# The use of MGI ATOPlex Technology for Complete Genome Sequencing and Identification of **SARS-CoV-2 Variants on DNBSEQ-G50* in Nigerian Institute of Medical Research**

## Highlight

- MGI ATOPlex Library Preparation Technology and Sequencing instruments at the Nigerian Institute of Medical Research show possibility of a robust and dynamic public health response for containment of the- COVID-19 pandemic in Nigeria, Africa and other parts of the world
- Linear amplification of libraries ensures low error rate and accurate base calling
- High depth, coverage and Q30 scores help accurate detection of different strains of SARS-CoV-2
- High mapping rate of sequences allows for efficient SARS-CoV-2 lineage assignment

## Introduction

MGI sequencing platforms continue to excel as a fast, flexible, and high-quality performing sequencing technology in the global sequencing market (1). This has been attributed to the novel DNA nanoballs (DNBs) and combinatorial Probe Anchor Synthesis (cPAS) technology employed by these platforms, enabling fast turnaround time, low cost per base, high loading density of the DNBs on flow cells, accurate base calling and a very low risk of signal duplication (1, 2, 3, 4, 5). The ongoing COVID-19 pandemic caused by SARS-CoV-2 continues to emerge as a global health problem of unprecedented impacts in the history of mankind. Here, we described MGI' s ATOPlex library workflow to obtain the SARS-CoV-2 genome libraries from randomly selected Nigerian patients that tested positive for COVID-19. We also evaluated the performance of the DNBSEQ-G50* sequencer in SARS-CoV-2 lineage characterization.

# Experimental Design

## Sample Selection and RNA Isolation

RNA of twenty-six (26) randomly selected samples were isolated from nasal/oropharyngeal mix swabs of COVID-19 positive patients using the QIAamp viral mini kit (Qiagen). The COVID-19 positive samples were diagnosed using reverse transcription real-time polymerase chain reaction (RT-PCR) with Genesig SARS-CoV-2 RT-PCR assay kit (PrimerDesign Ltd, UK), which targeted the orf1ab gene of the virus. The RNA isolations and SARS-CoV-2 detection were performed between March – December 2020.

## RNA Library Preparation, DNA Nanoball Generation, and Sequencing

A total of 26 Next Generation Sequencing (NGS) libraries were produced from RNA samples using the MGI ATOPlex RNA Universal Library Prep Module (Figure 1). The procedure was performed in three separate areas to avoid cross-contamination between PCR products. In the first area, the preparation of cDNA by reverse transcription was done. A first PCR amplification to target the SARS-CoV-2 genome plus a first cleanup step were done in the second area, and a second PCR amplification to introduce dual barcoded adaptors and amplify the libraries plus a second cleanup step were performed. The purified libraries were quantified using the Qubit dsDNA HS Assay kit (ThermoFisher Scientific), and pooled to a total of 400 ng in 48 µL. This was followed by the synthesis of single stranded circular DNA (ssCirDNA) using an MGIEasy dual barcode circularization kit. A cleanup step was performed on the ssCirDNA followed by quantification using the Qubit ssDNA Assay kit (ThermoFisher Scientific). The ssCirDNA libraries were converted into DNBs using DNB-making reagents from a DNBSEQ-G50RS* High-throughput Sequencing Set (FCL PE100). The DNBs were quantified using the Qubit ssDNA Assay kit (ThermoFisher Scientific), and balanced using an ATOPlex Dual Barcoded balanced library in a 3:1 ratio of sample DNB to balanced library DNB in 100 µL prior to sequencing. Paired-End 100 sequencing was performed on the DNBSEQ-G50* using a large Flow Cell (FCL) and the DNBSEQ-G50* high throughput PE100 sequencing reagents with an extra cPAS Barcode Primer 3 reagent required for PE sequencing. The DNBs were automatically loaded onto the flow cell by the DNBSEQ-G50* sequencing instrument. An overview of the workflow is provided in Figure 1.

