

结核分枝杆菌基因组学与基因组进化*

宛宝山¹⁾ 张秋芬¹⁾ 周爱萍¹⁾ 赵国屏²⁾ 姚玉峰^{1)**}

¹⁾上海交通大学医学院基础医学院病原生物学教研室, 上海 200025;

²⁾中国科学院上海植物生理与生态研究所合成生物学实验室, 上海 200032)

摘要 在后基因组时代,特别是在新的测序理论和设备大发展的背景下,一些重大传染性致病微生物基因组序列正在被逐一测定,并且随后的基因功能注释,蛋白质三维结构重建等工作也正在开展,以期对致病微生物的生物学特性、诊断策略和治疗方法等有突破性的认识。作为对人类健康一直存在严重威胁的结核分枝杆菌,其基因组在进化中所发生的各种遗传事件对其生物学性质、致病能力和抗药性等各方面有重要作用。本文旨在阐述结核分枝杆菌的起源及其基因组特征,论述其基因组进化的研究进展。

关键词 结核分枝杆菌, 基因组, 比较基因组学, 分子进化

学科分类号 Q93, R3

DOI: 10.3724/SP.J.1206.2011.00469

结核分枝杆菌(*Mycobacterium tuberculosis*)是重要的全球性致病菌,由其引起的结核病是世界上致死率最高的传染病之一。据世界卫生组织报告,每年约有 200 万人死于结核病,其中 98% 的死亡都发生在发展中国家^[1]。随着 HIV 病毒的流行,三分之一 HIV 感染者都患有结核病。某些地区,耐药性菌株的出现速度之快足以引起人们的警惕。结核病流行现状的彻底改变似乎还只是遥远的理想。本文将简述目前结核分枝杆菌基因组学研究的现状和发展趋势,介绍有关其基因组进化的研究成果。

1 结核分枝杆菌的基因组学

1.1 结核分枝杆菌的起源

结核分枝杆菌属于慢速生长型分枝杆菌。利用 16 S rRNA, *rpoB* 和 *hsp65* 等重要基因序列构建分枝杆菌的进化树发现结核分枝杆菌与其同属的 *M. africanum*、*M. bovis*、*M. bovis* (BCG)、*M. microti* 和 *M. caprae* 等在亲缘关系上更加相近,它们现在被归为结核分枝杆菌复合群(MTBC)^[2]。对结核分枝杆菌 H37Rv 菌株的测序结果表明,其基因组包含 20 个编码带有细胞色素 P450 的单氧化酶基因,这些酶可以催化疏水复合物,如长链脂肪酸的混合氧化,这种催化活性和土壤中自由生长的腐生菌有

密切联系,表明结核分枝杆菌的祖先菌可能属于土壤中生存的分枝杆菌^[3]。

1.2 结核分枝杆菌的基因组学研究

1998 年英国 Sanger 中心和法国 Pasteur 研究所合作完成了对结核分枝杆菌 H37Rv 菌株全基因组测序工作^[4]。结核分枝杆菌基因组大小为 4.4 Mb,预测含 4 411 个开放阅读框(ORF)。其中 3 924 个 ORF 被认为编码蛋白质,50 个基因编码稳定的 RNA。结核分枝杆菌基因组 G/C 含量高达 65.6%。基因组编码蛋白质的 10% 为 PE(脯氨酸和谷氨酸重复)和 PPE(脯氨酸、脯氨酸和谷氨酸重复)蛋白质家族。在初期的功能注释中,结核分枝杆菌基因表达产物中,40% 为有功能的蛋白质产物,另 44% 与基因组其他信息有关,这当中大多是“保守且功能假定的序列”(即它们在其他细菌中也存在但其功能未知),还有 16% 则完全未知且仅存在于结核分枝杆菌和其他分枝杆菌属中。

* 国家自然科学基金(31070114)和国家重点基础研究发展计划(973)(2009CB522605)资助项目。

** 通讯联系人。

Tel: 021-64671226, E-mail: yfyao@sytu.edu.cn

收稿日期: 2011-10-20, 接受日期: 2012-02-07

2002年, Camus等^[5]根据新的实验数据和序列比对信息, 对结核分枝杆菌 H37Rv 菌株的基因组进行重新分析并加以注释. 他们在原先基础上, 又发现 82 个能够编码多肽的新基因. 基于和已有基因组序列的比较以及来自其他文献的实验数据, 确定了 2058 个蛋白质的功能, 预测出 376 个蛋白质与已知蛋白质不具同源性, 是结核分枝杆菌所独有的.

随后, Fleischmann 等^[6]完成了临床分离得到的结核分枝杆菌 CDC1551 菌株的基因组测序并将其与 H37Rv 菌株进行了全基因组比较. 他们发现, H37Rv 和 CDC1551 菌株的基因组间有显著的不同. 在 CDC1551 和 H37Rv 菌株基因组中存在 1 075 个 SNP, 约 85% 的替换发生在编码区. 相对于 CDC1551 菌株基因组, H37Rv 菌株基因组中有 37 个插入片段(长于 10 bp). 而相对于 H37Rv 菌株基因组, CDC1551 菌株基因组则有 49 个影响着 ORF 的插入序列, 其中 14 个位于基因间. 所有这些插入事件除了串联重复、ORF 两末端序列增多外, 还导致增多了 17 个完整的 ORF. 研究表明, 相近的 IS6110 插入序列之间发生的同源重组事件可能导致基因组序列的缺失进而引起基因组的多样性^[7]. 研究人员在 CDC1551 菌株基因组中鉴定出 4 个拷贝的 IS6110 插入序列. 而在 H37Rv 菌株基因组中, 却有 16 个拷贝的 IS6110 插入序列.

1.3 结核分枝杆菌基因组的变异

1.3.1 结核分枝杆菌基因组变异与致病性的关系.

从历史进程上来说, 结核分枝杆菌 H37Rv 和 H37Ra 两菌株都是来源于它们共同的有毒亲本 H37, 该菌株于 1905 年从一个慢性肺结核病人体内被分离到^[8], H37Rv 菌株一直在实验室作为标准菌株使用并保持毒力, 而 H37Ra 菌株由于经过多次传代而丧失了毒力.

