

福氏 2a 志贺氏菌 301 株基因组水平的 缺失基因分析

张笑冰* 刘红 杨帆 杨剑 薛颖 董杰 孙立连 杨国威 朱俊萍
楚雍烈* 金奇†

(病毒基因工程国家重点实验室, 北京 100052. * 现工作于西安交通大学医学院微生物教研室, 西安, 710061.

† 联系人, E-mail: zdsys@sina.com.)

摘要 在完成福氏 2a 志贺氏菌 301 株全基因组序列测定工作的基础上, 以非致病性大肠杆菌 K12 MG1655 株的全基因组序列为参比对象, 进行比较基因组学分析. 结果显示, 与大肠杆菌基因组相比, 福氏 2a 志贺氏菌 301 株基因组中共有 136 个长度大于 1000 bp 的片段缺失, 总长为 717253 bp. 这些缺失片段中共包含 670 个可读框(ORF). 通过对 ORFs 编码产物的功能预测和分析, 发现这些缺失的 ORFs 主要编码与细菌代谢相关的酶类、结构蛋白、基因转录调控因子以及与细菌遗传物质水平转移相关的元件. 这不仅从全基因组水平揭示了两种具有不同生物学特性和表型的重要肠道细菌的分子差异, 而且为进一步探索其生理活动过程、致病机制以及肠道细菌间的进化等诸方面的问题提供了有价值的线索.

关键词 福氏 2a 志贺氏菌 大肠杆菌 K12 比较基因组学 缺失基因

细菌性痢疾具有发病时间短、感染率高、可引起爆发性流行等特点, 是严重威胁人类健康的主要传染病之一. 福氏 2a 志贺氏菌 301 株作为引起我国细菌性痢疾的主要病原体, 其染色体基因组全长为 4607203 bp, (G + C)含量为 0.5089 mol; 预测可读框总数为 4434 个, 编码区占到基因组总长的 80.4%. 与致病性密切相关的大质粒 pCP301 全长 221618 bp, 共编码 267 个 ORFs, 其中包括侵袭相关基因 *Ipa*, *virG* 和 *Mxi-Spa* 等. 志贺氏痢疾杆菌的主要致病特点是侵袭结肠黏膜上皮细胞, 虽然目前已知的侵袭性毒力基因都由大质粒编码, 但其表达则受染色体多个基因的调控^[1].

福氏 2a 志贺氏菌 301 株染色体与大肠杆菌 K12 MG1655 株染色体(4639221 bp)^[2]在长度上非常接近. 比较基因组学数据显示, 两者虽然有大约 3.9 Mb 的共有“骨架”序列, 但在福氏 2a 志贺氏菌 301 株的基因组中存在大量的 DNA 片段移位和倒置现象, 这使得两者在基因结构和组成上依然存在较大的差异. 与 K-12 MG1655 株相比, 福氏 2a 志贺氏菌具有 572 kb 的特异性序列, 而且福氏 2a 志贺氏菌 301 株基因组中有 717 kb 的片段是缺失的.

志贺氏菌被认为在起源和进化关系上与大肠杆菌极为密切, 甚至有报道推测志贺氏菌是大肠杆菌经过长期进化和演变的产物^[3,4]. 但是二者之间, 特

别是痢疾杆菌与非致病性大肠杆菌之间存在着明显的生物学性状和表型的差异, 这也反映了细菌在生存环境差异中所采取的适应性突变. 通过比较基因组学研究, 可以为揭示肠道细菌的致病机理、基因变异导致生物学性状和表型的改变以及进化关系等诸方面研究提供极有价值的线索. 本文在福氏 2a 志贺氏菌 301 株全基因组序列测定完成的基础上^[5], 将其与非致病性大肠杆菌 K-12 MG1655 株进行比较基因组学研究, 内容主要集中在分析福氏 2a 志贺氏菌 301 株的缺失基因, 目的在于探寻两者的基因组差异与生物学性状和表型差异之间的联系.

1 材料与方法

(i) 细菌菌株. 福氏 2a 志贺氏菌 301 株于 1984 年从北京市昌平区腹泻病患者的粪便标本中分离. 该菌株对氨苄青霉素、卡那霉素、四环素等抗生素敏感, 豚鼠角膜实验和细胞侵袭实验均为阳性, 并含有侵袭性基因 *ipaABCD* 和毒力调控基因 *virF* 等.

