

二代测序技术在结核分枝杆菌研究中的应用进展

徐鹏,甘明宇,高谦

复旦大学基础医学院教育部/卫生部医学分子病毒学重点实验室,上海 200032

摘要:结核病是由结核分枝杆菌引起的全球第二大传染病。二代测序技术为从基因组水平研究结核分枝杆菌提供了重要的研究方法。本文从结核病流行病学、结核分枝杆菌耐药和进化及相关生物信息学等方面,介绍二代测序技术在结核分枝杆菌研究中的应用进展。

关键词:结核分枝杆菌;二代测序;基因组

Current applications of next-generation sequencing technology in *Mycobacterium tuberculosis* research

XU Peng, GAN Ming-Yu, GAO Qian

Key Laboratory of Medical Molecular Virology of Ministries of Education and Health, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China

Abstract: Tuberculosis caused by *Mycobacterium tuberculosis* (*M. tuberculosis*) is the second important infectious disease in the world, and next-generation sequencing technology provides an effective research method to study the genome of *M. tuberculosis*. The current applications of next-generation sequencing technology in the study of *M. tuberculosis* from the aspects of epidemiology, drug resistance, evolution and related bioinformatics are reviewed.

Key words: *Mycobacterium tuberculosis*; Next-generation sequencing; Genome

结核病(tuberculosis, TB)是由结核分枝杆菌复合群(*Mycobacterium tuberculosis* complex)感染引起的慢性传染病。2012年^[1]全球新发结核病患者约860万例,死亡130万例。近年来,耐药结核病特别是耐多药结核病(multidrug-resistant TB, MDR-TB)、广泛耐药结核病(extensively drug-resistant TB, XDR-TB)及全耐药结核病(totally drug-resistant TB, TDR-TB)出现并日益增多,使结核病可能回到无药可治的时代。目前约3.7%的新发病例和20%的复治病例为MDR-TB^[2]; MDR-TB病例中,约10%为XDR-TB,全球已有84个国家报道了XDR-TB病例。对结核病特别是耐药结核病的

有效防控有赖于对结核分枝杆菌耐药机制、致病机制和传播规律的深入研究。基因组是结核分枝杆菌传播、致病、耐药和进化的遗传基础,对其全面认识是结核分枝杆菌研究的重要组成。但由于一代测序方法成本高、效率低,且错误较多,限制了结核病全基因组水平研究的全面开展。新一代高通量测序技术〔二代测序技术(next-generation sequencing, NGS)〕的发展为从全基因组水平认识结核病提供了准确、高效的方法,开启了从片面基因分型到全基因组研究的大门,为结核病的研究提供了新的角度和方法。

1998年全球第1株结核分枝杆菌标准菌株

通信作者:高谦

Corresponding author. GAO Qian, E-mail: qiangao@shmu.edu.cn

H37Rv 全基因组测序(whole genome sequencing, WGS)工作完成^[3],4年后临床分离株 CDC1551 全基因组测序数据公布^[4]。此后,2002~2009 年仅有 9 株结核分枝杆菌全基因组数据上传至美国国立生物技术信息中心(National Center for Biotechnology Information,NCBI)数据库。2005 年底,454 公司推出第 1 个基于焦磷酸测序原理的高通量基因组测序系统,标志着二代测序平台的诞生,这是核酸测序技术发展史上里程碑式的事件,使生物研究有望全面进入全基因组时代。随着高通量各种组学(omics)技术不断发展,生物学研究已进入大数据(big data)时代。2010 年以来,NCBI 数据库中结核分枝杆菌全基因组及二代测序数据迅速增长(图 1),为结核病研究提供了重要的数据。本文就二代测序方法在结核病分子流行病学及结核分枝杆菌耐药等方面的研究进展进行综述。