SARS-Cov-2 RNA

cDNA Synthesis by Reverse Transcription

1st PCR Amplification

Cleanup of 1st PCR product

2nd PCR Amplification

Cleanup of 2nd PCR product

Quantitation using sensitive dsDNA fluorometry Qubit Assay (ng/μL) and pooling

Circularization of pooled cDNA library (ssCirDNA synthesis)

Cleanup of the synthesized ssCirDNA

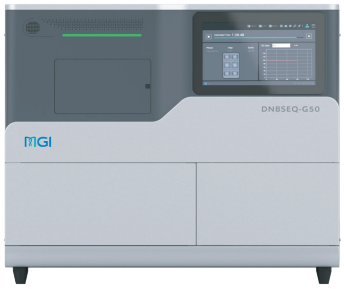
Quantitation using sensitive dsDNA fluorometry Qubit Assay (ng/μL)

Generation of DNB

Quantitation using sensitive dsDNA fluorometry Qubit Assay (ng/μL)

Balancing of Sample DNB with ATOPlex Balanced Library (3:1)

DNB Loading and Sequencing on DNBSEQ-G50* (FCL PE100)




MGI DNBSEQ-G50*

Fig. 1: SARS-CoV-2 Complete Genome Sequencing workflow on DNBSEQ-G50* using MGI ATOPlex RNA Library Prep Module and ATOPlex Dual Barcode circularization Kit.

The MGI ATOPlex SARS-CoV-2 bioinformatics pipeline (https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0) was used to perform the quality assessment of the demultiplexed reads in fastq format using FASTQC and Trimmonatic algorithms within the pipeline. The alignment of the filtered reads was performed against the SARS-CoV-2 reference genome (Wuhan-Hu-1, MN908947.3), followed by the assembly of the reads to generate consensus sequences in FASTA format. For all analyses, default parameters were used. The SAM files generated were further converted to BAM files as input for variant calling tools (results not provided). The complete genome sequences of each of the SARS-CoV-2 isolates were submitted to PANGOLIN (web-based) for lineage assignment.

## Results and Discussion

The cycle threshold values of SARS-CoV-2 RNA from COVID-19 positive results ranged from 15 to 25. Analyses of the complete genome datasets of the SARS-CoV-2 isolates enabled us to obtain 293,258 – 20,000,000 raw reads with 291,502 - 19,879,520 (99.17 -99.91%) mapping to the reference SARS-CoV-2 genome (Wuhan-Hu-1, MN908947.3) . The genome sizes of the SARS-CoV-2 isolates sequenced ranged from 29.8 – 29.9 kilobases with coverage depths of 388 to 55,453x. Further analysis revealed a median read coverage depth of 9,220x and Q30 percentage scores of 89.75 – 92.75 % (Figure 2; Figure 3; Figure 4). Lineage analysis revealed the occurrence of 9 distinct lineages among the SARS-CoV-2 isolates with lineage B.1.1 occurring most at 30.8%, followed by lineages B.1 at 19.2%. Other lineages seen in decreasing order of occurrences were B.1.438 and L.3 (11.5% each), B.1.22 (7.7%), B.1.8 and B.1.1.181 (3.8% each) (Figure 5). To date, NIMR has sequenced hundreds more SARS-CoV-2 genomes from COVID-19 positive samples using the DNBSEQ-G50*. These data will contribute to informing a robust and dynamic public health response for containment of the ongoing COVID-19 pandemic in Nigeria, Africa and other continents of the world.

