

# 志贺氏菌属各亚群菌株基因组“共有骨架”组成的分析

刘红<sup>\*†</sup> 彭俊平<sup>†</sup> 杨剑 孙立连 陈淑霞 金奇<sup>‡</sup>

(病毒基因工程国家重点实验室, 北京 100176. \*现工作于山东大学微生物技术国家重点实验室. †同等贡献, ‡联系人,  
E-mail: zdsys@sina.com)

**摘要** 不同原核生物在基因组组成上的差异是其生物学性状差异的基础. 然而, 进化中亲缘关系较近的菌群, 其基因组除了含有群或株特异的基因外, 还含有反映它们起源和进化痕迹的“共有骨架”结构, 而这些“共有骨架”恰恰是它们基本生存能力和共有生物学性状的基础. 志贺氏菌在起源和进化上与大肠杆菌极为密切, 目前倾向于将二者划为同一个种. 利用大肠杆菌 K-12 全基因组及 Sf301 特异性 ORFs 的芯片研究了志贺氏菌 4 个群间的基因组组成. 结果显示, 分别有 16% ~ 22% K-12 的 ORFs 序列没有在志贺氏菌的基因组中被检测出, 而志贺氏菌的基因组中包含至少 2800 个保守的 ORFs, 组成了其共有的“共有骨架”. 进一步分析提示, 这些“共有骨架”是维持肠道细菌正常生命生理活动所必需的基本组成. 此外, 只有 20% 的 Sf301 特异性 ORFs 同时存在于其他 3 个菌株中, 揭示了各菌株间基因组的异质性和菌属内的遗传多样性.

**关键词** 志贺氏菌属 大肠杆菌 K-12 全基因组微阵列 骨架序列

不同细菌基因组间的差异反映出菌株在不同自然选择压力下产生适应性变异的过程并赋予菌株特有的生物学性状. 因此, 同一种属内不同菌群或菌株间基因组差异的比较将有助于揭示细菌基因组在进化过程中获得与缺失基因的过程, 并揭示细菌生命活动多样性的遗传基础<sup>[1,2]</sup>. 基于 DNA-DNA 杂交的微阵列(Microarray)技术<sup>[3]</sup>为从比较基因组学角度快速研究微生物基因组组成提供了有力工具.

志贺氏菌属(*Shigella* spp)细菌通称为痢疾杆菌, 包括志贺氏(*S. dysenteriae*)、福氏(*S. flexneri*)、鲍氏(*S. boydii*)及宋内氏(*S. sonnei*) 4 个亚群, 即志贺氏菌 A, B, C 和 D 亚群, 临床感染均可以导致痢疾. 痢疾是世界上, 尤其是发展中国家的重要传染病之一<sup>[4]</sup>. 痢疾杆菌与大肠杆菌在许多方面非常相似, 基于持家基因的进化分析发现, 除个别血清型外痢疾杆菌所有血清型可在大肠杆菌内聚类为 3 个离散的簇. 目前普遍认为痢疾杆菌是从多个独立的大肠杆菌祖先起源后经历了 7 ~ 8 次趋同进化而形成的<sup>[5]</sup>.

本文以完成全基因组序列分析的志贺氏菌福氏 2a301 株(Sf301)<sup>[6~8]</sup>作为内参, 分别选取了其他 3 个亚群的代表株, 利用含有 Sf301 特异性和非致病性大肠杆菌 K-12 MG1655 株所有 ORFs (open reading frames, 可读框)的全基因组芯片检测其基因组组成, 重点分析它们之间共有的大肠杆菌基因, 即“骨架序列”(backbone sequences), 以期进一步确证二者的进

化关系, 初步揭示肠道菌基本生命与生理活动所必需的遗传基础, 同时为从根本上认识并控制痢疾杆菌的致病性提供理论依据.

## 1 材料与方法

(i) 菌株与培养. 志贺氏菌 Sf301、痢疾 1 型 CN51197 株(Sd51197)及鲍氏 5 型 CN51227 株(Sb51227)由中国疾病预防控制中心传染病预防控制研究所提供; 宋内 JD46 临床分离株(Ss46)由中国药物与生物制品鉴定所提供. 从常规刚果红培养基上挑取红色菌落, 并在不含抗生素的 LB 培养基中于 37℃振荡培养过夜, 用于分离基因组 DNA.

(ii) 微阵列. 大肠杆菌 K-12 MG1655 株的全基因组 ORFs 的引物由英国细菌科学研究会(BBSRC)的食品研究所提供, 包含 Sf301 特异性 ORFs 的微阵列(Eco-Sf301)由本实验室制备完成. 芯片经 80℃干烤固定, 1,2-二氯乙烷封闭后, 煮沸 2 min 使 DNA 变性. 用 Cy5-和 Cy3-dCTP 分别标记的 Sf301 与 MG1655 基因组与之共杂交检测点样效果.

(iii) 基因组 DNA 分离. 基因组 DNA 的提取使用 Wizard® Genomic DNA Purification Kit (Promega), 经过酚-氯仿抽提进一步纯化.