国家人类基因组南方中心于 2008 年完成了对 H37Ra 菌株的全基因组测序, 同时对 H37Rv 进行部分序列的测定, 运用比较基因组学方法, 试图解释 H37Ra 菌株毒力丢失的原因. 测序和比对结果表明, 结核分枝杆菌 H37Ra 与 H37Rv 菌株基因组之间有高度的保守性, 但由于存在 53 个插入和 21 个缺失事件, 前者比后者基因组多 8 445 bp. 研究表明, CDC1551 菌株更接近于 H37 菌株, 故以 CDC1551 菌株基因组为标准, 比对出 130 个 H37Ra 特异性的基因组变异, 共涉及到 57 个基因^[9].

鉴于结核分枝杆菌不同菌株基因组差异的存

在, 有研究者用动物模型来实际考察不同菌株的致病性差异. Claudia 等选取 CDC1551、HN60、HN878、H37Rv 和 Erdman 菌株, 构建小鼠吸入感染模型. 发现在感染后的早期生长阶段, 各个不同处理下, 除了 Erdman 菌株, 其他各菌株生长速率相似, 随后, CDC1551 生长速率相对下降. 相对于其他小鼠而言, 受 CDC1551 菌株感染的小鼠肺部产生的肉芽肿中有较高水平的 TNF- α 、IL-6、IL-10 和 IL-12 表达量, 并且 IFN- γ 的 mRNA 表达也有所提前. CDC1551 菌株感染的小鼠生存期较其他处理的小鼠明显要长, 在人类单核细胞中, CDC1551 菌株和 H37Rv 菌株生长速率相同, 可是前者在引起人类单核细胞表达更多的 TNF- α 、IL-6、IL-10 和 IL-12 等细胞因子. 从体内外的生长速率来看, CDC1551 菌株并不比其他菌株更具毒性, 但其确实能引起寄主更迅速且强烈的免疫反应^[10].

虽然 H37Ra 和 H37Rv 菌株基因组高度相似, 但二者表型完全不同, 这表明 H37Ra 菌株基因组发生了一些可以影响其毒力和生长特征的突变. 结核分枝杆菌 SigC 可以调控至少 38 个与细胞代谢过程有关的基因表达, sigC 基因的缺失使得 H37Rv 菌株对小鼠的毒力明显下降^[11]. 在 H37Ra 菌株基因组中发现预测的 sigC 启动子区域有 A-T 的颠换, qRT-PCR(实时荧光定量 PCR)研究表明, 在对数生长期的体外培养结核分枝杆菌 H37Ra 菌株中, 该基因表达上调, 但在巨噬细胞中, 该基因表达却下调. 在 H37Ra 菌株中, MazG 含一个错义突变, 导致在其高度保守的 ²¹⁸PAL(Pro-Ala-Leu)基序处发生 A219E 替换, 这一基序处于预测的一个保守的 α 螺旋结构内, 该结构加上其他 3 个螺旋结构, 共同构成了 MazG 的焦磷酸水解酶的活性部位^[12]. 随后我们发现, 在分枝杆菌属内的各菌株(无论是 H37Rv 菌株, 还是快速生长的耻垢分枝杆菌)中, MazG 是 NTP-PPase(焦磷酸水解酶), 可以水解各种 NTP 或 dNTP 以及致突变的 dUTP 和 8-氧代 -7, 8-二氢 -2-dGTP. A219E 的突变使 MazG 缺失水解活性. 在耻垢分枝杆菌中, mazG 的缺失能导致菌株对氧化压力响应能力的不足, 暗示 mazG 的突变可能与 H37Ra 的毒力丢失有关^[13].

phoP 是结核分枝杆菌已知的毒力基因^[14]. 由于单碱基突变, H37Ra 菌株的 PhoP 蛋白高度保守的 219 位丝氨酸突变为亮氨酸, 导致 PhoP 与 DNA 序列结合能力受损从而降低转录激活作用, 这也可

能和 H37Ra 菌株毒力丢失有关. H37Ra 菌株中, *pabB* 起始密码子上游 66 bp 处有 T-A 颠换, 可能导致该基因转录水平的改变, 这一基因编码对氨基苯甲酸合成酶组分 - I, 涉及 *p*-氨基苯甲酸(PABA)的合成, 在类鼻疽杆菌中该基因突变导致细菌毒力缺失^[5]. FtsH 是 ATP 依赖的膜结合锌金属蛋白酶, 负责一些胞质中的短寿命蛋白质和膜上未装配蛋白的降解. H37Ra 菌株基因组中, *fisH* 起始密码子上游 51 碱基处有 106 个碱基的丢失, 在上游 65 碱基处有 C-T 的颠换, 这都可能影响该基因的表达^[9].

1.3.2 结核分枝杆菌基因组变异与其耐药性的关系.

在结核病防治形势重新严峻的背景下, 结核分枝杆菌的耐药性问题重新引起了人们的重视. 耐药结核分枝杆菌尤其是耐多药结核分枝杆菌(MDR-TB)和广泛耐药结核分枝杆菌(XDR-TB)的出现和传播是导致结核病发病率升高的重要原因. 结核分枝杆菌一般无法通过质粒介导的途径从其他细菌获得耐药性, 因此染色体介导的耐药性是结核分枝杆菌产生耐药的主要分子基础. 由于有关结核分枝杆菌基因组变异与其异烟肼、利福平、链霉素等耐药性的综述文章已有较多发表, 这里就不再赘述. 主要谈一个最近有关吡嗪酰胺耐药性的进展.

吡嗪酰胺(PZA)需要细菌的 PZase(吡嗪酰胺酶)将其转化为有杀菌能力的物质 POA(吡嗪酸). PZase 由 *pncA* 编码. Mestdagh 等^[16]在收集了各地的结核分枝杆菌临床耐多药性分离株后, 以 H37Ra 菌株作为非耐药对照菌, 对 23 株 PZA 抗性菌株的 *pncA* 的 ORF 以及上游 85 个碱基进行测序, 结果有 17 株菌在该区域有突变, 6 株无突变. 仅 3 株有 PZase 酶活性, 其中 2 株该区域没突变. 可见, *pncA* 的突变是导致细菌对 PZA 产生抗性的重要因素. 张颖等曾鉴定出无 *pncA* 突变的 PZA 抗性菌株(DHM444), 暗示结核菌的 PZA 抗性可能有另外的机制^[17]. 进一步研究发现, DHM444 菌株的 RpsA 羧基端突变导致其和 POA 特异性结合能力丧失, 故此 POA 不能抑制由 RpsA 介导的反式翻译过程从而使细菌得以存活. 当 POA 与 RpsA 结合后可以阻止后者与 tmRNA 结合, 从而使因非正常翻译而滞留在 mRNA 3' 端的核糖体不能被释放, 最终导致细菌失去生存能力^[18].

