Obtaining the whole genome sequence of SARS-Cov-2

MGI Microbiological Detection Total Package

Highlight

Fast, Simple, Streamlined Workflow

- Realize the whole process management of sample input and output

Traceability of Viruses

- Speed up the identification of suspected pathogens

Simple, efficient localization platform

- Provide one-stop platform to realize the localization detection of pathogenic microorganisms

Introduction

Since the outbreak of COVID-19, nucleic acid detection technology based on real-time fluorescent has played an important role in the rapid identification of virus and confirming patients. As COVID-19 enters the normal phase of prevention and control, the whole genome information of the virus is required to trace the origin of the virus, monitor its mutation and predict its risk, which is inseparable from high-throughput sequencing and assembly of the virus sequence (FIG. 1).

Based on the MGI Microbiological Detection Total Package, an institute successfully assembled six cases of SARS-CoV-2 and constructed a phylogenetic tree. At the same time, they acquired the ability of autonomous sequencing and genome assembly of unknown pathogenic microorganisms in response to unknown infections.

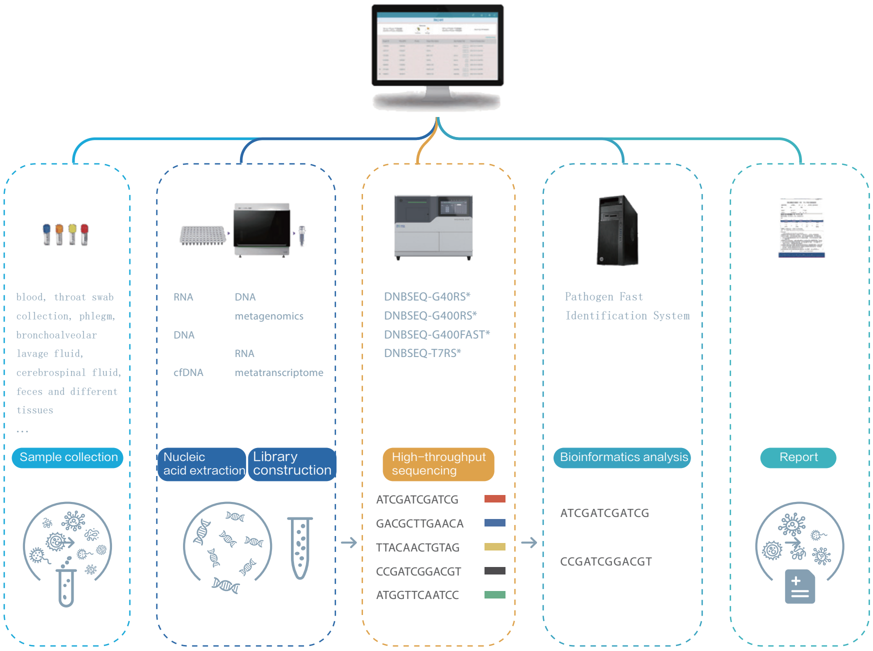

Figure 1. MGI Microbiological Detection workflow

Experimental Method

Library preparation

We apply the MGIEasy rRNA Depletion Kit (MGI, 32 RXN, Cat. No. 1000005953) to remove human rRNA if the Initial input is more than 50ng. We use the MGIEasy RNA library preparation reagent kit (16 RXN, Cat. No. 100006383) to construct the library. The libraries were obtained after a series of steps including reverse transcription, adapter ligation, PCR amplification, purification, etc. Then, it uses rolling circle amplification technology to prepare DNA nanoballs. These kits are suitable for automation system MGISP-100 (MGI, Cat. No. 900-000051-00) (FIG. 2).

Figure 2. DNA Sequencing Library Preparation System MGISP-100

Sequencing

DNBSEQ-G50* (Cat. No. 900-000354-00) is a compact and flexible benchtop genetic sequencer (FIG.3). With the design of two different flow cells, it empowers flexibility and creates a perfect balance between speed and affordability. FCS (100 M reads) allows short turnaround time for STAT samples and FCL (500 M reads) enables lower cost per sample. Based on DNBSEQ-G50* sequencer, the institute selected the PE100 sequencing reading length for high-depth sequencing to get whole genome of SARS-CoV-2. Besides, we also have DNBSEQ-G400*(Cat. No. 900-000170-00) and DNBSEQ-T7*(Cat. No. 900-000128-00) for instrument selection, and more for sequencing reads selection.

Figure 3. DNBSEQ-G50 sequencer

Data analysis

The MGI Pathogeny Fast Identification (PFI, Cat. No. 510-000164-00) software can identify more than 25000 microbes, also including SARS-CoV-2(Table 1). PFI software provides host filtering, pathogeny identification, identification of virulence factor and identification of resistance genes. It is a combination with both software and hardware, which includes the microbial database with analysis process and the whole process management of sample input and report output through ZLIMS.

