

全基因组测序技术在结核病流行病学调查中的应用

武洁 唐利红 杨崇广 严慧琴 孙华 沈鑫

200336 上海市疾病预防控制中心(武洁、沈鑫); 201101 上海市闵行区疾病预防控制中心(唐利红、严慧琴、孙华); 200032 上海, 复旦大学(杨崇广)

武洁、唐利红同为第一作者

通信作者: 沈鑫, Email: shenxin@scdc.sh.cn

DOI: 10.3760/cma.j.issn.0254-6450.2016.12.017

【摘要】 目的 评估全基因组测序技术在结核病分子流行病学调查中的应用。方法 对 2008—2012 年在上海市两家结核病定点医院发现 9 名耐多药患者中分离的结核分枝杆菌具有相同的可变数目串联重复序列, 本研究对此进行流行病学调查, 并对 9 株结核分枝杆菌进行全基因组测序, 分析其传播关系。结果 全基因组序列分析将 9 株结核分枝杆菌分为两个有传播关系的网络, 一个为包括 7 株结核分枝杆菌(5 例和 2 例患者分别来自不同的医院)的大簇, 一个为只有 2 株结核分枝杆菌的小簇。两个簇之间相差 15 个单核苷酸多态性(SNP)位点, 提示两个簇的遗传距离相对较远, 基于菌株 SNP 差异构建的传播链显示了每个簇内菌株的传播方向和耐药突变积累的过程。结论 基于全基因组测序数据研究耐药结核病的传播网络, 能准确判断传播路径和方向, 识别传染源和传播缺失环节。

【关键词】 结核病; 耐多药; 全基因组测序

基金项目: 国家科技重大专项(2013ZX10004903); 上海市第四轮公共卫生三年行动计划(15GWZK0801, GWTD2015S02)

Application of whole genome sequencing technology in the epidemiology of tuberculosis Wu Jie,

Tang Lihong, Yang Chongguang, Yan Huiqin, Sun Hua, Shen Xin

Shanghai Municipal Center for Disease Control and Prevention, Shanghai 100336, China (Wu J, Shen X);

Shanghai Minhang Center for Disease Control and Prevention, Shanghai 201101, China (Tang LH, Yan

HQ, Sun H); Fudan University, Shanghai 200032, China (Yang CG)

Wu Jie and Tang Lihong are the first authors who contributed equally to the article.

Corresponding author: Shen Xin, Email: shenxin@scdc.sh.cn

【Abstract】 **Objective** To delineate the application of whole genome sequencing technology in the epidemiology of tuberculosis. **Methods** From 2009 to 2012, nine *Mycobacterium tuberculosis* that sharing identical variable number of tandem repeats genotype (VNTR) patterns were reported from two TB cases designated hospitals. Both whole-genome sequencing analysis (WGS) and epidemiologic investigations were performed to describe the transmission patterns of these *Mycobacterium tuberculosis*. **Results** By WGS analysis, two genomic clusters including 7 and 2 *Mycobacterium tuberculosis* were noticed, respectively. The cluster of 2 cases possessed more than 15 single nucleotide polymorphisms (SNPs) when compared to the cluster of 7 cases and suggesting that the transmission route was independent. The transmission chain based on the SNPs difference showed the process of the propagation direction and the accumulation of drug resistance mutations in each cluster. **Conclusion** Using a WGS-based genomic epidemiologic approach, we were able to reconstruct the tuberculosis transmission network, tracing the putative source of the transmission and determining the transmission direction or the missing links.

【Key words】 Tuberculosis; Multidrug-resistance; Whole-genome sequencing

Fund programs: National Science and Technology Major Project of China (2013ZX10004903); The Fourth Round of Three-year Action Planning of Public Health in Shanghai (15GWZK0801, GWTD2015S02)