(ii) 细菌全基因组序列分析. 基因组 ORF 注释采用基因预测软件包 Glimmer 2.0 版本. 福氏 2a 志贺氏菌 301 株全基因组序列的 GenBank 登录号为 AE005674. 基因组分析中的参比菌株为非致病性大肠杆菌 K12 MG1655(GenBank 登录号: U00096).

(iii) 细菌代谢途径分析. 根据网站 EcoCyc (Encyclopedia of *Escherichia coli* genes and Metabo-

lism, 网址: <http://www.ecocyc.org/>)和 KEGG (Kyoto Encyclopedia of Genes and Genomes, 网址: <http://www.kegg.org/>)所提供的数据进行相关代谢途径的分析.

2 结果

全基因组序列比较结果显示, 大肠杆菌基因组中有 136 个 DNA 片段(长度均大于 1000 bp)在福氏 2a 志贺氏菌基因组中缺失, 缺失片段的总长为 717253 bp, 其中大于 5 kb 的缺失片段占 74%. 缺失片段中共

包含 670 个 ORFs, 其中有 40%左右的 ORFs 为功能未知的蛋白. 我们对功能明确的缺失基因进行分析, 发现这些基因主要包括代谢酶类基因、外膜结构蛋白基因、调控基因、转运系统和噬菌体基因等.

2.1 代谢相关基因

在福氏 2a 志贺氏菌基因组缺失的大肠杆菌已知功能基因中, 编码与细菌代谢相关的酶类基因占相当的比例, 这些酶多为细菌生长代谢过程中的关键酶, 包括水解酶、合成酶及异构酶(表 1).

表 1 福氏 2a 志贺氏菌基因组中缺失的大肠杆菌功能确定的代谢相关基因

基因	编号	编码产物功能	基因	编号	编码产物功能
<i>betA</i>	b0311	黄素蛋白	<i>eutB</i>	b2441	乙醇氨裂解酶重链
<i>betB</i>	b0312	NAD ⁺ -依赖的甜菜碱醛脱氢酶	<i>eutH</i>	b2452	乙醇氨合成反应
<i>BetI</i>	b0313	<i>bet</i> 基因转录抑制物	<i>eutG</i>	b2453	乙醇氨合成反应
<i>betT</i>	b0314	高亲和胆碱转运蛋白	<i>eutJ</i>	b2454	乙醇氨合成反应
<i>prpR</i>	b0330	prp 操纵子的调节蛋白	<i>eutE</i>	b2455	乙醇氨合成反应
<i>prpB</i>	b0331	磷酸酯变位酶	<i>eutI</i>	b2458	乙醇氨合成反应
<i>prpC</i>	b0333	柠檬酸合成酶	<i>sseA</i>	b2521	硫代硫酸盐巯基转移酶
<i>prpE</i>	b0335	丙酰-CoA 合成酶	<i>ascB</i>	b2716	6-磷酸-β-葡萄糖苷酶(隐匿的)
<i>codB</i>	b0336	胞嘧啶透性酶	<i>hycI</i>	b2717	蛋白酶
<i>codA</i>	b0337	胞嘧啶脱氨酶	<i>hycH</i>	b2718	蛋白酶
<i>cynR</i>	b0338	cyn 操纵子调节蛋白	<i>hycG</i>	b2719	氢化酶
<i>cynT</i>	b0339	碳脱水酶	<i>hycF</i>	b2720	铁硫蛋白氢化酶
<i>cynS</i>	b0340	氰酸盐氨基水解酶	<i>ygbL</i>	b2738	醛缩酶
<i>cynX</i>	b0341	氰酸盐转运蛋白	<i>yqeI</i>	b2847	传感蛋白
<i>lacA</i>	b0342	硫代半乳糖乙酰转移酶	<i>agaA</i>	b3135	N-酰基葡萄糖胺-6-磷酸脱乙酰基酶
<i>lacY</i>	b0343	半乳糖苷透性酶	<i>agaS</i>	b3136	塔格糖-6-磷酸醛缩异构酶
<i>lacZ</i>	b0344	β-D-半乳糖苷酶	<i>agaY</i>	b3137	塔格糖-二磷酸醛缩酶 2
<i>lac I</i>	b0345	lac 操纵子的抑制蛋白	<i>agaB</i>	b3138	PTS 系统; N-乙酰半乳糖胺-特异成分
<i>mhpR</i>	b0346	mhp 操纵子的调节蛋白	<i>lyxK</i>	b3580	L-木酮糖激酶(隐匿的)
<i>mhpA</i>	b0347	3-丙酸酯羟化酶	<i>sgbH</i>	b3581	3-己酮糖-6-磷酸合成酶
<i>mhpB</i>	b0348	2,3-二羟苯基丙酸盐 1,2-脱氢酶	<i>sgbU</i>	b3582	3-己酮糖-6-磷酸异构酶
<i>mhpC</i>	b0349	2-羟基-6-酮-2,4-双烯酸水解酶	<i>sgbE</i>	b3583	醛缩酶
<i>mhpD</i>	b0350	2-酮-4-环戊酸盐水解酶	<i>rfaL</i>	b3622	核心脂多糖合成
<i>mhpF</i>	b0351	乙醛脱氢酶	<i>rfaK</i>	b3623	己糖转移酶
<i>mhpE</i>	b0352	4-羟基-2-酮戊酸盐醛缩酶	<i>rfaZ</i>	b3624	核心脂多糖合成
<i>mhpT</i>	b0353	转运蛋白	<i>rfaY</i>	b3625	核心脂多糖合成
<i>adhC</i>	b0356	乙醇脱氢酶	<i>rfaJ</i>	b3626	UDP-D-葡萄糖: 脂多糖-葡萄糖基转移酶
<i>putA</i>	b1014	脯氨酸脱氢酶	<i>rfaI</i>	b3627	UDP-D-半乳糖: 脂多糖-α-1,3-D-半乳糖转移酶
<i>putP</i>	b1015	脯氨酸转移蛋白	<i>rfaB</i>	b3628	UDP-D-半乳糖: 脂多糖-1,6-D-半乳糖转移酶
<i>feaR</i>	b1384	苯乙氨代谢调控蛋白	<i>rfaS</i>	b3629	核心脂多糖合成
<i>feaB</i>	b1385	苯乙醛脱氢酶	<i>rfaP</i>	b3630	核心脂多糖合成
<i>tynA</i>	b1386	酪氨氧化酶	<i>phnP</i>	b4092	磷酸盐代谢
<i>maoC</i>	b1387	醛脱氢酶	<i>phnO</i>	b4093	phn operon 调控因子
<i>ansP</i>	b1453	L-天冬酰氨酸透性酶	<i>phnN</i>	b4094	ATP-结合的磷酸盐转运蛋白
<i>narV</i>	b1465	硝酸盐还原酶 2, γ亚单位(隐匿的)	<i>phnM</i>	b4095	磷酸盐代谢
<i>narW</i>	b1466	硝酸盐还原酶 2, Δ亚单位(隐匿的)	<i>phnL</i>	b4096	ATP-结合的磷酸盐转运蛋白
<i>wbbK</i>	b2032	葡萄糖转移酶	<i>phnK</i>	b4097	ATP-结合的磷酸盐转运蛋白
<i>wbbJ</i>	b2033	O-乙酰基转移酶	<i>phnJ</i>	b4098	磷酸盐代谢

