The Application of High-throughput Whole Genome Sequencing in Pathogenic Microbial Typing

Whole Genome Sequencing (WGS) based on DNBSEQ-G400 enables *Mycobacterium tuberculosis* Lineage Typing in Kashgar Prefecture of Xinjiang

A research team from the First People's Hospital of Kashgar, Xinjiang conducted a study of the identification and lineage typing of *Mycobacterium tuberculosis* lineage in Kashgar prefecture of Xinjiang based on the MGI's WGS product combination.

The related research was published in *BMC Infectious Diseases* with the title of "Distribution and identification of *Mycobacterium tuberculosis* lineage in Kashgar prefecture" in 2022[1].

Recommended applications: Pathogenic microorganism- *Mycobacterium tuberculosis*

Recommended model: DNBSEQ-G400RS

- ### Whole-genome SNP loci coverage

  The DNBSEQ-G400RS platform can support rapid detection and identification of microorganism lineages. This is an important prerequisite for mycobacterium lineage identification in Kashgar prefecture, providing a basis for the classification and diagnosis of tuberculosis in the region.

- ### Extremely high sequencing throughput

  The DNBSEQ-G400RS has extremely high sequencing throughput, generating 55~1440GB of data in each run, which meets the sequencing needs for large number of samples.

- ### Automatic operation compatible

  The automated extraction and library preparation equipment of MGI significantly save labor costs in high-throughput sequencing and improve processing efficiency.

Background

Mycobacterium tuberculosis, the causal agent of tuberculosis in humans, belongs to a class of strictly aerobic, positive acid-fast bacteria. Tuberculosis is a chronic infectious disease that poses a severe danger to human health. The World Health Organization (WHO) has reported that the number of patients with tuberculosis in China accounts for 8.4% (about 10 million cases) of the total number of cases worldwide [2]. *M. tuberculosis* has multiple lineages, which differ in disease prognosis, vaccine efficacy and drug resistance [3]. Genotyping among lineages is mainly based on single nucleotide polymorphism (SNP) [4]. Genotyping of *M. tuberculosis* allows for a better understanding of the local mainline branches, increases the ability to predict the prevalence trend, and also supports clinical isolation and diagnosis. In comparison, WGS can identify variants at the genomic level and can detect SNP loci more accurately and comprehensively.

Xinjiang is one of the most severely endemic regions for *M. tuberculosis* in China, with the incidence of TB in this region approaching 1 in 10,000, significantly higher than the national average of 1 in 100,000 for the rest of China[5]. Kashgar is bordered by Pakistan and India and is situated at an important hub of the Silk Road, which led to the introduction of lineages from other countries. In addition, *M. tuberculosis* was previously typed in this region using the conventional method, but the investigation of whole genome SNP was insufficient.

Study description

A research team from the First People's Hospital of Kashgar, Xinjiang performed WGS of 161 clinical samples in Kashgar prefecture based on the DNBSEQ sequencing platform and related library preparation products. They identified several lineage-specific SNPs and conducted the typing of *M. tuberculosis* lineages in Kashgar prefecture of Xinjiang. This research provides strong support for future screening of *M. tuberculosis* lineages and provides a basis for research on the diagnosis and treatment of *M. tuberculosis* in this region.

Materials and Methods

Sample collection

A total of 161 clinical specimens of *M. tuberculosis* were collected from sputum of lower respiratory tract in patients with tuberculosis in seven general hospitals including the First People's Hospital of Kashgar during 2018-2019.

Library preparation and sequencing

M. tuberculosis DNA was extracted and purified using the MGIEasy DNA Clean Beads, the specific operation procedures are described in the relevant instructions, and the nucleic acid concentration was quantified using Qubit 3.0.

The library was then prepared using the MGIEasy FS DNA Library Prep Set. Since the enzymatic digestion is used in this method, no additional equipment was required. Please refer to the relevant instructions for the specific operation procedures. The library was loaded on Agilent 2100 Bioanalyzer for size detection, mixed and sequenced on the DNBSEQ-G400 platform with the recipe of paired-end 100 bases (PE100). It is worth noting that the library preparation process can be completed on MGISP-100 and MGISP-960, the MGISP automated library preparation platforms developed by MGI, which can help save valuable time when the sample size is large.

