

高通量全基因组测序在病原微生物分型中的应用

基于华大智造MGISEQ-2000的WGS测序技术助力新疆喀什地区结核分枝杆菌谱系分型

新疆喀什市第一人民医院研究团队基于华大智造全基因组测序（WGS）产品组合对新疆喀什地区结核分枝杆菌的谱系进行了分型以及鉴定等研究工作。

具体研究成果已于2022年发表于*BMC Infectious Diseases*杂志，题为“Distribution and identification of *Mycobacterium tuberculosis* lineage in Kashgar prefecture”¹。

推荐应用：病原微生物 - 结核分枝杆菌

推荐机型：MGISEQ-2000RS

• 全基因组SNP位点覆盖

MGISEQ-2000平台能够支撑微生物的快速检测、不同位点的鉴定进而绘制更精细的系谱图，是区分喀什地区分枝杆菌谱系的重要前提，为该地区结核的分类和诊断提供依据。

• 极高的测序通量

MGISEQ-2000拥有极高的测序通量，每次运行产生55~1440GB数据，满足大量样本的测序需求。

• 适配自动化建库设备

测序仪搭配自动化建库平台MGISP-100和MGISP-960等，在大样本量时可确保高效，准确地建库。



背景介绍

结核分枝杆菌是人类结核病的病原体，是专性需氧的一类细菌，抗酸染色阳性。结核病是严重危害人类健康的慢性感染性疾病。世界卫生组织（WHO）的报告显示，我国的结核病患者占全球病例总数的 8.4%（约 1000 万例）²。结核分枝杆菌具有多种谱系，不同谱系在疾病预后、疫苗疗效和耐药性方面均有所不同³。谱系间的基因分型主要基于单核苷酸多态性（SNP）⁴。对结核分枝杆菌的基因分型可以进一步了解当地主流分支，预测流行趋势，为临床分离诊断提供支持。而全基因组测序（WGS）可以在基因组水平上识别变异，能够更加准确而全面的检测 SNP 位点。

新疆地区是国内结核分枝杆菌最严重的流行区域之一，该地区的结核病人的发病率接近 1/10,000，远高于中国其他地区的 1/100,000⁵。而喀什与巴基斯坦，印度等国家接壤，并且位于丝绸之路的重要枢纽，导致其他国家谱系的传入。另外，在此之前对于该地区结核分枝杆菌的分型主要依据传统方法，全基因组 SNP 的研究不充足。

研究描述

新疆喀什市第一人民医院研究团队采用华大智造 MGISEQ-2000 测序平台及相关建库产品，对喀什地区 161 个临床样本进行 WGS 测序，鉴定出多个谱系特异性 SNP，并据此对新疆喀什地区的结核分枝杆菌进行了谱系分型。该研究对未来结合分枝杆菌的谱系筛选提供了有力支持，并为该地区结核分枝杆菌的诊治提供研究依据。

实验方法

样本收集

2018-2019年间，在喀什市第一人民医院等共7家综合医院中通过采集结核病患者下呼吸道痰液，共收集161例患者的结核分枝杆菌临床样本。

文库制备及 WGS 测序

通过MGIEasy DNA 纯化磁珠试剂盒(货号1000005279)提取并纯化结核分枝杆菌DNA，具体操作步骤请查看相关说明书,并通过Qubit3.0定量核酸浓度。

之后使用 MGIEasy 酶切 DNA 文库制备试剂盒 V2.1(货号 1000005256)进行文库构建，该方法通过酶切打断，无需购置额外的设备，具体操作步骤请查看相关说明书。通过 Agilent 2100 生物分析仪检查文库片段大小。混合合格文库后，在 MGISEQ-2000 平台完成双端 100 碱基(PE100)的 WGS 测序。值得一提的是，该建库流程同时适配华大智造自主研发的 MGISP 自动化建库平台 MGISP-100、MGISP-960 等，在样本量大时可帮助节约宝贵的时间。

分析流程

将测序数据使用Fastqc toolkit (V0.11.8)进行质控后使用BWA软件比对至结核分枝杆菌基因组 H37Rv (NC_000962.3), 并使用GATK软件对SNP位点进行检测，所有检测出的SNPs位点使用ANNOVAR (V2.1.1)利用H37Rv基因组进行注释。

基于检测出的SNP位点对结核分枝杆菌进行分型并使用IQ-tree软件进行系统发育树的构建。随后从所有SNP中筛选出分支特异性SNP。最后将喀什地区结核分枝杆菌的谱系组成与周边省份和国家进行比较。

样本采集	文库制备和测序	生信分析	结果分析
161 例患者的结核分枝杆菌临床样本	 MGIEasy 酶切 DNA 文库制备试剂盒  MGISEQ-2000 基因测序仪	Fastqc toolkit BWA GATK ANNOVAR	进化树构建谱系和亚谱系分析

