

# 大肠杆菌 O54 O-抗原基因簇的破译及进化分析

杨静华 于 浩 任 一 王 磊 冯 露\*

(南开大学生命科学学院 天津 300071)

**摘 要** 破译了大肠杆菌 O54 O-抗原基因簇的序列,序列全长 14062bp。用生物信息学方法分析序列并鉴定基因,共确定 10 个基因,包括鼠李糖合成酶基因 BDA 和  $\alpha$  (*rmlBDA* 和 *rmlC*) 糖基转移酶基因, O-抗原转运酶基因, O-抗原聚合酶基因和合成磷酸丝氨酸侧链的基因及 1 个不能确定功能的开放阅读框。对 *rmlC* 的 (G + C)% 含量,稀有密码子含量及进化分析都表明大肠杆菌 O54 O-抗原基因簇是在近期通过 *rmlC* 介导的重组形成,而且大肠杆菌 O54 和鲍氏志贺菌 9 型的亲缘关系很近。对 UTP-葡萄糖-1-磷酸-尿苷转移酶基因(*galF*)和 6-磷酸葡萄糖脱氢酶基因(*gnd*)的进化分析揭示志贺菌属与大肠杆菌属在进化上属于同一个属。用 PCR 方法筛选出了针对大肠杆菌 O54 的特异基因,用于基因芯片或 PCR 方法对大肠杆菌 O54 的快速检测。

**关键词** 大肠杆菌 O54 O-抗原基因簇, *rmlC* 基因, 进化分析, 特异基因

中图分类号: Q93 文献标识码: A 文章编号: 0001-6209(2004)04-0445-07

O-抗原是大肠杆菌(*Escherichia coli*)脂多糖的最外层结构,它一般由重复的寡糖单位组成,每个单位有 3~6 个单糖,其中单糖种类的不同、单糖数量的不同、单糖间键的不同和寡糖单位之间连接键的不同构成了 O-抗原的多样性。大肠杆菌根据其 O-抗原分为 170 种不同的血清型<sup>[1]</sup>。负责合成 O-抗原的所有酶分子的基因一般都在染色体上相邻存在,形成一个基因簇<sup>[2]</sup>。在大肠杆菌和沙门氏菌(*Salmonella*)中, O-抗原基因簇通常位于 *galF* 和 *gnd* 基因之间<sup>[3]</sup>。O-抗原的多样性几乎完全取决于 O-抗原基因簇的变化。

O-抗原的多样性,使编码合成 O-抗原的基因簇成为在 DNA 分子水平上研究分子进化的最好材料之一。目前发现新的 O-抗原基因簇由两种方式产生。一种是由其他菌的多糖基因簇通过同源重组横向转移形成,如大肠杆菌和沙门氏菌的 O-抗原基因簇<sup>[3]</sup>;另一种方式是由转座因子介导的重组形成<sup>[4]</sup>。本文通过与已知功能的序列进行比较,鉴定了大肠杆菌 O54 O-抗原基因簇的鼠李糖合成酶基因。并对 *rmlC* 基因进行遗传和进化分析,揭示大肠杆菌 O54 O-抗原基因簇是在近期通过 *rmlC* 基因介导的重组形成。

大肠杆菌 O54 是一种致病菌,在 1986 年首先从

患腹泻和水肿的幼猪中分离到<sup>[5]</sup>。它可产生细胞毒素坏死因子(Cytotoxic necrotising factor, CNF),属于引起坏死的大肠杆菌(Necrotoxic *E. coli*, NTEC),潜伏在猪等家畜中,是侵袭型的病原菌。并且对人也有潜在的致病性,能引起动物和人腹泻综合症<sup>[6,7]</sup>,具有潜在的爆发性流行的危险。因此在农业、食品及卫生系统上急需一个可以快速、准确检测大肠杆菌 O54 的方法。本研究用 PCR 的方法筛选出了针对大肠杆菌 O54 的特异基因,可用于基因芯片或 PCR 方法对该菌的快速检测。

## 1 材料和方法

### 1.1 材料

**1.1.1 菌株** *E. coli* O54 标准菌株由澳大利亚悉尼大学 Reeves 教授提供,受体菌株(*E. coli* DH5 $\alpha$ )由本室保存。

**1.1.2 主要试剂** pGEM-T-Easy 购自 Promega 公司; Expand Long Template PCR 酶购自 Boehringer Mannheim 公司; PCR 纯化试剂盒、T4 DNA 连接酶、Wizard PCR Preps 纯化试剂盒购自 Promega 公司; *EcoR* I、DNase I、X-Gal、IPTG、氨苄青霉素购自上海生工生物技术有限公司。

