SARS-CoV-2 Sequencing Package (V3.1)

Highlight

High-quality technologies integration

For library preparation, the MGISP-100RS automated sample preparation system integrates the ATOPlex multiplex PCR technology, FAST PCR-FREE library preparation technology, and One Step DNB preparation technology. For sequencing and data processing, DNBSEQ-G50RS* genetic sequencer is used to perform SE100 rapid sequencing. After the sequencing is completed, MGI metargetCOVID can seamlessly perform automated data processing.

Extremely fast running time

The entire process from RNA to data processing can be completed within 24 hours for up to 16 samples simultaneously. The library preparation process takes about 9 hours, the sequencing process takes about 12 hours, and the data processing takes about less than 1 hour.

Ultra-high sequencing detection sensitivity

It is suitable for high-throughput sequencing of samples with Ct value of SARS-CoV-2 ≤ 35. The ATOPlex multiplex PCR technology makes it possible to obtain high-coverage sequencing data of the SARS-CoV-2 whole genome.

Automated experiments and data processing

With the MGISP-100RS automated sample preparation system, multiple steps in the entire experiment can be realized through automated operations. The MGI metargetCOVID software can seamlessly connect the sequencing data and the data processing, realize the automatic result generation, which vastly reduces the dependence on manual labor in the whole process.

Excellent data quality

DNB and cPAS technologies ensure the sequencing accuracy and efficiency, providing high-quality sequencing data.

Product Description

The SARS-CoV-2 Sequencing Package (V3.1) is based on self-developed compatible reagents, automated sample preparation systems, high-throughput sequencing platforms, and data processing systems for SARS-CoV-2 whole-genome sequencing, covering the processes from RNA to result generation. Parts of experimental steps and all data processing can realize automatic operation. This package can perform rapid, accurate, and comprehensive high-throughput sequencing of RNA samples that have been proven the positive of SARS-CoV-2, thereby providing essential references about variants detection, variants annotation, variant branch identification, and origin tracing of SARS-CoV-2 samples.

Based on ATOPlex multiplex PCR technology, the library preparation system provides a cost-effective, rapid, and convenient experimental method for the whole-genome sequencing of SARS-CoV-2. The system can amplify and enrich the extremely deficient virus genome millions of times, combined with the DNBSEQ-G50RS* desktop single flow cell genetic sequencer to obtain information from high-precision virus genome sequencing. The MGI metargetCOVID software aims at SARS-CoV-2 detection and variants analysis. Meanwhile, it provides results of clade assignment, allowing users to quickly identify whether samples contain SARS-CoV-2 and comprehensively analyze clade assignment of samples that are identified as positive.

Method

Use SARS-CoV-2 RNA standard (sequence information is identical with **SARS-CoV-2 isolate Wuhan-Hu-1**, GenBank: MN908947.3). Dilute viral RNA standard to reactions containing 400 and 4000 copies (named as S400 and S4000, respectively) each. Use ATOPlex RNA Multiplex PCR Amplification Set V3.1 (16 RXN, Cat. No. 940-000119-00) to perform reverse transcription and multiple PCR amplification. Use MGIEasy Fast PCR-FREE FS Library Prep Set (16 RXN, Cat. No. 940-000019-00), DNBSEQ One-Step DNB Prep Kit (OS-DB) (Cat. No. 1000026466) in the MGISP-100RS automated sample preparation system (Cat. No. 900-000206-00) for automated library and DNB preparation.

Use DNBSEQ-G50RS High-throughput Rapid Sequencing Set (FCS SE100)* (Cat. No. 1000019860) for SE100 sequencing.

Use MGI metargetCOVID software (Cat. No. 970-000228-00) to perform variants detection, variants annotation, variant branch identification relying on sequencing data.

Results

Data Overview

Table1 Data overview

Sample ID	Total Reads	Q30(%)	GC(%)	Mapping Reads (%)	SARS-CoV-2 Reads (%)	Cov 1x (%)	Prediction
S400	2067184	94.36	43.21	85.88	87.81	99.75	Positive
S4000	3195093	95.53	39.37	98.75	97.02	99.75	Positive

Quality control of high-throughput sequencing

Use MGI metargetCOVID software for quality control and filtering sequencing data from DNBSEQ-G50RS*. The SE100 Q30 is higher than 94%, demonstrating extremely high sequencing quality. See Table 2 for more detailed sequencing data.

