

## 卡介苗美国株全基因组测序缺口修补及序列

王志明, 潘元龙, 吴俊, 朱宝利\*

中国科学院微生物研究所, 中国科学院病原微生物与免疫学重点实验室, 北京 100101

**摘要:** 【目的】对卡介苗 (*Bacillus Calmette-Guerin*, BCG) 美国株 (BCG Tice) 进行基因组补缺口 (补洞) 工作, 以得到它的基因组完整序列。【方法】首先对 BCG Tice 进行高通量测序, 使用 SOAPdenovo 软件对得到的数据进行拼接。由于在高通量测序的过程中基因组某些区域测序覆盖度低, 测序质量差会使测序结果经拼接后形成众多的重叠群 (contig), 相邻的位置关系确定的 contig 形成一个 scaffold, contig 之间未测到的区域为缺口序列 (gap), 在 contig 末端设计引物进行 PCR 扩增, 得到连接相邻 contig 的 PCR 产物, 对 PCR 产物进行测序。通过优化 PCR 引物设计策略, 尝试不同的聚合酶进行聚合反应, 调整 PCR 反应条件并结合 PCR 产物构建克隆测序等方法, 补齐 contig 之间的缺口序列。【结果】完成了 BCG Tice 的全基因组测序, 得到了它的基因组完整序列, 序列已提交到美国国立生物技术信息中心 (NCBI) 的 GenBank 数据库。【结论】BCG 属于高 GC 含量的革兰氏阳性细菌, 其基因组 GC 含量高达 65.65%。本文以 BCG Tice 基因组补洞为例, 对高 GC 含量基因组补缺口过程中遇到的问题与采取的策略给予概述, 望给相关高 GC 含量基因组的物种全基因组测序补缺口工作提供一些借鉴。

**关键词:** 卡介苗, 高 GC, 补缺口, 序列分析

**中图分类号:** Q933      **文献标识码:** A      **文章编号:** 0001-6209 (2012) 10-1219-09

卡介苗 (*Mycobacterium bovis* BCG) 是由牛分支杆菌 (*Mycobacterium bovis*) 突变而来的, 它是唯一的一种用来预防结核病的疫苗。早在 1921 年 BCG 就被用于人类的预防接种, 至今已有近一个世纪的历史了。BCG 最初是由两位法国科学家通过将 *M. bovis* 连续传代而得到的。从 1924 年开始, BCG 被传播到世界各地。然而因为保存 BCG 菌株依然采取同祖先菌株 *M. bovis* 一样古老的传代培养方式, 所以 BCG 的基因组在传代过程中不断经历着突变, 直到 1961 年冷冻干燥菌株保存方法的出现才使得 BCG 的体外进化得以终止<sup>[1-3]</sup>。现今广泛使用的

BCG 是由 1921 年的 BCG 经历体外不断进化而产生的不同的子菌株。

BCG 的体外进化主要经历了两个阶段。第一个阶段是从 1908 年到 1921 年, 在这 13 年间, 法国两位科学家经过 230 代的连续传代培养得到了最初的结核病疫苗菌株 BCG, 第二个阶段是从 1924 年开始, 到 1966 年结束, 这个阶段 BCG 被传播到不同的国家, 开始了各自的进化历程, 产生了现今使用的 13 种不同的 BCG 子菌株。

BCG 的祖先菌株 *M. bovis* 的基因组在 2003 年测序完成<sup>[4]</sup>, 它的基因组含有 4345492 个碱基, 基因组

基金项目: 国家“973 项目” (2009CB522605, 2007CB513002)

\* 通信作者。Tel/Fax: +86-10-64807362; E-mail: zhubaoli@im.ac.cn

作者简介: 王志明 (1986-), 男, 山东省烟台市人, 硕士研究生, 微生物基因组学。E-mail: zhiming.w.com@163.com

收稿日期: 2012-04-20; 修回日期: 2012-05-09

的平均 GC 含量约为 65.6%。随后,在 2007 年与 2009 年,BCG 的巴斯德株 (BCG Pasteur) 与东京株 (BCG Tokyo) 的全基因组测序也相继完成。在 2011 年又有两株 BCG 子菌株墨西哥株 (BCG Mexico) 与巴西株 (BCG Moreau) 也完了全基因组测序<sup>[5-8]</sup>。

至今仍有 9 株 BCG 没有完成全基因组测序。其中 BCG Tice 是 Rosenthal 博士于 1933 年到 1934 年在巴斯德研究所培育得到的子菌株的后代,当时该子菌株叫做 Tice-Rosenthal,随后 Tice-Rosenthal 子菌株经过传代培养形成了 6 种不同的 BCG 菌株 (H, K, E, L, LH, BL)。1952 年, Rosenthal 保存并继续培养上述六种 BCG 菌株中预防结核病效果较好的一种 (BL 菌株)。而现今美国广泛使用的卡介苗是卡介苗美国菌株 (BCG Tice ATCC 35743),它是 BL 菌株的后代<sup>[9]</sup>。

我们使用 Illumina Genome Analyzer 测序仪对 BCG Tice 进行高通量测序,其基因组测序深度为 70.7 (测序深度指的是测序得到的碱基总量与基因组大小的比值,它是评价测序量的指标之一)。然后使用 SOAPdenovo 软件对得到的数据进行拼接并完成了基因组补缺口的初始工作,最终得到 25 个 scaffolds, 28 个 contigs, 即有 25 个物理缺口 (位置关系未知) 与 3 个序列缺口 (位置关系确定)。相关的结果已发表在 Journal of Bacteriology 上<sup>[10]</sup>。