(iv) 探针制备与杂交. 基因组 DNA 的 Cy5-或 Cy3-dCTP (Amersham)标记采用 Bioprime DNA labeling System(Gibco), 经 QIAGEN 公司的 Qia-quick PCR Purification Kit 纯化后, 在 65℃杂交过夜.

(v) 扫描与数据分析. 玻片先用  $2 \times \text{SSC}$ , 0.1% SDS 溶液,  $65^\circ\text{C}$  洗  $2 \times 5 \text{ min}$ , 再分别用  $1 \times$ ,  $0.2 \times \text{SSC}$  室温洗  $2 \times 5 \text{ min}$ . 扫描图像中荧光强度用 QuantArray<sup>®</sup> Quantitation 软件定量, 辅助以目测确证. 所有的杂交都重复 3 次, 信号强度有 2 次低于标准值的按缺失处理.

## 2 结果

### 2.1 信号标准的确定

为确定序列比对长度和同源性对于杂交信号的影响, 我们用 Sf301 和 MG1655 的基因组 DNA 与按照基因组注释设计制备的 Eco-Sf301 芯片进行了共杂交<sup>[6,9]</sup>. 根据获得的数据, 结合比较基因组学分析的结果发现, 只有当两条核苷酸序列的同源性大于 80%, 且参与杂交的序列长于 80 bp 时, 才能呈现阳性信号. 据此确定了阳性信号的标准是: 信噪比  $> 3$ ; 信号强度  $> 1200$ ; 信号点直径  $> 80\%$ .

通过与 Eco-Sf301 芯片进行 DNA-DNA 杂交, 比较了痢疾杆菌中 Sd51197, Sb51227 及 Ss46 菌株的基因组组成, 从杂交后的扫描图上即可直观地发现各菌株间存在明显差异(图 1).

### 2.2 菌株间基因组组成差异的总体分析

DNA-DNA 杂交分析结果显示, 除了 Sf301 特异性 ORFs, K-12 基因组的 ORFs, 在所检测的菌株间的

分布也存在明显差异(图 2). Ss46 与 MG1655 的染色体基因组成最为接近, 在其含有的 3816 个 ORFs 中有 3569 个是 K-12 基因组的, 下面依次是 Sf301, Sb51227 与 Sd51197, 它们所含有的 K-12 的 ORFs 分别为 3508, 3290 和 3285 个. 4 个基因组共有的 ORFs 为 2974 个, 组成了志贺氏菌属的“骨架序列”, 其中 2864 个为 K-12 保守的.

与 K-12 相比, 在 Ss46 株特异性的 102 个 ORFs 中, 绝大多数(84/102)编码功能未知产物(Function unknown, FUN), 功能鉴定的 17 个 ORFs 分别编码了参与多种物质代谢的 mhp 操纵子以及乙酰辅酶 A 亚基的基因簇; 在 77 个 Sf301 特异性 ORFs 中, 12 个有预测功能的 ORFs 组成了两个操纵子, 分别编码参与乙醛酸和二羧酸盐代谢的 glcCDFGB 操纵子及负责半乳糖胺转运的磷酸转移酶系统; 而 Sb51227 与 Sd51197 分别含有 23 和 27 个特异性 ORFs, 均编码 FUN 蛋白.

只有大约 20% 的 Sf301 特异性 ORFs 序列在其他 3 个基因组中被检测到. 值得注意的是, 这部分“骨架序列”中除了已知与沙门氏菌毒力相关的 *sitABCD*<sup>[10]</sup> (位于 Sf301 的 *ipaIV* 岛)外, 还包括一个完整的 *ipaV* 毒力岛, 而其他在 Sf301 中鉴别的可能的毒力岛均不存在于其他菌株中. 另外, 与其脂多糖 O-抗原合成相关的基因也证实是菌株/血清型特异性的.

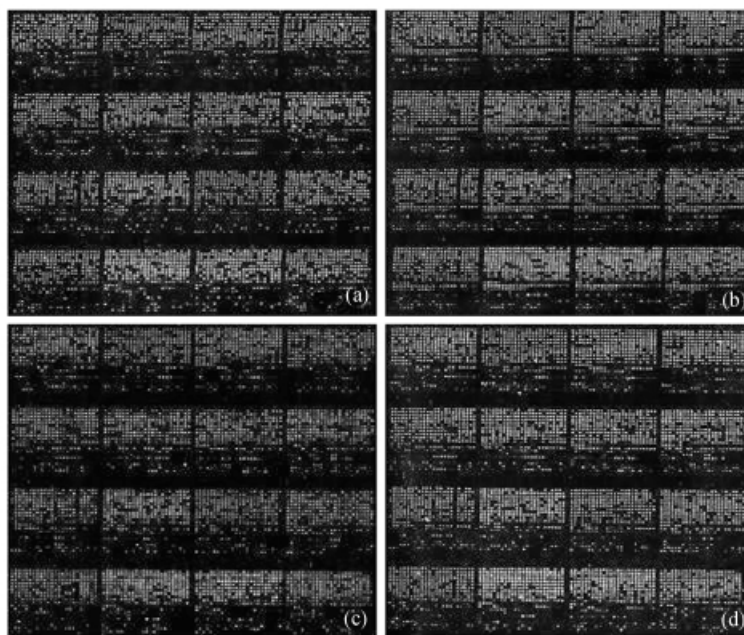


图 1 Cy5-dCTP 标记的各菌株基因组 DNA 与微阵列杂交后的信号强度  
(a) Sd51197, (b) Sf301, (c) Sb51227, (d) Ss46