Table 2 Sequencing quality control of DNBSEQ-G50RS* FCS SE100

Sample ID	S400		S4000	
Item	Raw Reads (fq1)	Clean Reads (fq1)	Raw Reads (fq1)	Clean Reads (fq1)
Read length	100	100	100	100
Total number of reads	2067184 (100.00%)	2057541 (100.00%)	3195093 (100.00%)	3192686 (100.00%)
Number of filtered reads	9643 (0.47%)	–	2407 (0.08%)	–
Reads with n rate exceed	0 (0.00%)	–	0 (0.00%)	–
Reads with low quality	0 (0.00%)	–	0 (0.00%)	–
Reads with adapter	9643 (100%)	–	2407 (100%)	–
Total number of bases	206718400 (100.00%)	205754100 (100.00%)	319509300 (100.00%)	319268600 (100.00%)
Number of filtered bases	964300 (0.47%)	–	240700 (0.08%)	–
Number of base A	59174333 (28.63%)	58901662 (28.63%)	97547856 (30.53%)	97476837 (30.53%)
Number of base C	43890194 (21.23%)	43692147 (21.24%)	61989049 (19.40%)	61944638 (19.40%)
Number of base G	45443655 (21.98%)	45184055 (21.96%)	63801596 (19.97%)	63741303 (19.96%)
Number of base T	58200122 (28.15%)	57966198 (28.17%)	96164702 (30.10%)	96099741 (30.10%)
Number of base N	10096 (0.00%)	10038 (0.00%)	6097 (0.00%)	6081 (0.00%)
Q20	98.25%	98.25%	98.31%	98.31%
Q30	94.36%	94.37%	95.53%	95.53%

Mapping information

Use SARS-CoV-2 isolate Wuhan-Hu-1, complete genome (GenBank: MN908947.3) as SARS-CoV-2 reference genome for further analysis. After qualified clean reads align with the reference, it can filter records and remain records whose clip ≤ 10 and mapping quality ≥20. Table 3 shows the relevant statistical information about qualified reads and reference genome after mapping, including the raw reads, SARS-CoV-2 reads percentage, and prediction results. Table 4 and 5 and figure 1 and 2 show the statistical information separately, such as coverage, depth of qualified reads of S400 and S4000 that distribute in the reference genome and their gene regions.

Table 3 Statistical information of qualified reads and reference genome after mapping

Sample ID	Reference (bp)	SARS-CoV-2 Reads	GAPDH Reads	SPIKE-IN CONTROL Reads	SARS-CoV-2 Ratio(%)	Identification	SARS-CoV-2 (copies/ml)
S400	29903	1422490	147028	197494	87.81	Positive	5920.62
S4000	29903	2992549	68292	91856	97.02	Positive	37898.51

Table 4 Statistical information of qualified reads of S400 that distribute in the reference genome and its gene regions

Gene Region	Avg Depth	Median	Coverage%	Cov 4x%	Cov 10x%	Cov 30x%	Cov 100x%
Whole genome	3948.87	2773.0	99.75	99.75	99.75	99.74	99.61
5'UTR	4296.58	5903.0	91.29	91.29	91.29	90.91	90.15
ORF1ab	3317.66	2472.0	100.00	100.00	100.00	100.00	99.85
S	4270.85	3206.0	100.00	100.00	100.00	100.00	100.00
ORF3a	2682.12	1986.0	100.00	100.00	100.00	100.00	100.00
E	2485.95	1442.0	100.00	100.00	100.00	100.00	100.00
M	5340.68	5072.0	100.00	100.00	100.00	100.00	100.00
ORF6	3107.80	3205.0	100.00	100.00	100.00	100.00	100.00
ORF7a	5869.19	6446.0	100.00	100.00	100.00	100.00	100.00
ORF8	7191.73	3376.0	100.00	100.00	100.00	100.00	100.00
N	11151.42	9917.0	100.00	100.00	100.00	100.00	100.00
ORF10	12724.68	12893.0	100.00	100.00	100.00	100.00	100.00
3'UTR	4485.14	1214.0	77.63	77.63	77.19	76.32	74.12

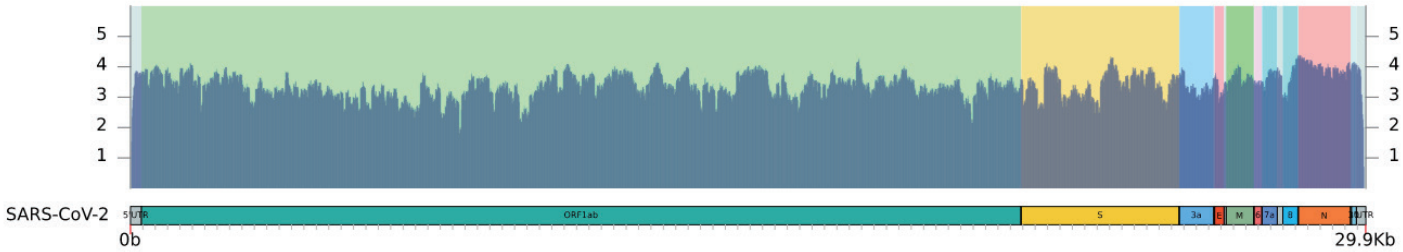

Figure 1 The qualified reads coverage of S400 that distribute in the reference genome and its gene regions.

Table 5 Statistical information of qualified reads of S4000 that distribute in the reference genome and its gene regions.