结果

结核分枝杆菌临床菌株的谱系和亚谱系分析

根据图 1 所展示的由结核分枝杆菌 SNP 构建的系统发育树，该地区的结核分枝杆菌共分为 3 个主要谱系，并根据分支特异性 SNP 进一步分成 11 个亚系。其中谱系 2 占 45.34% (73/161)，谱系 3 占 32.30% (52/161)，谱系 4 占 22.36% (73/161)。在各谱系的亚系中，亚系 2.2.1，亚系 3.3，亚系 4.5 分别占主要部分。

结核分枝杆菌临床菌株的特定 SNP

通过筛选获得了 136 个分支特异性 SNP。表 1 展示了各个谱系及亚系的 SNP 数量，这些 SNP 可以用作菌株分型的标记。在谱系 2 中共鉴定出 14 个分支特异性 SNP，其中包括 9 个新发现的 SNP；谱系 3 中鉴定出 14 个 SNP，其中 9 个是新发现的；谱系 4 中鉴定出 10 个 SNP，其中 7 个是新发现的。在各个亚系中，也共筛选出 72 个分支特异性 SNP。

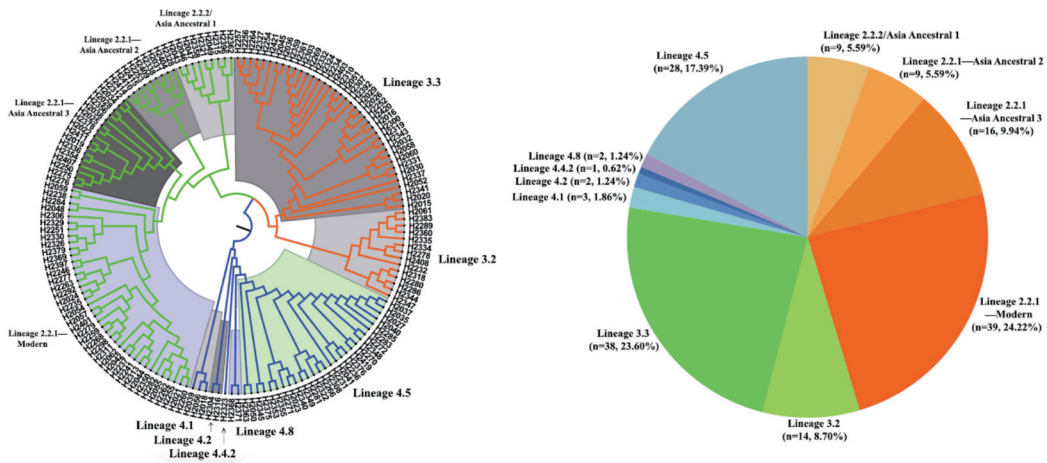


图1.喀什地区结核分枝杆菌的系统发育树及各谱系占比

Lineage	Sublineage	Name	N	Number of specific SNP	Number of specific SNPs in Coll's study	Common SNP	
2	2.1	East Asian	73	14	6	5(5/14, 35.71%)	
		Protobeijing	-	-	12	-	
	2.2	Beijing 2.2	-	-	5	-	
		2.2.1	Beijing 2.2.1	64	8	2	2(2/8, 25%)
	2.2.2	*	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	3.1	India and East Africa	52	14	9	5 (5/14, 35.71%)	
		-	-	-	-		
	3.1.1	-	-	3	-		
		3.1.2	-	-	2	-	
3.2	-	14	3	-	-		
3.3	-	38	3	-	-		
4	4.1	Euro-American	36	10	3	3 (3/10, 30.00%)	
			3	18	1	1 (1/18, 5.56%)	
			2	18	8	8 (8/18, 44.44%)	
	4.4	4.4.2	-	-	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%)		

表1.喀什地区结合分枝杆菌的分支特异性SNP

谱系 / 亚系的地理分布

图 2 展示了喀什地区及内部各县与周边区域结核分枝杆菌各谱系的分布情况，用以判断该地区结合分枝杆菌是从何处传入。喀什地区六个县的结核分枝杆菌谱系比例相似。与周边省份相比，有着更低的谱系 2 的比例和更高的谱系 3 和谱系 4 的比例。与中国其他地区相比，喀什地区结核分枝杆菌谱系更为复杂，并且存在一些周边国家特有的

谱系。由此推测，Lineage3 可能是从喀什地区周边国家传入的。

在亚系的组成上，2.2.1-Modern 亚系和谱系 4.5 是中国其他地区主要的优势菌株，谱系 3 在中国其他地区的亚系组成与喀什地区不同。进一步证明了该谱系可能是由周边国家传入。