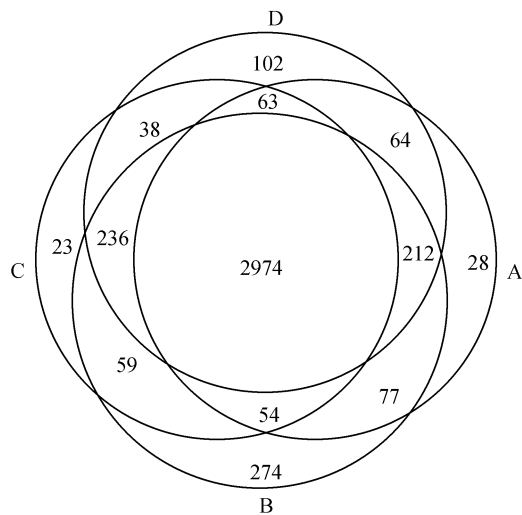


图 2 志贺氏菌各亚群菌株间基因组组成比较示意图  
A 示 Sd51197, B 示 Sf301, C 示 Sb51227, D 示 Ss46

2.3 共有“骨架序列”的基因组组成

如上所述, 本研究中所检测的志贺氏菌属源于大肠杆菌的“骨架序列”, 由 2864 个 ORFs 组成. 在 1917 个的编码产物功能明确的 ORFs 中, 有 992 个参与各种代谢途径, 另外 947 个编码 FUN 蛋白, 我们着重分析了代谢途径相关基因(<http://www.kegg.org>)及其他部分功能明确的基因. 共有“骨架序列”中所编码的基因组成了肠道杆菌的功能核心(function core), 其产物涵盖了菌群进行基本的生命生理活动所必需的物质. 表 1 列出了“骨架序列”中代谢途径的基因组组成及其与 K-12 的比较.

除上述代谢途径外, “骨架序列”中还包含其他一些基因簇, 其编码产物在细菌的生长过程中发挥重要作用, 如 DNA 聚合酶Ⅲ、RNA 聚合酶、抗性蛋白、转录调控蛋白、转录与翻译相关因子等. “骨架序列”中所有功能明确的基因均列于表 2, 对其中某些产物相关的基因进行了聚类.

3 讨论

DNA 微阵列技术在生命科学领域被广泛应用. 而基于微阵列的比较基因组研究, 可以揭示不同物种或同一物种不同个体间基因组组成的差异, 虽然在某些情况下, 微阵列也存在诸如通常不包括基因组中的基因间区域, 在仅利用 ORFs 作为探针进行基因组比较时, 不能检测出基因的移码突变、小片段的缺失、基因组重排以及多拷贝基因数目的不同等<sup>[11]</sup>不足之处, 但是利用该技术进行的比较基因组研究仍然获

表 1 “骨架序列”中代谢途径基因组组成及与 K-12 的比较

| 代谢途径及功能类群   | 志贺氏菌属 | <i>E. coli</i> K-12 |
|-------------|-------|---------------------|
| 能量代谢        | 104   | 141                 |
| 复合脂类代谢      | 48    | 68                  |
| 脂代谢         | 66    | 95                  |
| 核苷酸代谢       | 104   | 125                 |
| 碳水化合物代谢     | 186   | 268                 |
| 复合碳水化合物代谢   | 93    | 124                 |
| 辅因子与维生素代谢   | 147   | 164                 |
| 氨基酸代谢       | 222   | 278                 |
| 其他氨基酸代谢     | 64    | 79                  |
| 次级代谢产物的生物合成 | 16    | 24                  |
| 异生素的生物降解    | 48    | 72                  |
| Ⅱ型分泌系统      | 8     | 26                  |
| 核糖体         | 51    | 55                  |
| 蛋白输出        | 17    | 17                  |
| 磷酸转移酶系统     | 36    | 52                  |
| 细菌趋化性       | 6     | 20                  |
| 鞭毛组装        | 18    | 36                  |
| 信号转导        | 74    | 87                  |
| ABC 转运系统    | 154   | 192                 |

得了许多极有价值的信息<sup>[12~14]</sup>. 我们在完成 Sf301 全序列测定的基础上, 制备了第一个痢疾杆菌的全基因组微阵列, 并开展了一系列菌属内遗传多样性比较及表达谱等功能基因组学的研究.

当细菌生长在一个相对稳定的环境中时, 还原性进化将导致在该环境中生存非必需基因发生丢失, 而保留那些生命活动必需的关键代谢途径以及与环境之间进行物质和能量转换的基因. 本研究显示, 志贺氏菌属的遗传骨架即“核心基因组”包括大约 3000 个 ORFs, 其中保守的大肠杆菌 ORFs 为 2879 个, 包括了除高温生存和铜转运必需的 *htrC* 和 *lnt* 外所有已经证实的大肠杆菌在不同环境下生存所必需的基因(<http://www.genome.wisc.edu>).