Gene Region	Avg Depth	Median	Coverage%	Cov 4x%	Cov 10x%	Cov 30x%	Cov 100x%
Whole genome	8329.67	7877.0	99.75	99.75	99.75	99.75	99.74
5'UTR	8119.38	10749.0	91.29	91.29	91.29	91.29	91.29
ORF1ab	7920.10	7443.0	100.00	100.00	100.00	100.00	100.00
S	8250.37	7927.0	100.00	100.00	100.00	100.00	100.00
ORF3a	7149.55	7434.0	100.00	100.00	100.00	100.00	100.00
E	5729.41	6260.0	100.00	100.00	100.00	100.00	100.00
M	10979.08	10996.0	100.00	100.00	100.00	100.00	100.00
ORF6	8383.91	8682.0	100.00	100.00	100.00	100.00	100.00
ORF7a	10457.10	10607.0	100.00	100.00	100.00	100.00	100.00
ORF8	11071.37	9011.0	100.00	100.00	100.00	100.00	100.00
N	13145.21	13008.0	100.00	100.00	100.00	100.00	100.00
ORF10	14800.66	14730.0	100.00	100.00	100.00	100.00	100.00
3'UTR	5207.87	2365.0	77.63	77.63	77.19	77.19	76.32

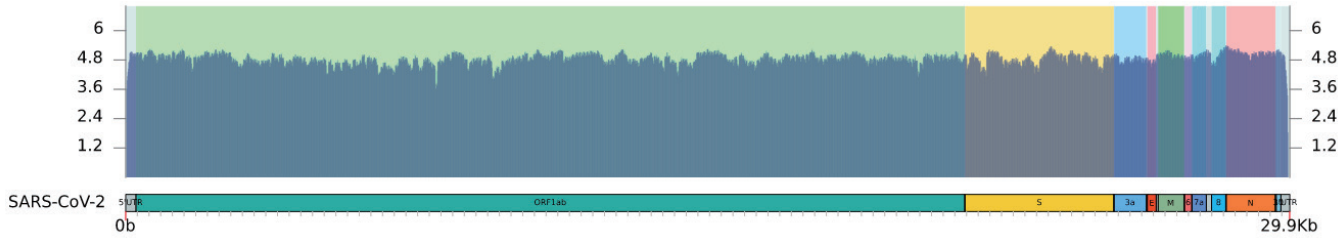

Figure 2 The qualified reads coverage of S4000 that distribute in the reference genome and its gene regions.

Variants Calling

The metargetCOVID is able to deeply unveil single nucleotide polymorphisms (SNPs) and insert and deletion (Indel). This method uses FreeBayes to call SNPs and Indels in mapping sequences. After obtaining SNPs or Indels, these variants will be annotated. However, considering that SARS-CoV-2 RNA standards used in the experiment were synthesized stemming from SARS-CoV-2 isolate Wuhan-Hu-1 (MN908947.3, complete genome), there is no SNP or Indel detected in this test.

Clade Assignment

Based on the consensus sequences after mapping the SARS-CoV-2 reference, assign them to the SARS-CoV-2 phylogenetic tree, which has been completed clades identification, attaining their clade ID information. At the same time, use Pangolin to calculate the SARS-CoV-2 lineage, and the relevant statistics are shown in Table 6.

Table 6 Clade assignment results

Sample ID	Assemble Size(bp)	Num Ns	Num SNPs	Num INSs	Num DELs	Clade ID	Lineage ID
S400	29903	75(0.25%)	0	0	0	19A	B
S4000	29903	75(0.25%)	0	0	0	19A	B

Summary

Equipped with MGI's self-developed reagents, automated sample preparation system, high-throughput sequencing platform, and MGI metargetCOVID software, the SARS-CoV-2 Sequencing Package (V3.1) enables to perform high-quality SARS-CoV-2 whole-genome sequencing quickly, blasting, variants calling and clade assignment, etc. Therefore, This package is ideal for SARS-CoV-2 whole genome sequencing and origin tracing, which can help customers achieve their research goals faster, more conveniently, and more accurately.

Ordering Information

Product	Specification	Cat. No.
Automated Sample Preparation System and Sequencer		
MGISP-100RS Automated Sample Prep System	Standard Configuration	900-000206-00
Genetic Sequencer DNBSEQ-G50RS*	Configuration 2	900-000354-00
Library and DNB Prep Reagent		
ATOPlex RNA Multiplex PCR-based Library Preparation Set V3.1	16 RXN	940-000132-00
DNBSEQ OneStep DNB Make Reagent Kit (OS-DB)	4 RXN	1000026466
Sequencing Reagent		
DNBSEQ-G50RS High-throughput Rapid Sequencing Set*	FCS SE100	1000019860
CPAS Barcode Primer 4 Reagent Kit	3.5 ml	1000014048
Data processing system		
MGI metargetCOVID software	V1.4	970-000228-00

Contact Us

MGI Tech Co.,Ltd
Address: Building 11,Beishan Industrial Zone, Yantian District, Shenzhen, CHINA 518083
Email: MGI-service@mgi-tech.com
Website: www.en.mgi-tech.com
Tel: +86 400-096-6988
Version： April 2022 | MGPA051810202-07

https://www.linkedin.com/com pany/mgi-bgi

https://twitter.com/MGI_BGI

*Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, USA, Spain, UK, Hong Kong, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland and Portugal.