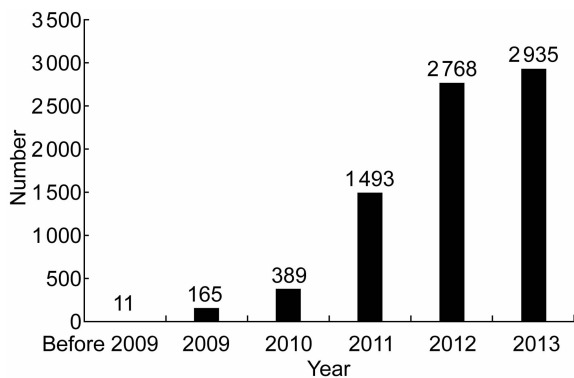


图 1 NCBI 数据库中结核分枝杆菌全基因组及二代测序数量

Fig. 1 The number of *Mycobacterium tuberculosis* genome and next-generation sequencing data in NCBI database

1 结核病流行病学研究

传统流行病学调查和分子流行病学分型是研究结核病传播规律的重要工具。尽管在过去 20 年中,基于不同分子标记的基因型分型技术极大加深了对结核病传播的认识,但 Niemann 等^[5]对 2 株北京基因型菌株,应用 IS6110 限制性片段长度多态性(IS6110 restriction fragment length polymorphism, IS6110 RFLP)分析、24 位点可变数目串联重复(variable number tandem repeat, VNTR)和间隔区寡核苷酸分型(spoligotyping)技术等流行病学分型方法与全基因组单核苷酸多态性(single nucleotide polymorphism, SNP)分析比较发现,即便是传统基因分型定义的成簇菌株,相互之间仍存在较大的差

异。另一方面,尽管传统基因型分型方法可鉴定传播链,但很难准确鉴别传播发生的路径、顺序及是否存在多个传播源,而通过全基因组数据分析可解决此类问题。例如,在加拿大对 34 株结核病社区暴发分离株进行流行病学研究^[6],全基因组序列分析发现, VNTR 分型相同的 34 株仍可分为基因型不同的 2 个簇,存在不同的传播源和传播路径。这些研究显示,在结核病流行病学研究领域,全基因组技术较先前的基因分型方法有优势。MacLean 等^[7]认为,全基因组测序将来可能成为分子流行病学研究的金标准。Walker 等^[8]也认为,全基因组测序可取代传统的流行病学调查,2 个菌株基因组差异 ≤ 5 SNP 便可认为存在传播关系。Medini 等^[9]同样认可全基因组测序的这些优势,其有能力取代目前标准的细菌基因分型方法。

借助生物信息学和计算机分析,可在二代测序数据与传统基因分型之间建立联系。Coll 等^[10]编写程序,通过全基因组数据分析结核分枝杆菌 spoligotype 基因型,实现了二代测序或全基因组数据向传统基因分型结果的转换。IS6110 RFLP 分型方法曾被认为是结核分枝杆菌分子流行病学研究的金标准,但其操作复杂,费时费力。借助二代测序技术,Reyes 等^[11]建立了 IS6110 RFLP 高通量分型方法,可同时对几百株结核分枝杆菌进行分型,极大提高了分型效率。

全基因组测序可确定传播链中各菌株间的变异和传播源。在 IS6110 RFLP、spoligotype 或 VNTR 分析基因型相同的情况下,全基因组分析也可提供更高的分辨率,更准确地区别近期传播或相关菌株的复发。该方法目前已在多项结核病流行病学研究中使用^[12,13],显示了其优势。

2 耐药结核分枝杆菌研究

耐药结核病,特别是 MDR-TB、XDR-TB,是结核病防控工作中亟待解决的问题。虽然在结核分枝杆菌耐药研究方面已取得较大进展,发现了许多与耐药相关的基因及点突变^[14],但许多耐药现象仍有待深入阐释。

人们借助二代测序和比较基因组学等方法,大大提高了发现耐药基因和突变位点的效率。Zhang 等^[15]对 161 株中国分离株进行基因组比较发现, Rv0026 等 72 个新基因、Rv0010c-Rv0011c 等 28 个基因间隔区(intergenic region, IGR),以及 11 个非同义 SNP 和 10 个 IGR SNP 可能与耐药相关。