由于细菌中存在多种功能互补的代谢网络, 所以参与某种代谢的基因数目在不同细菌间可能有所不同, 如表 1 所示, 尽管“骨架序列”中所有代谢途径相关基因数目均少于 K-12 中的, 但参与基本生命代谢途径的基因组组成却都是完整的. 另外, “骨架序列”中还保留了基本的胁迫应答相关基因.

研究结果发现, 本应作为基本代谢元件而保留在“骨架序列”中的核糖体蛋白基因在不同的菌株中均有 2~6 个不等的缺失, 而 *rpsK* 与 *rplX* 基因在所有 4 个菌株中都没有被检测出, 由此提示在菌株特异性功能未知的基因中必定存在作为代替或补偿的相关基因, 这还有待于进一步的功能研究予以证实.

表2 志贺氏菌“骨架序列”中功能明确基因

| 基因                                                                                | 产物                    | 基因                                                        | 产物              |
|-----------------------------------------------------------------------------------|-----------------------|-----------------------------------------------------------|-----------------|
| <i>hrpB</i> , <i>hrpA</i> , <i>hepA</i> , <i>yoeA</i>                             | ATP 依赖的解旋酶            | <i>yeiO</i>                                               | 糖外排转运蛋白         |
| <i>ruvAB</i>                                                                      | Holliday 结合部解旋酶亚基     | <i>fepE</i>                                               | 铁肠菌素转运蛋白        |
| <i>dnaB</i>                                                                       | 复制型 DNA 解旋酶           | <i>feoAB</i>                                              | 亚铁转运蛋白          |
| <i>dnaC</i>                                                                       | 染色体复制蛋白               | <i>narK</i> , <i>nirC</i>                                 | 亚硝酸盐转运蛋白        |
| <i>ruvC</i>                                                                       | Holliday 结合部核酸酶       | <i>molR</i> , <i>modE</i>                                 | 钼酸盐代谢/转运调节蛋白    |
| <i>hns</i> , <i>fis</i> , <i>hupA</i> , <i>hupB</i>                               | DNA-结合蛋白              | <i>dcuBA</i>                                              | 厌氧的 C4-二羧酸盐转运蛋白 |
| <i>priABC</i> , <i>dnaAGT</i>                                                     | 引发体蛋白                 | <i>fucP</i>                                               | 岩藻糖转运蛋白         |
| <i>umuC</i> , <i>umuD</i>                                                         | DNA 聚合酶 V             | <i>gltS</i>                                               | 钠/谷氨酸盐共转运载体蛋白   |
| <i>topAB</i> , <i>parCE</i>                                                       | DNA 拓扑异构酶             | <i>uidB</i>                                               | 葡萄糖酸载体蛋白        |
| <i>gyrAB</i>                                                                      | DNA 旋转酶               | <i>galSR</i>                                              | 半乳糖利用的转录抑制剂     |
| <i>recAR</i>                                                                      | DNA 链交换和重组蛋白          | <i>ygaG</i>                                               | 自主诱导产生蛋白        |
| <i>recON</i> , <i>radA</i>                                                        | DNA 修复蛋白              | <i>nac</i>                                                | 氮同化转录阻遏物        |
| <i>mutLHS</i>                                                                     | 甲基指导的错配修复蛋白           | <i>tyrR</i>                                               | 芳族氨基酸合成转录调节剂    |
| <i>uvrABC</i>                                                                     | 切除修复酶                 | <i>gcvR</i>                                               | 甘氨酸分裂的转录阻遏物     |
| <i>recF</i>                                                                       | 缺口修复蛋白                | <i>treR</i>                                               | 海藻糖吸收的转录阻遏物     |
| <i>alkA</i> , <i>ada</i>                                                          | 烷基化 DNA 修复系统蛋白        | <i>deoR</i>                                               | 核苷酸分解的转录阻遏物     |
| <i>rpoSDEHN</i>                                                                   | RNA 聚合酶因子             | <i>rbsR</i>                                               | 核糖代谢的转录阻遏物      |
| <i>slyA</i> , <i>hypF</i> , <i>rtcR</i> , <i>nadR</i> , <i>rfaH</i> , <i>cytR</i> | 转录调控蛋白                | <i>metR</i>                                               | 甲硫氨酸生物合成的转录调节物  |
| <i>greAB</i>                                                                      | 转录延长因子                | <i>argR</i>                                               | 精氨酸合成转录阻遏物      |
| <i>nusAB</i>                                                                      | N 利用底物蛋白              | <i>rob</i>                                                | 抗生素、有机溶剂及重金属抗性  |
| <i>rho</i>                                                                        | 转录终止因子                |                                                           | 转录激活剂           |
| <i>nusG</i>                                                                       | 抗转录终止成分               | <i>lysR</i>                                               | 赖氨酸生物合成转录激活剂    |
| <i>rnd</i> , <i>rnhAB</i> , <i>rnb</i> , <i>rnc</i> , <i>rph</i> , <i>rnpA</i> ,  | 核糖核酸酶                 | <i>melR</i>                                               | 蜜二糖分解代谢转录激活剂    |
| <i>vacB</i> , <i>rnt</i> , <i>rnb</i>                                             |                       | <i>uidR</i>                                               | 葡萄糖酸酯代谢转录阻遏物    |
| <i>nth</i> , <i>nfo</i> , <i>nfi</i> , <i>nei</i> , <i>endA</i>                   | 核酸内切酶 III, IV, V, VII | <i>gntR</i>                                               | 葡萄糖酸盐利用的转录阻遏物   |
| <i>sbcBCD</i> , <i>recBCD</i> , <i>xseAB</i> , <i>xthA</i> , <i>xni</i>           | 核酸外切酶                 | <i>trpR</i>                                               | 色氨酸生物合成的转录阻遏物   |
| <i>rimIJK</i> , <i>prmA</i> , <i>yjiT</i> ,                                       | 核糖体蛋白修饰蛋白             | <i>arsR</i>                                               | 砷抗性转录阻遏物        |