本研究基于上述发表的结果继续进行基因组的补缺口工作,最终得到了 BCG Tice 的基因组完整序列,BCG Tice 全基因组序列已经提交至 NCBI 的 GenBank 数据库,登录号为 CP003494。本文提供了高 GC 含量的基因组补缺口的策略,并阐述了我们在此过程中所遇到的困难以及所采取的解决方案。希望能为高 GC 含量基因组的物种的补缺口工作提供一些参考。此外,我们还对 BCG Tice 基因组中相对于结核分枝杆菌 (*Mycobacterium tuberculosis*, *M. tb*) 的基因组片段缺失进行了确证,并将 BCG Tice 同其他 BCG 的毒力相关基因进行了比较分析。

## 1 材料和方法

### 1.1 材料

**1.1.1 主要试剂和仪器:** LA DNA 聚合酶、高保真 DNA 聚合酶、PMD19-T 载体、PSIMPLE-EcoR V/BAP 载体购自宝生物工程有限公司;大肠杆菌

(*Escherichia coli*) DH5 $\alpha$  感受态细胞购自全式金生物技术有限公司;凝胶回收试剂盒购自 OMEGA 公司;PCR 产物纯化试剂盒、异丙基- $\beta$ -D-硫代吡喃半乳糖苷、X-Gal 购自生工生物工程 (上海) 有限公司。SensoQuest Labcycler PCR 仪、Thermo 高速离心机、凝胶成像仪等。

**1.1.2 菌株和培养基:** BCG Tice (ATCC 35743) 购买自美国模式培养物集存库 (American type culture collection, ATCC)。Middlebrook 7H9 培养基 (硫酸铵 0.5 g, 磷酸二氢钾 1.0 g, 磷酸氢二钠 2.5 g, 柠檬酸钠 0.1 g, 硫酸镁 0.05 g, 氯化钙 0.0005 g, 硫酸锌 0.001 g, 硫酸铜 0.001 g, L-谷氨酸 0.5 g, 柠檬酸铁铵 0.04 g, 维生素 B6 0.001 g, 生物素 0.0005 g, 加入 100 mL 蒸馏水。)

### 1.2 细菌的培养与基因组 DNA 的提取

BCG 培养在含 10% ADC 的 Middlebrook 7H9 培养基中。基因组 DNA 提取使用的是 QIAGEN Genomic-tip 100/G kit (Qiagen)<sup>[11]</sup>。

### 1.3 选择参考基因组序列

由于 *M. bovis*, BCG Pasteur, BCG Tokyo, BCG Mexico, 以及 BCG Moreau 的基因组序列都已经完成全基因组测序,对这 5 株菌进行多序列比对分析,以寻找这些菌株序列之间的同源性。通过观察分析比对结果我们选择了上述 5 株菌中保守性较好的 BCG Pasteur 作为参考基因组序列。

### 1.4 确定 scaffolds 的前后位置关系

通过对 BCG Tice 25 个 scaffolds 与参考基因组 BCG Pasteur 进行多序列比对,初步确定这 25 个 scaffolds 的前后位置关系。

### 1.5 引物的设计与合成

**1.5.1 方案一:** 根据确定的 contig 的前后位置关系,利用相邻 contig 末端序列,使用 Primer Premier 5.0 进行引物设计。

**1.5.2 方案二:** 将缺口对应于参考序列上的区域设定为引物设计模板,在模板上设计多对扩增引物使它们的扩增产物产生重叠,重叠区约为 500 个碱基。

**1.5.3 引物合成:** 在六合华大基因公司进行。

### 1.6 PCR 反应

采用 1.5% 琼脂糖凝胶电泳分析 PCR 产物扩增的情况。

**1.6.1 LA DNA 聚合酶的反应体系:** 30  $\mu$ L 中包括双蒸水 6.5  $\mu$ L、模板 DNA 1  $\mu$ L、上下游引物各

1.2  $\mu\text{L}$  (10 mmol/L)、dNTP 4.8  $\mu\text{L}$ 、GC Buffer 15  $\mu\text{L}$ 、LA DNA 聚合酶 0.3  $\mu\text{L}$ 。PCR 反应条件为: 94 $^{\circ}\text{C}$  5 min; 94 $^{\circ}\text{C}$  1 min, 55 $^{\circ}\text{C}$  (60 $^{\circ}\text{C}$  或是 50 $^{\circ}\text{C}$ ) 1 min, 72 $^{\circ}\text{C}$  4 min, 30 个循环; 72 $^{\circ}\text{C}$  5 min。

**1.6.2 PrimeSTAR 酶反应体系:** 30  $\mu\text{L}$  包括双蒸水 8.9  $\mu\text{L}$ , 模板 DNA 1  $\mu\text{L}$ , 上下游引物各 1.2  $\mu\text{L}$  (10 mmol/L), dNTP 4.8  $\mu\text{L}$ , PrimeSTAR GC Buffer 15  $\mu\text{L}$ , PrimeSTAR DNA 聚合酶 0.3  $\mu\text{L}$ 。PCR 反应条件为 98 $^{\circ}\text{C}$  1 min; 98 $^{\circ}\text{C}$  1 min, 68 $^{\circ}\text{C}$  4–8 min 进行 30 个循环。

### 1.7 将 LA 酶 PCR 扩增产物构建 T-A 克隆

LA DNA 聚合酶扩增得到的 PCR 产物连接到 T 载体。首先将 PCR 扩增产物进行琼脂糖凝胶电泳对目的回收产物进行鉴定; 然后将经鉴定大小正确的条带切胶回收; 接着将回收产物与 T 载体相连并转化入大肠杆菌感受态细胞, 在含有 IPTG, X-Gal 以及氨苄青霉素 (100 mg/mL) 的 LB 平板上挑取白色的菌落; 随后将挑取的菌落接种在含有氨苄青霉素的液体 LB 培养基中进行扩大培养; 最后对扩大培养的菌液进行 PCR 鉴定, 鉴定正确后进行测序。