在 MDR-TB 和 XDR-TB 全基因组数据的基础上,通过直系同源簇(cluster of orthologous group, COG)数据库, Wu 等^[16]发现北京家族菌株在细胞膜、脂质运输和防御机制等方面的相关基因存在大量突变。对刚上市或正在临床评估的新药,二代测序技术在其耐药机制研究方面也发挥重要作用。通过人工筛选的耐药菌株与敏感菌株全基因组序列对比,可有效发现耐药突变和耐药基因,进而对其耐药机制进行研究。基于该方法,人们发现 *mmpL3* 基因突变可对抗结核新药 SQ109 产生耐药^[17],而 *rplC* T460C 突变与耐利奈唑胺有关^[18]。结核分枝杆菌耐药主要是由于耐药点突变,传统上认为这些突变主要是基因的非同义突变和启动区的突变。基因的同义突变由于不产生氨基酸改变,容易被忽视;但同义突变可影响蛋白质翻译效率和准确率^[19]。另外, Wang 等^[20]研究发现,结核分枝杆菌可能存在大量选择性优势的同义突变。近期研究^[21]发现, *Rv3792* 基因的同义 SNP 突变可增加下游基因 *embC* 表达,从而引起对乙胺丁醇的耐药性增高,扩展了人们对结核分枝杆菌耐药机制的认识。

仅依靠传统分子流行病学方法和药敏试验,很难确定耐药表型相同的成簇菌株耐药的产生是由于化学药物治疗产生的获得性耐药,还是耐药菌株传播产生的原发性耐药。在研究南非地区近期暴发的耐药结核病传播中, Ioerger 等^[22]通过全基因组比较发现,当地 XDR-TB 是通过逐步获得耐药性而形成的,并非 XDR-TB 传播。在中国, Lin 等^[23]对 2 株 XDR-TB 全基因组分析后也得出相似结论,认为广泛耐药菌株产生的原因是单个药物耐药逐步累积的结果。这些研究提示, XDR-TB 在获得耐药能力的同时,可能存在相对较低的适应性,在疾病传播方面处于劣势。人们普遍认为耐药突变可增加结核分枝杆菌适应成本(fitness cost),但耐药结核病特别是北京家族 MDR-TB 的传播优势一直很难解释,直至 Comas 等^[24]发现利福平耐药菌株的补偿性突变(compensatory mutation)。该突变可弥补或减少耐药突变对结核分枝杆菌造成的适应成本。另外,结核分枝杆菌基因组在体内的进化情况一直是人们较为关心的问题。Saunders 等^[25]对 1 例从全敏感到耐多药病例前后间隔 12 个月的菌株全基因组比较发现,仅存在 *katG* S315T 和 *rpoB* D516Y 耐药突变,显示了结核分枝杆菌基因组的稳定。但 Sun 等^[26]通过对 3 例患者的 7 株不同阶段结核分枝杆菌样本深度测序,发现结核分枝杆菌存在较大的

群体多样性,且不同耐药突变菌株间可能存在相互竞争,最终耐药水平高、适应成本小的基因型被固定下来,这加深了对结核分枝杆菌群体多样性的认识。

3 结核分枝杆菌进化与基因组变异研究

结核分枝杆菌是放线菌目(Actinomycetales)分枝杆菌属(*Mycobacterium*)的成员,其起源与进化过程一直是人们探究的问题之一。McGuire 等^[27]比较了分枝杆菌及其相关放线菌全基因组,发现与其他放线菌相比,结核分枝杆菌的脂类代谢、DNA 修复等基因存在较大差异,推测这些基因变异与其从环境菌向致病菌进化相关。光滑结节细菌(smooth tubercle bacilli, STB)与结核分枝杆菌同属于分枝杆菌属,可在培养基上呈现特征性的光滑菌落。Supply 等^[28]对结核分枝杆菌和 STB 全基因组进行系统发育分析发现, STB 起源更早,基因组更大,有较高的基因重组和基因变异,较少的分子瘢痕(molecular scar)和明显的 CRISPR-Cas 系统。动物实验显示, STB 与结核分枝杆菌相比,持续感染能力和毒力减小。尽管有这些差异, STB 与结核分枝杆菌存在共同的保守核心基因,因此作者认为结核分枝杆菌起源于 STB-like 菌株,并在进化过程中获得持续感染能力和更高毒力。在排除结核分枝杆菌起源于牛分枝杆菌(*M. bovis*)后^[29],结核病的起源问题一直困扰学界多年。Comas 等^[30]通过分析全球 259 株结核分枝杆菌复合群全基因组多样性,构建结核分枝杆菌复合群的进化史,结果表明结核分枝杆菌早在 7 万年前就存在于非洲的现代人类中,并伴随人类走出非洲而传播至全球,在此过程中细菌继续进化以适应人口密度的变化。在分析 21 株全球分离的结核分枝杆菌复合群全基因组过程中, Comas 等^[31]发现必需基因的保守性较非必需基因高,特别是 491 个人 T 细胞表位更为保守。这种纯化选择提示,在进化过程中被 T 细胞识别有利于结核分枝杆菌感染。结核分枝杆菌在进化过程中获得感染人类的能力后,不同进化分支的差异也是人们关注的研究议题。Schürch 等^[32]研究发现,北京家族菌株在编码调控网络的基因中存在非同义核苷酸突变,推测这些改变增加了北京型菌株对环境的适应性,使其成为主要流行菌株。Kato-Maeda 等^[33]通过比较基因组学和豚鼠实验,发现北京家族 RD207 亚型致病性更高,可导致严重肺损伤,并诱导调节性 T 细胞高水平表达和肺内效应 T 细胞显著减少。基因组分析显示其存在与致病性相关的突

