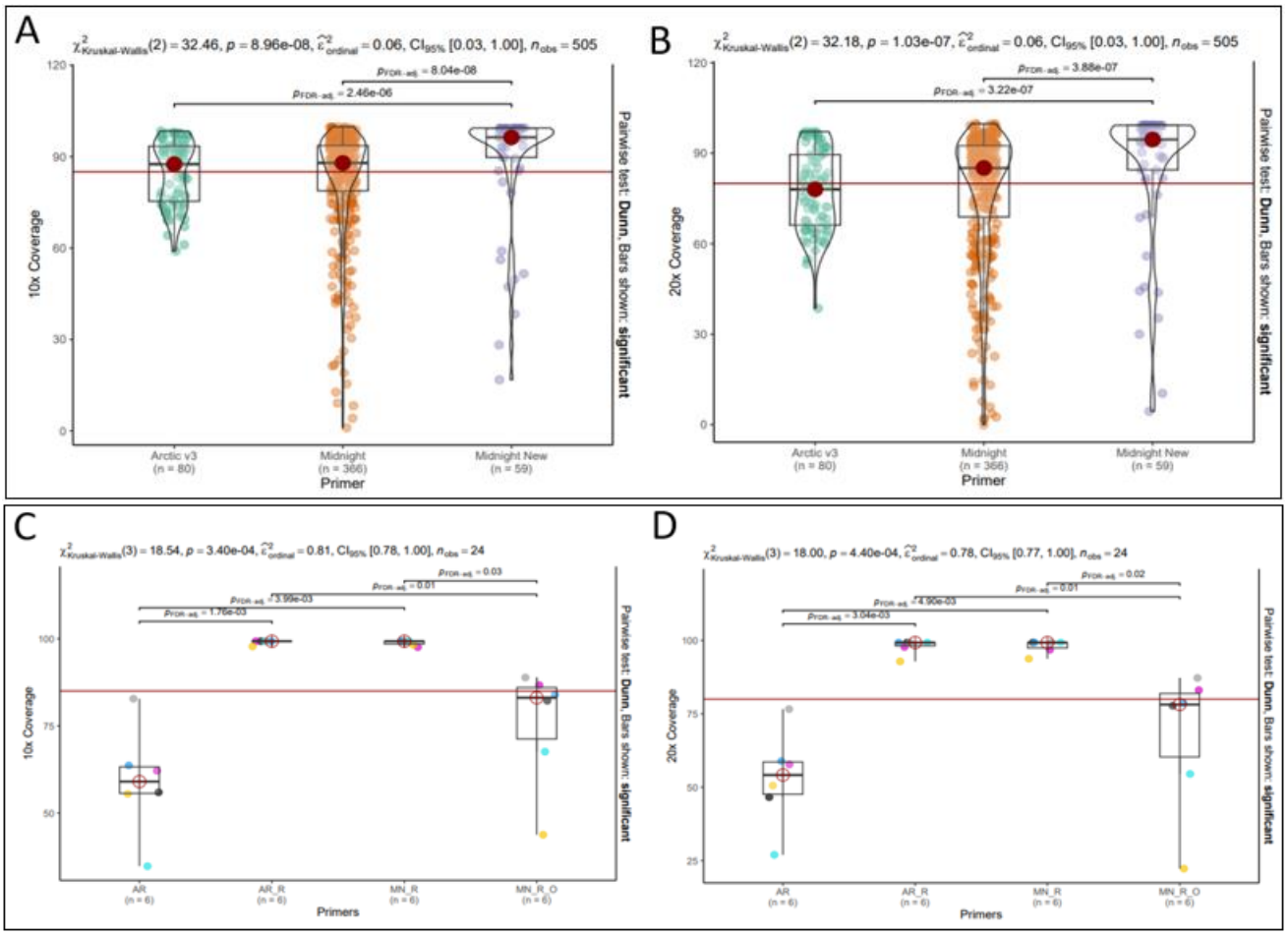

Fig 3. **Arctic v3 Primers are sensitive to DNA loss per wash**. Comparison of the 10x Coverage values (A) and 20x coverage values (B) per primer pair used (Arctic Version 3, Midnight (Genotypic) and Midnight (Oxford Nanopore Technologies). The Kruskal - Wallis test was applied to identify significant differences between groups (p = 8.86e - 08 for A and 1.03e - 07 for B), and the Dunn test showed pairwise comparisons. The Kruskal - Wallis test detected significant differences among the groups (p = 3.40e - 04 for C and 4.40e - 04 for D), and the Dunn test showed pairwise comparisons. Primers used were AR (Arctic v3 primers with Native barcoding protocol), AR - R (Arctic v3 primers with rapid barcoding protocol), MN - R (Midnight (Oxford Nanopore Technologies) with rapid barcoding protocol) and MN - R - O (Midnight (Genotypic)) with rapid barcoding protocol. In (A) and (C), a red line intersects the y - axis at the 85 mark and in (B) and (D) at the 80 mark. In all cases, The X - axis represents the primer pair used and the Y–axis represents either 10x or 20x coverage.