### 1.8 平末端克隆的建立

该方法针对的是通过 PrimeStar DNA 酶扩增得到的 PCR 产物, 该酶扩增得到的 PCR 产物为平末端, 需要连接到平末端载体, 我们使用的是 PSILMPLE-EcoR V/Bap 载体, 同样, 将酶扩增得到的 PCR 产物进行切胶回收, 回收并经鉴定的产物与 PSILMPLE-EcoR V/Bap 载体进行连接与转化。连接产物经 PCR 鉴定正确后进行测序。PCR 产物与菌液的测序工作在六合华大基因公司进行。

### 1.9 测序结果的比对与拼接

利用 blast-2.2.22 软件对将测序得到的序列与参考序列进行比对分析。利用 phrap-phred-consed 软件<sup>[12]</sup>进行序列的拼接。利用 BioEdit 软件对序列进行编辑整理。

### 1.10 基因组注释与比较基因组分析

基因组注释使用 glimmer 3.0 软件对基因的编码区进行预测, 使用 t-RNAscan-SE 进行 tRNA 的预测。注释的结果与 KEGG 数据库, COG 数据库, 以及 NCBI 的蛋白质数据库进行比对。利用 Mauve 2.3.1 进行多序列比对分析。

## 2 结果

### 2.1 参考序列基因组的选择

通过对 5 株 BCG 子菌株 (包括 BCG Tice, BCG Pasteur, BCG Tokyo, BCG Mexico 与 BCG Moreau) 进行多序列比对, 我们发现这 5 株菌的基因组之间差别不大, 基因组序列之间比较保守, 图 1 显示了 5 株 BCG 子菌株进行多序列比对的部分结果。从比对结果可以看出, *M. bovis* 相对其他的 BCG 子菌株基因组存在较大差异, 而其他 BCG 子菌株之间的基因组差异不大, 我们从 BCG Tice 以外的 4 株 BCG 中选择了基因组相对最为保守的 BCG Pasteur 作为参考序列基因组。

### 2.2 确定顺序未知的 25 个 scaffolds 的前后位置关系

通过对 BCG Tice 的 25 个 scaffolds 和参考基因组进行序列比对, 初步确定 scaffolds 的前后位置关系为: scaffold 20-6-11-16-10-15-21-23-4-18-25-5-17-8-14-7-13-22-1-3-19-12-24-2-9-20。

### 2.3 BCG Tice 基因组概况

BCG Tice 全基因组大小为 4334064 个碱基, 全基因组的 GC 含量为 65.65% (图 2), 预测得到 4038 个可编码蛋白质的基因, 45 个 tRNA 基因以及 3 个 rRNA 基因 (表 1)。在预测得到的 4038 个蛋白质编码基因中, 负责铁离子获取与代谢的基因最少, 只有 1 个; 负责运动与趋化性、潜伏与孢子形成的基因各有 2 个; 芳香族化合物代谢的基因有 8 个; 辅助因子, 维生素, 辅基, 色素相关的基因 299 个; 细胞壁与荚膜基因 37 个; 毒力、致病性与保护性的基因 42 个; 钾离子代谢, 辅助功能, 膜运输, RNA 代谢的基因分别为 15 个, 129 个, 24 个, 114 个; 核苷与核苷酸基因 78 个; 蛋白质代谢, 细胞分裂与细胞周期, 调控与细胞信号, DNA 代谢的基因分别为 150 个, 29 个, 75 个, 130 个; 参与脂肪酸, 脂类以及异戊烯化合物代谢相关的基因有 197 个; 氮代谢基因 21 个; 呼吸, 压力反应基因分别有 91 个和 80 个; 氨基酸与氨基酸衍生物代谢相关基因高达 353 个; 硫代谢基因为 15 个; 磷代谢代谢基因 47 个, 碳水化合物代谢基因为 245 个, 功能未知基因为 1198 个 (图 3)。

