

大肠杆菌 O23 O-抗原基因簇序列的破译及 UDP-N-乙酰葡萄糖-C4 异构酶的鉴定

程剑松¹ 王 威¹ 王 荃¹ 高春绪¹ 王 磊^{1,2*}

(¹南开大学生命科学学院 天津 300071)

(²南开大学泰达学院功能基因组学研究中心 天津 300457)

摘 要 采用鸟枪法破译大肠杆菌 O23 标准株的 O-抗原基因簇序列,并用生物信息学的方法进行了基因注释和分析;采用基因缺失和互补的方法鉴定了 O23 的 UDP-GlcNAc C4 异构酶(*Gne*);用同源建模的方法构建了 O23 *Gne* 的高级结构并对其活性位点进行了分析;分析了不同血清型大肠杆菌 O-抗原基因簇中 *gne* 基因的多样性;根据 O23 O-抗原基因簇中的特异基因筛选出了可用于大肠杆菌 O23 快速检测的特异 DNA 序列。

关键词 大肠杆菌;O-抗原基因簇;基因缺失;高级结构

中图分类号:Q78 文献标识码:A 文章编号:0001-6209(2006)05-0702-07

大肠杆菌是动物和人类肠道中占主导地位的革兰氏阴性菌,部分菌株为条件致病菌,按其致病机制分为肠致病性大肠杆菌(Enteropathogenic *E. coli*, EPEC)、肠侵袭性大肠杆菌(Enteroinvasive *E. coli*, EIEC)、肠产毒性大肠杆菌(Enterotoxigenic *E. coli*, ETEC)、肠出血性大肠杆菌(Enterohaemorrhagic *E. coli*, EHEC)和肠聚集性大肠杆菌(Enterocoagulative *E. coli*, EAEC)等^[1]。大肠杆菌 O23 是一株重要的 ETEC 致病菌,能产生热稳定肠毒素(heat-stable toxin, ST),引起儿童和家畜的腹泻、皮下组织水肿和神经系统病,在畜牧业生产上危害很大^[2~4]。

O-抗原是一类多糖抗原,一般由 3~8 个单糖重复单位组成,存在于几乎所有革兰氏阴性菌的表面,与细菌的致病性关系密切,在致病性、信号识别、黏附、免疫逃避等过程中发挥重要作用。根据 O-抗原不同,大肠杆菌可分为 166 种血清型。O-抗原是自然界中最具多样性的分子之一,成为研究遗传多样性和进化机理的理想材料^[5]。

本实验破译了大肠杆菌 O23 的 O-抗原基因簇全序列,并鉴定了 UDP-GlcNAc C4 异构酶(*Gne*),通过构建并比较其高级结构,分析了 O23 *Gne* 可能的活性位点。筛选出了对大肠杆菌 O23 高度特异的引物,可用于对该菌的快速检测。另外,对来自不同

血清型大肠杆菌的 13 个 *gne* 基因的遗传学分析表明它们属于 3 个不同的进化来源。

1 材料和方法

1.1 材料

1.1.1 菌种和质粒 大肠杆菌 O23、O55 标准菌株、O55 的 *gne* 缺失菌株^[6]及 Red 辅助质粒 pKD20^[7]由悉尼大学 Reeves 教授实验室提供;大肠杆菌 (*Escherichia coli*)DH5 α 由本实验室保存;pGEM-T-easy 载体购自 Promega 公司,质粒 pKK232-8 和表达载体 pTRC99A 购自 Pharmacia 公司。

1.1.2 酶和试剂 Expand Long Template PCR 酶购自 Boehringer Mannheim 公司,PCR 纯化试剂盒、DNA 胶回收试剂盒、T4 DNA 连接酶、Wizard PCR Preps 纯化试剂盒购自 Promega 公司, *Eco*RI、*Nco*I、*Bam*HI、*Sac*I、DNaseI、X-gal、IPTG、氨苄青霉素和氯霉素购自上海 Sangon 公司。

1.2 基因组 DNA 的提取

采用 Bastin 法^[8]提取细菌染色体 DNA。

1.3 O23 O-抗原基因簇的获得

根据 *galF* 基因序列设计上游引物(5'-ATTGTGGCTGCAGGGATCAAAGAAAT-3'),根据 *gnd* 基因设计下游引物(5'-TAGTCGCGTGNGCCTGGATT

基金项目 国家 863 计划(2002AA2Z2051);国家杰出青年科学基金资助项目(30125001)

* 通讯作者。Tel:86-22-66229584; Fax:86-22-66229596; E-mail:wanglei@nankai.edu.cn

作者简介 程剑松(1976-)男,山西省阳泉市人,博士研究生,主要从事微生物基因组学、细菌表面抗原的分子遗传学研究。

E-mail:jiansong_cheng@yahoo.com.cn

其他作者 彭 霞^{1,2}

收稿日期 2005-11-18 接受日期 2005-12-28 修回日期 2006-04-10

AAGTTCGC-3')。以大肠杆菌 O23 细菌基因组为模板,用长 PCR 方法扩增 O-抗原基因簇。PCR 反应程序如下:94℃ 2min;94℃ 10s,60℃ 30s,68℃ 15min,共 30 个循环,68℃ 延伸 7min,得到 PCR 产物。合并 6 管长 PCR 产物,用 Wizard PCR Preps 纯化试剂盒纯化 PCR 产物^[9]。