To investigate the effect of different primers on the quality of sequencing, the overall 10x and 20x coverage values were compared to the primers used in the 507 sequenced samples. It was shown that Midnight Primers (Oxford Nanopore Technologies) performed significantly better than Midnight Primers (Genotypic) and Arctic Primers using native barcoding (Fig **3A, 3B**). To investigate primer efficacy, SARS - CoV - 2 samples underwent Extraction followed by sequencing using Midnight Primers (Genotypic) (MN - O), Midnight Primers (Oxford Nanopore Technologies) (MN - N), Arctic Primers v3 with native barcoding (AR) and Arctic Primers v3 with rapid barcoding kits. Samples undergoing AR had a significantly decreased 10x and 20x coverage and consensus sequence coverage, and none of the 8 samples crossed the 85% and 80% coverage cutoffs, respectively (**Fig 3C, 3D**; **S. Fig 3A, 3B**). This can be attributed to the significant loss of DNA during washing steps and library preparation, which has plagued this protocol in previous attempts by the laboratory. In contrast, all 6 samples analyzed, which underwent sequencing using AR - R or with the MN - N, had coverage values which crossed the cutoff values. Only 2 out of 6 samples using MN - O crossed the 85% and 80% coverage cutoffs. Statistical testing using the Kruskal Wallis test [24] followed by pairwise comparisons via Dunn testing ([25]) showed that AR - N and MN - N primers performed significantly better in terms of 10x and 20x coverage than AR and MN - O primers.

4. Discussion

The results showed a strong surge in COVID cases across Goa between March and June brought about by the XBB 1.16 line of variants. This temporal pattern aligns with a similar rise in cases observed across India and other neighboring states ([26; 17]). XBB 1.16 strain has outcompeted all other strains in Goa and is the dominant strain, with 54% of all sequenced cases driven by it. The XBB 1.16 is shown to have an effective reproductive number (Re) 1·22 - fold higher than the parental XBB.1 and 1·13 - fold higher than XBB.1.5; another independently evolved strain spreading in India at the time ([27; 28]). This higher Re suggests that the XBB 1.16 variant is more contagious, contributing significantly to its rapid spread and dominance. Only 3 cases

of XBB 1.15 and its descendent sub - variants were observed in the sequence data, which fits with the overall decrease in the variant worldwide ([28]). This sheds essential light on the landscape of COVID - 19 infections in Goa and shows that the occurrence and spread of variants in Goa coincide with the rest of India. The other subvariant variant observed was XBB 2.3 and its descendants. This is concerning as studies have shown that XBB 2.3 possesses a higher transmission rate than XBB.1.16 and exhibits a greater evasive capacity of immune - generated antibodies and vaccines, which could have a profound impact on breakthrough infection rates ([29]). However, as of August 2023, overall COVID - 19 cases remain low in Goa. With regards to breakthrough infection, future studies could look into the rate of breakthrough infections caused by the new variants in Goa to analyze the efficacy of previous doses in combating these variants. Unfortunately, such data was not accurately collected at the North Goa District Hospital, so variants could not be linked to vaccination status.

Whilst praised for their portability and ease of use, ONT - based machines have always been criticized for their relatively low coverage. From previous studies, it is shown, that although these technologies can be applied for COVID sequencing and variant analysis and offer a less labor - intensive solution to produce results quickly, the genomic coverage is inferior to that of short - read sequencers such as the Illumina MiSeq ([30; 31; 15]). Our study observed that 37.46% (190) of the 507 samples sequenced could not reach the >85% 10x coverage genome cut - off. Furthermore, of the 317 samples that did cross the cut - off, 10 could not be assigned to a variant using Pangolin. Although samples with 10x coverage values above 95% were obtained, all of the sequenced samples received a bad QC rating on Nextclade. However, New kits are coming into the market, and the recently developed R10 flowcells used in conjunction with Q20+ Kits have shown promising results with the advent of duplex sequencing. But, they prove to be incredibly expensive for the lab and may not be accessible to other smaller institutes due to the cost of entry.