Data analysis

After the sequencing data were quality-controlled with FastQC toolkit (V0.11.8) and mapped to the *M. tuberculosis* genome H37Rv (NC_000962.3) with BWA software, the SNP loci were detected with GATK software, and all detected SNP loci were annotated with ANNOVAR (V2.1.1) with the H37Rv genome reference.

M. tuberculosis was typed based on the detected SNP loci and a phylogenetic tree was constructed with IQ-tree software. Subsequently, the branch-specific SNP was screened from all SNPs. Finally, the lineage composition of *M. tuberculosis* in Kashgar was compared with the surrounding provinces of China and countries.

Sample collection	Library preparation and sequencing	Bioinformatics analysis	Result analysis
Clinical samples of *M. tuberculosis* from 161 patients	 MGIEasy FS DNA Library Prep Set  DNBSEQ-400 Genetic Sequencer	FastQC toolkit BWA GATK ANNOVAR	Lineage and sublineage analysis by phylogenetic tree construction

Results

Lineage and sublineage analysis of clinical strains of *M. tuberculosis*

According to the phylogenetic tree constructed from *M. tuberculosis* SNPs presented in Figure 1, *M. tuberculosis* in the region was divided into three major lineages and further divided into 11 sublineages based on branch-specific SNP. Among them, lineage 2 accounts for 45.34% (73/161), lineage 3 accounts for 32.30% (52/161), and lineage 4 accounts for 22.36% (73/161). Among the sublineages of each lineage, sublineage 2.2.1, sublineage 3.3 and sublineage 4.5 account for the major part, respectively.

Specific SNPs of clinical strains of *M. tuberculosis*

A total of 136 branch-specific SNPs were obtained by screening. Table 1 shows the number of SNPs for each lineage and sublineage that can be used as markers for strain typing. A total of 14 branch-specific SNPs were identified in lineage 2, including 9 newly discovered SNPs; 14 SNPs were identified in lineage 3, 9 of which were newly discovered; and 10 SNPs were identified in lineage 4, 7 of which were newly discovered. A total of 72 branch-specific SNPs were also screened in each sublineage.

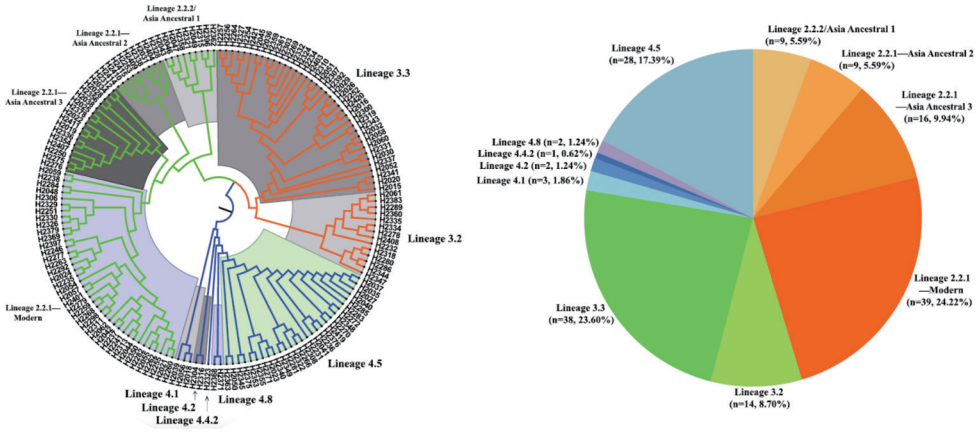

Figure 1. Phylogenetic tree of *M. tuberculosis* in Kashgar prefecture and the proportion of each lineage