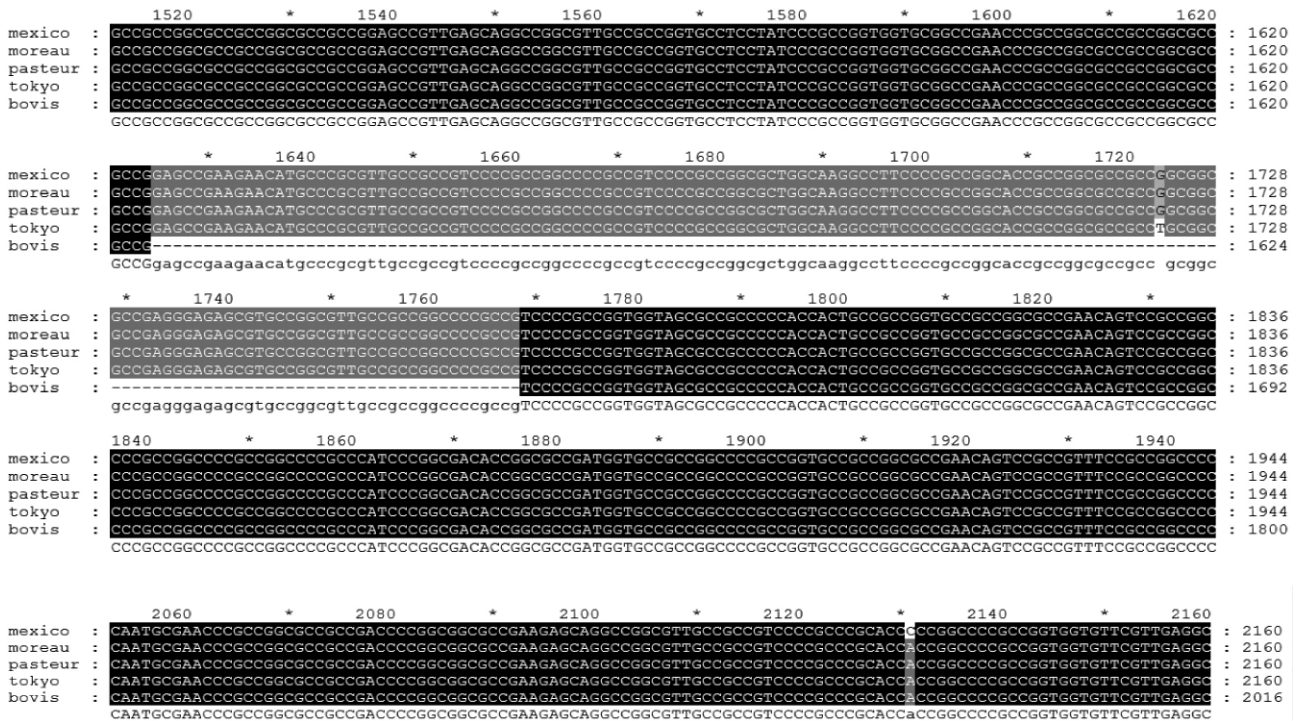


图 1 BCG mexico, BCG Moerau, BCG Pasteur, BCG Tokyo, M. bovis 多序列比对部分结果

Fig. 1 Partial results of multi-sequence alignment among BCG Mexico, BCG Moreau, BCG Pasteur, BCG Tokyo and *M. bovis* The sequence that are consensus among four BCG substains and *M. bovis* are shown in black, while those that are not consensus are shown in gray.

2.4 BCG Tice 与其他 BCG 基因组比较分析

2.4.1 BCG 子菌株之间的比较: 通过将 BCG Tice 同 BCG Mexico, BCG Moreau, BCG Pasteur 以及 BCG Tokyo 的全基因组比对 (图 4), 我们发现它们之间基因组相似性非常高, 只存在少数插入缺失的差异, 在进化上非常保守。

2.4.2 相对于结核分支杆菌基因组大片段缺失的验证: 根据相关的文献报道<sup>[17]</sup>, 相对于 *M. tb*, BCG Tice 基因组中存在 RD1 (缺失区 1), RD2 (缺失区 2), nRD18 (缺失区 18) 区域。这些基因组的差异是通过 DNA 芯片以及基因组芯片的信息得到的, 我们将 BCG Tice 的基因组同结核分枝杆菌 (*Mycobacterium tuberculosis*, *M. tb*) 的基因组进行全基因组多序列比对, 证明了这些缺失区域确实存在于 BCG Tice 中, 对芯片的结果进行了确证。

2.4.3 PE, PPE, PE-PRGS 基因家族的比较分析: PE, PPE, PE-PRGS 基因家族与 BCG 同宿主免疫细胞的相互作用直接相关<sup>[13]</sup>。这些基因家族在不同的 BCG 子菌株之间存在较大差异。在 BCG Tice 中 PE 基因家族的编码基因有 18 个, PPE 基因家族的编码基因有 74 个, 而 PE-PRGS 基因家族的

编码基因有 80 个。在 BCG Pasteur 中, PE 基因家族, PPE 基因家族, PE-PRGS 基因家族的数目分别是 32 个、60 个和 62 个。在 BCG Tokyo 中, PE 基因家族编码基因为 32 个, PPE 基因家族基因 60 个, PE-PRGS 家族基因 68 个。在 BCG Mexico 中, 有 31 个 PE 基因家族的编码基因, 61 个 PPE 基因家族编码基因, 49 个 PE-PRGS 基因家族的编码基因。BCG Moreau 中, PE, PPE, PE-PRGS 的编码基因分别是 32 个、56 个和 56 个。表明在 BCG 的体外进化过程中, 这 3 个基因家族在 BCG 子菌株之间产生较大的差异 (表 2)。

3 讨论

高 GC 含量基因组补缺口工作一直以来都是全基因组测序工作的困难所在, 我们通过引物设计上的调整, 结合构建克隆测序的方法, 最终完成了 BCG Tice 的全基因组测序工作, BCG Tice 全基因组序列的获得为我们寻找 BCG 菌株之间差异, 研究 BCG 的毒力与进化都提供了更多的信息。