The lab successfully ran the Midnight protocol and consistently got successful runs and high amounts of DNA after the protocol's different washing steps. The process was quick and gave consistent results. However, the lab successively used the Arctic protocol, using the Arctic primers v3, to run successfully only twice out of 5 attempts. In all other attempts, DNA post - washing during library preparation resulted in the consequent loss, leading to insufficient DNA concentrations for optimum sequencing. This loss of DNA also affected Arctic results using native Arctic barcoding kits in the comparison study (Fig 3) while comparing primer sets.

In contrast, it was shown that Arctic v3 primers paired with rapid barcoding kits showed better results in the study. A more thorough comparison of the coverage values obtained from each primer set should be done in the future with a larger number of samples. Although statistical testing was conducted, due to the small number of samples per group (6), the statistical test results cannot be taken at face value. They should be repeated in the future with a higher sample size. Furthermore, due to the large loss of DNA during washing steps, this experiment is not representative of the performance of Arctic primers with their native barcoding kit, as in the case of the two successful Arctic runs where coverage values were incredibly high, having a mean 10x coverage of 84.6 and a 20x coverage of 77.4 which was comparable to those of the genotypic Midnight primers (82.2% and 76.9% respectively), but rather to emphasize the protocol's vulnerability to DNA loss and the utmost care that needs to be taken to prevent this loss. Furthermore, this also highlights how it is simpler instead to get consistent results with lesser DNA loss while using the adapted midnight protocol.

Due to the complex nature of the bioinformatic tools used for data analysis and the lack of specialized bioinformatics - based training, a lot of open - source software such as Epi2me labs and Interarctic were inaccessible and proprietary GUI - based software such as Commander ngswas chosen instead. This led to a significant increase in lab expenditure in acquiring an institutional license. However, this made the post - sequencing analysis steps incredibly intuitive for a small - scale non - bioinformatics laboratory at the expense of the flexibility and accessible documentation provided by open - source tools.

5. Conclusion

This report highlights the surge in COVID cases and has shown that the driver of these cases is the spread of new sub - variants such as XBB 1.16, XBB 2.3 and their descendants. The study validated the use of the ONT Minion in a low - resource area to monitor variants in the local population while also showing the limitations of using long - read sequencing technologies.

References

[1] Cucinotta, D. and Vanelli, M. (2020). WHO Declares COVID - 19 a Pandemic. [online] Acta bio - medica: AteneiParmensis. Available at: https: //pubmed. ncbi. nlm. nih. gov/32191675/.

[2] Mallah, S. I., Ghorab, O. K., Al - Salmi, S., Abdellatif, O. S., Tharmaratnam, T., Iskandar, M. A., Sefen, J. A. N., Sidhu, P., Atallah, B., El - Lababidi, R. and Al - Qahtani, M. (2021). COVID - 19: breaking down a global health crisis. Annals of Clinical Microbiology and Antimicrobials, [online] 20 (35). doi: https: //doi. org/10.1186/s12941 - 021 - 00438 - 7.

[3] Gertz, A. H., Pollack, C. C., Schultheiss, M. D. and Brownstein, J. S. (2022). Delayed medical care and underlying health in the United States during the COVID - 19 pandemic: A cross - sectional study. Preventive Medicine Reports, 28, p.101882. doi: https: //doi. org/10.1016/j. pmedr.2022.101882.

[4] Sharma, J., Mahajan, A., Bakhshi, S., Patil, V., Verma, N., Radhakrishnan, V., Singh, A., Kayal, S., Seth, R., Pushpam, D. and Arora, R. S. (2021). The impact of COVID-19 pandemic on access to treatment for children with cancer in India and treating center practices. Cancer, 128 (3), pp.579–586. doi: https: //doi. org/10.1002/cncr.33945.

[5] 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. and Deveson, I. W. (2020). Analytical validity of nanopore sequencing for rapid SARS - CoV - 2 genome analysis. Nature Communications, [online] 11 (1), p.6272. doi: https: //doi. org/10.1038/s41467 - 020 - 20075 - 6.