Lineage	Sublineage		Name	N	Number of specific SNP	Number of specific SNPs in Coll's study	Common SNP
2			East Asian	73	14	6	5(5/14, 35.71%)
	2.1		Protobeijing	–	–	12	–
	2.2		Beijing 2.2	–	–	5	–
		2.2.1	Beijing 2.2.1	64	8	2	2(2/8, 25%)
		*	Asia Ancestral 2	9	4	–	–
		*	Asia Ancestral 3	16	1	–	–
		*	Modern Beijing	39	1	–	–
		2.2.2	Asia Ancestral 1	9	6	6	5 (5/6, 83.33%)
3			India and East Africa	52	14	9	5 (5/14, 35.71%)
	3.1			–	–	–	–
		3.1.1		–	–	3	–
		3.1.2		–	–	2	–
	3.2			14	3	–	–
	3.3			38	3	–	–
4			Euro-American	36	10	3	3 (3/10, 30.00%)
	4.1			3	18	1	1 (1/18, 5.56%)
	4.2			2	18	8	8 (8/18, 44.44%)
	4.4			–	–	2	–
		4.4.2		1	20	11	11 (11/20, 55.00%)
	4.5			28	13	8	6 (6/13, 46.15%)
	4.8			2	3	1	1 (1/3, 33.33%)

Table 1. Branch-specific SNPs of *M. tuberculosis* in Kashgar prefecture

Geographical distribution of lineages/sublineages

Figure 2 shows the distribution of each lineage of *M. tuberculosis* in the Kashgar prefecture, its inner counties and the surrounding regions, to facilitate the determination of the source of the introduction of *M. tuberculosis* in the region. The proportions of *M. tuberculosis* lineages is similar in six counties of Kashgar prefecture. Compared to the surrounding provinces, Kashgar prefecture has a lower proportion of lineage 2 and a higher proportion of lineage 3 and 4. The *M. tuberculosis* lineage is more complex in Kashgar prefecture than in other regions of China, and some neighboring countries specific lineages are present. It is assumed that lineage 3 may have been introduced from the neighboring countries surrounding the Kashgar prefecture.

In terms of sublineage composition, sublineage 2.2.1-Modern and lineage 4.5 are the main dominant strains in other regions of China, and lineage 3 has a different sublineage composition in other regions of China than in Kashgar prefecture. It is further demonstrated that the lineage may have been introduced from the surrounding countries.