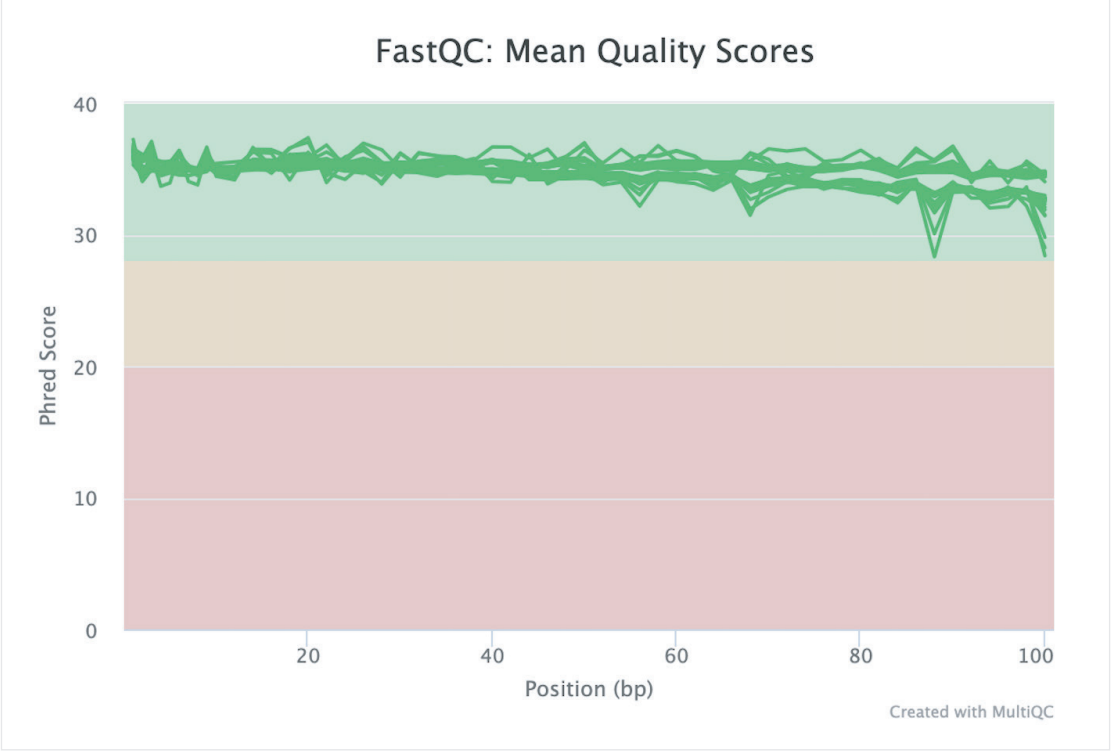



Fig. 2: FastQC histogram plot report of sequence raw reads of 26 SARS-CoV-2 complete genome datasets prior to trimming.

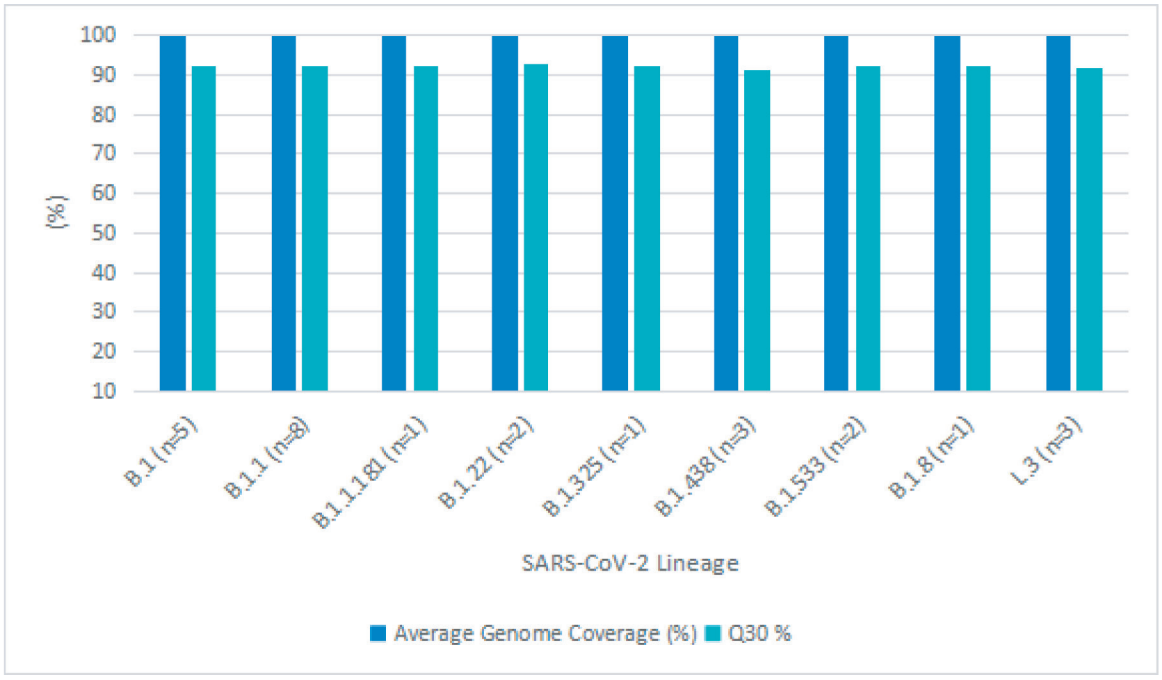



Fig. 3: Average Genome coverage and Q30 % quality scores for different lineages

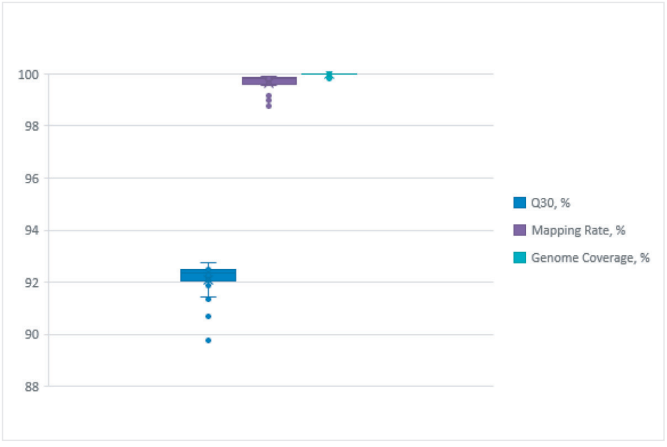



Fig. 4: Key quality metrics for sequencing over all samples (n= 26)