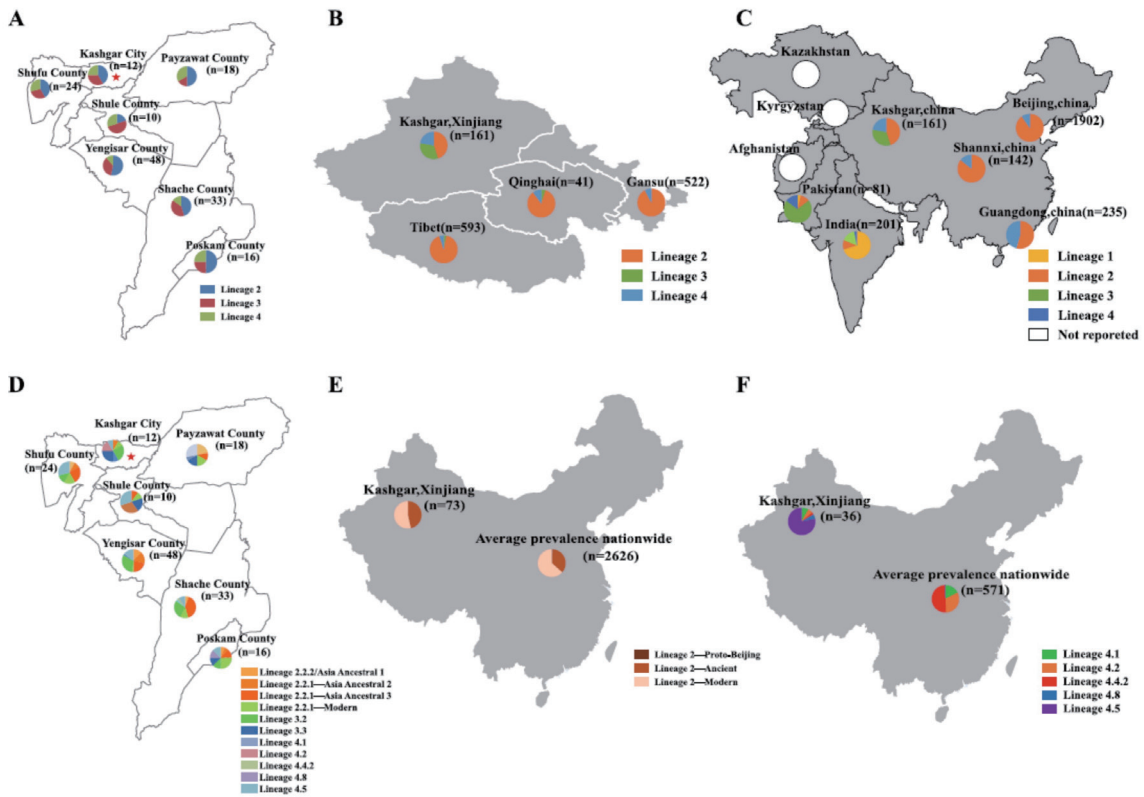


图2.结核分枝杆菌谱系和亚系的地理分布

总结

在本应用指南中，我们介绍华大智造组合产品在肺结核病原识别中的应用。该研究团队通过对喀什地区 161 株结核分枝杆菌的全基因组测序，将其分为了 3 个谱系和 11 个亚系，并鉴定出区域特异性的 SNP 位点。通过比较分析周边地区结合分枝杆菌的地理分布，认为该地区谱系 3 结核分枝杆菌主要由邻国传入。

在这一过程中，华大智造产品助力该项研究。通过 MGIEasy 酶切 DNA 文库制备套装可以快速构建 WGS 测序所需文库，结合 MGISEQ-2000 测序仪，可以同时大量样本完成测序。加之华大智造特有的 DNBSEQ 测序技术，减少测序过程中可能出现的错误，从而使 SNP 的识别更加准确。



基因测序仪 MGISEQ-2000RS

参考文献

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* .

推荐订购信息

产品类型	产品名称	产品货号
仪器	基因测序仪 MGISEQ-2000RS	900-000035-00
	MGISP-100RS 自动化样本制备系统	900-000070-00
	MGISP-960RS 自动化样本制备系统	900-000100-00
软件	MegaBOLT 生信分析加速器(工作站式服务器)	970-000085-00
建库试剂	MGI Easy 酶切 DNA 文库制备试剂套装(16 RXN)	1000006987
测序试剂	MGISEQ-2000RS 高通量测序试剂套装(FCL PE100)	1000012554

深圳华大智造科技股份有限公司

深圳市盐田区北山工业区综合楼11栋

☎ 4000-688-114

🌐 www.mgi-tech.com

✉ MGI-service@mgi-tech.com

股票简称：华大智造

股票代码：688114



仅供研究使用

版权声明：本手册版权属于深圳华大智造科技股份有限公司所有，未经本公司书面许可，任何其他个人或组织不得以任何形式将本手册中的各项内容进行复制、编辑或翻译为其他语言。本手册中所有商标或标识均属于深圳华大智造科技股份有限公司及其提供者所有。

版本：2022年10月版

撰稿：陈春伶 秦耀旭

责任编辑：王其伟

审稿：江遥