(续表 1)

基因	编号	编码产物功能	基因	编号	编码产物功能
<i>wbbI</i>	b2034	转移酶	<i>phnI</i>	b4099	磷酸盐代谢
<i>wbbH</i>	b2035	O-抗原聚合酶	<i>phnH</i>	b4100	磷酸盐代谢
<i>glf</i>	b2036	UDP-吡喃型半乳糖变位酶	<i>phnG</i>	b4101	磷酸盐代谢
<i>rfbX</i>	b2037	O-抗原转运蛋白	<i>phnF</i>	b4102	转录调节蛋白
<i>rfbC</i>	b2038	dTDP-6-脱氧-D-葡萄糖-3,5-变构酶	<i>phnE</i>	b4104	Pn 转运的膜通道蛋白
<i>wcaE</i>	b2055	M 抗原合成转移酶	<i>cadA</i>	b4131	赖氨酸脱羧酶
<i>atoS</i>	b2219	AtoC 蛋白反应传感器	<i>cadC</i>	b4133	转录激活物
<i>atoC</i>	b2220	ato 调节蛋白	<i>idnR</i>	b4264	L-idonate 转录调节子
<i>atoD</i>	b2221	乙酰-CoA:乙酰乙酰基-CoA 转移酶 α 亚单位	<i>idnT</i>	b4265	L-idonate 转运蛋白
<i>atoA</i>	b2222	乙酰-CoA:乙酰乙酰基-CoA 转移酶 β 亚单位	<i>idnO</i>	b4266	5-酮-D-葡萄糖酸盐-5-还原酶
<i>atoE</i>	b2223	转运蛋白	<i>idnD</i>	b4267	L-idonate 脱氢酶
<i>atoB</i>	b2224	乙酰-CoA 乙酰转移酶	<i>idnK</i>	b4268	葡萄糖酸盐激酶
<i>xapR</i>	b2405	xapA 调节蛋白	<i>mcrD</i>	b4344	McrE 5-甲基胞嘧啶限制系统抑制物
<i>xapB</i>	b2406	黄嘌呤透性酶	<i>mcrC</i>	b4345	McrBC 5-甲基胞嘧啶限制系统组分
<i>xapA</i>	b2407	黄嘌呤磷酸化酶	<i>mcrB</i>	b4346	McrBC 5-甲基胞嘧啶限制系统组分
<i>eutC</i>	b2440	乙醇氨裂解酶轻链	<i>mrr</i>	b4351	限制腺嘌呤甲基化