1.3.3 结核分枝杆菌耐药性研究前景.

人们已经发现, 莽草酸代谢途径对于细菌、寄

生虫和植物合成芳香族氨基酸是必不可少的^[19-20], 分枝杆菌通过该途径合成分枝杆菌素、泛醌和芳香族氨基酸, 而哺乳动物细胞内却没有类似的代谢途径. 以该途径组分为靶点的新抗结核药物因此具有良好的前景. Zhang 等^[21]研究了结核分枝杆菌中莽草酸脱氢酶的化学性质. 该酶由 *aroE* 基因编码, 从序列上看, 该酶有 G-X-(N/S)-V-(T/S)-X-PX-K 这一莽草酸脱氢酶家族的保守肽段. 通过圆二色谱研究, 该酶 α 螺旋、 β 折叠、 β 转角和无规则卷曲结构分别占到 29.2%、9.3%、32.7% 和 28.8%. 研究者希望找到该酶的抑制剂来进一步探索出可用的抗结核新药.

Sebastian 等研究了莽草酸途径另外一个酶:

3-脱氧-D-阿拉伯庚酮糖-7-磷酸合成酶(DAH7PS). DAH7PS 催化莽草酸途径的第一步反应以合成芳香族组分. 该酶催化磷酸烯醇式丙酮酸与赤藓糖-4-磷酸反应, 生成 DAH7P. 经研究发现, 该酶有一个具有构型优势的四面体的催化活性中间状态, 研究人员根据这一模型构建 DA7PHS 的两种结构类似物作为酶活性的抑制剂并取得结晶体, 二者互为旋光异构体, 对 DAH7PS 的催化有很强抑制作用, 可以进一步加以研究以取得抗结核病的新药物^[22].

喹诺酮类抗生素主要是针对细菌 DNA 促旋酶亚单位 A (GyrA). 但随着结核分枝杆菌对这类抗生素抗性的增强, 人们开始将目光瞄准促旋酶亚单位 B (GyrB) 来寻找潜在的抗菌药物. Sidharth 等选用新生霉素和氨基苯并咪唑-1(AB-1)作为 GyrB 的抑制剂来测试二者的抗菌活性. 新生霉素是已知的 GyrB 抑制剂, AB-1 则针对 GyrB 的 ATP 结合位点. 经研究, 无论是是否耐药, AB-1 对结核菌都具有时间依赖性杀菌活性, 包括氟喹诺酮抗性菌株. AB-1 对于低氧环境下缓慢生长的结核菌也有有效的杀菌能力. AB-1 与异烟肼和利福平都没有交互作用且与喹诺酮类抗生素没有交叉耐药性现象. 大鼠肺结核模型研究显示, 在低剂量用药下 AB-1 能显著降低大鼠肺部存活的菌落数. 多种结果表明, AB-1 是潜在的抗结核一线药物^[23].

1.3.4 结核分枝杆菌多菌株间遗传变异的分析.

结核分枝杆菌几种菌株全基因组测序的完成使得通过比较基因组学方法来研究在结核分枝杆菌种内发生遗传变异的本质及其结果成为可能. Cubillos-Ruiz 等^[24]通过对几种结核分枝杆菌的临床菌株(CDC1551、F11、Haarlem 和 C)以及实验用菌

株(H37Rv 和 H37Ra)进行基因组比对, 获得了关于结核分枝杆菌对寄主适应性机制的新见解. 在上述 6 株菌株中发现的遗传变异并未涉及重大的基因组重排而大多来自插入序列和 *PE/PPE* 基因家族相关缺失和转座事件. 对 6 株结核分枝杆菌基因组比对应后, 研究人员鉴定出了多态性的分布模式及其频率, 指明了传播中的结核分枝杆菌正处于活跃的选择压力下, 从而通过基因缺失或失活导致变异的发生. 对基因的鉴定也显示出一些菌株特异的多态性, 将其数量外推至更多的菌株基因组, 可见不同菌株含有有限数目的独有的基因组多态性.

为了检验全球各主要实验室所独自保有的结核分枝杆菌 H37Rv 菌株有无基因组的变异, Ioerger 等^[25]从不同的实验室取了 6 株 H37Rv 菌株样本, 实施全基因组测序, 以评估各实验室所使用的 H37Rv 菌株基因组序列的一致性以及它们从起始菌株异化生长的程度. 在与 H37Rv 参考序列的对比中, 研究人员发现有 72 个多态性现象是所有 6 株 H37Rv 菌株变种以及 H37Ra 菌株所共有的. 这 72 个多态性包括 57 个 SNPs 和 15 个 indels(插入和缺失). 所有的 SNPs 中, 6 个位于非编码区域, 21 个属于同义突变, 30 个属于非同义突变. *Rv0197* 和 *pstA1* 基因被截短, *pks3* 和 *Rv1783* 的终止密码子发生有义突变从而延长了 ORF. 在编码区域发生 indels 导致移码突变的有 *Rv0197*、*PPE7*、*Rv0907*、*Rc1046c*、*Rv1575*、*Rv2251*、*Rv3655c* 和 *sigM*. 在非编码区发生的 15 个突变中, 有 6 个在预测的调控区域内. Spoligotyping(间隔区寡核苷酸多态性)分型结果显示, 有 3 个被测菌株 H37RvLP、H37RvJO 和 H37RvCO 分型结果与 H37Rv 参考序列相匹配, 即缺失第 20、21 和第 33 到 36 个间隔子序列. 另外 3 个菌株 H37RvAE、H37RvMA 和 H37RvHA 则另外还缺失第 26 个间隔子. 这些结果暗示, 在系统发育关系基础上这些菌株可以被分为不同的簇. 这表明在一个可控的环境中, 结核分枝杆菌在持续进化.