变,提示结核分枝杆菌不同亚型存在致病性和临床表型的差异,可影响其在人群中的传播能力。

结核分枝杆菌基因组突变速率直接影响其进化和适应能力。对同一菌株不同时间的分离株,应用比较全基因组学方法可估算结核分枝杆菌的突变速率。Yang等^[34]对前后间隔30多年的结核分枝杆菌进行研究,显示微卫星位点变异较小,推测在潜伏结核感染(latent tuberculosis infection, LTBI)状态细菌复制水平较低。传统理论认为,潜伏感染与活动期相比,结核分枝杆菌的复制水平较低直接导致突变率降低。但Ford等^[35]通过猕猴实验,证明潜伏感染的结核分枝杆菌与活动期相比存在相同的突变率。同样,研究结核分枝杆菌微进化过程,比较人体与传播链中2个阶段的突变水平,发现结核分枝杆菌突变率无差异^[36]。这些研究提示,结核分枝杆菌的突变速率较为恒定。目前报道显示^[13,37],结核分枝杆菌的SNP进化速率约为每年0.5 SNP/基因组。但Ford等^[38]研究发现,结核分枝杆菌东亚分支比欧美分支更易产生耐药,认为该现象并不是由于适应抗生素压力的能力增强,而可能是突变率更高所致。结核分枝杆菌不同基因家族的突变率是否存在差异,有待进一步研究。

人们普遍认为,结核分枝杆菌基因组多样性主要表现为SNP和小片段缺失,大片段改变较少见。但Namouchi等^[39]分析24个结核分枝杆菌复合群基因组,检测水平基因转移(horizontal genetic transfer, HGT),认为结核分枝杆菌复合群中主要存在约50 bp短DNA片段的重组,同时也有>150 kb DNA片段的重组痕迹。近期研究显示^[40, 41],结核分枝杆菌基因组中3.7 Mb位置附近存在约300 kb大片段DNA重排现象,认为可能是该位置不稳定和(或)存在选择性优势所致。

4 生物信息学在全基因组分析中的应用

二代测序方法测序深度较高,往往产生百倍甚至千倍于基因组的数据,因此对这些结果的分析 and 研究离不开计算机软件的辅助分析和生物信息学的应用。目前运用较广泛的组装软件有Velvet、AbySS和SOAPdenovo,这些软件都基于de Bruijn图(de Bruijn graph)的算法^[42]。Velvet对基因组相对较小(<40 Mb)物种的组装效果较好^[43]。而基于BWT数据转换(Burrows-Wheeler transform)算法发展起来的一系列比对软件,如BWA、Bowtie、SOAPaligner/SOAP2,为大规模测序且与参考序列

进行高效、精确的比对提供了解决方案。目前,这些软件均在结核分枝杆菌的全基因组分析组装中发挥重要作用。例如,对于结核分枝杆菌Erdman菌株, Miyoshi-Akiyama等^[44]用MAQ软件对其基因组进行组装;而对北京家族CCDC5079和CCDC5080菌株,则采用Velvet方法^[45]。