2.2 转运系统及结构蛋白相关的缺失基因

在福氏 2a 志贺氏菌基因组缺失,而在大肠杆菌中存在的功能已知基因中,还包括编码 PTS 系统(phosphotransferase system, 磷酸转移酶系统)、GSP (general secretion pathway, 一般分泌系统)蛋白输出系统、柠檬酸依赖的铁离子运输系统等的组分. 另外,一些细菌鞭毛蛋白、部分药物抗性相关蛋白的基因在福氏 2a 志贺氏菌基因组中也发生缺失. 表 2 为 *E. coli* K12 具有,但在福氏 2a 志贺氏菌基因组中发生缺失的大肠杆菌转运系统组分的基因及一些结构基因.

2.3 与遗传物质水平转移有关的元件

- (i) 原噬菌体基因. 在微生物基因组中通常都会有一些温和噬菌体以原噬菌体的形式整合到细菌基因组中,福氏 2a 志贺氏菌基因组缺失片段中也含有一些 K12 基因组中存在的原噬菌体基因,如: *CP4-6*, *e14*, *CPS-53*, *CPZ-55*, *CP4-57*, *DLP12*, *Rac* 和 *Qin*.
- (ii) 基因重排元件(rearrangement hotspots elements, rhs 元件). *E. coli* K12 基因组中存在的基因重排元件 *rhsA*, *rhsB*, *rhsC*, *rhsD* 和 *rhsE* 在福氏 2a 志贺氏菌中缺失.
- 3 讨论
- 病原微生物的致病机制及其进化过程等一直是人们关注的重点. 从毒力相关基因^[6]和致病性大质粒的发现^[7],到毒力岛^[8]、基因水平转移^[9]以及趋同性

进化(convergent evolution)^[10]和适应性致病突变(pathoadaptive mutation)^[11]等概念的提出,表明人们已从单一分析某一性状或功能相关的基因,转向到从整个基因组水平去认识病原微生物的生物学性状及其致病特点. 微生物基因组工程的开展和生物信息学的应用,为我们提供了丰富的信息资源和有效的分析工具.

对于福氏 2a 志贺氏菌的进化,有许多种推测. 目前倾向于将志贺氏菌(*Shigella*)与埃希氏大肠杆菌(*Escherichia coli*)视为同一属,属内的菌株之间能够通过接合、转导和基因重排等机制进行高效率的基因重组^[12,13]. 如果将福氏志贺菌的毒力基因转移进非致病性大肠杆菌中形成杂交体,其表现出的致病性状类似于肠侵袭性大肠杆菌(EIEC). 虽然已知福氏 2a 志贺氏菌的致病性与其菌体内存在的侵袭性大质粒有关,但是通过比较基因组学分析,发现两者在染色体上也存在明显的差别,大肠杆菌 K12 染色体上约有 717 kb 的片段是福氏 2a 志贺氏菌染色体所没有的,占到 K12 染色体总长的 15.3%.

以大肠杆菌的代谢途径为参照,我们对福氏 2a 志贺氏菌缺失片段上的 ORFs 进行分析,发现这些缺失基因所参与的代谢途径约有 20 个,涉及 4 种类型的代谢,即化合物降解、合成代谢、中间产物代谢以及能量代谢(表 3).