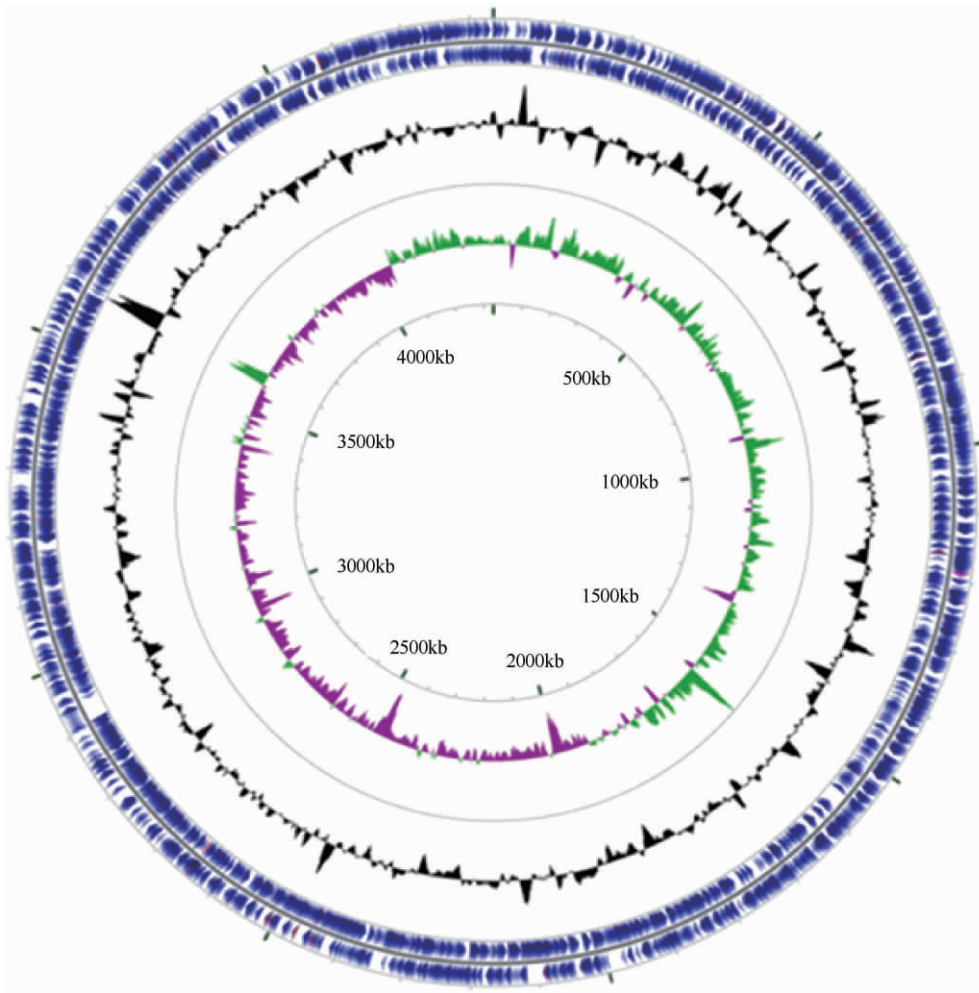


图 2 BCG Tice 的全基因组序列

Fig. 2 complete genome sequence of BCG Tice. Protein-encoding genes, tRNA genes and rRNA genes are shown in the first two outer blue ring, which indicates genes coded on the forward and complementary strand, respectively. The next inner ring indicated the average GC ratio, out emerge representing higher GC content than the average and inner emerge representing lower. The fourth ring indicated GC skew. The innermost ring represented the location of genome, Kbp unit of measurement.

表 1 五株 BCG (BCG Tice, BCG Pasteur, BCG Tokyo BCG Mexico, BCG Moreau) 基因组信息比较

Table 1 Genomic comparative information of five BCG substrains (including BCG Tice, BCG Pasteur, BCG Tokyo BCG Mexico and BCG Moreau)

Strains	Genome size (bp)	G + C content (%)	Coding sequence	No. of tRNA	No. of rRNA
BCG Tice	4334064	65.65	4038	45	3
BCG Pasteur	4374522	65.64	3954	47	3
BCG Tokyo	4371711	65.64	3950	45	3
BCG Mexico	4350386	65.7	3904	45	3
BCG Moreau	4340116	65.64	3945	45	3

### 3.1 BCG Tice 基因组补缺口策略

在整个实验过程中高 GC 含量的基因组片段扩增一直是基因组补缺口的瓶颈,以前关于高 GC 含量基因组片段扩增与测序的方法仅有较少的相关报道<sup>[14-16]</sup>。例如 Moreau 等曾报道通过反向 PCR 的

方法扩增得到大小为 1.9 Kb,GC 含量为 74% 的基因组片段<sup>[14]</sup>。而我们在实验过程中遇到的最大缺口约为 12 Kb,GC 含量则高达 80%。

理论上两个 contig 只要位置关系确定,在 contig 的末端设计 PCR 引物,并进行 PCR 扩增与 PCR 产

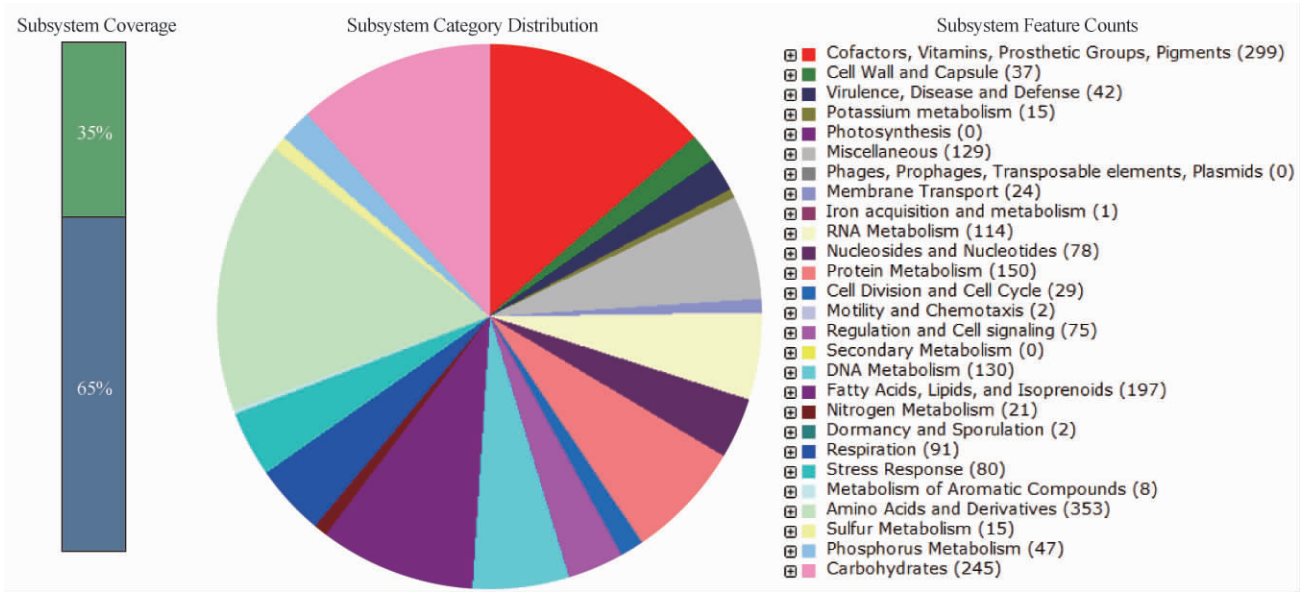


图3 直系同源蛋白聚类

Fig. 3 Cluster of Orthologous Groups of proteins (COG) functional categories.