在全基因组数据的基础上对结核分枝杆菌深入研究离不开对基因结构和功能的分析;同时,全基因组数据公布的不断增加及生物信息学方法的应用也促进了人们对结核分枝杆菌基因研究的深入。对基因组的分析和预测,是随着认识的提高及不断纠正错误而逐步走向完善的。2010年结核分枝杆菌基因再注释工程启动,目标是实质性改进和完善结核分枝杆菌基因组的注释^[46]。基因组注释是在基因组水平确定基因及其他元件的位置和结构,并赋予这些基因和元件生物功能的过程。结核分枝杆菌H37Rv是重要的实验室参考株。DeJesus等^[47]对其989个开放读码框架(open reading frame, ORF)重新进行全面的生物信息学分析,纠正先前转录起始位的错误注释。而Doerks等^[48]对H37Rv未知或假定功能的基因进行分析和预测,增加了对其基因组和蛋白组的了解。Xu等^[49]通过文献计量分析方法(bibliometric approach)在基因组水平鉴定了H37Rv的必需基因。按Stein^[50]的观点,基因组的注释分为3个层次:核酸水平、蛋白质水平及代谢水平。利用蛋白质基因组学(proteogenomics)可有效在蛋白质水平对基因组注释。“蛋白质基因组学”一词由Jaffe等于2004年首次提出,是指利用蛋白质组学数据,结合基因组数据(DNA)、转录组数据(RNA)来研究基因组注释问题^[51]。Kelkar等^[52]应用高分辨质谱研究结核分枝杆菌的蛋白质基因组学,发现H37Rv共有3 176个蛋白,其中41个为未知基因编码的新蛋白质。除H37Rv以外,Erdman菌株、43-16836菌株等临床分离株的全基因组注释工作已报道。结核分枝杆菌标准菌株和临床分离株基因组注释工作的完成和不断优化,为结核分枝杆菌的功能基因组学等相关研究提供了基础。

在基因组注释的基础上,基因组水平蛋白质功能预测和分析也是生物信息学运用的重要领域。生物信息学工具可用于寻找结核分枝杆菌的潜在药物靶点。Mazandu等^[53]对结核分枝杆菌功能基因组分析,鉴定其生存、生长和毒力相关的必需基因,以及潜在的药物靶点。Zvi等^[54]对H37Rv的3 989个ORF产物抗原性进行分析,寻找到45个最有可能

成为疫苗的蛋白质。蛋白质结构是推测蛋白质功能的重要依据, Anand 等^[55]通过计算机分析方法, 在基因组水平注释蛋白的结构, 建立了基因与蛋白质结构的联系。

5 结语

二代测序技术是高通量测序方法, 虽然存在一些缺点, 如数据分析复杂、结核分枝杆菌的 PE/PPE 等重复基因仍需进行传统方法的测序; 但与一代测序相比, 准确率提高, 成本显著降低。随着二代测序技术的普遍应用, 以及三代测序技术的逐步发展和完善, 全基因组测序有望实现低成本与高效率兼得。通过这些技术而获得的基因组信息, 不仅为结核分枝杆菌的基因组学研究提供重要的研究对象, 还可为下游各种组学研究提供重要信息, 增加对结核病及结核分枝杆菌认识的广度和深度。