| <i>frr</i> , <i>rmf</i> , <i>rbfA</i>                                             | 核糖体调节因子               | <i>ascG</i>                                               | 碳代谢转录阻遏物        |
| <i>ppiABCD</i>                                                                    | 肽酰-脯氨酸-顺反式异构酶         | <i>sdiA</i>                                               | 细胞分裂转录激活剂       |
| <i>pth</i>                                                                        | 肽基-tRNA 水解酶           | <i>iclR</i>                                               | 乙醛酸旁路阻遏物        |
| <i>infBC</i>                                                                      | 肽链起始因子                | <i>fur</i>                                                | 铁摄取调控蛋白         |
| <i>fusA</i> , <i>tufAB</i> , <i>tsf</i> , <i>efp</i>                              | 肽链延伸因子                | <i>crl</i>                                                | 纤毛基因转录激活蛋白      |
| <i>prfABCH</i>                                                                    | 肽链释放因子                | <i>ilvY</i>                                               | 异亮氨酸/缬氨酸合成转录    |
| <i>pepET</i> , <i>yfhI</i>                                                        | 肽酶                    |                                                           | 激活剂             |
| <i>def</i>                                                                        | 肽脱酰基酶                 | <i>asnC</i>                                               | 天冬酰胺生物合成调控蛋白    |
| <i>pmbA</i>                                                                       | 肽成熟蛋白                 | <i>chaAB</i>                                              | 阳离子转运调节剂        |
| <i>dsBEDG</i>                                                                     | 硫醇：二硫化物交换蛋白           | <i>fadR</i>                                               | 脂脂肪酸代谢调节物       |
| <i>dsbA</i>                                                                       | 周质蛋白二硫化物异构酶           | <i>nlp</i> , <i>malIT</i> , <i>sfsA</i>                   | 麦芽糖发酵调控蛋白       |
| <i>deadD</i> , <i>dbpA</i> , <i>srmB</i>                                          | ATP 依赖 RNA 解旋酶        | <i>gutM</i> , <i>srlR</i>                                 | 山梨糖醇利用调控蛋白      |
| <i>clpX</i> , <i>clpP</i> , <i>clpB</i> , <i>lon</i>                              | ATP-依赖的蛋白水解酶          | <i>thiCFGHJ</i>                                           | 硫胺生物合成酶亚基       |
| <i>htrA</i> , <i>yaeL</i> , <i>ptrB</i> , <i>ptr</i> , <i>sohB</i>                | 蛋白酶                   | <i>nipBCD</i> , <i>spr</i>                                | 脂蛋白/前体          |
| <i>rfe</i> , <i>wzzE</i> , <i>rffAC</i> , <i>wzxE</i> , <i>wecF</i>               | 脂多糖生物合成蛋白             | <i>yeaA</i> , <i>msrA</i>                                 | 肽蛋氨酸亚砷还原酶       |
| <i>ompAWX</i> , <i>tolABCR</i> , <i>phoE</i> , <i>pal</i> ,                       | 外膜蛋白/整合/脂蛋白           | <i>ordL</i> , <i>ucpA</i>                                 | 氧化还原酶           |
| <i>lolAB</i> , <i>slyB</i> , <i>blc</i>                                           |                       | <i>mdoGH</i>                                              | 周质葡聚糖生物合成蛋白     |
| <i>aqpZ</i>                                                                       | 水孔蛋白 Z                | <i>ymdD</i>                                               | 葡萄糖生物合成蛋白       |
| <i>tolQ</i>                                                                       | 跨膜的内膜蛋白               | <i>himAD</i>                                              | 整合宿主因子亚基        |
| <i>acrABDF</i>                                                                    | 吡啶黄抗性蛋白               | <i>xerDC</i>                                              | 位点特异性酪氨酸重组酶     |
| <i>marABCR</i>                                                                    | 多抗生素抗性蛋白              | <i>cutCF</i>                                              | 铜动态平衡蛋白         |
| <i>emrABDKRY</i> , <i>cmr</i> , <i>mdl</i> , <i>mdlB</i> ,                        | 多药抗性蛋白                | <i>htgA</i> , <i>dnaK</i> , <i>htpG</i> , <i>hslJUV</i> , | 热激蛋白            |
| <i>msbA</i> , <i>ydhE</i> , <i>cydCD</i> , <i>acrR</i> , <i>pmrD</i>              |                       | <i>hscA</i> , <i>grpE</i> , <i>ibpA</i>                   |                 |
| <i>rarD</i>                                                                       | 氯霉素抗性蛋白               | <i>uidC</i>                                               | 膜关联蛋白           |
| <i>ampH</i> , <i>mrdA</i> , <i>dacABC</i> , <i>pbpG</i>                           | 青霉素结合蛋白               | <i>sppA</i>                                               | 内肽酶 IV          |
| <i>mepA</i>                                                                       | 青霉素不敏感的胞壁质 DD-内肽酶     | <i>pyrH</i>                                               | 尿苷酸激酶           |
| <i>bcr</i>                                                                        | 双环霉素抗性蛋白              | <i>ung</i>                                                | 尿嘧啶-DNA-糖苷化酶    |
| <i>tehAB</i>                                                                      | 亚碲酸盐抗性蛋白              | <i>purR</i>                                               | 嘌呤核苷酸合成阻遏物      |
| <i>sspAB</i>                                                                      | 严紧型饥饿蛋白               | <i>zntA</i>                                               | 铅、镉、锌、汞转运 ATP 酶 |
| <i>asr</i>                                                                        | 酸休克蛋白                 | <i>hlyE</i>                                               | 溶血素 E           |