2 结核分枝杆菌的基因组进化研究

2.1 比较基因组学与结核分枝杆菌分子进化的研究

Mulder 等^[26]通过比对不同结核分枝杆菌基因组, 来研究它们独有的和扩增的基因. 他们发现结核分枝杆菌在演化中获得耐药性的能力至少是部分

地与基因扩增的机制有关. 在结核分枝杆菌等病原菌中, 基因扩增在其进化中担当一定的角色, 就结核分枝杆菌而言, 52%蛋白质来源于与基因扩增事件有关的蛋白质大家族的一部分^[27]. 对病原菌来说, 基因扩增提供了一个简单的机制来产生为数众多的必需基因产物, 或者进化出对快速适应生存环境十分重要的新基因.

在约 15 000 年前, 结核分枝杆菌就从分枝杆菌属细菌的共同祖先分异到今天的类型. 在结核分枝杆菌中发现低水平的遗传变异表明, 在发生过进化的瓶颈后, 在大约 35 000 到 15 000 年前之间, 结核分枝杆菌发生着正常的克隆扩增而产生一定规模的群体^[28-29]. 细菌遗传变异的发生有 3 种情况: 核酸序列较小位点的改变、基因组内部重排和从其他菌体中获得 DNA 序列^[30]. 第 3 种是细菌获得新蛋白质最普遍的途径. 基因水平转移在细菌中是普遍的, 在进化过程中也被认为是广泛分布的^[31].

MTBC 包括: *M. tuberculosis*、*M. canettii*、*M. africanum*、*M. bovis*、*M. microti* 等^[32]. 后来 Ernst 等^[33]又补充了 *M. pinnipedii* 和 *M. caprae* 等组成了总共 7 个种或亚种. MTBC 成员有相同的 16 S rRNA 序列(除了 *M. canettii*)和超过 99.9%的核酸水平的同一性^[6, 34]. 在该群体中较少有同时代的成员间发生水平基因转移的证据^[35-38]. 而 *M. canettii* 的持家基因是嵌入的, 它们的自有片段在其他成员中也被发现, 这就为在祖先菌株间发生水平基因转移提供了证据. 在代表菌株中, 连续而单向的 DNA 缺失揭示出 MTBC 成员从共同祖先进化的顺序. 从系统发育关系上看, *M. canettii* 最接近祖先菌株, 而 *M. tuberculosis* 和 *M. bovis* 却最为年轻, 其余的 5 种居中间位置^[39-40].

2.2 结核分枝杆菌的人类 T 细胞抗原表位在进化上的保守性

Ernst 等^[39]认为, 尽管结核分枝杆菌多个不同菌株的全基因组已完成测序, 结核病的研究正日益深入, 但人们依旧有诸多生物学问题未解决. 人类在感染结核分枝杆菌后, 机体产生可识别 ESAT-6 多肽的记忆和效应 T 细胞. 有证据表明, 在小鼠感染结核分枝杆菌的慢性阶段, 编码 ESAT-6 蛋白的基因表达量有所降低^[41]. 若该基因表达水平降低明显, 则 ESAT-6 专一性的 T 细胞可能在起始感染阶段丧失部分免疫效用.

Comas 等选取 MTBC 中 6 个主要世系且在全

球都具多样性的 21 个菌株, 采用新一代 DNA 测序技术, 构建了基因组尺度上的系统发育关系图. 通过比较基因组学研究发现, 重要基因在进化上比非重要基因要保守得多, 但结核分枝杆菌基因组中编码人类 T 细胞抗原表位的序列却十分保守, 其变异速率低于重要基因, 这就暗示出结核分枝杆菌可能是通过人类 T 细胞的识别而获得在人体内长期存活. 通过比较 DNA 序列, 人们发现 MTBC 起源于非洲, 经历过菌群的扩增. 在古代非洲的结核分枝杆菌携带者移民到他国后, 结核分枝杆菌经历了多样化发展, 随即引起 MTBC 三个主要世系通过旅行、贸易和战争在全球传播并传回非洲^[42]. 人们还发现, MTBC 的遗传多样性与人类移民状况相一致, 不同的世系与不同的人群共发展^[43]. 人类、非人灵长类和小鼠对结核分枝杆菌的免疫依靠 T 细胞^[44]. 在人类 T 细胞中, 主要是 CD4⁺ T 细胞参与对 MTBC 的免疫. 通过对 21 个菌株测序并分析了 491 个实验查实的人类 T 细胞抗原表位, 实验人员发现, 已知的人类 T 细胞抗原表位高度保守, 这就说明 MTBC 与人类寄主之间的关系不同于经典的国与国之间的军备竞赛的演变, 即并不是相关的双方都在发展进化, 而其中一方保持进化上的相对停滞. 那么, 治疗结核病的新方法势必要考虑特定的人类免疫反应对 MTBC 的生存有利的可能性^[45].

2.3 高度可变和有免疫原性的 *PPE38* 基因区的快速进化

结核分枝杆菌基因组含有的两大基因家族, 即 *PE* 和 *PPE* 家族, 可能来源于快速生长的分枝杆菌种系, 继而在特定的慢速生长种系, 特别是 *M. ulcerans*, *M. marinum* 和 MTBC 中经历了大量的扩增和多样化过程. 尽管有报道说这两个基因家族编码的蛋白质中有些与细胞表面有关且能被寄主免疫系统所识别^[46-49], 但整体上, 人们对这两大基因家族所编码的蛋白质功能依旧未知. *PPE* 蛋白家族在它们高度保守的大约 180 个氨基酸氨基端结构域的第 7 到 9 位的氨基酸残基分别是脯氨酸、脯氨酸和谷氨酸. 而 *PE* 蛋白家族在它们高度保守的氨基端, 有包含约 110 个氨基酸残基的结构域, 该结构域的第 8 和 9 位分别是脯氨酸和谷氨酸. 这两个蛋白质家族的羧基端结构域在长度和序列上都高度可变且 *PE* 和 *PPE* 基因家族常包含重复 DNA 序列, 在不同基因中这些重复的拷贝数不同^[4]. 虽然 *PE* 和 *PPE* 蛋白家族的氨基端结构域都很保守, 但它

们之间在氨基端却很少有同源性. 已有的研究表明, *PPE-MPTR* 亚家族蛋白如 Rv1917c(*PPE34*)可能与细胞壁有关. 特定的 *PE-PGRS* 蛋白是细菌细胞表面的组分, 影响细菌的细胞结构和菌落形态. *PPE36* 和 *PPE68* 也和细菌的细胞壁有关^[50].