耐多药(MDR)结核病给全球结核病控制带来严重的挑战^[1-2]。目前我国耐药结核病控制形势十

分严峻, 耐药率及广泛耐药率均高于全球的平均水平^[3-4], MDR 病例的治愈率低、死亡率高^[5]。结核病

的潜伏期较长,传统的流行病学方法对其传染源的追踪和传播规律的认识有一定的局限。随着高通量测序技术的发展,应用全基因组测序技术为传染病的传播机制研究提供了新的检测手段^[6]。该方法目前已在多个结核病流行病学研究中使用^[7],显示了其独特的优势。2008—2012 年某结核病定点医院(A 医院)7 名患者与另一结核病定点医院(B 医院)2 名患者分离的结核分枝杆菌 VNTR 基因型(9+3 位点组合)具有高度同源性^[8],本研究对这 9 名患者分离的结核分枝杆菌进行全基因组测序,探讨全基因组测序技术在流行病学研究中的重要作用。

对象与方法

1. 流行病学调查:获得患者的知情同意后,对 2008—2012 年 A 医院 7 名患者与 B 医院 2 名患者进行调查,采集患者基本信息,包括姓名、年龄、住址、病史等,询问患者密切接触者情况以及日常活动范围。

2. 全基因组测序:在生物安全二级实验室中采用十六烷基三甲基溴化铵法从 9 名患者分离的结核分枝杆菌阳性培养物中提取基因组 DNA,80 °C 水浴 30 min 灭活,纯化后送至上海派森诺生物科技有限公司,依照 Illumina 公司双侧测序的标准流程构建插入长度约 300 bp 的测序文库。基于 Illumina Hiseq 2000 平台,实际测序深度为平均 100 倍以上,双侧测序数据采用 Bowtie2 和参考序列 H37Rv (GenBank AL123456) 比对,采用 SAMtools/BCFtools 比对分析得到菌株的单核苷酸多态性(SNP),分析菌株间的遗传差异时,排除基因组上高度重复区域(如 PE/PPE 家族基因)和耐药相关基因突变(受药物选择压力),簇病例间差别少于 12 个 SNP 位点以内定义为近期传播^[7]。基于 SNP 数据,采用 Median-joining 法构建可能的传播网络示意图。

结果

1. 基本情况:通过基因型分析发现有 9 株结核分枝杆菌具有相同的可变数目串联重复序列(VNTR)基因型,提示可能存在传播关系。资料显示 9 名患者均为上海市人,男性占 89%(8/9),其中 6 名为初治患者,3 名为复治患者。9 名患者中 3 名患者死亡,其中 1 名患者因其他疾病死亡(表 1)。

为了进一步研究这些病例之间是否存在接触或传播关系,本研究对患者进行流行病学问卷调查,调查结果显示,病例 1、2、3、4、5、6、8 目前就诊于 A 医院,病例 7、9 就诊于 B 医院。9 名患者确诊结核病前

表 1 9 名结核病患者现场流行病学调查资料

病例	性别	年龄	治疗史	医疗机构	利福平	乙胺丁醇	异烟肼	链霉素
1	男	67	初	A	耐药	敏感	耐药	耐药
2	男	41	初	A	耐药	敏感	耐药	耐药
3	男	59	初	A,C	耐药	耐药	耐药	耐药
4	男	57	初	A,C	耐药	耐药	耐药	耐药
5	男	54	复	A	耐药	敏感	耐药	耐药
6	女	61	初	A	耐药	敏感	耐药	耐药
7	男	41	复	B	耐药	耐药	耐药	耐药
8	男	47	初	A	耐药	敏感	耐药	耐药
9	男	84	复	B	耐药	耐药	耐药	耐药

注:病例 3 和 4 来自 C 医院

均因其他类疾病而分别入住于其他不同的医院,其中病例 3 与病例 4 来自于同一医院(C 医院)同一病房,病例 6 和病例 8 来自于 A 医院不同病房,以上两组患者存在流行病学联系。其余的 5 名患者分别来自 5 家不同的医疗机构,且这些医疗机构分属不同的行政辖区,未找到病例间明显的接触证据。从诊断时间来看(图 1)病例 2 最先发病,提示可能为传播事件的传染源。