1.4 O23 O-抗原基因文库的构建

用 DNaseI 消化 PCR 纯化产物,收集大小集中在 1~3kb 之间的片段,经过补平,尾部补加碱基 A,与 pGEM-T-easy 载体在 T4 DNA 连接酶作用下 16℃ 连接 24h,取连接产物 2μL 电转化感受态大肠杆菌 DH5α,涂布在含有氨苄青霉素、X-gal 和 IPTG 的 LB 固体培养基上,37℃ 过夜培养,筛选白色克隆,以碱裂解法用 96 孔平板提取质粒,EcoRI 酶切鉴定插入片段大小,挑选插入片段在 1kb 以上的克隆。采用 T7/Sp6 公用引物,用 ABI3730 型 DNA 自动测序仪完成测序,用英国剑桥 MRC (Medical Research Council) 分子生物学实验室出版的 Staden package 软件包的 Pregap4 和 Gap4 软件拼接和编辑所有序列。

1.5 生物信息学方法进行序列分析

用英国 Sanger 中心的 Artemis 软件发现开放阅读框(Orfs),用 BLAST 系列软件与 GenBank 中的基因比较初步确定这些开放阅读框的功能,将相似性较差的基因搜索 COG, Pfam 等二级数据库,进一步确定其功能,并用 Artemis 完成基因注释。用 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) 和 SMART (<http://smart.embl-heidelberg.de/>) 进行蛋白质跨膜片段的分析;用 ClustalW 软件精确比对

DNA 和蛋白质序列;用 MEGA 3.0 软件构建系统进化树;用 Accelrys 公司的分子模拟软件 InsightII 以及 Swiss-pdb Viewer V3.7 构建并分析 Gne 的三维结构。

1.6 基因缺失

以 pKK232-8 为模板,用大肠杆菌 O23 *gne* 缺失引物扩增氯霉素基因(引物见表 1),该引物 5' 端为根据 O23 *gne* 基因上、下游 DNA 设计的用于同源重组的 40bp DNA 片段,3' 端则根据氯霉素基因上、下游设计,用于扩增完整的氯霉素基因。预期大小的 PCR 产物用低熔点胶回收,试剂盒纯化。将纯化产物电转化含有 pKD20 并经 IPTG 诱导的大肠杆菌 O23 感受态细胞。转化细胞在 37℃、200r/min 诱导 1h,涂布于含有氯霉素的 LB 固体培养基上,37℃ 过夜培养,pKD20 为温度敏感型辅助质粒,温度高于 30℃ 培养即自动消除^[7]。挑取单克隆,用缺失检测引物进行 PCR 检测,筛选大肠杆菌 O23 的 *orf7* 缺失株^[7]。

1.7 体内互补

分别扩增大肠杆菌 O23 的 *orf7* 和 O55 的 *gne* 基因(引物见表 1),用 *NcoI/BamHI*,*SacI/BamHI* 分别进行双酶切,克隆至表达载体 pTRC99A,构建重组质粒,并分别电转化 O55 *gne* 缺失株和 O23 *orf7* 缺失株的感受态细胞,涂布于含有氨苄青霉素和氯霉素的 LB 固体培养基上,37℃ 过夜培养。筛选转化子转接于 20mL 含有氨苄青霉素和氯霉素的 LB 液体培养基,加入 IPTG 诱导过夜。提取脂多糖(LPS),用 SDS-PAGE 和银染方法^[10]检测互补情况。

表 1 大肠杆菌 O23 *gne* 基因缺失和体内互补所用引物

Table 1 Primers used for mutation and complementation tests of *gne* in *E. coli* O23

Primers	Oligonucleotide sequences(5' ~ 3')
Primer pairs used to delete <i>gne</i> in <i>E. coli</i> O23	AATAATCTAATATATAGATTAACATCTACCGGAGTGTTTTATGAGAAAAAATCACTGG AAATACCATTGAGCATTTCATGCTTTTATATAACCGTTCGGTAAAAAATTCACGCCCC
Primer pairs used to confirm the mutation of <i>gne</i> in <i>E. coli</i> O23	CAATTGATGATAATCCTGGTG TCAGCACCTTGTGCGCTTGC
Primer pairs used to clone <i>gne</i> of <i>E. coli</i> O23	TGTGCCATGGCTATTCTTGTACTGGC GTTGGATCCTGTGACGGTGATACTGCG
Primer pairs used to clone <i>gne</i> of <i>E. coli</i> O55	ATATAGAGCTCATGAACGATAACGTTTTGCTC CGGGATCCTTACTCAGACAAAAATGCTAT