Table 1. Microbial database

Classification of Microbes	Number of Species	Number of Family
Bacteria	8335	1925
Archaea	361	130
Fungi	9461	2606
Virus	9031	1050
Parasite	642	361

Sequence assembly

After automatically extracting SARS-CoV-2 reads from all sequences, the MGAP microbial genome analysis software (MGI, Cat. No.970-000109-00) starts the assembly work. Finally, the full-length sequence of the SARS-CoV-2 genome was successfully obtained, and the phylogenetic tree was further constructed, providing reference for subsequent detection and research.

Results

Identification of SARS-COV-2

After sequencing, DNBSEQ-G40 sequencer generated 19.48 Gb data with a total of 97.4 M metagenome sequencing reads. Combined with the Pathogen Fast Identification software, a total of 279, 992 SARS-CoV-2 reads were identified (Table 2).

Table 2. Identification results of virus

No.	Species	Reads Number	Relative Abundance (%)
1	*2019-nCoV*	279,992	72,184
2	*Escherichina virus FV3*	131	0.034
3	*Escherichina virus slur09*	14	0.004
4	*Escherichina phage APCEc02*	13	0.003
5	*Escherichina virus V18*	12	0.003
6	*Escherichina virus V5*	12	0.003
7	*Proteus phage VB PmiS-Isfahan*	12	0.003
8	*Escherichina virus FFH2*	11	0.003
9	*Escherichina virus FJES2013*	11	0.003
10	*shigella virus Pss1*	10	0.003

Assembly of SARS-COV-2

The analysis software automatically extracted 279, 992 SARS-CoV-2 reads from all the sequences. The MGI MGAP software with high assembly efficiency was used, and a full-length 29.86 kb genome sequence was obtained, with 1 scaffold article number and 38.01% GC content, which is very close with the known coronavirus reference sequence (MN908947.3, length of 29.90 KB, GC content of 37.9%). Then, the phylogenetic tree was built (table 3, figure 4).

Table 3. Statistical table of the viral genome sequence assembly

Genome Size (Kb)	GC (%)	Scaffold num	Gap number (bp)
29.86	38.01	1	0

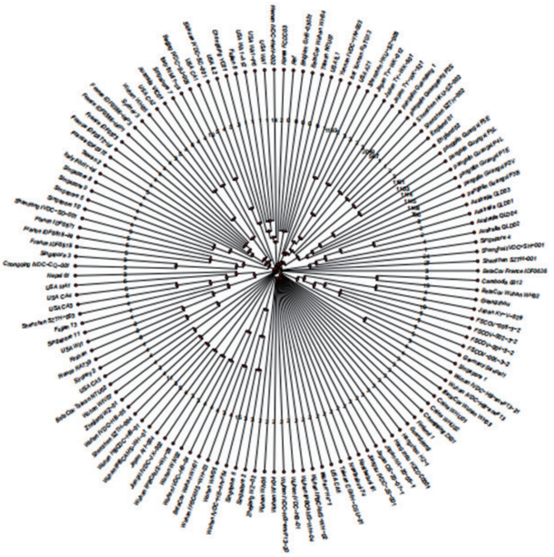

Figure 4. Shows the evolutionary tree

Ordering information

Equipment

Product Name	Catalog No.	Configuration
DNA Sequencing Library Preparation System MGISP-100	900-000051-00	RUO
Genetic Sequencer DNBSEQ-G50RS*	900-000354-00	RUO
Genetic Sequencer DNBSEQ-G400RS*	900-000170-00	RUO
Genetic Sequencer DNBSEQ-T7RS*	900-000128-00	RUO

Reagents

Product Name	Catalog No.	Configuration
MGIEasy RNA Library Prep Set	1000006383	16 RXN, RUO
MGIEasy RNA Library Prep Set	1000006384	96 RXN, RUO
MGIEasy rRNA Depletion Kit	1000005253	32 RXN, RUO

Analysis

Product Name	Catalog No.	Configuration
Fast identification of pathogenic infections (PFI, with server)	510-000164-00	Server + software, RUO
Fast identification of pathogenic infections (PFI)	057-000060-00	software, RUO
Microbial Genome Analysis Pipeline (MGAP)	970-000109-00	software, RUO

MGI Tech Co.,Ltd | Building 11, Beishan Industrial Zone, Yantian District, Shenzhen, CHINA, 518083
en.mgi-tech.com | MGI-service@mgi-tech.com

*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.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.