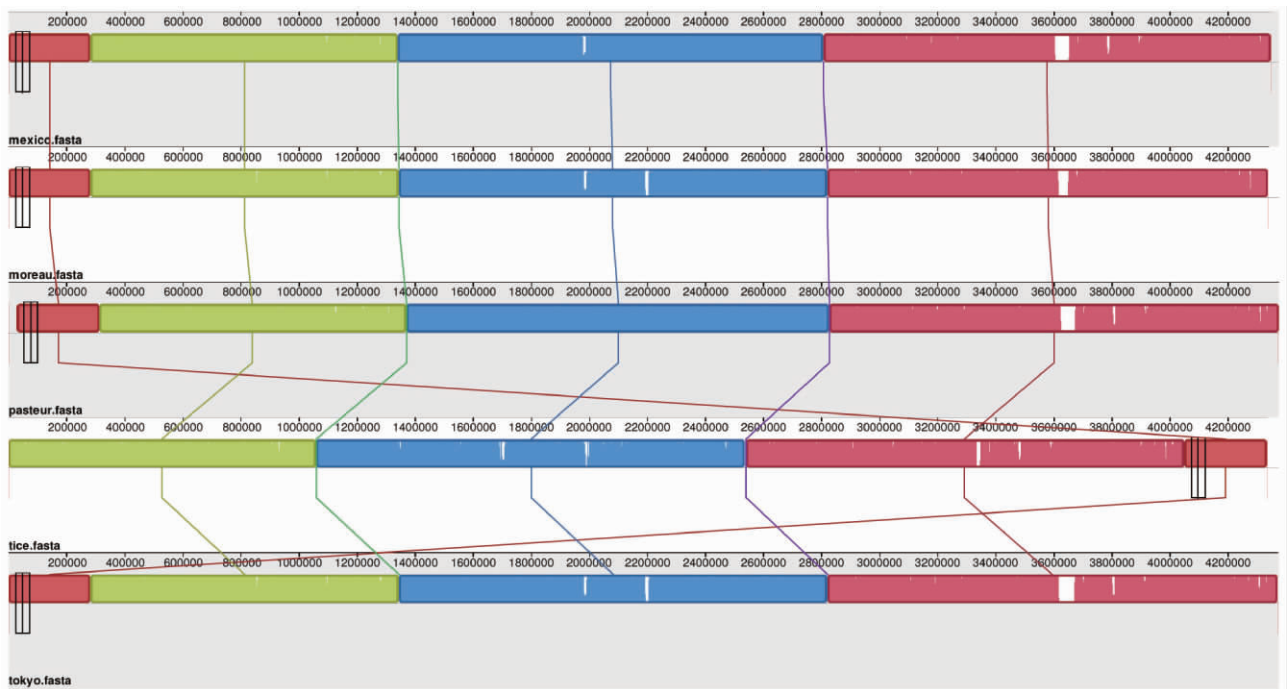


图4 BCG 墨西哥株, 巴西株, 巴斯德株, 美国株, 与东京株多序列比对的结果

Fig. 4 Multi-sequence alignment of BCG Mexico, BCG Moreau, BCG Pasteur, BCG Tice and BCG Tokyo. The white spot above indicate differences among five BCG substains.

物测序就可以将 contig 连接起来, 最终拼接成完整的基因组。这种方法对于 GC 含量适中并且不存在较为复杂结构的基因组较为适用, 但不能完全适用于基因组高 GC 含量的革兰氏阳性细菌。通过将 contig 同参考序列进行比对, 我们发现缺口序列的

GC 含量都高于 65%, 有的缺口较大, 超过 10 Kb。因此我们仅仅利用末端设计引物, PCR 扩增测序的策略无法将所有的缺口补齐。对于高 GC 含量的缺口, 长度大于 5 Kb 的缺口我们所采取的策略是以参考序列作为引物设计模板, 设计能扩增产生重叠区



表 2 五株 BCG 中 PE, PPE, PE-PGRS 基因家族的比较

Table 2 Comparison of PE, PPE and PE-PGRS gene family between five BCG substrains

Strains	No. of PE family protein	No. of PPE family protein	No. of PE-PGRS family protein
BCG Tice	18	74	80
BCG Pasteur	32	60	62
BCG Tokyo	32	60	68
BCG Mexico	31	61	49
BCG Moreau	32	56	56

域的引物。实验结果也表明, 这种策略能够克服基因组 GC 含量高, 缺口片段长而产生的难扩增, 难测序的问题。

在 BCG Tice 基因组补缺口的过程中, 同样由于扩增产物 GC 含量高而经常出现测序失败的情况。我们通过将 PCR 产物连接到测序载体上, 并利用载体上的引物进行测序, 大大提高了测序反应的成功率。

对无完成图参考的高 GC 含量菌株的拼接, 或者高 GC 含量且与参考基因组序列差异大的菌株的拼接, 可有多种 PCR 为基础的补缺口技术作为参考。反向 PCR 的方法, 即通过将 DNA 片段环化以及在 DNA 片段另外的位点去环化使目的 DNA 片段反向, 在已知序列的末端设计引物扩增, 得到已知序列之外的未知序列。多重 PCR 的方法, 即将多对 PCR 引物放在一个 PCR 反应体系中进行扩增反应, 如果 PCR 产物进行凝胶电泳有条带出现, 则需进行进一步的实验以确定该产物是由哪一对引物配对所产生。此外还有接头 PCR, 限制性位点 PCR, 捕获 PCR 等方法<sup>[17]</sup>都被成功的运用到参考序列未知的基因组补缺口工作中。而对于高 GC 的 DNA 片段难以扩增的问题有长片段扩增 PCR 技术加以解决。这些技术在基因组补缺口工作中都显示了各自的优势。