先前有研究认为在 *PPE* 基因家族其他成员(非 *PPE38*)中发现的高水平的多态性可能代表了抗原变异的手段. McEvoy 等研究了在系统进化关系中主要世系的结核分枝杆菌临床分离菌株和 MTBC 中祖先菌株 *M. canettii* 的 *PPE38* 基因区域的遗传变异, 发现 *M. canettii* 的 *PPE38* 区域包含两个同源的 *PPE* 基因(*PPE38* 和 *PPE71*), 二者被两个 *esat-6* (*esx*)样基因隔开. 这一结构来源于 *esx/esx/PPE* 复制体, 该复制结构存在于结核分枝杆菌和 *M. marinum* 的共同祖先中, 并且在结核分枝杆菌中, 由于 IS6110 插入序列的整合, 这一抗原编码区域(*PPE38* 基因区)高度可变. 研究人员发现, 许多不同的突变都被固定在特定的世系中, 这表明这一基因区正经历着快速的分子进化并且随这种世系限定的突变事件而来的将是更深入的世系特异性的进化扩增和多样性发展. 由于这些突变多导致基因功能的缺失, 研究人员推测它们并不会导致细菌适应性显著地下降, 而因为 *PPE38* 基因区有高度的抗原性, 这些突变可能有助于细菌的生存^[51].

2.4 结核分枝杆菌北京型菌株的进化道路

2.4.1 北京型菌株的产生. 结核分枝杆菌的各种基因型中, 在地域上广为传播又具有较高毒力的一员就是北京型菌株. Van Soolingen 等^[52]在 1992 年到 1994 年期间, 收集了中国大陆和蒙古的结核病人并对这些病人所携带的结核分枝杆菌基因组做了分析, 发现相比于更远地区, 来自东亚, 特别是北京地区的结核分枝杆菌有特异的 IS6110 RFLP(限制性片段长度多态性)指纹图谱, 故而认为这些菌株发源自北京地区并向其他地区传播, 而相应的菌株则因此被认为属于北京型. 基因组特征上, 在直接重复序列(DR)区域, 北京型菌株缺失 1 到 34 个间隔区序列, 而在第 35 到第 43 的 9 个间隔序列中至少含 3 个^[52-54]. 比较基因组学研究表明, 北京型菌株主要通过 IS6110 插入序列的转座、染色体区域的缺失(相关的区域被命名为: RD, regions of deletion)和突变等机制, 从祖先菌株进化而来. 通过 SNP 分析表明, 北京型菌株被认为是来源于东亚, 从结核分枝杆菌的一个祖先群体中进化而来, 后者首先出现于印度次大陆^[55].

2.4.2 北京型菌株基因组的特征. 通过长序列多态性(LSP)分析可见北京型菌株已经进化到不同的分支, 这通过 RD 序列的特异性缺失可以鉴定. 所有的北京型菌株都可见 RD105 区域的缺失, 故可认为该区域可作为鉴定该家族菌株的分子标记. 其他的长序列多态性, 如 RD142、RD150 和 RD181, 可将北京菌株家族再细分到 4 个亚组, 对应着 4 个不同的菌落结构^[53]. Rindi 等^[56]从意大利的结核病人中采样, 经过 Spoligotyping 分析, 共得到 73 个北京型菌株. 通过分析长序列多态性, 如上述 RD105、RD181、RD150 和 RD142 及 putative mutator 基因, 包括 *mutT4* 和 *mutT2* 等编码 DNA 修复酶基因的单碱基替换, 提出了北京家族的系统发育的重建模型. 根据该模型, 北京型菌株从结核分枝杆菌的一个祖先菌株进化而来, 该祖先菌株发生了 RD207 区域的缺失, 导致在直接重复序列中, 北京型菌株缺失 1 到 34 个间隔序列. 随后, RD105、RD181 和 RD150 序列的缺失直接导致北京家族几个主要世系的进化发展. 在最古老的进化世系中, *mutT4* 和 *mutT2* 是野生型. 在祖先菌株中, 随着 RD181 区域的缺失, *mutT4* 发生突变, 引起了相应一个亚系的产生, 该亚系含有 *mutT2* 的突变体. 在北京家族菌株的主要分支中, 直接重复序列间的其他间隔序列的缺失导致不同的 Spoligotyping 分型结果, 这对应着不同的亚系. Zhou 等^[57]针对西安地区近年传播的结核分枝杆菌临床菌株进行基于 Spoligotyping 分型和总共涉及 21 个位点的 MIRU-VNTR 分型, 发现在 195 株临床菌株中, Spoligotyping 分型结果表明在西安地区最盛行的结核分枝杆菌属于北京家族菌株, 包括 ST1、ST190、ST265、ST269、ST585 和 ST1364 等基因型. MIRU-12 分型将北京家族菌株分为 62 个不同的基因型, 而 MIUR-15 法则将北京家族菌株细分为 155 个不同的基因型. 这些结果表明, 在西安地区, 北京家族菌株作为主要的肺结核致病菌, 具有不同基因型. 另外, 虽然同属于北京型菌株, 但在世界各地分离到的结核分枝杆菌临床菌株中, 有些菌株出现的频率明显高, 称之为典型的北京型菌株. 该型菌株在染色体的 NTF(noise transfer function)区域有 IS6110 序列的插入, 在 putative mutator 基因区也存在突变. 相对应地, 另外被称之为非典型的北京型菌株, 则缺乏在 putative mutator 基因区的突变, 也没有染色体 NTF 区的 IS6110 插入事件^[58]. 该菌群细菌在临床上较为少

见, 在系统发育关系上被认为更加接近该世系的早期菌株, 即可能比典型的北京型菌株古老些^[59].