1.8 特异基因的筛选

用 PCR 方法筛选针对大肠杆菌的特异基因。分别针对大肠杆菌 O23 O-抗原基因簇中的 *wzx* 和 *wzy* 基因设计两对引物(表 2),以 192 株大肠杆菌和

志贺氏菌的基因组为模板分别进行 PCR(表 3),PCR 扩增条件:94℃ 2min;94℃ 15s,退火温度因引物的差异而不同(表 2),退火时间 50s,72℃ 2min,共 30 个循环,72℃ 7min,反应体系为 25μL。

表 2 大肠杆菌 O23 O-抗原基因簇中寡糖单位处理基因及其中的引物和 PCR 数据

Specific genes	Oligonucleotide sequence(5' ~ 3')	Length of the PCR fragment(bp)	Annealing temperature(°C)
<i>wzx</i>	TTTGTATTATCTTGGCGC TTTTTCTTATGCTGCC	423	50
	TAAATATACTTGGCCCGG TATTCATCTGCCATGGC	386	56
<i>wzy</i>	TGCATAACTCTTCAAGCGC GCTTGTAAGGTACGCTG	402	53
	TTAGGCAGTCAATTTCCG AATCTAGCCAATGATGG	460	52

表 3 用于筛选特异基因的 192 株大肠杆菌和志贺氏菌及其来源*

Pool No.	Strains of which chromosomal DNA included in the pool	Source
1	<i>E. coli</i> type strains O1 ,O2 ,O3 ,O4 ,O5 ,O7 ,O12 ,O13 ,O14 ,O15 ,O16 ,O17 ,O19ab ,O20 ,O21 ,O22 ,O23 ,O24 ,O59	IMVS
2	<i>E. coli</i> type strains O25 ,O26 ,O27 ,O28 ,O29 ,O30 ,O32 ,O33 ,O35 ,O36 ,O37 ,O38 ,O40 ,O41 ,O42 ,O43	IMVS
3	<i>E. coli</i> type strains O44 ,O45 ,O46 ,O48 ,O49 ,O50 ,O51 ,O52 ,O54 ,O55 ,O56 ,O57 ,O58 ,O60 ,O61 ,O62	IMVS
4	<i>E. coli</i> type strains O63 ,O65 ,O66 ,O69 ,O70 ,O71 ,O74 ,O75 ,O76 ,O77 ,O78 ,O79 ,O80 ,O81 ,O82 ,O83	IMVS
5	<i>E. coli</i> type strains O84 ,O85 ,O86 ,O87 ,O88 ,O89 ,O91 ,O92 ,O98 ,O99 ,O101 ,O102 ,O103 ,O104 ,O106	IMVS
6	<i>E. coli</i> type strains O107 ,O108 ,O109 ,O110 ,O111 ,O112ab ,O112ac ,O113 ,O115 ,O116 ,O118 ,O120 ,O123 ,O125 ,O126 ,O128	IMVS
7	<i>E. coli</i> type strains O129 ,O130 ,O131 ,O132 ,O133 ,O134 ,O135 ,O136 ,O137 ,O138 ,O139 ,O141 ,O142 ,O143 ,O144 ,O145	IMVS
8	<i>E. coli</i> type strains O146 ,O147 ,O148 ,O150 ,O152 ,O154 ,O156 ,O157 ,O158 ,O159 ,O160 ,O161 ,O163 ,O164 ,O165 ,O166	IMVS/SSI
9	<i>E. coli</i> type strains O168 ,O169 ,O170 ,O171 ,O172 ,O173 <i>Shigella dysenteriae</i> type strains D1 ,D2 ,D3 ,D4 ,D5 ,D6 ,D7 ,D8 ,D9 ,D10 ,D11 ,D12	a IEM
10	<i>Shigella boydii</i> type strains B1 ,B2 ,B3 ,B4 ,B6 ,B7 ,B8 ,B9 ,B10 ,B11 ,B12 ,B13 ,B14 ,B15 ,B16 ,B17 ,B18	IEM
11	<i>Shigella flexneri</i> type strains F1a ,F1b ,F2a ,F2b ,F3 ,F4b ,F5 (v 4) ,F5 (v 7) ,F6 ,FX ,FY <i>Shigella sonnei</i> type strains DS ,DR	IEM
12	<i>E. coli</i> type strains O3 ,O11 ,O39 ,O59 ,O64 ,O73 ,O96 ,O95 ,O100 ,O114 ,O151 ,O155 ,O124	IMVS
13	<i>E. coli</i> type strains O1 ,O2 ,O3 ,O4 ,O5 ,O7 ,O12 ,O13 ,O14 ,O15 ,O16 ,O17 ,O19ab ,O20 ,O21 ,O22 ,O24 ,O59	IMVS

* A total of 13 pools of DNA were made, the pool of No. 13 contains same strains with the pool of No. 1 except for *E. coli* O23. IMVS: Institute of Medical and Veterinary Science, Anlaide, Australia. SSI: Statens Serum Institute, Copenhagen, Denmark. IEM: Institute of Epidemiology and Microbiology, Chinese Academy of Preventive Medicine. a: O172 and O173 from Statens Serum Institute, Copenhagen, Denmark, others from IMVS.