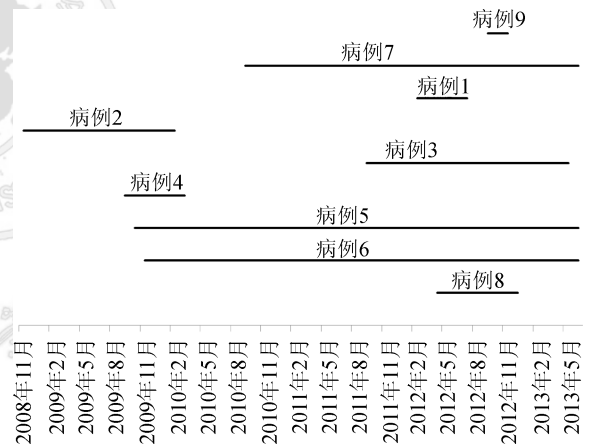


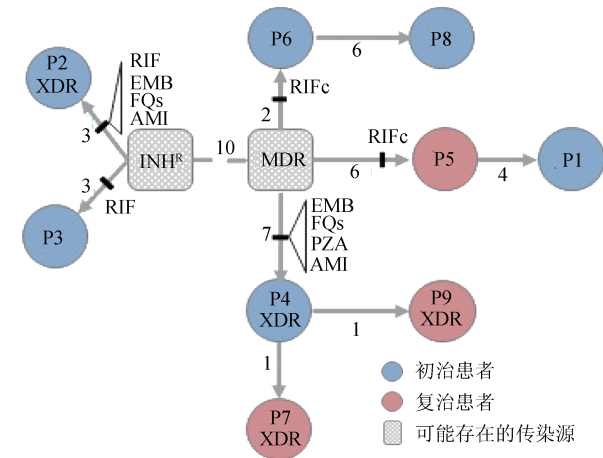
图 1 9 名结核病患者发病及治疗结束时间

2. 基于全基因组测序的相互关系:对 VNTR 分型成簇的 9 株结核分枝杆菌进行全基因组测序,根据菌株的 SNP 差异构建相互关系图,结果显示,9 株结核分枝杆菌分为两个相对独立的传播关系网络,其中 5 名 A 医院病例和 2 名 B 医院病例分离的结核分枝杆菌在一个大簇,病例 2 和病例 3 分离的结核分枝杆菌在另一个小簇中,两簇遗传距离最近的菌株之间遗传距离相差 15 个 SNP(图 2)。

进一步分析发现,大簇的 7 株结核分枝杆菌中,任一菌株与同簇其他菌株的最近遗传距离 < 9 个 SNP 位点,且 7 株结核分枝杆菌具有相同的异烟肼、利福平和链霉素相关耐药基因突变,提示其存在相互传播关系,且传染源是 MDR 患者。7 株结核分枝杆菌共形成了 3 个分支,表明存在至少 3 次传播事

件。病例 2 和病例 3 分离的结核分枝杆菌形成另一个小簇,与上述 7 株结核分枝杆菌不同的是,尽管这两株结核分枝杆菌均从 MDR 患者中分离,但由于利福平的耐药突变不同,提示传染源可能不是 MDR 患者,而是耐异烟肼的患者。

在传播过程 MDR 患者在不断地积累新的耐药,不同分支均有新的耐药突变发生,部分菌株发展为广泛耐药菌株,包括病例 4、7 和 9。且病例 4 有可能直接将广泛耐药结核(XDR)传染给病例 7 和 9(图 2)。



注:线上的数字代表 SNP; XDR 为广泛耐药结核; RI 为利福平; RIFc 表示 RIF 补偿性突变; EMB 为乙氨丁醇; FQs 为氟喹诺酮; AMI 为非典型结核分枝杆菌感染; INH^R 为异烟肼; PZA 为吡嗪酰胺; MDR 为耐药药

图 2 结核分枝杆菌全基因组测序结果

讨 论

通过结合全基因组测序技术和分子流行病学方法,本研究回顾性描述了一起 MDR 菌株传播事件。基于全基因组测序数据构建的传播网络,能比较准确鉴别传播途径和方向,识别传染源或传播过程中可能的缺失环节。根据本研究对基因组成簇(差异 ≤ 12 个 SNPs)的定义,9 株 VNTR 成簇的结核分枝杆菌经过全基因组测序分析分成了两个传播网络(7+2)。同时,传播网络呈星状分布,缺少“源菌株”,提示可能存在未发现的传染源。此外,在传播过程中部分菌株积累了新的耐药突变。