在福氏 2a 志贺氏菌缺失基因所涉及的一些代谢途径中,有些反应的代谢底物为丙酮酸盐、琥珀酸盐

表 2 福氏 2a 志贺氏菌基因组中缺失的大肠杆菌转运蛋白及结构相关的基因

基因	编号	编码产物功能	基因	编号	编码产物功能
<i>betA</i>	b0311	黄素蛋白	<i>eutH</i>	b2452	乙醇氨合成反应
<i>betB</i>	b0312	NAD ⁺ -依赖的甜菜碱醛脱氢酶	<i>eutG</i>	b2453	乙醇氨合成反应
<i>betI</i>	b0313	<i>bet</i> 基因转录抑制物	<i>eutJ</i>	b2454	乙醇氨合成反应
<i>betT</i>	b0314	高亲和胆碱转运蛋白	<i>eutE</i>	b2455	乙醇氨合成反应
<i>prpR</i>	b0330	prp 操纵子的调节蛋白	<i>eutI</i>	b2458	乙醇氨合成反应
<i>prpB</i>	b0331	磷酸酯变位酶	<i>sseA</i>	b2521	硫代硫酸盐巯基转移酶
<i>prpC</i>	b0333	柠檬酸合成酶	<i>ascB</i>	b2716	6-磷酸-β-葡萄糖苷酶(隐匿的)
<i>prpE</i>	b0335	丙酰-CoA 合成酶	<i>hycI</i>	b2717	蛋白酶
<i>codB</i>	b0336	胞嘧啶透性酶	<i>hycH</i>	b2718	蛋白酶
<i>codA</i>	b0337	胞嘧啶脱氨酶	<i>hycG</i>	b2719	氢化酶
<i>cynR</i>	b0338	<i>cyn</i> 操纵子调节蛋白	<i>hycF</i>	b2720	铁硫蛋白氢化酶
<i>cynT</i>	b0339	碳脱水酶	<i>ygbL</i>	b2738	醛缩酶
<i>cynS</i>	b0340	氰酸盐氨基水解酶	<i>yqeI</i>	b2847	传感蛋白
<i>cynX</i>	b0341	氰酸盐转运蛋白	<i>agaA</i>	b3135	N-酰基葡萄糖胺-6-磷酸脱乙酰基酶
<i>lacA</i>	b0342	硫代半乳糖乙酰转移酶	<i>agaS</i>	b3136	塔格糖-6-磷酸醛缩酶异构酶
<i>lacY</i>	b0343	半乳糖苷透性酶	<i>agaY</i>	b3137	塔格糖-二磷酸醛缩酶 2
<i>lacZ</i>	b0344	β-D-半乳糖苷酶	<i>agaB</i>	b3138	PTS 系统; N-乙酰半乳糖胺-特异成分
<i>Lac I</i>	b0345	lac 操纵子的抑制蛋白	<i>lyxK</i>	b3580	L-木酮糖激酶(隐匿的)
<i>mhpR</i>	b0346	mhp 操纵子的调节蛋白	<i>sgbH</i>	b3581	3-己酮糖-6-磷酸合成酶
<i>mhpA</i>	b0347	3-丙酸酯羟化酶	<i>sgbU</i>	b3582	3-己酮糖-6-磷酸异构酶
<i>mhpB</i>	b0348	2,3-二羟苯基丙酸盐-1,2-脱氢酶	<i>sgbE</i>	b3583	醛缩酶
<i>mhpC</i>	b0349	2-羟基-6-酮-2,4-双烯酸水解酶	<i>rfaL</i>	b3622	核心脂多糖合成
<i>mhpD</i>	b0350	2-酮-4-环戊酸盐水解酶	<i>rfaK</i>	b3623	己糖转移酶
<i>mhpF</i>	b0351	乙醛脱氢酶	<i>rfaZ</i>	b3624	核心脂多糖合成
<i>mhpE</i>	b0352	4-羟基-2-酮戊酸盐醛缩酶	<i>rfaY</i>	b3625	核心脂多糖合成
<i>mhpT</i>	b0353	转运蛋白	<i>rfaJ</i>	b3626	UDP-D-葡萄糖: 脂多糖-葡萄糖基转移酶
<i>adhC</i>	b0356	乙醇脱氢酶	<i>rfaI</i>	b3627	UDP-D-半乳糖: 脂多糖-α-1,3-D-半乳糖转移酶

(*mhp*)(*ato*)、6-磷酸葡萄糖酸盐(*idn*)、尿嘧啶(*cod*)、丙酸盐(*prp*)和乙酰 coA, 而这些化合物在整个的代谢网络或者局部的代谢网络中起着枢纽的作用^[14~17]. 它们可以从多条途径中产生, 又与多个代谢途径相连, 缺失基因只与其中的一条途径相关, 因此我们推断, 即使这些基因缺失, 也不会对细菌的生理活动造成太大的影响. 相反, 在另一些代谢途径中, 缺失基因编码的酶所催化的反应是这一代谢网络中的惟一途径, 当这些基因发生缺失时, 很可能会导致相关代谢通路被阻断, 如: 福氏 2a 志贺氏菌基因组中与细菌利用氰酸盐水解生成 CO₂ 和水的代谢途径相关的 *cyn* 操纵子发生缺失, 导致相应的代谢途径被阻断. 类似的情况还出现在 *lac*, *fea*, *tyn*, *lyx*, *phn*, *rfb*, *glf*, *cad*, *bet* 等代谢途径中. 在经典的生化实验如乳糖发酵(*lac*)、木糖发酵(*lyx*)、赖氨酸脱羧酶等实验(*cad*)