2 结果和分析

2.1 大肠杆菌 O23 O-抗原基因文库的构建与其基因簇的序列分析

通过鸟枪法构建大肠杆菌 O23 O-抗原基因文库并测序,得到大肠杆菌 O23 O-抗原基因簇完整序列。该序列全长 9502bp,编码 7 个基因,所有基因的转录方向一致,均从 *galF* 基因到 *gnd* 基因。O23 O-抗原基因簇平均(G + C)含量为 35.79%,显著低于大肠杆菌基因组的平均(G + C)含量(50%)。目前普遍认为 GC 差异反映了 DNA 片段的外源性^[11]。*galF* 和 *gnd* 均为大肠杆菌的持家基因,不负责 O-抗原基因簇的合成。大肠杆菌 O23 O-抗原的每个重复单位包括 6 个单糖,分别为 GalNAc, Gal, 2 个 GlcNAc 和 2 个 Glc^[12]。由于转移第一个单糖 GlcNAc 或 GalNAc 的糖基转移酶基因位于 O-抗原基因簇之外^[13],我们推测 O23 O-抗原合成还需要 5 个糖基转移酶,O23 O-抗原基因簇编码 4 个糖基转移酶基因,分别为 *orf2*, *orf4*, *orf5* 和 *orf6*,另外 1 个糖基转移

酶基因可能位于 O23 O-抗原基因簇之外。此外,O23 O-抗原基因簇还包含 1 个 UDP-N-乙酰葡萄糖-C4 异构酶基因以及 2 个寡糖单位处理基因 *wzx* 和 *wzy*(表 4)。*Wzx* 和 *Wzy* 均为膜蛋白,分别具有 11 个和 9 个潜在的跨膜区段,*Wzy* 含有一个约 60 个氨基酸的亲水环,是 O-抗原聚合酶的典型特征。

2.2 大肠杆菌 O23 特异基因的鉴定

自上世纪 30 年代,O-抗原一直被用作目标分子用于病原菌的血清学鉴定。但传统的血清学方法耗时长,灵敏度低。本研究以大肠杆菌 O23 O-抗原基因簇为基础实现了对 O23 的分子分型。最近的研究表明 O-抗原基因簇中的寡糖单位处理基因(*wzx*, *wzy*)通常为血清型特异基因^[13]。分别针对 *wzx*, *wzy* 设计 2 对 PCR 引物(表 2)并对 192 株大肠杆菌和志贺氏菌进行 PCR 反应,结果在第 1 组中均产生了预期的 PCR 产物,片段大小在 250 ~ 500bp 之间,在其它 12 组细菌中都没有扩增出可见的 PCR 产物,因此说明根据 *wzx* 和 *wzy* 基因设计的引物对大肠杆菌 O23 是高度特异的,可用于快速准确检测人

体和环境中大肠杆菌 O23。

表 4 大肠杆菌 O23 O-抗原基因簇的基因功能预测

Table 4 Summary of genes in *E. coli* O23 O-antigen gene cluster

Gene	Location in O-antigen cluster	Conserved domain(s) (Pfam No. /E Value)	Similar protein(s), strain(s) (GenBank accession no.)	Identical/Similar % (total no. of aa)	Putative function
<i>wzy</i>	1202 ~ 2227		Wzy <i>Salmonella enterica</i> (AAB49386)	25/49 (342)	O-antigen polymerase
<i>orf2</i>	2224 ~ 3093	Glycosyltransferase (PF00535/4 × e ⁻³⁰)	Sugar transferase <i>Lactococcus lactis</i> subsp. <i>Lactis</i> (AAK04314)	34/55 (269)	Glycosyltransferase
<i>wzx</i>	3105 ~ 4271		Hypothetical chloroplast RF81 <i>Mesostigma viride</i> (AAF43857)	26/48 (182)	O unit flippase
<i>orf4</i>	4261 ~ 4917	Glycosyltransferase sugar-binding region (PF04488/7.7 × e ⁻¹⁶)	CpsIVN <i>Streptococcus agalactiae</i> (AAK43611)	40/56 (202)	Glycosyltransferase
<i>orf5</i>	4914 ~ 5684	Glycosyltransferase (PF00535/1.1 × e ⁻²⁵)	WbgO <i>Escherichia coli</i> (AAL67559)	37/58 (255)	Glycosyltransferase
<i>orf6</i>	5876 ~ 6994	Glycosyltransferases group 1 (PF00534/2.3 × e ⁻³⁴)	Rfp <i>Shigella dysenteriae</i> (AAA26537)	49/67 (373)	Glycosyltransferase
<i>gne</i>	7025 ~ 8044	NAD dependent epimerase/dehydratase (PF01370/1.4 × e ⁻¹⁵⁶)	Gne <i>Yersinia enterocolitica</i> (type O 8) (AAC60777)	61/75 (336)	UDP-galactose 4-epimerase

