

结核分枝杆菌功能基因组研究的方法学

谢建平^{1,2*} 乐 军¹ 王洪海^{1**}

(复旦大学生命科学院 上海 200433)¹

(西南师范大学生物科学学院 重庆 400715)²

摘要: DNA 芯片、基因中断、基因互补法、差示荧光诱导法、生物信息学、蛋白质组、DNA 库免疫、细菌人工染色体库是结核分枝杆菌研究的主要方法。综合利用上述技术可能是我们开展结核分枝杆菌功能基因组学的最佳切入点。

关键词: 结核分枝杆菌, 耐药机制, 功能基因组学, 细菌人工染色体库, 蛋白质组

中图分类号: Q93 **文献标识码:** A **文章编号:** 0253-2654 (2001) 05-0092-06

结核病 (Tuberculosis-TB) 仍然是全球人类健康的主要威胁。全球被结核分枝杆菌 (*Mycobacterium tuberculosis* 以下简称 MTB) 感染者约 17~20 亿, 结核病人约 2000 万; 每年至少 3000 万人感染 MTB, 约 800 万新结核病患者, 约 200 万人死于结核病; 耐药 (Drug resistance DR) 和耐多药 (Multidrug resistance, MDR) MTB 迅速扩散, 难治疗结核病的范围在扩大, 爱滋病泛滥等使问题更加严峻, 唯一的预防性疫苗 (卡介苗) 的效果不稳定。为数甚少的药物多是 30 年前发现, 且 MTB 对所有一线药物均已产生了不同程度的耐药性。全球急需对付 MTB 的有效措施, 包括预防性和治疗性疫苗、新型药物。我国的问题尤其严峻, 因为中国是世界上除印度以外的第二大结核病重灾区。

结核病的发生、发展是 MTB 和人之间不断相互作用的结果。其相互作用研究至少

* 微生物学博士研究生

** 联系作者

收稿日期: 2000-04-26 修回日期: 2000-10-08

应包括两个方面：致病菌的毒力因子、逃避宿主免疫和耐药的机制、宿主的防御和防御失败机制。这也是世界研究的热点和难点。最终根除结核病很大程度上依赖于方法学的突破。本文综述近年来 MTB 研究的方法学进展。

1 基因组概况^[2,3]

MTB (H37Rv) 全基因组序列已经测定, 这是继上世纪发现结核致病菌 MTB 后又一里程碑。MTB 全基因组 4411529bp, G + C% 65.6%。(GTG 为起始密码的占 35%, 而枯草芽孢杆菌 9%, 大肠杆菌 14%; 相关性分析表明富含 GC 的密码子编码的氨基酸, 如丙氨酸, 甘氨酸, 脯氨酸, 精氨酸, 色氨酸较多; 富含 AT 的密码子编码的氨基酸, 如天冬氨酸, 异亮氨酸, 赖氨酸, 苯丙氨酸, 酪氨酸较少; tRNA 反密码子第一位较少出现腺嘌呤); 约 3924 个开放阅读框, 占整个基因组编码容量的 91%, 其中大部分用于编码脂类代谢酶 (250 种, 大肠杆菌仅 50 种), 10% 编码两类大的、不相关的富含甘氨酸的酸性蛋白质, 可能与产生抗原变异, 逃避免疫有关 (PE, PPE 家族); 40% 蛋白质有功能, 44% 可能有功能, 16% 是孤儿蛋白质 (Orphan protein: 指首次发现, 蛋白质数据库中尚无相应位置的蛋白质)。另外两个临床分离株的序列也正在测定。

2 MTB 功能基因组研究方法学

2.1 DNA 芯片

美国 1998 年 6 月 29 日宣布正式启动 “BIOCHIP” 计划。DNA 芯片的基本原理是 DNA 碱基配对和序列互补。用机器人在较小的固相支持物 (硅片、玻片、聚丙烯、尼龙膜等) 上密集成千上万的特异寡核苷酸探针, 抽提不同条件下样品 mRNA, 标记 (荧光、同位素), 与探针杂交, 共聚焦荧光显微镜扫描, 记录杂交结果, 分析杂交位点及其信号强弱, 找出差异表达基因 (是否表达和表达量)。用途: 基因诊断、设计基因药物、优化药物先导物、研究药物作用机理。