中福氏 2a 志贺氏菌呈现阴性而大肠杆菌中为阳性的结果也证实了以上基因缺失导致相关代谢途径被阻断的推测^[18,19].

在福氏 2a 志贺氏菌缺失基因中, 还有一些大肠杆菌的隐匿基因, 如与 L-木糖代谢、硝酸盐代谢和甘露醇运输有关的 *lyxK*, *narVW* 和 *cmtAB* 等基因, 它们编码单一代谢途径相关的酶^[20,21]. 通过同源比较及已知 COGs 库所提供的参比信息可以看出, 在福氏 2a 志贺氏菌中并不存在这些酶的同工酶或者相似功能的酶, 因此我们推断福氏 2a 志贺氏菌没有这几条代谢途径. 这意味着福氏 2a 志贺氏菌不具备利用 L-木糖和硝酸盐的代谢途径.

福氏 2a 志贺氏菌所缺失的 ORFs 中, 有一些是编码与细菌表面鞭毛、菌毛形成有关的结构蛋白和外膜蛋白(表 2), 其中包括细菌对外界机械性压力作出反

表 3 福氏 2a 志贺氏菌基因组中发生缺失的大肠杆菌代谢途径

缺失基因	相关代谢途径	反应底物
化合物降解		
<i>AgaASYB</i>	半乳糖醇代谢	3-磷酸-甘油醛; 磷酸二羟基丙酮; 1, 6-二磷酸塔格糖
<i>mhpRABCDEFT</i>	3-羧苯基丙酸盐降解	3-(2,3-二羧苯基)丙酸盐; 乙酰-CoA
<i>AtoSCDAEB</i>	短链脂肪酸降解	乙酰乙酸; 乙酰乙酰辅酶 A; 乙酰-CoA
<i>cynRTSX^{a)}</i>	氰酸盐代谢	氰酸盐; 二氧化碳; H ₂ CO ₃
<i>lacAYZ^{a)}</i>	乳糖代谢	乳糖; β-D-葡萄糖; β-D-半乳糖
<i>feaRB^{a)}</i>	苯乙胺降解	苯乙胺; 乙酸苯酯
<i>tynA^{a)}</i>	苯乙胺降解	苯乙胺; 乙酸苯酯
	苏氨酸代谢	氨基丙酮; 甲基乙二醛
<i>lyxK^{a)}</i>	L-木糖代谢	L-木酮糖; L-5-磷酸木酮糖
<i>IdnRODK</i>	酮葡萄糖酸代谢	5-酮葡萄糖; 葡萄糖酸盐; 6-磷酸葡萄糖酸盐; 5-酮葡萄糖酸(D); L-idonate
<i>phnEFGHIJKLMNOP^{a)}</i>	磷酸盐利用	
合成代谢		
<i>codAB^{a)}</i>	嘧啶核苷酸/核苷代谢	胞嘧啶; 尿嘧啶
<i>rfbXC^{a)}</i>	O-抗原生物合成	dTDP-4-脱氢-6-脱氧-D-葡萄糖; dTDP-鼠李糖
<i>glf^{a)}</i>	O-抗原生物合成	UDP-半乳糖; UDP-D-半乳糖苷-1,4-呋喃糖
<i>CadAC</i>	赖氨酸及二氨基庚酸盐代谢	L-赖氨酸; 尸胺
中间产物代谢		
<i>betABIT^{a)}</i>	甜菜碱合成	胆碱; 甜菜碱
<i>XapRBA</i>	核苷代谢	黄嘌呤; 黄苷
<i>PrpRBCE</i>	丙酸盐代谢/甲基柠檬酸循环	丙酸盐; 2-甲基柠檬酸盐; 2-甲基乌头酸盐; 丙酮酸盐; 琥珀酸盐
能量代谢		
<i>narVW^{a)}</i>	厌养呼吸	NO ₃ ⁻ + 还原受体; 柠檬酸 + 受体
<i>HycIHGF</i>	厌养呼吸	H ⁺ ; H ₂
<i>AdhC</i>	发酵	乙醛; 乙醇

a) 示单一代谢途径

应的效应基因 *mscL*^[22], 和一些细菌耐药性有关的膜蛋白基因 *dac*, *acr*.