2.3 大肠杆菌 O23 中 Gne 的功能鉴定

野生型大肠杆菌 O23 和 O55 均可产生正常的 O-抗原,其脂多糖(LPS)表现为光滑型,LPS 的聚丙烯酰胺凝胶电泳显示类脂 A 和 O-抗原的典型带谱(图 1)。采用 Datsenko 等^[7]建立的基因缺失方法用氯霉素基因替换大肠杆菌 O23 的 *orf7* 以构建 O23 *orf7* 的缺失株。大肠杆菌 O23 的 *orf7* 缺失株以及大肠杆菌 O55 的 *gne* 缺失株均不能合成正常的 O-抗原,其 LPS 只包含类脂 A,表现为粗糙型,LPS 聚丙烯酰胺凝胶电泳只显示类脂 A 而缺少 O-抗原的典型带谱(图 1)。将大肠杆菌 O23 的 *orf7* 转化大肠杆菌

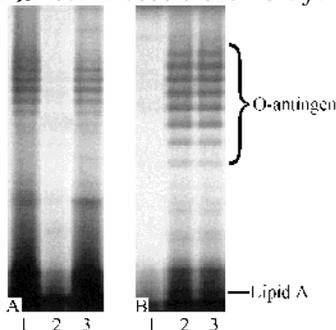
图 1 大肠杆菌 O23 *gne* 的缺失及互补株的 LPS 聚丙烯酰胺凝胶电泳分析

Fig.1 Deletion and complementation analysis of *E. coli* O23 *gne* gene. Membrane extracts were run on SDS-PAGE gels and stained by silver staining. A: 1. *E. coli* O55 type strain; 2. *E. coli* O55 missing its *gne* gene; 3. *E. coli* O55 missing its *gne* gene and complemented with *orf7* of *E. coli* O23. B: 1. *E. coli* O23 missing its *gne* gene; 2. *E. coli* O23 type strain; 3. *E. coli* O23 missing its *gne* gene and complemented with the *gne* gene of *E. coli* O55.

菌 O55 的 *gne* 缺失株或将大肠杆菌 O55 的 *gne* 转化大肠杆菌 O23 的 *orf7* 缺失株后,均可使粗糙型 LPS 恢复为光滑型 LPS,即大肠杆菌 O23 的 *orf7* 和 O55 的 *gne* 基因可以实现功能互补(图 1)。因大肠杆菌 O55 的 Gne 已经过实验鉴定,所以可进一步确定大肠杆菌 O23 的 *orf7* 为 UDP-GlcNAc C4 异构酶基因,并命名为 *gne*。

2.4 构建大肠杆菌 O23 中 Gne 的高级结构

大肠杆菌 O23 中的 Gne 与 299 位氨基酸突变的大肠杆菌 UDP-Glu C4 异构酶有较高相似性(70%)。在此,我们将 299 位氨基酸发生突变 UDP-Glu C4 异构酶称为 GalE*,GalE* 的 299 位氨基酸由原来的 Try 突变为 Cys,导致底物结合位点空间变大。突变后,GalE* 对 UDP-Glu 的催化活性下降了 5 倍,但对

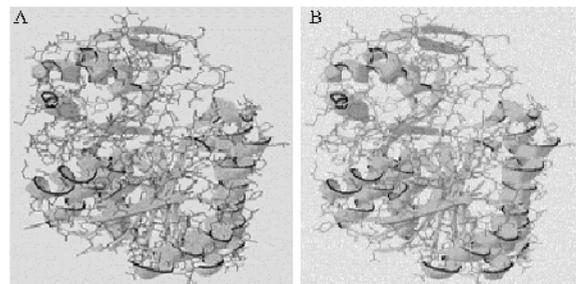


图 2 用同源建模方法构建的大肠杆菌 O23 Gne 的高级结构

Fig.2 Structures of *E. coli* O23 Gne generated by homology modeling method. A: The predicted structure of *E. coli* O23 Gne based on the structure of GalE* complexed with NADH and UDP-Glc; B: The predicted structure of *E. coli* O23 Gne based on the structure of GalE* complexed with UDP-Glc.

UDP-GlcNAc 的催化活性提高了 230 倍^[14]。GalE * 的高级结构已知,我们分别采用与 UDP-Glu 和 UDP-GlcNAc 结晶的 GalE * 为模板构建了大肠杆菌 O23 Gne 的高级结构(图 2)。