试验所用 DNA 芯片有 4896 个点, 代表 MTB H37Rv 全基因组 3924 ORFs 的 99.4% (3902ORFs)。H37Rv 和 BCG 比较, 提供了 3756 ORFs (96%) 的具有可比性的数据。*M. bovis* 较 MTB H37Rv 缺失 11 个区 (包括 91 个 ORFs), 同时也比 MTB H37Rv 多 5 个区 (包括 38ORFs), 而这 5 个区恰为所有 BCG 菌株所缺失, 这 5 个区都不属毒力因子, 同源分析表明它们功能多样化。这也表明: BCG 菌株一直在演进^[2]。与 H37Rv 基因组比较, 1927 年后的 BCG 菌株 (RD2 中的 Rv1985c) BCG-Pasteur (RD14 中的 Rv1773c), BCG-Moreau (RD16 中的 Rv3405c) 比之前 BCG 菌株缺失了大量调控元件。这些基因家族的功能尚不十分确定, 但可能与适应环境有关。与实验室培养不同, 体内感染时适应功能必不可少。为适应实验室条件, BCG 丢失了部分调控元件, 但是 BCG 在宿主体内生存和刺激产生持久免疫反应的能力却因此下降。MTB 比 *M. bovis* (包括 BCG) 多 9 个区 (61 ORFs), 二者表型差异可能与此有关。*M. bovis* 和 MTB 所致肺病的临床、放免、病理等方面没有差异。

M. bovis 在人与人之间的传染性较弱。这可能和 *M. bovis* 缺失 3 个磷脂酶基因 (*plcA*, -B, and -C) 有关 (因为铜绿假单胞菌磷脂酶 C 是条件性肺炎的毒力因子, MTB *plcA*, -*plcC* 基因簇可能具有相似功能。)

当然以 H37Rv 序列为参照分析种间差异也有局限性。还需进行生化、遗传研究; 同时 H37Rv 序列为参照不能研究 *M. bovis* 特异性毒力因子。

可用 DNA 芯片寻找新的药物靶子及抑制这些靶子的新化合物。用含 3924ORFs97% 的 DNA 芯片监测 MTB 对 INH 反应^[4]发现被 INH 诱导表达的基因包括：一个操纵子基因簇（共 5 个基因，编码 II 型脂肪酸合成酶及编码海藻糖二分支菌酸酰转移酶的 *fbpC*），*efpA*，*fadE23*，24，*ahpC*。生物合成途径未受到明显影响。

2.2 体内表达技术^[5] 基本原理：建立 MTB 的 DNA 库，将其中的 DNA 片段与缺失启动子、宿主体内生存必须基因融合，DNA 片段有启动子活性，菌才能生存。筛选有启动子活性的 DNA 片段，可以找出：感染不同阶段被诱导表达的致病基因，该基因在致病中的作用，宿主如何控制其表达。正研究 MTB 转录机器对宿主胁迫的反应，如调控铁代谢的 *IdeR* 与其靶子间的相互作用，Sigma 因子的差异表达。

2.3 签名标记转座子中断 标记为带不同特征的转座子（带标记的 DNA 片段称为 tag，一般约 80bp，分两部分：臂和可变区。臂位于两端，一般固定，供设计 PCR 引物；可变区位于中间，为 ATGC4 种碱基随机组成，理论上具 10^{12} 个不同序列）。用带 tag 的转座子或质粒构建 MTB 突变子，所获突变子应各有一特异 tag。用一组带不同 tag 的突变 MTB 接种动物模型（通常是鼠），筛选那些突变后毒力减弱或消失的 MTB。操作程序一般是：筛 96 个带不同 tag 的转座子或质粒 → 分别用其突变 MTB，构建一个突变子库（96 个，带不同 tag，构建 96 个是为便于在 96 孔酶标板上培养。）→ 复制酶标板上的 96 个突变子到两张尼龙膜上培养，供下一步杂交 → 动物实验（用酶标板上 MTB 突变子库攻击动物）→ 比较杂交实验（两种探针：输入探针—抽酶标板上 MTB 的 DNA 经 PCR 扩增的 tag 和输出探针—从经 MTB 突变子库攻击后，收集发病动物体内 MTB，抽提 DNA，PCR 扩增 tag），两种探针分别与尼龙膜上菌落杂交 → 挑选与输入探针杂交呈阳性，与输出探针杂交呈阴性的 MTB 菌落，供下一步研究（毒力减弱或消失的 MTB 不能从动物体内重新分离出来）。