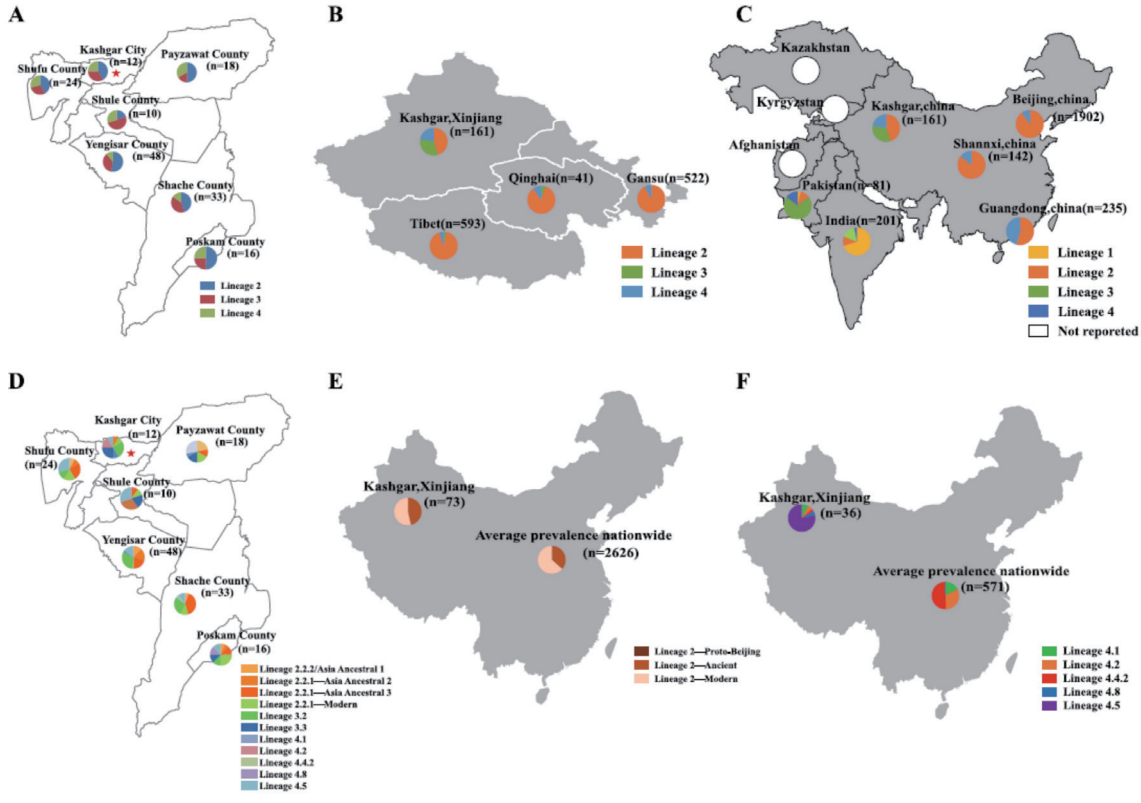

Figure 2. Geographical distribution of *M. tuberculosis* lineages and sublineages

Conclusion

In this application note, we introduced the application of MGI product combination in the identification of tuberculosis pathogens. The team sequenced the whole genome of 161 strains of *M. tuberculosis* in Kashgar prefecture, classified them into 3 lineages and 11 sublineages, and identified the region-specific SNP loci. By comparing and analyzing the geographical distribution of *M. tuberculosis* in the surrounding regions, it was concluded that lineages 3 *M. tuberculosis* in the region were mainly introduced from neighboring countries.

MGI products played a vital role in the research. The library for WGS can be quickly prepared with the MGIEasy FS DNA Library Prep Set, and in combination with the DNBSE-G400 genetic sequencer, a large number of samples can be sequenced simultaneously. Leveraging MGI's unique DNBSEQ sequencing technology, possible errors in the sequencing process are reduced, leading to more accurate identification of SNPs.

DNBSEQ-G400RS Genetic Sequencer

Reference

1. Zheren, Z., Shraddha, S., Bart, C., Dennis, C. & E., R. D. Mutational meltdown of putative microbial altruists in Streptomyces coelicolor colonies *Nature Communications* 13 (2022).
2. López, B. *et al.* A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clinical and experimental immunology* 133, 30-37, doi:10.1046/j.1365-2249.2003.02171.x (2003).
3. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5, 4812, doi:10.1038/ncomms5812 (2014).
4. Chakaya, J. *et al.* Global Tuberculosis Report 2020 - Reflections on the Global TB burden, treatment and prevention efforts. *Int J Infect Dis* 113 Suppl 1, S7-s12, doi:10.1016/j.ijid.2021.02.107 (2021).
5. 2010 年全国第五次结核病流行病学抽样调查报告 . 中国防痨杂志 .

Recommended Ordering Information

Category	Product	Cat. NO.
Instruments	DNBSEQ-G400RS Genetic Sequencer	900-000170-00
	MGISP-100RS Automated Sample Preparation System	900-000206-00
	MGISP-960RS Automated Sample Preparation System	900-000146-00
Software	MegaBOLT Bioinformatics Analysis accelerator (workstation server)	900-000555-00
Library Prep reagents	MGIEasy FS DNA Library Prep Set V2.1 (16 RXN)	1000006987
	MGIEasy DNA Clean Beads	1000005279
Sequencing Reagents	DNBSEQ-G400RS High-throughput Sequencing Set (FCL PE100)	1000016950

MGI Tech Co.,Ltd

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

+86-4000-688-114

en.mgi-tech.com

MGI-service@mgi-tech.com

*1. For StandardMPS and CoolMPS: Unless otherwise informed, StandardMPS and CoolMPS sequencing reagents, and sequencers for use with such reagents are not available in Germany, Spain, UK, Sweden, Belgium, Italy, Finland, Czech Republic, Switzerland, Portugal, Austria and Romania. Unless otherwise informed, StandardMPS sequencing reagents, and sequencers for use with such reagents are not available in Hong Kong. No purchase orders for StandardMPS products will be accepted in the USA until after January 1, 2023.

2. For HotMPS sequencers: This sequencer is only available in selected countries, and its software has been specially configured to be used in conjunction with MGI's HotMPS sequencing reagents exclusively.

3. For HotMPS reagents: This sequencing reagent is only available in selected countries.