O23 的 Gne 与 GalE * 的高级结构极为相似,虽然它们表面电势分布有较大差别,但在底物结合位点附近具有相似的结构和电势分布,前者底物结合袋状区的入口略大于后者(图 3)。当我们比较大肠杆菌 O23 Gne 高级结构(以与 UDP-Glu 结晶的 GalE * 为模板构建)和与 UDP-Glu 结晶的 GalE * 的高级结构时,发现两者具有相同的活性位点,而比较大肠杆菌 O23 Gne 高级结构(以 GalE * 为模板构建)和与 UDP-GlcNAc 结晶的 GalE * 的高级结构时,我们发现,两者在与底物 UDP-GlcNAc 及辅酶 NADH 相邻的 25 个氨基酸中,仅有一个氨基酸的差异,即 83 位氨基酸在 O23 Gne 和 GalE * 中分别为 Try 和 Leu(图 4)。因此,我们推测大肠杆菌 O23 的 Gne 具有 UDP-GlcNAc C4 异构酶和 UDP-Glu C4 异构酶的双重活性。在第一组来源相同的 11 个 Gne 中,来自小肠结肠炎耶尔森菌 O8 的 Gne 也具有 UDP-GlcNAc C4 异构酶和 UDP-Glu C4 异构酶的双重活性^[15]。在这 11 个 Gne 中除大肠杆菌 O86 和 O127 的 Gne 在 299 位置的氨基酸为 Ser 外,其它 Gne 在 299 位置的氨基酸均为 Cys, Ser 和 Cys 结构相似,且均为不带电荷的极性氨基酸。因此,我们推测第一组 Gne 可能均有 UDP-GlcNAc C4 异构酶和 UDP-Glu C4 异构酶的双重活性。

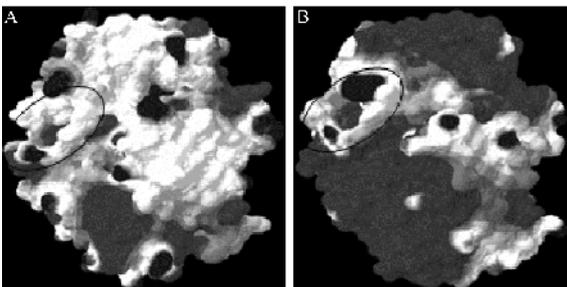


图 3 大肠杆菌 O23 Gne 和 GalE * 的表面静电势分布

Fig.3 Electrostatic potential maps of *E. coli* O23 Gne and GalE * . A : The distribution of the electric charge at the molecular surface of *E. coli* O23 Gne ; B : The distribution of the electric charge at the molecular surface of GalE * . The molecular surfaces in A and B indicate Van der Waals surfaces which were calculated based on the Van der Waals radius of each atom by Swiss-pdb Viewer V3.7. White represents neutral potential, black represents positive and negative potentials. The cavities circled by black lines were the binding sites of substrates.

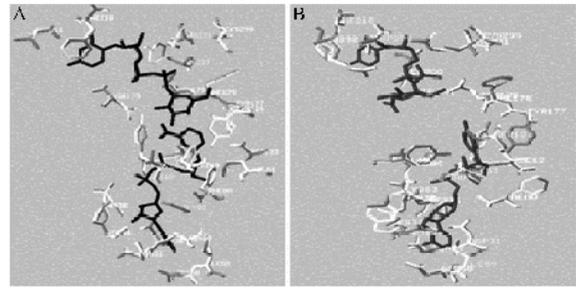


图 4 大肠杆菌 O23 Gne 和 galE * 活性位点的比较

Fig.4 Active site differences between Gne of *E. coli* O23 and GalE * . A :The amino acid residues of Gne of *E. coli* O23 (gray) and GalE * (white) located within approximately 3.2 Å of atoms of the NADH (black) and UDP-glucose (black) molecules ; B :The amino acid residues of Gne of *E. coli* O23 (gray) and GalE * (white) located within approximately 3.2 Å of atoms of the NADH (black) and UDP-glcNAc (black) molecules .

2.5 gne 的多样性分析

UDP-GlcNAc C4 异构酶基因,可催化 UDP-GlcNAc 形成 UDP-GalNAc。在此,我们将所有 UDP-GlcNAc C4 异构酶基因通称为 *gne*。目前,已知 10 种血清型大肠杆菌和 3 种血清型志贺氏菌的 O-抗原基因簇中存在 *gne*,与相应菌株 O-抗原中含有 GlcNAc 结果一致。将这些 *gne* 与相似性较高的基因构建系统进化树,发现它们有 3 个不同的进化来源(图 5)。其中,包括来自大肠杆菌 O23 的大部分 *gne* 属于第一组,与另一组旁系同源基因 *galE* 具有一定的相似性。*GalE* 为 UDP-Glc C4 异构酶,催化 UDP-Glc 形成 UDP-Gal。*galE* 在大肠杆菌中为持家基因,与半乳糖激酶基因(*galK*),半乳糖-1-磷酸尿苷酰转移酶基因(*galT*)组成半乳糖操纵元^[16]。*gne* 不是持家基因,在大肠杆菌和志贺氏菌中与 O-抗原中的单糖合成相关。系统进化树显示的第一组 *gne* 和第二组 *gne* 均位于相应 O-抗原基因簇的内部,第三组 *gne* 则位于持家基因 *galF* 前面,与相应 O-抗原基因簇相邻。O-抗原基因簇中存在频繁的基因横向转移,序列比较的结果显示这种转移主要由于同源重组或转座子转座导致^[17]。由此推测,在大肠杆菌 O-抗原基因簇中存在不同来源的 *gne* 可能由于基因的横向转移造成,这种横向转移进一步造成了大肠杆菌 O-抗原的多样性,同时可以使宿主在较短时间内获得更好的适应性。