细胞壁一直被怀疑是 MTB 的毒力因子，但缺乏有说服力的证据。PDIM（Phthiocerol dimycocerosate 结核菌醇二分支菌酸）是细胞壁组分，结构复杂，仅分布于致病分支杆菌。最近，用签名中断突变获得 3 个 MTB 突变株：2 个中断参与 PDIM 合成的基因（*ppsA-E* 编码多亚基的聚酮化合物合酶，*fadD28* 编码分支菌酰基转移酶），1 个为中断编码 PDIM 亚细胞定位所需跨膜蛋白质的基因（*mmpL*）。这 3 个突变子在体外、肝和脾里的繁殖未受影响（肺部除外），提示合成、运输 PDIM 仅是 MTB 在肺部繁殖所需。首次获得脂类生物合成与 MTB 毒力、菌落形态有关^[10] 的遗传证据。

从 1927 个签名转座突变子中筛到 16 个的毒力减弱，其中大部分与脂代谢或跨膜运输相关^[7]，能在染色体上定位；4 个突变子与毒力基因相关（参与结核菌醇衍生物合成）。中断嘌呤生物合成基因构建的营养缺陷型突变子不能在鼠骨髓衍生的巨噬细胞中繁殖，动物体内毒力也减弱，对气胶态有毒株 MTB 攻击的保护效果与 BCG 疫苗一样^[8]。

带 Ts-sacB（Temperature sensitive 温度敏感）反选择标记的自杀性质粒构建突变子。该方法兼具 sacB 的反选择特性和 MTB 的温度敏感型特性，在 39℃ 蔗糖培养基上能够高效反选择丢载体插入的突变子^[9]。该方法构建的转座突变子库的容量为 10^6 ，大大超过理论要求值。筛到了结核菌 *lysA*（编码间-二氨基庚二酸脱羧酶，催化赖氨酸生物合成最后一步），该营养缺陷型突变子对赖氨酸的需求比野生型菌株高 25 倍且依赖

Tween-80, 可作候选疫苗或研究细胞壁生物合成及氨基酸代谢^[6]。

利用基因组序列从粘粒库克隆 BCG 寡肽透性酶操纵子 *oppBCAD*, 以卡那霉素、链霉素抗性为选择标记, *oppD* 基因中断的突变子的肽转运蛋白缺陷, 不能在以 2~60 个氨基酸的短肽为唯一 C/N 源培养基上生长, 抗谷胱甘肽毒害能力强 (通常 2mmol/L 即可抑制野生型菌株生长, 该突变株可以抵抗 10mmol/L)。中断 MTB85 复合物表明: 85 复合物是纤连蛋白 (fibronectin), 也是毒力因子, 具枝菌酸酰基转移酶活性, 可能参与细胞壁生物合成。

2.4 基因互补法 将毒力株基因随机克隆到同种或近种无毒、弱毒株, 分析后者毒力是否改变以及改变程度, 筛选使后者毒力增强的基因。步骤: 构建有毒株基因库→(电)转化无毒、弱毒株 (耻垢分枝杆菌 1~2C/LR222, 在细胞株或动物中一般不易存活)→转化子感染人巨噬细胞样细胞株或动物→从细胞株或动物中重新分离菌株, 平板划线培养→鉴定使后者毒力增强的基因。

该方法现已找到 3 个基因: 仅分布于 MTB H37Rv, *M. bovis*, 非致病菌 *M. smegmatis* 中不存在^[11]的基因 EIS (Enhance Intracellular Survival), 编码 42kD 蛋白; 有助于 MTB 在溃疡和肉芽等厌氧环境生存的厌氧氮还原酶基因^[12]; 可使耻垢分枝杆菌在人 THP-1 中存活率明显上升的谷氨酸合成酶 A 基因^[13]。

2.5 差示荧光诱导法 可鉴定被宿主特定细胞 (如巨噬细胞)、特定条件诱导的细菌基因及其启动子。通常以 GFP 基因 (Green Fluorescent Protein 绿色荧光蛋白) 为报告基因。该方法优点是: 自动化、高通量, 不涉及代谢营养要求。