### 3.2 BCG Tice 基因组缺失区域的验证

现已有大量文献报道了关于不同 BCG 菌株基因组之间差异的研究结果, 这些研究成果大都以 *M. tb* 为参照菌株。1999 年, Behr 等<sup>[18]</sup>通过 DNA 芯片的方法分析了 BCG 菌株同 *M. tb* 的基因组差异, 他们使用的 DNA 芯片的分辨率为仅为 2Kb (即 BCG 基因组的 1/2000), 只找到了基因组中大片段插入缺失的差异, DNA 芯片结果显示在 BCG Tice 中存在基因组缺失区 1 与缺失区 2 (RD1, RD2, 相对于 *M. tb*)。但是由于 DNA 芯片分辨率的限制, 因此无

法反映出基因组之间 2Kb 以下的微小差异。缺失区 1 (RD1) 是所有的 BCG 菌株都存在的缺失区域, 而缺失区 2 (RD2) 则存在于 BCG Tice 等在 1927 年到 1931 年传播出巴斯德研究所的 BCG 菌株中, 这两个区域都与 BCG 的毒力相关。因此基因组中 RD1 与 RD2 区域的缺失致使 BCG 毒力减弱。

Serge<sup>[19]</sup>等通过基因芯片的方法企图找到 BCG 菌株基因组之间的更微小的差异, 基因芯片分辨率要比 DNA 芯片高 10 倍。在 BCG Tice 中除了发现 Behr 等找到的 RD1, RD2 缺失区外, 还发现了另外的大片段缺失区 (nRD18 区域), 该区域包含 1547 个碱基, 它的缺失使 BCG Tice 基因组中失去了 3 个开放阅读框。

我们通过对 BCG Tice 同 *M. tb* 进行全基因组多序列比对, 证明相对于 *M. tb*, BCG Tice 中确实存在 RD1, RD2, nRD18 区域的缺失。

### 3.3 毒力基因突变的生物学意义

PE, PPE, PE-PGRS 基因家族与 BCG 的毒力相关。它们的作用可能是参与宿主与细菌之间的相互作用, 这些基因在细菌中表达的蛋白, 通过诱导 T 细胞与 B 细胞产生效应, 引发宿主的免疫反应<sup>[13]</sup>。他们还参与细胞表面的相互作用以及与巨噬细胞的相互作用<sup>[20]</sup>, 这些基因所表达的蛋白在结构上具有一定的保守性, 即 N 端为 110 个氨基酸组成的保守区, 并且具有甘氨酸组成的保守结构域。我们通过比较不同 BCG 菌株之间 PE, PPE, PE-PGRS 基因, 发现这些基因家族在基因个数上存在较大差异, 这可能是在 BCG 体外进化过程中不同的选择压力导致的结果。

BCG Tice 全基因组序列的获得为我们发掘更多的 BCG 菌株之间的差异, 研究 BCG 的体外进化历程与减毒机制, 设计更好的预防结核病的疫苗, 都提供了更多的参考与思路。

### 参考文献

- [1] Moskwa Z. History of anti-tuberculosis vaccination and the development of BCG vaccine. *Wiadomosci Lekarskie*, 1983, 36 (1): 81-85.
- [2] Crispin R. History of BCG and its substrains. *Progress in Clinical and Biology Research*, 1989, 310: 35-50.
- [3] Liu J, Tran V, Leung AS, Alexander DC, Zhu B. BCG vaccines: their mechanisms of attenuation and impact on safety and protective efficacy. *Human vaccines*, 2009, 5 (2): 70-78.