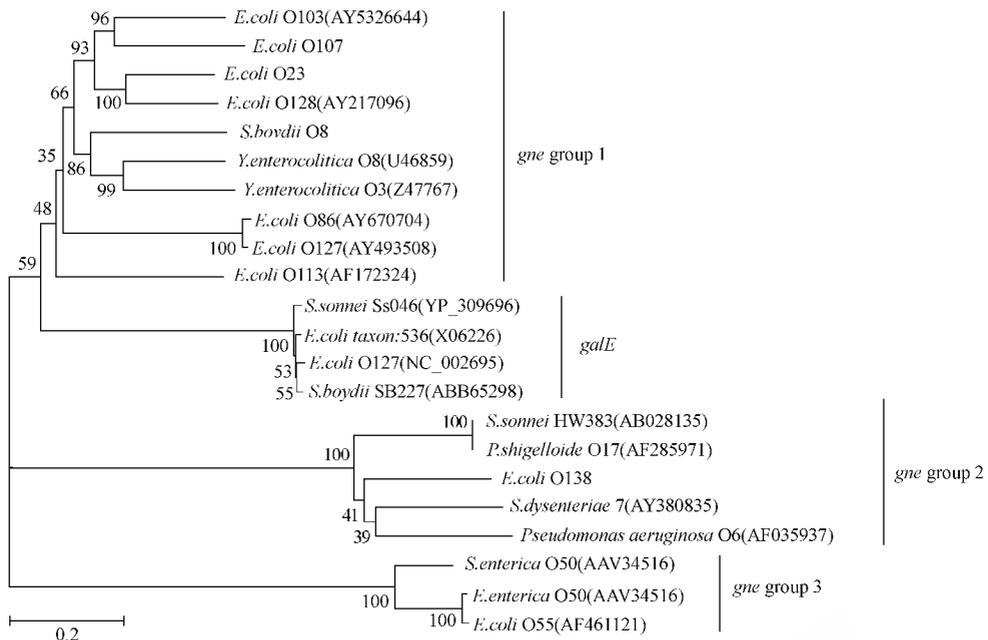


图5 用邻位加入法构建的 *gne* 和 *galE* 的系统进化树

Fig. 5 The phylogenetic tree for *gne* and *galE* generated by the neighbor-joining method.

Numbers in brackets represent accession numbers of the sequences in Genbank.

3 讨论

以表面多糖为目标的血清学免疫反应自上世纪30年代以来一直被用于对细菌的分型和鉴定,是鉴定致病菌的重要手段。这种诊断方法需要大量的抗血清,而抗血清一般种类不全,数量不足,大量的抗血清在制备和储存中也存在一些困难。另一方面此法耗时长、灵敏度低、漏检率高、准确性差,这种传统的血清学检测方法将逐渐为现代分子生物学方法取代。我们根据大肠杆菌 O23 的 O-抗原特异基因设计的4对引物,对大肠杆菌 O23 有很好的特异性,因此可利用筛选出的特异 DNA 片段用于基因芯片或 PCR 方法对大肠杆菌 O23 的快速检测。目前,利用特异基因对大肠杆菌进行分子分型的方法已被广泛研究^[18]。

大肠杆菌 O-抗原是自然界中最富多样性的生物物质之一,O-抗原多样性的遗传学研究成为近年来的一个热点领域^[13]。O-抗原基因簇的横向转移是促成 O-抗原多样性的重要原因之一,不同血清型大肠杆菌 O-抗原基因簇中的 *gne* 基因存在3个不同的进化来源,为 O-抗原基因簇中基因的横向转移提供了很好的例证。在大肠杆菌3种不同来源的 *gne* 中,只有 O55 的 *gne* 经过鉴定^[6]。在本研究中,我们结合基因缺失和三维建模详细分析了 O23 中 *Gne* 的功能。

参 考 文 献

- [1] Morris JA, Sojka WJ. *Escherichia coli* as a pathogen in animals. In: Sussman M. ed. The Virulence of *Escherichia coli*: Reviews and Methods. London: Academic Press, 1985, 47-77.
- [2] Blanco J, Cid D, Blanco JE, et al. Serogroups, toxins and antibiotic resistance of *Escherichia coli* strains isolated from diarrhoeic lambs in Spain. *Vet Microbiol*, 1996, **49**(3-4): 209-217.
- [3] Garabal JI, Gonzalez EA, Vazquez F, et al. Serogroups of *Escherichia coli* isolated from piglets in Spain. *Vet Microbiol*, 1996, **48**(1-2): 113-123.
- [4] Peruski LF Jr, Kay BA, El-Yazeed RA, et al. Phenotypic diversity of enterotoxigenic *Escherichia coli* strains from a community-based study of pediatric diarrhea in periurban Egypt. *J Clin Microbiol*, 1999, **37**(9): 2974-2978.
- [5] Goldman RC, Hunt F. Mechanism of O-antigen distribution in lipopolysaccharide. *J Bacteriol*, 1990, **172**: 5352-5359.
- [6] Wang L, Huskic S, Cisterne A, et al. The O-antigen gene cluster of *Escherichia coli* O55:H7 and identification of a new UDP-GlcNAc C4 epimerase gene. *J Bacteriol*, 2002, **184**(10): 2620-2625.
- [7] Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci*, 2000, **97**(12): 6640-6645.
- [8] Bastin DA, Romana LK, Reeves PR. Molecular cloning and expression in *Escherichia coli* K12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. *Mol Microbiol*, 1991, **5**: 2223-2231.