细菌基因组通常是由一个在大多数菌株中普遍存在的“核心基因”(core genes)和株特异性的可变“基因池”(flexible gene pool)所组成. 前者是维持细菌正常生理活动所必须的基本组成, 是细菌遗传信息的骨架, 后者则保证了细菌对环境的适应性^[23]. 由于环境的不同, 不同菌株在进化过程中获取或丢失基因的根本原因在于为细菌的生长繁殖创造有利的微环境. 本研究中 2 个细菌基因组比较的结果显示, 福氏 2a 志贺氏菌基因组所缺失的主要是代谢相关基因与表面结构蛋白基因, 这些基因显然属于可变“基因池”中的组分, 所有由于基因缺失导致阻断的代谢途径均属于细菌生长繁殖非必需或可通过其他代谢途径进行补偿的途径. 而福氏 2a 志贺氏菌在进化过程中所获取基因中的大部分往往属于和毒力相关的基因, 这些基因大多呈簇状排列, 构成多个毒力岛. 此

外, 福氏 2a 志贺氏菌还通过缺失毒力抑制基因而导致毒力增强, 如 *cad* 基因的缺失导致福氏 2a 志贺氏菌毒力增强就是一例^[24]. Maurelli 等人^[24]发现, *E coli* 染色体上存在的赖氨酸脱羧酶基因 *cad* 操纵子所在的大约 90 kb 的片段在福氏 2a 中是缺失的, 如果将 *cad* 基因在福氏 2a 的菌体中表达, 福氏 2a 的肠毒素活性将丧失, 这主要是因为赖氨酸脱羧酶使赖氨酸脱羧后变成尸胺, 而尸胺能直接抑制福氏 2a 志贺氏菌所产生的肠毒素的作用. 因此, 对福氏 2a 志贺氏菌缺失基因分析将有助我们揭示其毒力进化的规律.

肠道致病菌间存在着极为活跃的基因转移, 这与肠道致病菌基因组中存在大量的噬菌体插入序列和基因重排元件有关. 在艾希氏菌属上存在一类 *rhs* 元件, 既高频重组元件. 迄今为止发现的 5 类 *rhs* 元件共同具有的结构特征是: 一个高(G+C)含量的核心 ORF, 紧接着的是下游 ORF(downstream ORF), 而在下游 ORF 的右侧会出现一个或多个 IS 序列, *Rhs* 元

件所编码的产物为处于细胞表面具有结合功能的蛋白. 有关的文献显示, 从该元件的结构来看, 它在细菌的遗传物质的进化中起重要作用, 因为它不仅可以相互组合形成新的 Rhs 元件^[25], 而且还具有组织和包含插入序列的多重功能. 然而, 福氏 2a 志贺菌染色体中没有发现 rhs 元件的存在, 但是在其基因组中却存在大量的、多种类型的重复序列, 这或许提示福氏 2a 志贺菌具有与大肠杆菌不同的水平转移机制.

病原微生物所表现出的多种生理特征及毒力特性反映了不同的基因组分导致细菌不同的生活方式. 单个菌株基因组序列分析所得到的信息不能完整代表其他亲缘关系相近成员的遗传信息. 因此, 多个亲缘关系相近菌株的“整体基因组研究”(whole genome approaches)发展迅猛, 特别是包含一个种内所有菌株基因的“种基因组概念”(species genome)的提出将相关研究推向一个新水平. 随着不同菌株基因组信息的不断增多, 大肠杆菌和志贺氏菌以及相关菌属中其他成员间的进化关系和菌株特异性基因对生物性状的影响等将逐渐被阐明.

致谢 本课题由病毒基因工程国家重点实验室及北京大学肝病研究所、复旦大学医学院教育部分子病毒学重点实验室、国家疾病控制与预防中心卫生部分子细菌学重点实验室、国家人类基因组北方研究中心和华北制药集团等共同完成. 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: Z19-02-05-01), 北京市创新工程(批准号: 955020700)及华北制药集团资助项目.