2.6 生物信息学 比较 MTB 基因组与其他细菌海藻糖生物合成途径的同源性发现 MTB 海藻糖生物合成途径可能有 3 个: (1) Tres 从麦芽糖起始, (2) TreY-TreZ 从糖原样 α (1>4) 连接葡萄糖聚合物起始, (3) OtsA-OtsB 从葡萄糖-6-磷酸, UDP-葡萄糖起始。实验分析 MTB 的 DNA 支持上述推论, 同时也证明: 胞浆中的游离二糖海藻糖在 MTB 的生理及病理中具有重要的作用^[14], 既是胞壁糖脂组分, 也与 MTB 感染所致组织损伤有关。

致病菌分泌蛋白往往是宿主免疫反应的主要靶子。几个程序分析 MTB 的 3924ORF, 搜索到 52 个有-N 端分泌信号肽但无膜锚定成分的蛋白序列, 利用 *E. coli* PhoA 亚细胞定位标记与其中 10 个序列融合, 发现生物信息学的准确率 90%^[15]。

致病过程也是被精密调控的信号传导过程。MTB 基因组编码的调控因子有: 13 个 sigma 因子, 调控蛋白 100 个, 二元件调控系统 (组氨酸激酶+反应调节子) 11 个, 激酶及其调控基因, 真核样丝氨酸/苏氨酸激酶系统, 可能与休眠、细胞分裂^[2]控制有关的磷酸化基因 (磷酸中继系统的一部分)。

分析 cNMP-结合蛋白及核苷环化酶超家族。MTB 与其他原核生物相同的 cNMP-结合蛋白有: cNMP 依赖的激酶、cNMP 门控通道、cAMP 鸟苷交换子、细菌 cAMP 依赖的转录因子; MTB 特有: ABC-转运蛋白亚基、转位酶、酯化酶。据活性位点保守序列的微小差异, 环化酶分 5 类: 多细胞真核腺苷环化酶、多细胞真核受体型鸟苷环化酶、多细胞真核可溶鸟苷环化酶、单细胞真核腺苷环化酶、单细胞原核鸟苷环化酶。分析单细胞和多细胞的系统演进发现: MTB 基因组编码的 cNMP 远比其他原核多。胞内致病菌 MTB 的运输蛋白种类明显比游离生存细菌少。

2.7 蛋白质组研究 比较两无毒株 (*M. bovis* BCG-Chicago/Copenhagen) 与两有毒株 (*M. tuberculosis*-H37Rv/Erdman) 的蛋白质组: MTB 的细胞或上清液 2D-电泳 → 银染 → MS 分析 (特征性蛋白) → 与动态 2D-电泳数据库 (www.mpiib-birlin.mpg.de/2D-PAGE) 比较 → 细胞或上清液分别含 1800 点和 800 点, 有 263 点为特征性 (54 点来自上清液)。蛋白质组研究有助于设计新的防治 MTB 方案。该方法在其他生物的运用已有综述, 在此略过。

2.8 DNA 库免疫 预防性和治疗性 DNA 疫苗在人抗结核免疫中有较好的应用前景。其关键之一是获得适当的保护性抗原基因。DNA 库免疫 (也称基因组疫苗库 Genomic vaccine library, Expression Library Immunization, ELI 表达库免疫) 是获得保护性抗原基因的高效途径。原理: 机体免疫系统自身会选择保护性抗原, 没有必要逐个筛选。步骤: (1) 构建 DNA 库, 抽提致病菌 DNA, 随机打断, 克隆进不同质粒; (2) 免疫动物, 筛选保护性质粒, 将重组质粒免疫动物, 随后用致病菌攻击动物, 不发病动物即是受到保护的; 将免疫该动物的 1000 个质粒每 100 个分成一组, 重复上述实验, 直到找出有保护作用的一个或几个质粒。短时间内即可系统地对致病菌相关抗原进行分类研究 (9 个月到一年)。正对结核分枝杆菌的 4000 个基因进行筛选。

2.9 细菌人工染色体库 细菌人工染色体库 (Bacterial artificial chromosome library 简称 BAC) 是后基因组研究的重要平台, 对难进行遗传操作、危险、生物学背景知识缺乏的原核生物尤其有效, 是现在和今后对它们进行功能基因组学研究的主要资源。