参考文献

- [1] Eurosurveillance Editorial Team. WHO publishes global tuberculosis report 2013 [J]. *Euro Surveill*, 2013, 18(43):20615.
- [2] Hartkoorn RC, Sala C, Uplekar S, Busso P, Rougemont J, Cole ST. Genome-wide definition of the SigF regulon in *Mycobacterium tuberculosis* [J]. *J Bacteriol*, 2012, 194(8): 2001-2009.
- [3] Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence [J]. *Nature*, 1998, 393(6685): 537-544.
- [4] Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs WR Jr, Venter JC, Fraser CM. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains [J]. *J Bacteriol*, 2002, 184(19): 5479-5490.
- [5] Niemann S, Köser CU, Gagneux S, Plinke C, Homolka S, Bignell H, Carter RJ, Cheetham RK, Cox A, Gormley NA, Kokko-Gonzales P, Murray LJ, Rigatti R, Smith VP, Arends FP, Cox HS, Smith G, Archer JA. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints [J]. *PLoS One*, 2009, 4(10): e7407.
- [6] Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak [J]. *New Engl J Med*, 2011, 364(8): 730-739.
- [7] MacLean D, Jones JD, Studholme DJ. Application of 'next-generation' sequencing technologies to microbial genetics [J]. *Nat Rev Microbiol*, 2009, 7(4): 287-296.
- [8] Walker TM, Monk P, Smith EG, Peto TE. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing [J]. *Clin Microbiol Infect*, 2013, 19(9): 796-802.
- [9] Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, Falkow S, Rappuoli R. Microbiology in the post-genomic era [J]. *Nat Rev Microbiol*, 2008, 6(6): 419-430.
- [10] Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, Martin N, Clark TG. SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences [J]. *Bioinformatics*, 2012, 28(22): 2991-2993.
- [11] Reyes A, Sandoval A, Cubillos-Ruiz A, Varley KE, Hernández-Neuta I, Samper S, Martín C, García MJ, Ritacco V, López L, Robledo J, Zambrano MM, Mitra RD, Del Portillo P. IS-seq: a novel high throughput survey of in vivo IS6110 transposition in multiple *Mycobacterium tuberculosis* genomes [J]. *BMC Genomics*, 2012, 13: 249.
- [12] Schürch AC, Kremer K, Daviena O, Kiers A, Boeree MJ, Siezen RJ, van Soolingen D. High-resolution typing by integration of genome sequencing data in a large tuberculosis cluster [J]. *J Clin Microbiol*, 2010, 48(9): 3403-3406.
- [13] Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study [J]. *Lancet Infect Dis*, 2013, 13(2): 137-146.
- [14] Almeida Da Silva PE, Palomino JC. Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs [J]. *J Antimicrob Chemother*, 2011, 66(7): 1417-1430.
- [15] Zhang H, Li D, Zhao L, Fleming J, Lin N, Wang T, Liu Z, Li C, Galwey N, Deng J, Zhou Y, Zhu Y, Gao Y, Wang T, Wang S, Huang Y, Wang M, Zhong Q, Zhou L, Chen T, Zhou J, Yang R, Zhu G, Hang H, Zhang J, Li F, Wan K, Wang J, Zhang XE, Bi L. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance [J]. *Nat Genet*, 2013, 45(10): 1255-1260.
- [16] Wu W, Zheng H, Zhang L, Wen Z, Zhang S, Pei H, Yu G, Zhu Y, Cui Z, Hu Z, Wang H, Li Y. A genome-wide analysis of multidrug-resistant and extensively drug-resistant strains of *Mycobacterium tuberculosis* Beijing genotype [J].