续表 1

| 基因                         | 产物                                                        | 基因          | 产物                      |
|----------------------------|-----------------------------------------------------------|-------------|-------------------------|
| <i>osmBCEY</i>             | 渗透诱导蛋白                                                    | <i>hha</i>  | 溶血素表达调节蛋白               |
| <i>damX</i>                | 损伤诱导蛋白                                                    | <i>thdF</i> | 噻吩和呋喃氧化蛋白               |
| <i>pspABCDE</i>            | 噬菌体休克蛋白                                                   | <i>arsB</i> | 砷泵                      |
| <i>cspCDEH</i>             | 冷休克蛋白                                                     | <i>arsC</i> | 砷酸盐还原酶                  |
| <i>hycCDE, hypA</i>        | 甲酸氢化酶亚基                                                   | <i>mtgA</i> | 生物合成的肽聚糖转糖酶             |
| <i>fhIA</i>                | 甲酸氢化酶转录激活剂                                                | <i>birA</i> | 生物素合成阻遏物                |
| <i>hycA</i>                | 甲酸氢化酶转录阻遏物                                                | <i>bioC</i> | 生物素生物合成酶                |
| <i>moaABCDE</i>            | 钼喋呤辅因子生物合成蛋白                                              | <i>bisC</i> | 生物素亚砷还原酶                |
| <i>ahpF</i>                | 烷基过氧化物还原酶亚基                                               | <i>yaeS</i> | 十一萆酰焦磷酸盐合成酶             |
| <i>moeAB, mog</i>          | 钼喋呤生物合成                                                   | <i>ppha</i> | 丝氨酸/苏氨酸蛋白磷酸酶            |
| <i>hyaDEF</i>              | 氢化酶 1 操纵子蛋白                                               | <i>clpA</i> | 丝氨酸蛋白酶 ATP 结合亚基         |
| <i>hybABDEFG</i>           | 氢化酶 2 操纵子蛋白                                               | <i>degQ</i> | 丝氨酸内蛋白酶                 |
| <i>hyfBCDEFGHI, yffE</i>   | 氢化酶 4 亚基                                                  | <i>csrA</i> | 碳储存调节蛋白                 |
| <i>hypBCDE</i>             | 氢化酶同工酶形成蛋白                                                | <i>cstA</i> | 碳饥饿蛋白                   |
| <i>hyfR</i>                | 氢化酶转录激活剂                                                  | <i>slp</i>  | 碳饥饿诱导的外膜蛋白              |
| <i>napHGDF, ydiT, ygcO</i> | 铁氧还蛋白                                                     | <i>yabM</i> | 糖外排蛋白                   |
| <i>trxAC</i>               | 硫氧还蛋白                                                     | <i>sgaE</i> | 糖异构酶                    |
| <i>fdx</i>                 | [2Fe-2S]铁氧还蛋白                                             | <i>sodB</i> | 铁超氧化物歧化酶                |
| <i>ybhRSF, uup, yjjK</i>   | ABC 超家族转运蛋白                                               | <i>ftn</i>  | 铁蛋白                     |
| <i>yfhS</i>                | 3-苯丙酸转运蛋白                                                 | <i>fhuF</i> | 铁氧脂酸盐转运                 |
| <i>ynhD</i>                | ATP-依赖的转运蛋白                                               | <i>phrB</i> | 脱氧核糖嘧啶光裂合酶              |
| <i>dctA</i>                | C4-二羧酸转运蛋白                                                | <i>cbpA</i> | 弯曲 DNA 结合蛋白             |
| <i>chaA</i>                | Ca <sup>2+</sup> /H <sup>+</sup> 逆转运蛋白                    | <i>ahpC</i> | 烷基过氧化物还原酶 C22 亚基        |
| <i>mobAB</i>               | 钼喋呤-鸟喋呤二核苷酸生物合成蛋白 AB                                      | <i>cycA</i> | D-丙氨酸/D 丝氨酸/甘氨酸转运蛋白     |
| <i>kefB</i>                | K <sup>+</sup> : H <sup>+</sup> 逆转运蛋白                     | <i>btuB</i> | 维生素 B12 受体前体            |
| <i>rhaT</i>                | L-鼠李糖: H <sup>+</sup> 共转运蛋白                               | <i>csiE</i> | 稳定期诱导蛋白                 |
| <i>corA</i>                | Mg <sup>2+</sup> /Ni <sup>2+</sup> /Co <sup>2+</sup> 转运蛋白 | <i>selB</i> | 硒代半胱氨酸酰基转移酶因子           |