参 考 文 献

1 Hale T L. Genetic basis of virulence in *Shigella* species. Microbiol Rev, 1991, 55: 206~224

2 Plattner F R, Plunkett G III, Bloch C A, et al. The complete genome sequence of *Escherichia coli* K12. Science, 1997, 277, 1453~1474

3 Brenner D J, Fanning G R, Johnson K E, et al. Polynucleotide sequence relationships among members of Enterobacteriaceae. J Bacteriol, 1969, 98(2): 637~650

4 Ochman H, Whittam T S, Caugant DA, et al. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J Gen Microbiol, 1983, 129 (Pt9): 2715~2726

5 Jin Q, Yuan Z H, Xu J G, et al. Genome sequence of *Shigella flexneri* 2a: Insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucl Acid Res, 2002, 30(20): 4432~4441

6 Sasakawa C, Komatsu K, Tobe T, et al. Eight genes in region 5 that form an operon are essential for invasion of epithelial cells by *Shigella flexneri* 2a. J Bacteriol, 1993, 175(8): 2334~2346

7 Venkatesan M M, Goldberg M B, Rose D J, et al. Complete DNA sequence and analysis of the large virulence plasmid of *Shigella*

flexneri. Infect Immun, 2001, 69(5): 3271~3285

8 Al-Hasani K, Rajakumar K, Bulach D, et al. Genetic organization of the she pathogenicity island in *Shigella flexneri* 2a. Microb Pathog, 2001, 30(1): 1~8

9 Eisen J A. Horizontal gene transfer among microbial genomes: New insights from complete genome analysis. Curr Opin Genet Dev, 2000, 10(6): 606~611

10 Gulietta M P, Ruiting L, Peter R R. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci USA, 2000, 97(19): 10567~10572

11 Day W A Jr, Fernandez R E, Maurelli A T. Pathoadaptive mutations that enhance virulence: Genetic organization of the cadA regions of *Shigella* spp. Infect Immun, 2001, 69(12): 7471~7480

12 Levine M M, Woodward W E, Formal S B. Studies with a new generation of oral attenuated *shigella* vaccine: *Escherichia coli* bearing surface antigens of *Shigella flexneri*. J Infect Dis, 1977, 136(4): 577~582

13 Falkow S, Formal S B. Restriction in genetic crosses between *Escherichia coli* and *Shigella flexneri*. J Bacteriol, 1969, 100(1): 540~541

14 Ingledew W J, Poole R K. The respiratory chains of *Escherichia coli*. Microbiol Rev, 1984, 48(3): 222~271

15 Teichmann S A, Rison S C, Thornton J M, et al. Small-molecule metabolism: an enzyme mosaic. Trends Biotech, 2001, 19(12): 482~486

16 Viswanathan A, You H J, Doetsch P W. Phenotypic change caused by transcriptional bypass of uracil in nondividing cells. Science, 1999, 284(5411): 159~162

17 London R E, Allen D L, Gabel S A, et al. Carbon-13 nuclear magnetic resonance study of metabolism of propionate by *Escherichia coli*. J Bacteriol, 1999, 181(11): 3562~3570

18 Ito H, Kido N, Arakawa Y, et al. Possible mechanisms underlying the slow lactose fermentation phenotype in *Shigella* spp. Appl Environ Microbiol, 1991, 57: 2912~2917

19 Ewing W H. The genus *Escherichia*. In: Edwards and Ewing's Identification of the Enterobacteriaceae. 4th ed. New York: Elsevier Science, Publishing, 1986

20 Blasco F, Pommier J, Augier V, et al. Involvement of the narJ or narW gene product in the formation of active nitrate reductase in *Escherichia coli*. Mol Microbiol, 1992, 6(2): 221~230

21 Sprenger G A. Two open reading frames adjacent to the *Escherichia coli* K12 transketolase (tkt) gene show high similarity to the mannitol phosphotransferase system enzymes from *Escherichia coli* and various gram-positive bacteria. Biochim Biophys Acta, 1993, 1158(1): 103~106

22 Sukharev S I, Blount P, Martinac B, et al. Mechanosensitive channels of *Escherichia coli*: The MscL gene, protein, and activities. Annu Rev Physiol, 1997, 59: 633~657

23 Dobrindt U, Hacker J. Whole genome plasticity in pathogenic bacteria. Curr Opin Microbiol, 2001, 4: 550~557

24 Maurell A T, Fernandez R E, Bloch C A, et al. "Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *shigella* spp and enteroinvasive *E coli*. Proc Natl Acad Sci USA, 1998, 95: 3943~3948

25 Zhao S, Hill C N. Reshuffling of Rhs components to create a new element. J Bacteriol, 1995, 177(5): 1393~1398

(2002-09-15 收稿, 2003-02-11 收修改稿)