BAC 基于大肠杆菌的 F 因子。F 因子的复制受到严紧控制, 该因子在每个细胞中最多 1~2 拷贝, 外源 DNA 片段重组几率小, 因此外源 DNA 片段能稳定存在; 同时, 也避免了外源 DNA 片段致死性高表达对宿主的危害。BAC 已经广泛用于克隆不同真核生物 DNA, 如植物、动物、真菌。

覆盖结核分枝杆菌 4.4Mb 基因组仅需 68 个克隆^[1]。结核分枝杆菌基因组的 BAC 库是 BAC 在原核生物基因组首次运用。该库含有 5000 个 BAC 克隆, 插入片段大小为 25~104kb, 平均 70kb, 是结核分枝杆菌基因组容量 (4.4Mb) 的 70 倍。该库作用体现在: (1) 帮助全基因组测序数据正确拼接: 由于粘粒库的代表性不够, 也不能完整拼接染色体, 而 BAC 库可以用最小 68 个克隆完整拼接出整个染色体, 因此, 它是结核分枝杆菌全基因组测序数据拼接的有利工具。当初首次构建原核生物 BAC 库也是基于弥补粘粒库和 YAC 的不足。(2) 比较基因组研究。(3) 结核分枝杆菌代谢途径研究: 以前由于受到缺乏遗传操作系统、生长缓慢和安全因素制约, 结核分枝杆菌代谢途径等生物学基础研究往往利用快生型同属菌 *M. smegmatis* 作模式系统, 由于 BAC 系统的建立和它的忠实性, 能够克隆大片段 DNA, 因此可很方便进行生物合成途径、分泌系统、毒力岛等复杂位点的研究。(4) 结核分枝杆菌耐药机理研究: 耐药结核分枝杆菌不能在普通实验室培养, 构建其 BAC 库后, 生物危害性大大降低, 一般实验室也可研究, 大肠杆菌生长快, 可加快进度。我们实验室正在开展这方面工作。

3 展望

综合利用以上方法和免疫学手段将有助于理解 MTB 基因组功能, 识别药物作用靶子、保护性抗原、宿主体内生存和致病所需基因, 构建新机理的新药筛选模型, 开发新型疫苗、药物, 控制和最终根除结核病。

参 考 文 献

- [1] Gordon B R. *Infect. immun.*, 1998, 66: 2221 ~ 2229.
- [2] Cole S T. *Nature*, 1998, 393: 537 ~ 544.
- [3] www.sanger.ac.uk/Projects/M-tuberculosis/Gene-list.
- [4] Hilson M, DeRisi J, Kristensen H H, *et al.* *P. N. A. S.* 1999, 96 (22): 12833 ~ 12838.
- [5] Smith I, Dussurget O, Rodriguez G M, *et al.* *Tuber Lung Dis*, 1998, 79 (2): 91 ~ 97.
- [6] Pavelka M S Jr, Jacobs W R Jr. *J. bacteriol.*, 1999, 181 (16): 4780 ~ 4789.
- [7] Camacho L R, Ensergueix D, Perez E. *Mol Microbiol*, 1999, 34 (2): 257 ~ 267.
- [8] Jackson M, Phalen S W. *Infect. immun*, 1999, 67 (6): 2867 ~ 2873.
- [9] Pelicic V, Jackson M, Reyrat J M. *P. N. A. S.* 1997, 94 (20): 10955 ~ 10960.
- [10] Jeffery S Cox, Bing Chen. *Nature*, 1999 402: 79 ~ 83.
- [11] Wei J, Dahl J L, Moulder J W, *et al.* *J. Bacteriol*, 2000, 182 (2): 377 ~ 384.
- [12] Weber I, Fritz C, Ruttkowshi S. *Mol Microbiol*, 2000, 35 (5): 1017 ~ 1025.
- [13] Miller B H, Shinnick T M. *Infect. immun*, 2000, 68 (1): 387 ~ 390.
- [14] Desmet K A, Weston A, Broun I N. *Microbiol*, 2000, 126 (Pt1): 199 ~ 208.
- [15] Gomez M, Johnson S. *Infect. immun*, 2000, 68 (4): 2323 ~ 2327.