| <i>nhaAB</i>               | Na <sup>+</sup> /H <sup>+</sup> 逆转运蛋白                     | <i>div</i>  | 细胞分裂蛋白                  |
| <i>gabP</i>                | γ 氨基丁酸转运蛋白                                                | <i>fic</i>  | 细胞丝状蛋白                  |
| <i>araE</i>                | 阿拉伯糖: 质子共转运蛋白                                             | <i>mesJ</i> | 细胞周期蛋白                  |
| <i>amtB</i>                | 铵转运蛋白                                                     | <i>acpP</i> | 酰基载体蛋白                  |
| <i>galP</i>                | 半乳糖: 质子共转运蛋白                                              | <i>mutY</i> | 腺嘌呤 DNA 糖基化酶            |
| <i>pheP</i>                | 苯氨基丙酸转运蛋白                                                 | <i>dps</i>  | 胁迫应答 DNA 结合蛋白           |
| <i>fadL</i>                | 长链脂肪酸转运蛋白                                                 | <i>ispB</i> | 辛异戊二烯基二磷酸盐合酶            |
| <i>pitA</i>                | 低亲和性磷酸盐转运蛋白                                               | <i>bolA</i> | 形态发生途径催化剂               |
| <i>aroP</i>                | 芳族氨基酸转运蛋白                                                 | <i>bglJ</i> | 熊果甙/水杨甙利用激活剂            |
| <i>proP</i>                | 脯氨酸/甜菜碱转运蛋白                                               | <i>hemY</i> | 血红素合成蛋白                 |
| <i>glpT</i>                | 甘油-3-磷酸转运蛋白                                               | <i>ccmD</i> | 血红素输出蛋白 D               |
| <i>kefC</i>                | 谷胱甘肽调控的 K <sup>+</sup> 外排逆转运蛋白                            | <i>cysQ</i> | 亚硫酸盐合成的 3'-磷酸腺苷-5'-磷酸硫酸 |
| <i>exuT</i>                | 几糖醛酸盐转运蛋白                                                 | <i>mdaB</i> | 药物活性调节剂 B               |
| <i>pnuC</i>                | 核苷/喋呤/嘧啶转运蛋白                                              | <i>fucU</i> | 岩藻糖操纵子蛋白                |
| <i>nupCG</i>               | 核苷转运蛋白                                                    | <i>nrdG</i> | 厌氧核糖核苷还原酶激活蛋白           |
| <i>trkH</i>                | 钾转运蛋白                                                     | <i>eutC</i> | 乙醇胺裂解酶轻链                |
| <i>tyrP</i>                | 酪氨酸特异性转运蛋白                                                | <i>accB</i> | 乙酰 CoA 羧化酶生物素羧基载体       |
| <i>uraA</i>                | 尿嘧啶转运蛋白                                                   | <i>tldD</i> | 抑制 CsrA 的抑制活性           |
| <i>mtr</i>                 | 色氨酸特异性转运蛋白                                                | <i>IMP</i>  | 有机溶剂抗性                  |
| <i>sdaC</i>                | 丝氨酸转运蛋白                                                   | <i>spy</i>  | 原生质球形成蛋白                |
| <i>tdcC</i>                | 苏氨酸转运蛋白                                                   | <i>intB</i> | 原噬菌体 P4 整合酶             |
| <i>csgD</i>                | csgAB 操纵子转录调控蛋白                                           | <i>sseB</i> | 增强丝氨酸敏感性                |
| <i>cysB</i>                | Cys 调控子的的转录激活剂                                            | <i>brnQ</i> | 支链氨基酸转运系统 II 载体蛋白       |
| <i>exuR</i>                | exu 调控子的转录调控剂                                             | <i>plsX</i> | 脂肪酸/磷脂合成蛋白              |
| <i>yfhT</i>                | hca 簇转录激活剂                                                | <i>lipA</i> | 脂酸合成酶                   |
| <i>fucR</i>                | L-岩藻糖利用的转录激活剂                                             | <i>mukF</i> | 致死因子                    |
| <i>lrhA</i>                | NADH 脱氢酶转录阻遏物                                             | <i>torT</i> | 周质蛋白前体                  |
| <i>pspF</i>                | psp 操纵子的转录激活剂                                             | <i>degS</i> | 周质的丝氨酸内蛋白酶              |
| <i>cutA</i>                | 周质二价阳离子耐受蛋白                                               | <i>ebgR</i> | β-半乳糖苷利用的转录阻遏物          |