结核分枝杆菌结构基因组的多态性水平较低, 对于能够全球传播的结核分枝杆菌, 其自然选择优势可能与其基因组内的重复序列有关. 从系统进化的时间尺度上, 北京株被认为是较年轻的变异体. Surikova 等^[60]认为, 可以通过使用与进化事件发生的速率水平相一致的遗传标记来重构基因组进化. 例如, 区分近期散播的菌株就需要有高进化速率的遗传标记. 而北京株在遗传上较为保守, 为此, 他们寻找满足条件的 VNTR(数目可变的串联重复序列)位点, 用 ETR-VNTR 分子分型方法以及 IS6110 插入序列、DR(直接重复)序列、SNP 等其他方法, 针对 99 株结核分枝杆菌和参考菌株来鉴定北京家族菌株. 结果, 所有菌株中, 51 个被鉴定为北京株, 且都有 ETR-A、-C、-E 基因型. 研究人员认为, VNTR 位点的多态性程度可能反映一些功能的差异, 如已有多多样性的保持. 另外, 其与修复系统的状态和对环境的适应也有联系. 基于 ETR 分型, 研究人员根据遗传关系将所有菌株分为两组——遗传分异组和近亲的北京菌株家族组, 并且比较了这两组菌株中每个 VNTR 位点的重复序列数目, 发现相比其他菌株, 在几乎所有研究的 VNTR 位点中, 北京株有更多的重复序列数目. 与其他菌株明显不同的是, 北京菌株家族的基因组在 QUB26 和 QUB18 位点有 7、8 或 9 个串联重复拷贝. QUB18 和 QUB11B 位点位于 *PPE* 基因内部, 后者可能与抗原变异性有关从而使致病菌具有逃避宿主免疫应答的潜力.

2.4.3 北京型菌株广泛传播的特性与其基因组进化的关系. 北京型菌株虽然曾是地方性流行病原菌, 但目前已经在全世界广泛传播并引起了耐多药结核病的大爆发. 目前就全球来看, 结核病 10% 以上由感染北京型菌株引起, 就东南亚地区而言, 这一比例则高达 50% 以上^[61-62]. 在中国台湾, Chang 等^[63]收集了从 2004 年到 2007 年间的共 338 株北京型菌株. 研究者们通过 10 个同义 SNPs 位点分析, 将这些临床菌株分为 9 个亚型. 通过分析染色体 NTF 区 IS6110 序列插入情况, RD105、RD207、RD181、RD142 和 RD150 等 LSP 以及 *mutT2*、*mutT4* 和 *ogt* 基因的突变等, 发现 9 个亚型中仅有 ST11 和 ST26 基因型没有 RD181 缺失且 NTF 区没有 IS6110 序列的插入, 故被认为是较古老的北京型菌株. 其他较现代的亚型中, ST10 包括了最多

的临床菌株,但这当中,79%属于非耐药性菌株。在所有古老的北京型菌株中未见 *fbpA* 或 *fbpB* 基因的突变,而在较现代的北京型菌株中, *fbpB* 基因却有 238 位密码子的一个 SNP。在所有较现代的菌株中,有关 DNA 修复酶的编码基因,如 *mutT2*、*mutT4*、*ogt12* 和 *ogt37* 等基因均有变异;而在古老的北京型菌株中,这些基因全都未见变异,但这些菌种,MDR 菌株比例却最高。由此可见,耐药性强且较古老的北京型菌株在感染人群中数量较少。相反,由于 DNA 修复酶基因的突变使得细菌基因组容易产生变异,由此产生的较现代的北京型菌株不仅传播力强,且更易于适应各个地区的人群从而实现广泛传播。北京型菌株易传播的特性与菌株本身的耐药性并无直接关系。瑞典 Solomon 等收集了从 1994 年到 2008 年间的结核分枝杆菌临床菌株,经过分子流行病学调查研究,发现 536 株耐药菌株中有 70 株为北京型菌株,且有耐药基因的突变。但发现被这些耐药的北京型菌株感染的病人并非瑞典当地人,而是来自北京株盛行的国家,如俄罗斯等。可见,即使有致病特性,北京型菌株在瑞典并未传播开,这可以解释为是该基因型并未适应斯堪的纳维亚人群^[64]。

3 结 语

以上综述了最近几年有关结核分枝杆菌及其近亲菌种基因组的最新认识及研究进展,回答了在结核分枝杆菌基因组进化过程中的几个主要问题。从中可以总结出结核分枝杆菌基因组的进化趋势:即可能向着有利于细菌本身对寄主的适应和对不良环境的耐受方向进化,其结果是结核分枝杆菌不断地产生变异以抵抗包括药物等各种不利因素从而能长期而广泛地感染并寄生于人类等寄主上,发挥着其对寄主的有害作用。故此除了做好日常的流行病防治工作外,结核分枝杆菌致病性与人群迁徙的相关性也越来越多地被研究。这些研究以地区为界限,以人群为单位研究结核分枝杆菌的流行状况。而个性化医疗概念的提出,使人们开始注意到一个人群内部不同个体间的差异,使医疗手段更具针对性。对结核病的科研成果向临床的转化环节也是我们防治结核病工作的重要环节。相信随着人们的不断努力,对结核分枝杆菌的生理生化性质的认识会逐渐接近其本质,从而步步为营,直到攻克结核病。