- [9] 郭宏杰,冯 露,张 淳,等.大肠杆菌 O150 O-抗原基因簇的破译和 dTDP-鼠李糖合成酶基因的鉴定. *微生物学报*, 2004, **44**(1): 34-40.
- [10] Wang L, Reeves PR. Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica* *rfbP* gene in O-antigen subunit processing. *J Bacteriol*, 1994, **176**: 4348-4356.
- [11] Lan R, Reeves PR. Gene transfer is a major factor in bacterial evolution. *Molecular Biology and Evolution*, 1996, **13**: 47-55.
- [12] Bartelt M, Shashkov AS, Kochanowski H, *et al.* Structure of the O-specific polysaccharide of the O23 antigen (LPS) from *Escherichia coli* O23 K7:H16. *Carbohydr Res*, 1993, **248**: 233-240.
- [13] Samuel G, Reeves P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research*, 2003, **338**(23): 2503-2519.
- [14] Thoden JB, Henderson JM, Fridovich-Keil JL, *et al.* Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase. *J Biol Chem*, 2002, **277**(30): 27528-27534.
- [15] Bengoechea JA, Pinta E, Salminen T, *et al.* Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O:8. *J Bacteriol*, 2002, **184**(15): 4277-4287.
- [16] Lemaire HG, Muller-Hill B. Nucleotide sequences of the *galE* gene and the *galT* gene of *E. coli*. *Nucleic Acids Res*, 1986, **14**(19): 7705-7711.
- [17] Pacinelli E, Wang L, Reeves PR. Relationship of *Yersinia pseudotuberculosis* O Antigens IA, IIA, and IVB: the IIA gene cluster was derived from that of IVB. *Infect Immun*, 2002, **70**(6): 3271-3276.
- [18] Wang L, Liu B, Kong Q, *et al.* Molecular markers for detection of pathogenic *Escherichia coli* strains belonging to serogroups O138 and O139. *Vet Microbiol*, 2005, **111**: 181-190.

Molecular characterization of O-antigen gene cluster of *Escherichia coli* O23 reference strain and identification of UDP-N-acetylglucosamine 4-epimerase

CHENG Jian-song¹, WANG Wei¹, WANG Quan¹, GAO Chun-xu¹, WANG Lei^{1,2*}

(¹ College of Life Science, Nankai University, Tianjin 300071, China)

(² Center for Functional Genomics Research, TEDA College, Nankai University, Tianjin 300457, China)

Abstract: Lipopolysaccharide (LPS) is one of the major components of the outer membrane of gram-negative bacteria. It is an amphipathic molecule composing of lipid A, a core oligosaccharide and an O-specific antigen. O-antigen, which is a repeat unit polysaccharide, is a major contribution to the antigenic variability of the bacterial cell surface. Genes involved in O-antigen biosynthesis are generally found to be clustered between the housekeeping genes *galF* and *gnd* on the chromosome of *E. coli*. *E. coli* O23 is one of the enterotoxigenic *E. coli* causing pediatric diarrhea in the developing world. The O-antigen gene cluster of *E. coli* O23 type strain was amplified by long-range PCR using primers based on *galF* and *gnd* and then sequenced. Except for *galF* and *gnd*, seven open reading frames were identified and assigned functions on the basis of their similarity to those from available databases. The seven genes include a UDP-N-acetylglucosamine 4-epimerase gene (*gne*), the O-antigen polymerase gene (*wzy*), the O-antigen transferase gene (*wzx*) and four glycosyltransferase genes (*orf2*, *orf4*, *orf5*, *orf6*). The UDP-N-acetylglucosamine 4-epimerase (Gne) was identified by mutation and complementation complement tests. The structure of Gne was predicted by the homology modeling method, and the active sites were also analyzed. The phylogenetic and structural analysis showed that the Gne derived from the common ancestor with *E. coli* O23 Gne were UDP-GlcNAc/UDP-GalNAc epimerases. The specific DNA used for rapid molecular genotyping for *E. coli* O23 was also identified.

Keywords: *Escherichia coli*; O-antigen gene cluster; Specific gene; Protein structure

Foundation item: Chinese National Programs for High Technology Research and Development (2002AA2Z2051); Chinese National Science Fund for Distinguished Young Scholars (30125001)

* Corresponding author. Tel: 86-22-66229584; Fax: 86-22-66229596; E-mail: wanglei@nankai.edu.cn

Other author: PENG Xia^{1,2}

Received: 18 November 2005/Accepted: 28 December 2005/Revised: 10 April 2006