[6] Deval, H., Nyayanit, D. A., Shailendra Kumar Mishra, Yadav, P., Zaman, K., Shankar, P., Brij Ranjan Misra, Sthita Pragnya Behera, Kumar, N., Kumar, A., Bhardwaj, P., Gaurav Raj Dwivedi, Singh, R., Shete, A. M., Pandit, P., Pandey, A., Girijesh Kumar Yadav, Gupta, S. K., Kumar, M. and Asif Kavathekar (2022). Genome Sequencing Reveals a Mixed Picture of SARS - CoV - 2 Variant of Concern Circulation in Eastern Uttar Pradesh, India. Frontiers, [online] 8. doi: https: //doi. org/10.3389/fmed.2021.781287.

[7] Ionut Dragos Capraru, Marcel Laurentiu Romanescu, Flavia Medana Anghel, Cristian Oancea, Marian, C., Ioan Ovidiu Sirbu, Aimee Rodica Chis and Paula Diana Ciordas (2022). Identification of Genomic Variants of SARS - CoV - 2 Using Nanopore Sequencing. Medicana, 58 (12), pp.1841–1841. doi: https: //doi. org/10.3390/medicina58121841.

[8] Mannsverk, S., Bergholm, J., Palanisamy, N., Ellström, P., Kaden, R., Lindh, J. and Lennerstrand, J. (2022). SARS - CoV - 2 variants of concern and spike protein mutational dynamics in a Swedish cohort during 2021, studied by Nanopore sequencing. Virology Journal, 19 (1). doi: https: //doi. org/10.1186/s12985 - 022 - 01896 - x.

[9] Sarkar, A., Chakrabarti, A. K. and Dutta, S. (2021). Covid - 19 Infection in India: A Comparative Analysis of the Second Wave with the First Wave. Pathogens, 10 (9), p.1222. doi: https: //doi. org/10.3390/pathogens10091222.

[10] Choudhary, O. P., Priyanka, Singh, I. and Rodriguez - Morales, A. J. (2021). Second wave of COVID - 19 in India: Dissection of the causes and lessons learnt. Travel Medicine and Infectious Disease, 43, p.102126. doi: https: //doi. org/10.1016/j. tmaid.2021.102126.

[11] World Health Organization (2021). Classification of Omicron (B.1.1.529): SARS - CoV - 2 Variant of Concern. [online] www.who. int. Available at: https: //www.who. int/news/item/26 - 11 - 2021 - classification - of - omicron - (b.1.1.529) - sars - cov - 2 - variant - of - concern.

[12] Fan, Y., Li, X., Zhang, L., Wan, S., Zhang, L. and Zhou, F. (2022). SARS - CoV - 2 Omicron variant: recent progress and future perspectives. Signal Transduction and Targeted Therapy, [online] 7 (1), pp.1–11. doi: https: //doi. org/10.1038/s41392 - 022 - 00997 - x.

[13] Lancet, T. (2021). Genomic sequencing in pandemics. The Lancet, [online] 397 (10273), p.445. doi: https: //doi. org/10.1016/S0140 - 6736 (21) 00257 - 9.

[14] MacKenzie, M. and Argyropoulos, C. (2023). An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. Micromachines, [online] 14 (2), p.459. doi: https: //doi. org/10.3390/mi14020459.

[15] Tshiabuila, D., Giandhari, J., Pillay, S., Ramphal, U., Ramphal, Y., Maharaj, A., Anyaneji, U. J., Naidoo, Y., Tegally, H., San, E. J., Wilkinson, E., Lessells, R. J. and de Oliveira, T. (2022). Comparison of SARS - CoV - 2 sequencing using the ONT GridION and the Illumina MiSeq. BMC Genomics, 23 (1). doi: https: //doi. org/10.1186/s12864 - 022 - 08541 - 5

[16] Selvanathan, E. A., Jayasinghe, M. and Selvanathan, S. (2021). International Tourism and Infectious Disease Transmission Nexus: A Cross - Country and Regional Study. Journal of Travel Research, 61 (8), p.004728752110489. doi: https: //doi. org/10.1177/00472875211048932.

[17] Tarun Kumar Suvvari, Mohanty, A., Charmi Dedeepya Boddapalli, Venkata Ramana Kandi, Padhi, B. K. and Sah, R. (2023). Emergence of XBB.1.16—a new recombinant COVID - 19 variant driving the surge of COVID - 19 cases in India. The Egyptian Journal of Internal Medicine, 35 (1). doi: https: //doi. org/10.1186/s43162 - 023 - 00229 - w.