参 考 文 献

- [1] Narayanan S, Gagneux S, Hari L, *et al.* Genomic interrogation of ancestral *Mycobacterium tuberculosis* from south India. *Infect Genet Evol*, 2008, **8**(4): 474–483
- [2] Devulder G, Perouse de Montclos M, Flandrois J P. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int J Syst Evol Micr*, 2005, **55**(Pt 1): 293–302
- [3] Cole S T. Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS Lett*, 1999, **452**(1–2): 7–10
- [4] Cole S T, Brosch R, Parkhill J, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 1998, **393**(6685): 537–544
- [5] Camus J C, Pryor M J, Medigue C, *et al.* Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*, 2002, **148**(Pt 10): 2967–2973
- [6] Fleischmann R D, Alland D, Eisen J A, *et al.* Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol*, 2002, **184**(19): 5479–5490
- [7] Fang Z, Doig C, Kenna D T, *et al.* IS6110-mediated deletions of wild-type chromosomes of *Mycobacterium tuberculosis*. *J Bacteriol*, 1999, **181**(3): 1014–1020
- [8] Steenken W Jr, Gardner L U. History of H37 strain of tubercle bacillus. *Am Rev Tuberc*, 1946, **54**: 62–66
- [9] Zheng H, Lu L, Wang B, *et al.* Genetic basis of virulence attenuation revealed by comparative genomic analysis of *Mycobacterium tuberculosis* strain H37Ra versus H37Rv. *PLoS One*, 2008, **3**(6): e2375
- [10] Manca C, Tsenova L, Barry 3rd C E, *et al.* *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response *in vivo* and *in vitro*, but is not more virulent than other clinical isolates. *J Immunol*, 1999, **162**(11): 6740–6746
- [11] Sun R, Converse P J, Ko C, *et al.* *Mycobacterium tuberculosis* ECF sigma factor SigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol Microbiol*, 2004, **52**(1): 25–38
- [12] Moroz O V, Murzin A G, Makarova K S, *et al.* Dimeric dUTPases, HisE, and MazG belong to a new superfamily of all-alpha NTP pyrophosphohydrolases with potential "house-cleaning" functions. *J Mol Biol*, 2005, **347**(2): 243–255
- [13] Lu L D, Sun Q, Fan X Y, *et al.* Mycobacterial MazG is a novel NTP pyrophosphohydrolase involved in oxidative stress response. *J Biol Chem*, 2010, **285**(36): 28076–28085
- [14] Perez E, Samper S, Bordas Y, *et al.* An essential role for *phoP* in *Mycobacterium tuberculosis* virulence. *Mol Microbiol*, 2001, **41**(1): 179–187
- [15] Cuccui J, Easton A, Chu K K, *et al.* Development of signature-tagged mutagenesis in *Burkholderia pseudomallei* to identify genes important in survival and pathogenesis. *Infect Immun*, 2007, **75**(3):

- 1186–1195
- [16] Mestdagh M, Fonteyne P A, Realini L, *et al.* Relationship between pyrazinamide resistance, loss of pyrazinamidase activity, and mutations in the *pncA* locus in multidrug-resistant clinical isolates of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 1999, **43**(9): 2317–2319
- [17] Scorpio A, Lindholm-Levy P, Heifets L, *et al.* Characterization of *pncA* mutations in pyrazinamide-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*, 1997, **41**(3): 540–543
- [18] Shi W, Zhang X, Jiang X, *et al.* Pyrazinamide inhibits trans-translation in *Mycobacterium tuberculosis*. *Science*, 2011, **333**(6049): 1630–1632
- [19] Anton I A, Coggins J R. Sequencing and overexpression of the *Escherichia coli aroE* gene encoding shikimate dehydrogenase. *Biochem J*, 1988, **249**(2): 319–326
- [20] Roberts F, Roberts C W, Johnson J J, *et al.* Evidence for the shikimate pathway in apicomplexan parasites. *Nature*, 1998, **393**(6687): 801–805
- [21] Zhang X, Zhang S, Hao F, *et al.* Expression, purification and properties of shikimate dehydrogenase from *Mycobacterium tuberculosis*. *J Biochem Mol Biol*, 2005, **38**(5): 624–631
- [22] Reichau S, Jiao W, Walker S R, *et al.* Potent inhibitors of a shikimate pathway enzyme from *Mycobacterium tuberculosis*: combining mechanism- and modeling-based design. *J Biol Chem*, 2011, **286**(18): 16197–16207
- [23] Chopra S, Matsuyama K, Tran T, *et al.* Evaluation of gyrase B as a drug target in *Mycobacterium tuberculosis*. *J Antimicrob Chemother*, 2012, **67**(2): 415–421
- [24] Cubillos-Ruiz A, Morales J, Zambrano M M. Analysis of the genetic variation in *Mycobacterium tuberculosis* strains by multiple genome alignments. *BMC Res Notes*, 2008, **1**: 110
- [25] Ioegeer T R, Feng Y C, Ganesula K, *et al.* Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *J Bacteriol*, 2010, **192**(14): 3645–3653
- [26] Mulder N, Rabi H, Jamieson G, *et al.* Comparative analysis of microbial genomes to study unique and expanded gene families in *Mycobacterium tuberculosis*. *Infect Genet Evol*, 2009, **9**(3): 314–321
- [27] Tekaia F, Gordon S V, Garnier T, *et al.* Analysis of the proteome of *Mycobacterium tuberculosis* in silico. *Tuber Lung Dis*, 1999, **79**(6): 329–342
- [28] Brosch R, Gordon S V, Marmiesse M, *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA*, 2002, **99**(6): 3684–3689
- [29] Gutierrez M C, Brisse S, Brosch R, *et al.* Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog*, 2005, **1**(1): e5
- [30] Arber W. Elements for a theory of molecular evolution. *Gene*, 2003, **317**(1–2): 3–11
- [31] Lorenz M G, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev*, 1994, **58**(3): 563–602
- [32] Arnold C. Molecular evolution of *Mycobacterium tuberculosis*. *Clin Microbiol Infect*, 2007, **13**(2): 120–128
- [33] Ernst J D, Trevejo-Nunez G, Banaiee N. Genomics and the evolution, pathogenesis, and diagnosis of tuberculosis. *J Clin Invest*, 2007, **117**(7): 1738–1745
- [34] Huard R C, Fabre M, de Haas P, *et al.* Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. *J Bacteriol*, 2006, **188**(12): 4271–4287
- [35] Hirsh A E, Tsolaki A G, DeRiemer K, *et al.* Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci USA*, 2004, **101**(14): 4871–4876
- [36] Smith N H, Dale J, Inwald J, *et al.* The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc Natl Acad Sci USA*, 2003, **100**(25): 15271–15275
- [37] Supply P, Warren R M, Banuls A L, *et al.* Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol Microbiol*, 2003, **47**(2): 529–538
- [38] Smith N H, Gordon S V, de la Rua-Domenech R, *et al.* Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat Rev Microbiol*, 2006, **4**(9): 670–681
- [39] Brosch R, Gordon S V, Marmiesse M, *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA*, 2002, **99**(6): 3684–3689
- [40] Veyrier F J, Dufort A, Behr M A. The rise and fall of the *Mycobacterium tuberculosis* genome. *Trends Microbiol*, 2011, **19**(4): 156–161
- [41] Shi L, North R, Gennaro M L. Effect of growth state on transcription levels of genes encoding major secreted antigens of *Mycobacterium tuberculosis* in the mouse lung. *Infect Immun*, 2004, **72**(4): 2420–2424
- [42] Hershberg R, Lipatov M, Small P M, *et al.* High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol*, 2008, **6**(12): e311
- [43] Gagneux S, DeRiemer K, Van T, *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*, 2006, **103**(8): 2869–2873
- [44] North R J, Jung Y J. Immunity to tuberculosis. *Annu Rev Immunol*, 2004, **22**: 599–623
- [45] Comas I, Chakravarti J, Small P M, *et al.* Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet*, 2010, **42**(6): 498–503
- [46] Sampson S L, Lukey P, Warren R M, *et al.* Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. *Tuberculosis (Edinb)*, 2001, **81**(5–6): 305–317
- [47] Banu S, Honore N, Saint-Joanis B, *et al.* Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?. *Mol Microbiol*, 2002, **44**(1): 9–19
- [48] Le Moigne V, Robreau G, Borot C, *et al.* Expression,