全基因组测序能在基因组水平揭示细菌在传播过程中的遗传变异,准确判定传播方向。比如加拿大哥伦比亚区的结核病暴发研究表明^[7],相同 VNTR 的结核分枝杆菌经全基因组测序分析显示是两起独立的传播事件。在本研究中, VNTR 成簇提示 9 名病例间存在传播关系,但流行病学调查发现除了病例 3 与病例 4、病例 6 和病例 8 之外,其他病例之间均不存在流行病学直接联系。由于结核病是一

种呼吸道传染病,患者之间存在偶然传播的可能性,不能排除在医院外或医疗场所的偶然接触传播导致的感染发病,而传统流行病学很难通过调查获得偶然传播证据,而常用的 VNTR 分型方法提供的信息量有限,不能进一步分析传播方向。全基因组测序分析则进一步提供了这些患者存在传播关系的证据,同时帮助判断传播方向和可能的传染源。此外,流行病学调查显示病例 3 与病例 4 之间、病例 6 与病例 8 之间存在直接联系,但是从全基因组测序的分析结果来看,病例 6 与病例 8 确属同一传播链;病例 3 与病例 4 不存在传播联系。

另外,本研究中不管是较大的传播链(簇),还是 2 例患者的小簇,通过全基因组测序分析,均未发现明确的“源”病例,说明传染源还未找到、传播链不够完整,另一种解释是源病例有较长时间的诊断延迟,在未被发现前已经传染多人,菌株在不同个体内积累突变,造成“源”病例缺失假象。同时,本研究也发现,在不同个体内,耐药菌株分别获得了不同类型的新耐药位点突变,提示个体的治疗可能存在不恰当或依从性不好等问题。

通过结合全基因组测序技术和分子流行病学方法,本研究从基因组和流行病学两方面描述一起 MDR 结核病传播事件。相比于传统的基因型分型方法,二代测序技术能够提供更全面的病原菌基因组特征,监测传播过程中遗传特征的变化,从而更准确鉴定传播途径。

利益冲突 无

参 考 文 献

- [1] 中国防痨协会结核病临床专业委员会. 结核病临床诊治进展年度报告(2013)(第一部分结核病临床诊断)[J]. 中国防痨杂志, 2014, 36(8): 714-740. DOI: 10.3969/j.issn.1000-6621.2014.08.023.
- [2] Chinese Antituberculosis Association of Clinic Society. Annual report on clinical diagnosis and treatment progress of tuberculosis (2013) (Part 1 clinical diagnosis)[J]. Chin J Antituberc, 2014, 36(8): 714-740. DOI: 10.3969/j.issn.1000-6621.2014.08.023.
- [3] Centers for Disease Control and Prevention. 2006. Emergence of *Mycobacterium tuberculosis* with extensive resistance to second-line drugs-worldwide, 2000-2004 [J]. MMWR Morb Mortal Wkly Rep. 2006, 55(11): 301-305.
- [4] World Health Organization. Global tuberculosis report 2013[M]. Geneva: World Health Organization, 2013.
- [5] 中华人民共和国卫生部. 全国结核病耐药性基线调查报告(2007-2008年)[M]. 北京:人民卫生出版社, 2010.
- [6] Ministry of Health, PRC. National tuberculosis drug resistance baseline survey report (2007-2008) [M]. Beijing: People's Medical Publishing House, 2010.
- [7] Cohn DL, Bustreo F, Raviglione MC. Drug-resistant tuberculosis: review of the worldwide situation and the WHO/IUATLD global surveillance project. International union against tuberculosis and lung disease [J]. Clin Infect Dis, 1997, 24 Suppl 1: S121-130.
- [8] Relman DA. Microbial genomics and infectious diseases [J]. N Engl J Med, 2011, 365(4): 347-357. DOI: 10.1056/NEJMr1003071.
- [9] Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study [J]. Lancet Infect Dis, 2013, 13(2): 137-146. DOI: 10.1016/S1473-3099(12)70277-3.
- [10] Luo T, Yang CG, Pang Y, et al. Development of a hierarchical variable-number tandem repeat typing scheme for *Mycobacterium tuberculosis* in China [J]. PLoS One, 2014, 9(2): e89726. DOI: 10.1371/journal.pone.0089726.

(收稿日期: 2016-06-29)
(本文编辑: 万玉立)