无鞭毛、无动力是区分痢疾杆菌和其他肠道菌的主要指标之一,但本实验所用的4株痢疾杆菌的骨架序列中仍然保留了大部分鞭毛组装相关基因。另外,我们在细菌表达谱的研究中也检测到该部分基因在感染过程中的表达变化。已知部分鞭毛蛋白同时作为Ⅲ型分泌系统的成分参与蛋白质的分泌与降解,所以除了鞭毛形成以外,这些结构蛋白基因是否以及如何在其他生命过程发挥作用尚待进一步研究。

除了表面抗原构造与生化反应特性的不同,痢疾杆菌亚群之间在致病力及流行病学方面均存在差异:志贺氏痢疾杆菌通常散在流行,但病死率高;福氏和宋内氏分别是发展中和发达国家痢疾流行的优势群,发病率高;鲍氏的流行则很少被报道。本工作所揭示的各群代表株基因组组成的差异将可以部分地阐明这些现象。例如,与其他3株不同, Sd51197基因组中完全缺失了可作为Ⅲ型分泌系统组分的那部分鞭毛合成基因;而在 Sb51227菌株中也没有检测出大质粒编码的侵袭蛋白基因 *ipaABCD* 及部分Ⅲ型分泌系统基因。所以基于福氏痢疾杆菌研究所建立的Ⅲ型分泌系统的组成能否代表所有的痢疾杆菌还有待证实。

基因的水平转移与基因还原代表了原核生物基因组进化中“飞跃性量变”的两种机制。在还原性进化保持了细菌在特定环境中生存必需的基因组成的基础上,外源 DNA 片段的获得在新物种、亚种及不同毒力型的发育中发挥了重要作用<sup>[14]</sup>。比较基因组学分析已经证实 Sf301 特异性的 ORFs 大多以毒力岛的形式存在,保证了该菌对于微环境的适应性,也是影响其致病力的原因之一<sup>[7]</sup>。其中 *sitABCD* 基因簇及毒力岛 *ipaV* 存在于所有被检测的菌株中,推测该序列可能也存在于痢疾杆菌其他血清型的菌株中,部分证实了痢疾杆菌所经历的趋同进化。而 Sf301 特异性 ORFs 中的 80% 在其他菌株中都没有检测出,则证实了各菌株间基因组的异质性和该属细菌的遗传多样性。上述结果将为研究志贺氏菌属各群及血清型间的表型差异提供遗传基础。

随着更多微生物基因组和功能基因组计划的完成及新技术的不断涌现和应用,将使我们最终阐明微生物的起源、进化以及繁殖和致病机制。

**致谢** 感谢于军教授和 Jay C. D. Hinton 博士为本工作提供建议,及 Sacha Lucchini, Matthew Rolfe 博士对于微阵列技术的支持。本工作为国家重点基础研究发展规划(批准号: G1999054105)、国家高技术研究发展计划(批准号: 2001AA223011)资助项目及国际科技合作重点项目(批准号: 2001AA223116)和北京市科技项目(合同号: H0102-10360119)。

## 参 考 文 献

- 1 Ochman H, Jones I B. Evolutionary dynamics of full genome content in *Escherichia coli*. EMBO, 2000, 19: 6637~6643
- 2 Ochman H, Moran N A. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science, 2001, 292: 1096~1098
- 3 Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, 270: 467~470
- 4 Hale T L. Genetic basis of virulence in *Shigella* species. Microbiol Rev, 1991, 55: 206~224
- 5 Pupo G M, Lan R, Reeves P R. Multiple independent of origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci USA, 2000, 97: 10567~10572
- 6 Jin Q, Yuan Z H, Xu J G, et al. Genome sequence of *Shigella flexneri* 2a: Insight into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucl Acid Res, 2002, 30(20): 4432~4441
- 7 刘红, 杨帆, 张笑冰, 等. 痢疾杆菌全基因组序列及基因组岛的分析. 中国工程科学, 2002, 4(10): 40~47
- 8 张笑冰, 刘红, 杨帆, 等. 福氏 2a 志贺氏菌 301 株基因组水平的缺失基因分析. 科学通报, 2003, 48(6): 587~592
- 9 Plattner F R, Plunkett G III, Bloch C A, et al. The complete genome sequence of *Escherichia coli* K-12. Science, 1997, 277: 1453~1474
- 10 Janakiraman A, Slauch J M. The putative iron transport system SitABCD encoded on SPI1 is required for full virulence of *Salmonella typhimurium*. Mol Microbiol, 2000, 35(5): 1146~1155
- 11 Schoolnik G K. Functional and comparative genomics of pathogenic bacteria. Curr Opin Microbiol, 2002, 5: 20~26
- 12 Behr M A, Wilson M A, Gill W P, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science, 1999, 284: 1520~1523
- 13 Alm R A, Ling L S, Moir D T, et al. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature, 1999, 397: 176~180
- 14 Dobrindt U, Agerer F, Michaelis K, et al. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. J Bacteriol, 2003, 185(6): 1831~1840

(2003-06-03 收稿, 2003-11-18 收修改稿)