- immunochemical characterization and localization of the *Mycobacterium tuberculosis* protein p27. *Tuberculosis* (Edinb), 2005, **85**(4): 213–219
- [49] Cascioferro A, Delogu G, Colone M, *et al.* PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. *Mol Microbiol*, 2007, **66**(6): 1536–1547
- [50] Delogu G, Pusceddu C, Bua A, *et al.* Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol*, 2004, **52** (3): 725–733
- [51] McEvoy C R, van Helden P D, Warren R M, *et al.* Evidence for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis* PPE38 gene region. *BMC Evol Biol*, 2009, **9**: 237
- [52] van Soolingen D, Qian L, de Haas P E, *et al.* Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *J Clin Microbiol*, 1995, **33**(12): 3234–3238
- [53] Tsolaki A G, Gagneux S, Pym A S, *et al.* Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol*, 2005, **43**(7): 3185–3191
- [54] Kremer K, Glynn J R, Lillebaek T, *et al.* Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. *J Clin Microbiol*, 2004, **42**(9): 4040–4049
- [55] Filliol I, Motiwala A S, Cavatore M, *et al.* Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: Insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol*, 2006, **188**(2): 759–772
- [56] Rindi L, Lari N, Cuccu B, *et al.* Evolutionary pathway of the Beijing lineage of *Mycobacterium tuberculosis* based on genomic deletions and mutT genes polymorphisms. *Infect Genet Evol*, 2009, **9**(1): 48–53
- [57] Zhou A, Nawaz M, Xue X, *et al.* Molecular genotyping of *Mycobacterium tuberculosis* in Xi'an, China, using MIRU-VNTR typing. *Int J Tuberc Lung Dis*, 2011, **15**(4): 517–522
- [58] Hanekom M, Gey van Pittius N C, McEvoy C, *et al.* *Mycobacterium tuberculosis* Beijing genotype: A template for success. *Tuberculosis* (Edinb), 2011
- [59] Hanekom M, van der Spuy G D, Streicher E, *et al.* A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. *J Clinical Microbiology*, 2007, **45**(5): 1483–1490
- [60] Surikova O V, Voitech D S, Kuzmicheva G, *et al.* Efficient differentiation of *Mycobacterium tuberculosis* strains of the W-Beijing family from Russia using highly polymorphic VNTR loci. *Eur J Epidemiol*, 2005, **20**(11): 963–974
- [61] Brudey K, Driscoll J R, Rigouts L, *et al.* *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*, 2006, **6**: 23
- [62] Filliol I, Driscoll J R, van Soolingen D, *et al.* Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *J Clinical Microbiology*, 2003, **41**(5): 1963–1970
- [63] Chang J R, Lin C H, Tsai S F, *et al.* Genotypic analysis of genes associated with transmission and drug resistance in the Beijing lineage of *Mycobacterium tuberculosis*. *Clin Microbiol Infect*, 2011, **17**(9): 1391–1396
- [64] Ghebremichael S, Groenheit R, Pennhag A, *et al.* Drug resistant *Mycobacterium tuberculosis* of the Beijing genotype does not spread in Sweden. *PLoS One*, 2010, **5**(5): e10893

Genomics and Genome Evolution of *Mycobacterium tuberculosis**

WAN Bao-Shan¹⁾, ZHANG Qiu-Fen¹⁾, ZHOU Ai-Ping¹⁾, ZHAO Guo-Ping²⁾, YAO Yu-Feng^{1)**}

(¹⁾ Department of Medical Microbiology and Parasitology, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China;

²⁾ Laboratory of Synthetic Biology, Institute of Plant and Ecology, Shanghai Institute for Biological Sciences,
Chinese Academy of Sciences, Shanghai 200032, China)

Abstract The endless emergence of new theories and scientific instruments aiming to genome sequencing boosts the large ship of history advance of biology in the post-genomic era. With the genome sequences of major infectious disease-causing microorganisms being determined one by one, the subsequent annotation of gene function and the reconstruction work about protein functions are being carried out. All the efforts are to get a breakthrough understanding of biological characteristics, diagnostic strategies and treatment methods of pathogenic microorganisms. *Mycobacterium tuberculosis* has been a serious threat to human health globally. All genetic events occurring in the genome evolution play an important role in all aspects such as *M. tuberculosis*'s biological properties, pathogenicity and drug resistance. This review will summarize the origins and genomic features of *M. tuberculosis*, and discuss the current progress about its genome evolution.

Key words *Mycobacterium tuberculosis*, genome, comparative genomics, molecular evolution

DOI: 10.3724/SP.J.1206.2011.00469

* This work was supported by grants from The National Natural Science Foundation of China (31070114) and National Basic Research Program of China (2009CB522605).

**Corresponding author.

Tel: 86-21-64671226, E-mail: yfyao@sjtu.edu.cn

Received: October 20, 2011 Accepted: February 7, 2012