- Mol Genet Genomics, 2013, 288(9): 425-436.
- [17] Tahlan K, Wilson R, Kastrinsky DB, Arora K, Nair V, Fischer E, Barnes SW, Walker JR, Alland D, Barry CE 3rd, Boshoff HI. SQ109 targets MmpL3, a membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis* [J]. *Antimicrob Agents Chemother*, 2012, 56(4): 1797-1809.
- [18] Beckert P, Hillemann D, Kohl TA, Kalinowski J, Richter E, Niemann S, Feuerriegel S. rplC T460C identified as a dominant mutation in linezolid-resistant *Mycobacterium tuberculosis* strains [J]. *Antimicrob Agents Chemother*, 2012, 56(5): 2743-2745.
- [19] Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS [J]. *PLoS Genet*, 2008, 4(12): e1000304.
- [20] Wang TC, Chen FC. The evolutionary landscape of the *Mycobacterium tuberculosis* genome [J]. *Gene*, 2013, 518(1): 187-193.
- [21] Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, McNeil M, Peterson S N, Chatterjee D, Fleischmann R, Alland D. Evolution of high-level ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-beta-D-arabinose biosynthetic and utilization pathway genes [J]. *Nat Genet*, 2013, 45(10): 1190-1197.
- [22] Ioerger TR, Feng Y, Chen X, Dobos KM, Victor TC, Streicher EM, Warren RM, Gey van Pittius NC, Van Helden PD, Sacchettini JC. The non-clonality of drug resistance in Beijing-genotype isolates of *Mycobacterium tuberculosis* from the Western Cape of South Africa [J]. *BMC Genomics*, 2010, 11: 670.
- [23] Lin N, Liu Z, Zhou J, Wang S, Fleming J. Draft genome sequences of two super-XDR isolates of *M. tuberculosis* from China [J]. *FEMS Microbiol Lett*, 2013, 347(2): 93-96.
- [24] Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes [J]. *Nat Genet*, 2012, 44(1): 106-110.
- [25] Saunders NJ, Trivedi UH, Thomson ML, Doig C, Laurenson IF, Blaxter ML. Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background [J]. *J Infect*, 2011, 62(3): 212-217.
- [26] Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE 3rd, Mei J, Gao Q. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients [J]. *J Infect Dis*, 2012, 206(11): 1724-1733.
- [27] McGuire AM, Weiner B, Park ST, Wapinski I, Raman S, Dolganov G, Peterson M, Riley R, Zucker J, Abeel T, White J, Sisk P, Stolte C, Koehrsen M, Yamamoto RT, Iacobelli-Martinez M, Kidd MJ, Maer AM, Schoolnik GK, Regev A, Galagan J. Comparative analysis of *Mycobacterium* and related Actinomycetes yields insight into the evolution of *Mycobacterium tuberculosis* pathogenesis [J]. *BMC Genomics*, 2012, 13: 120.
- [28] Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Loch C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Médigue C, Parkhill J, Cruveiller S, Brosch R. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis* [J]. *Nat Genet*, 2013, 45(2): 172-179.
- [29] Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D, Cole ST. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex [J]. *Proc Natl Acad Sci USA*, 2002, 99(6): 3684-3689.
- [30] Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans [J]. *Nat Genet*, 2013, 45(10): 1176-1182.
- [31] Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, Ernst JD, Gagneux S. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved [J]. *Nat Genet*, 2010, 42(6): 498-503.
- [32] Schürch AC, Kremer K, Warren RM, Hung NV, Zhao Y, Wan K, Boeree MJ, Siezen RJ, Smith NH, van Soolingen D. Mutations in the regulatory network underlie the recent clonal expansion of a dominant subclone of the *Mycobacterium tuberculosis* Beijing genotype [J]. *Infect Genet Evol*, 2011, 11(3): 587-597.
- [33] Kato-Maeda M, Shanley CA, Ackart D, Jarlsberg LG, Shang S, Obregon-Henao A, Harton M, Basaraba RJ, Henao-Tamayo M, Barrozo JC, Rose J, Kawamura LM, Coscolla M, Fofanov VY, Koshinsky H, Gagneux S, Hopewell PC, Ordway DJ, Orme IM. Beijing sublineages of *Mycobacterium tuberculosis* differ in pathogenicity in the guinea pig [J]. *Clin Vaccine Immunol*, 2012, 19(8): 1227-1237.
- [34] Yang Z, Rosenthal M, Rosenberg NA, Talarico S, Zhang L, Marrs C, Thomsen VØ, Lilleback T, Andersen AB. How dormant is *Mycobacterium tuberculosis* during latency? A study integrating genomics and molecular epidemiology [J]. *Infect Genet Evol*, 2011, 11(5): 1164-1167.
- [35] Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM. Use of whole genome

- sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection [J]. *Nat Genet*, 2011, 43(5): 482-486.
- [36] Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, García-de-Viedma D. Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*; potential impact on the inference of tuberculosis transmission [J]. *J Infect Dis*, 2014, 209(1): 98-108.
- [37] Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rüscher-Gerdes S, Supply P, Kalinowski J, Niemann S. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study [J]. *PLoS Med*, 2013, 10(2): e1001387.
- [38] Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, Johnston JC, Gardy J, Lipsitch M, Fortune SM. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis [J]. *Nat Genet*, 2013, 45(7): 784-790.
- [39] Namouchi A, Didelot X, Schöck U, Gicquel B, Rocha EP. After the bottleneck: Genome-wide diversification of the *Mycobacterium tuberculosis* complex by mutation, recombination, and natural selection [J]. *Genome Res*, 2012, 22(4): 721-734.
- [40] Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, Plikaytis BB, Posey JE, van Helden PD, Gey van Pittius NC, Koehrsen M, Sisk P, Stolte C, White J, Gagneux S, Birren B, Hung D, Murray M, Galagan J. Independent large scale duplications in multiple *M. tuberculosis* lineages overlapping the same genomic region [J]. *PLoS One*, 2012, 7(2): e26038.
- [41] Domenech P, Kolly GS, Leon-Solis L, Fallow A, Reed MB. Massive gene duplication event among clinical isolates of the *Mycobacterium tuberculosis* W/Beijing family [J]. *J Bacteriol*, 2010, 192(18): 4562-4570.
- [42] Lee HC, Lai K, Lorenc MT, Imelfort M, Duran C, Edwards D. Bioinformatics tools and databases for analysis of next-generation sequence data [J]. *Brief Funct Genomics*, 2012, 11(1): 12-24.
- [43] Paszkiewicz K, Studholme DJ. De novo assembly of short sequence reads [J]. *Brief Bioinform*, 2010, 11(5): 457-472.
- [44] Miyoshi-Akiyama T, Matsumura K, Iwai H, Funatogawa K, Kirikae T. Complete annotated genome sequence of *Mycobacterium tuberculosis* Erdman [J]. *J Bacteriol*, 2012, 194(10): 2770.
- [45] Zhang Y, Chen C, Liu J, Deng H, Pan A, Zhang L, Zhao X, Huang M, Lu B, Dong H, Du P, Chen W, Wan K. Complete genome sequences of *Mycobacterium tuberculosis* strains CCDC5079 and CCDC5080, which belong to the Beijing family [J]. *J Bacteriol*, 2011, 193(19): 5591-5592.
- [46] Brennan PJ, Brosch R, Birren B, Sobral B. TBCAP; tuberculosis annotation project [J]. *Tuberculosis (Edinb)*, 2013, 93(1): 1-5.
- [47] DeJesus MA, Sacchetti JC, Ioerger TR. Reannotation of translational start sites in the genome of *Mycobacterium tuberculosis* [J]. *Tuberculosis (Edinb)*, 2013, 93(1): 18-25.
- [48] Doerks T, van Noort V, Minguéz P, Bork P. Annotation of the *M. tuberculosis* hypothetical orfome: adding functional information to more than half of the uncharacterized proteins [J]. *PLoS One*, 2012, 7(4): e34302.
- [49] Xu G, Liu B, Wang F, Wei C, Zhang Y, Sheng J, Wang G, Li F. High-throughput screen of essential gene modules in *Mycobacterium tuberculosis*; a bibliometric approach [J]. *BMC Infect Dis*, 2013, 13: 227.
- [50] Stein L. Genome annotation: from sequence to biology [J]. *Nat Rev Genet*, 2001, 2(7): 493-503.
- [51] Renuse S, Chaerkady R, Pandey A. Proteogenomics [J]. *Proteomics*, 2011, 11(4): 620-630.
- [52] Kelkar DS, Kumar D, Kumar P, Balakrishnan L, Muthusamy B, Yadav AK, Shrivastava P, Marimuthu A, Anand S, Sundaram H, Kingsbury R, Harsha HC, Nair B, Prasad TS, Chauhan DS, Katoch K, Katoch VM, Kumar P, Chaerkady R, Ramachandran S, Dash D, Pandey A. Proteogenomic analysis of *Mycobacterium tuberculosis* by high resolution mass spectrometry [J]. *Mol Cell Proteomics*, 2011, 10(12): M111.011627.
- [53] Mazandu GK, Mulder NJ. Generation and analysis of large-scale data-driven *Mycobacterium tuberculosis* functional networks for drug target identification [J]. *Adv Bioinformatics*, 2011, 2011: 801478.
- [54] Zvi A, Ariel N, Fulkerson J, Sadoff JC, Shafferman A. Whole genome identification of *Mycobacterium tuberculosis* vaccine candidates by comprehensive data mining and bioinformatic analyses [J]. *BMC Med Genomics*, 2008, 1: 18.
- [55] Anand P, Sankaran S, Mukherjee S, Yeturu K, Laskowski R, Bhardwaj A, Bhagavat R; OSDD Consortium, Brahmachari SK, Chandra N. Structural annotation of *Mycobacterium tuberculosis* proteome [J]. *PLoS One*, 2011, 6(10): e27044.

(收稿日期:2014-03-10)