[18] Harris, E. (2023). XBB.1.16 Deemed COVID - 19 'Variant of Interest'. JAMA, [online] 329 (20), p.1731. doi: https: //doi. org/10.1001/jama.2023.7768.

[19] De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long - read sequencing data. Bioinformatics, 34 (15), pp.2666–2669. doi: https: //doi. org/10.1093/bioinformatics/bty149.

[20] O'Toole, Á., 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. and Rambaut, A. (2021). Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. Virus Evolution. doi: https: //doi. org/10.1093/ve/veab064.

[21] Aksamentov, I., Roemer, C., Hodcroft, E. and Neher, R. (2021). Nextclade: clade assignment, mutation calling and quality control for viral genomes. Journal of Open Source Software, [online] 6 (67), p.3773. doi: https: //doi. org/10.21105/joss.03773.

[22] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V. and Takahashi, K. (2019). Welcome to the Tidyverse. Journal of Open Source Software, [online] 4 (43), p.1686. doi: https: //doi. org/10.21105/joss.01686.

[23] Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. Journal of Open Source Software, 6 (61), 3167, doi: 10.21105/joss.03167

[24] Kruskal, W. H. and Wallis, W. A. (1952). Use of Ranks in One - Criterion Variance Analysis. Journal of the American Statistical Association, 47 (260), pp.583–621.

[25] Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. Technometrics, 6 (3), pp.241–252. doi: https: //doi. org/10.1080/00401706.1964.10490181.

[26] Rajesh Karyakarte, Das, R., Mansi Rajmane, Sonali Dudhate, Agarasen, J., Pillai, P., Priyanka Chandankhede, Rutika Labhshetwar, Yogita Gadiyal, Kulkarni, P. P., SafanahNizarudeen, Joshi, S., KrishanpalKarmodiya and Varsha Potdar (2023). Chasing SARS - CoV - 2 XBB.1.16 Recombinant Lineage in India and the Clinical Profile of XBB.1.16

Cases in Maharashtra, India. Cureus. doi: https: //doi. org/10.7759/cureus.39816.

[27] Daichi Yamasoba, Keiya Uriu, Arnon Plianchaisuk, Yusuke Kosugi, Pan, L., Zahradnik, J., Ito, J. and Sato, K. (2023). Virological characteristics of the SARS - CoV - 2 omicron XBB.1.16 variant. Lancet, 23 (6). doi: https: //doi. org/10.1016/s1473 - 3099 (23) 00278 - 5.

[28] Samal, J., Arjun Bhugra, Varun Suroliya, Gautam, P., Agarwal, R., Bihari, C. and Gupta, E. (2023). The emergence of the Omicron XBB.1.5 variant in India: a brief report on clinical presentation of a few cases. Frontiers in virology, 3. doi: https: //doi. org/10.3389/fviro.2023.1158703.

[29] Cóbar, O. M. and S. Cóbar (2023). XBB.2.3, Prevalence, Structural, Genomic, and Pathogenic Properties. Medical research archives, 11 (7.2). doi: https: //doi. org/10.18103/mra. v11i7.2.4137.

[30] BroňaBrejová, Kristína Boršová, Viktória Hodorová, Viktória Čabanová, Askar Gafurov, Dominika Fricova, Neboháčová, M., Tomáš Vinař, Klempa, B. and Nosek, J. (2021). Nanopore sequencing of SARS - CoV - 2: Comparison of short and long PCR - tiling amplicon protocols. PLOS ONE, 16 (10), pp. e0259277–e0259277. doi: https: //doi. org/10.1371/journal. pone.0259277.

[31] UlykbekKairov, Amanzhanova, A., Daniyar Karabayev, Saule Rakhimova, Akbota Aitkulova, Samatkyzy, D., Ruslan Kalendar, Ulan Kozhamkulov, AskhatMolkenov, Aidana Gabdulkayum, Sarbassov, D. D. and Ainur Akilzhanova (2022). A high scale SARS - CoV - 2 profiling by its whole - genome sequencing using Oxford Nanopore Technology in Kazakhstan. Frontiers in Genetics, [online] 13. doi: https: //doi. org/10.3389/fgene.2022.906318.

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

[33] Vashishtha, V. M. and Kumar, P. (2023). Preliminary clinical characteristics of Pediatric Covid - 19 cases during the ongoing Omicron XBB.1.16 driven surge in a north Indian city. medRxiv (Cold Spring Harbor Laboratory). doi: https: //doi. org/10.1101/2023.04.18.23288715.