- [4] Garnier T, Eiglmeier K, Camus JC, Medina N, Mansoor H, Pryor M, Duthoy S, Grondin S, Lacroix C, Monsempe C, Simon S, Harris B, Atkin R, Doggett J, Mayes R, Keating L, Wheeler PR, Parkhill J, Barrell BG, Cole ST, Gordon SV, Hewinson RG. The complete genome sequence of *Mycobacterium bovis*. *Proceeding of the National Academy of Sciences of the United States of America*, 2003, 100 (13) : 7877-7882.
- [5] Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, Dos Santos S, Duthoy S, Lacroix C, Garcia-Pelayo C, Inwald JK, Golby P, Garcia JN, Hewinson RG, Behr MA, Quail MA, Churcher C, Barrell BG, Parkhill J, Cole ST. Genome plasticity of BCG and impact on vaccine efficacy. *Proceeding of the National Academy of Sciences of the United States of America*, 2007, 104 (13) : 5596-5601.
- [6] Seki M, Honda I, Fujita I, Yano I, Yamamoto S, and Koyama A. Whole genome sequence analysis of *Mycobacterium bovis* bacillus Calmette-Guerin (BCG) Tokyo 172: a comparative study of BCG vaccine substrains. *Vaccine*, 2009, 27 (11) : 1710-1716.
- [7] Orduna P, Cevallos MA, de Leon SP, Arvizu A, Hernandez-Gonzalez IL, Mendoza-Hernandez G, Lopez-Vidal Y. Genomic and proteomic analyses of *Mycobacterium bovis* BCG Mexico 1931 reveal a diverse immunogenic repertoire against tuberculosis infection. *Biomed Central Genomics*, 2011, 12: 493.
- [8] Gomes LH, Otto TD, Vasconcellos EA, Ferrao PM, Maia RM, Moreira AS, Ferreira MA, Castello-Branco LR, Degraive WM, Mendonca-Lima L. Genome sequence of *Mycobacterium bovis* BCG Moreau, the Brazilian vaccine strain against tuberculosis. *Journal of bacteriology*, 2011, 193 (19) : 5600-5601.
- [9] Grzybowski S. BCG Vaccine: Tuberculosis-Cancer. *Canadian Medical Association Journal*, 1981, 125 (4) : 365-366.
- [10] Pan Y, Yang X, Duan J, Lu N, Leung AS, Tran V, Hu Y, Wu N, Liu D, Wang Z, Yu X, Chen C, Zhang Y, Wan K, Liu J, Zhu B. The whole genome sequence of four BCG vaccine strains. *Journal of bacteriology*, 2011, 193 (12) : 3152-3153.
- [11] Leung AS, Tran V, Wu Z, Yu X, Alexander DC, Gao GF, Zhu B, and Liu J. Novel genome polymorphisms in BCG vaccine strains and impact on efficacy. *Biomed Central Genomics*, 2008, 9: 413.
- [12] Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Research*, 1998, 8 (3) : 195-202.
- [13] Delogu G, Brennan MJ. Comparative immune response to PE and PE \_ PGRS antigens of *Mycobacterium tuberculosis*. *Infection and Immunity*, 2001, 69 (9) : 5606-5611.
- [14] Moreau A, Wang DS, Forget S, Duez C, Dusart J. GC-rich template amplification by inverse PCR. DNA polymerase and solvent effects. *Methods in Molecular Biology*, 2002, 192: 75-80.
- [15] Frey UH, Bachmann HS, Peters J, Siffert W. PCR-amplification of GC-rich regions: 'slowdown PCR'. *Nat Protoc*, 2008, 3 (8) : 1312-1317.
- [16] Sahdev S, Saini S, Tiwari P, Saxena S, Singh Saini K. Amplification of GC-rich genes by following a combination strategy of primer design, enhancers and modified PCR cycle conditions. *Molecular and Cellular Probes*, 2007, 21 (4) : 303-307.
- [17] Rogers YC, Munk AC, Meincke LJ, Han CS. Closing bacterial genomic sequence gaps with adaptor-PCR. *Biotechniques*, 2005, 39 (1) : 31-34.
- [18] Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, 1999, 284 (5419) : 1520-1523.
- [19] Mostowy S, Tsolaki AG, Small PM, Behr MA. The in vitro evolution of BCG vaccines. *Vaccine*, 2003, 21 (27) : 4270-4274.
- [20] Flores J, Espitia C. Differential expression of PE and PE \_ PGRS genes in *Mycobacterium tuberculosis* strains. *Gene*, 2003, 318: 75-81.



# Complete genome sequencing and sequence analysis of BCG Tice

Zhiming Wang, Yuanlong Pan, Jun Wu, Baoli Zhu\*

Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

**Abstract:** [Objective] The objective of this study is to obtain the complete genome sequence of Bacillus Calmette-Guerin Tice (BCG Tice), in order to provide more information about the molecular biology of BCG Tice and design more reasonable vaccines to prevent tuberculosis. [Methods] We assembled the data from high-throughput sequencing with SOAPdenovo software, with many contigs and scaffolds obtained. There are many sequence gaps and physical gaps remained as a result of regional low coverage and low quality. We designed primers at the end of contigs and performed PCR amplification in order to link these contigs and scaffolds. With various enzymes to perform PCR amplification, adjustment of PCR reaction conditions, and combined with clone construction to sequence, all the gaps were finished. [Results] We obtained the complete genome sequence of BCG Tice and submitted it to GenBank of National Center for Biotechnology Information (NCBI). The genome of BCG Tice is 4334064 base pairs in length, with GC content 65.65%. [Conclusion] The problems and strategies during the finishing step of BCG Tice sequencing are illuminated here, with the hope of affording some experience to those who are involved in the finishing step of genome sequencing. The microarray data were verified by our results.

**Keywords:** BCG, high GC, gap finish, sequence analysis

(本文责编:王晋芳)

Supported by the Key Project of China National Programs for Fundamental Research and Development (2009CB522605, 2007CB513002)

\* Corresponding author. Tel/Fax: +86-10-64807362; E-mail: zhubaoli@im.ac.cn

Received: 20 April 2012 / Revised: 9 May 2012

## 科学出版社新书推介(2012年8月)

### 免疫的细胞社会生态学原理

吴克复; ISBN: 9787030351999; 定价: 98; 开本: 16; 装帧: 平装; 页码: 456; 专业分类: 生物学、农学、医学

**内容简介:** 从细胞生态学和进化论的观点介绍与免疫相关的一些重要问题的最新进展, 探讨它们在 21 世纪的发展方向和线索。《免疫的细胞社会生态学原理》还阐述了进化论的主要现代观点和方法; 从超有机体的视角分析细胞社会; 讨论免疫系统在细胞社会中的作用; 用景观生态学观点探讨局部免疫; 不仅有理论探讨和免疫机制的分子生物学、细胞生物学和系统生物学研究, 也有紧密结合临床实际的移植免疫和白血病的免疫治疗探讨。



订购方式 1: 网上购书

淘宝商城科学出版社旗舰店: <http://kxcbs.tmall.com/>

卓越亚马逊: <http://www.amazon.cn/>

当当网: <http://www.dangdang.com/>

京东图书: <http://book.360buy.com/>

订购方式 2: 电话购书

联系人: 科学出版社 贾海涛 13501022258 010-64017321

订购方式 3: 邮件购书

生物分社: [lifescience@mail.sciencep.com](mailto:lifescience@mail.sciencep.com); [jiahaitao@mail.sciencep.com](mailto:jiahaitao@mail.sciencep.com)