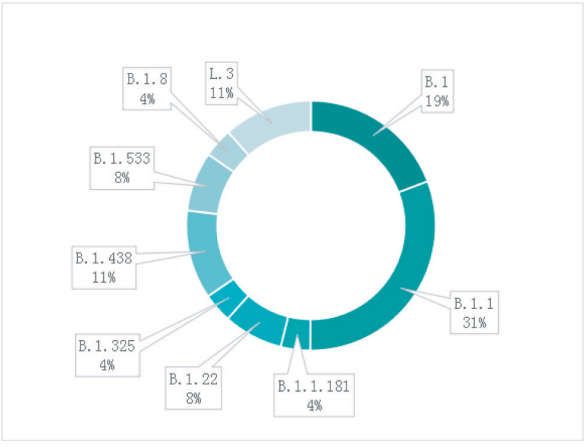



Fig. 5: Overview of SARS-Cov-2 Pangolin Lineage Frequency Observed

# Ordering information

## Equipment

| Product Name | Catalog No. | Configuration |
|---|---|---|
| Genetic Sequencer DNBSEQ-G50RS* (Config 2) | 900-000354-00 | RUO |

## Reagents

| Product Name | Catalog No. | Configuration |
|---|---|---|
| ATOPlex SARS-CoV-2 Full Length Genome Panel | 1000021178 | RUO |
| ATOPlex Dual Barcode Primer Module (01-96) V1.0 | 1000021626 | RUO, 96 RXN |
| ATOPlex RNA Library Prep Set | 1000027431 | RUO, 96 RXN |
| MGIEasy Dual Barcode Circularization Module | 1000018650 | RUO, 96 RXN |
| CPAS Barcode Primer 3 Reagent Kit | 1000020834 | RUO |
| DNBSEQ-G50RS* High-throughput Sequencing Set (FCL PE100) | 1000019859 | RUO |

## Analysis

| Product Name | Pipeline Access | Configuration |
|---|---|---|
| SARS-CoV-2_Multi-PCR_v1.0 | GitHub - MGI-tech-bioinformatics/SARS-CoV-2_-Multi-PCR_v1.0: SARS-CoV-2 analysis pipeline for multiplex-PCR MPS(Massive Parrallel Sequencing) data | RUO |

# References


1. Jeon SA, Park JL, Kim JH, Kim JH, Kim YS, Kim JC, Kim SY. Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing. Genomics Inform. 2019 Sep;17(3):e32.
2. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78–81
3. Kim, H. M., Jeon, S., Chung, O., Jun, J. H., Kim, H. S., Blazyte, A., Lee, H. Y., Yu, Y., Cho, Y. S., Bolser, D. M., & Bhak, J. (2021). Comparative analysis of 7 short-read sequencing platforms using the Korean Reference Genome: MGI and Illumina sequencing benchmark for whole-genome sequencing. GigaScience, 10(3), giab014. https://doi.org/10.1093/gigascience/giab014
4. Li Q, Zhao X, Zhang W, Wang L, Wang J, Xu D, Mei Z, Liu Q, Du S, Li Z, et al. Reliable multiplex sequencing with rare index mis-assignment on DNB-based NGS platform. BMC Genom. 2019;20(1):215.
5. Jeon SA, Park JL, Park SJ, Kim JH, Goh SH, Han JY, Kim SY. Comparison between MGI and Illumina sequencing platforms for whole genome sequencing. Genes Genomics. 2021 Genes Genomics. 2021;43:713-724.



**MGI Tech Co.,Ltd.**

Building 11, Beishan Industrial Zone, Yantian District,
Shenzhen, CHINA, 518083

400-096-6988
mgi-tech.com
MGI-service@mgi-tech.com




*Unless otherwise informed, all sequencers and sequencing reagents are not available in Germany, USA, Spain, UK, Hong Kong, Sweden, Belgium and Italy.