基金项目 国家 863 基因芯片专项资助项目(2002AA2Z2051)、国家自然科学基金资助项目(30270029)

\* 通讯作者。Tel 86-22-66229592; Fax 86-22-66229596; E-mail: lfeng@tedamail.nankai.edu.cn

作者简介 杨静华(1970-),女,内蒙古呼和浩特人,讲师,博士。主要从事致病细菌的遗传进化及快速检测的研究。E-mail: jinghuayang@vip.sina.com

收稿日期 2003-12-04,修回日期 2004-03-31

### 1.2 基因组 DNA 的提取

采用 Bastin<sup>[8]</sup>法提取细菌基因组 DNA

### 1.3 O-抗原基因簇的获得

大肠杆菌 O-抗原基因簇位于 *galF* 基因和 *gnd* 基因之间。根据 *galF* 基因设计上游引物(5'-ATT-GTGGCTGCAGGGATCAAAGAAATC-3'),根据 *gnd* 基因设计下游引物(5'-TAGTCGCGCTGNGCCTGGATT-AAGTTCGC-3')。以大肠杆菌 O54 的基因组为模板,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℃ 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。

### 1.4 O-抗原基因簇文库的构建

通过鸟枪法建库。用 DNase I 消化 PCR 纯化产物,随后经酚:氯仿:异戊醇(25:24:1)抽提和无水乙醇沉淀,酶切产物回溶到水中,补平末端,加 dA 尾,然后与 pGEM-T-Easy 载体在 T4 DNA 连接酶作用下 16℃ 连接 24h。取连接产物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 孔平板大规模提取质粒,用 *Eco*RI 酶切鉴定其中的插入片段的大小,挑选插入片段在 1000bp 以上的克隆用于测序。

### 1.5 序列测定和拼接

利用双脱氧终止法,采用 T7-sp6 公用引物,用本实验室的 ABI3730 型 DNA 自动测序仪对克隆中的插入片段进行测序,保证序列达到 3 倍以上的覆盖率。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的序列。

### 1.6 生物信息学进行序列分析

用美国国家生物技术信息学中心(The National Center for Biotechnology Information,NCBI)的 Orf finder 发现基因,用 Blast 系列软件与 GenBank 中的基因比较以发现这些开放的阅读框的功能并确定它们是什么基因,用英国 Sanger 中心的 Artemis 软件完成基因注释。用 Clustal W 2.0([http://www.ebi.ac.uk/Clustal W/](http://www.ebi.ac.uk/ClustalW/))做 DNA 序列的精确比对,用 Block make(<http://www.blocks.fhcrc.org/blockmkr/make-blocks.html>)寻找 DNA 序列中的保守基序。用 TMHMM 2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)和 HMMTOP(<http://www.enzim.hu/hmmtop1.1/server/submit.html>)分析蛋白质的疏水性以鉴定蛋白质。同时分析了 *rmlC* 基因的(G+C)%含量和稀有密码子的含量,并通过 Mega2 软件产生 *rmlC* 基因的进化树及 *galF* 和 *gnd* 基因的进化树。

### 1.7 特异基因的筛选

用 PCR 方法筛选大肠杆菌 O54 的特异基因。针对大肠杆菌 O54 的 O-抗原基因簇中的 *wzx* 和 *wzy* 基因,在每个基因内部各设计两对引物(表 1),所有 166 株大肠杆菌和 43 株志贺氏菌被分成 26 组<sup>[9]</sup>,在每一组中含有 6~10 个菌,第 9 组含有大肠杆菌 O54 作为正对照,第 27 组是由去除大肠杆菌 O54 的第 9 组中的菌组成,作为负对照。将每一组中所有菌的基因组混合后作模板进行 PCR。PCR 条件:94℃ 2min 94℃ 15s,退火温度因引物的不同而不同(表 1),退火 50s,72℃ 2min,共 30 个循环;72℃ 10min。反应体系是 25μL。

表 1 大肠杆菌 O54 的 O-抗原基因簇的 *wzx*、*wzy* 基因及其中的引物及 PCR 数据

Table 1 PCR testing of the specificity of <i>E. coli</i> O54 genes						
Gene	Function	Position of gene	Position of positive primer	Position of negative primer	Length of PCR product/bp	Annealing temperature/℃
<i>wzy</i>	O-antigen Polymerase	4059~5117	4289~4306	5000~5017	729	60
			4391~4408	4695~4712	322	61
<i>wzx</i>	O-antigen flippase	8961~10211	9019~9036	9591~9608	590	57
			9127~9144	10040~10057	931	60

## 2 结果和讨论

### 2.1 O-抗原基因簇的获得及测序

大肠杆菌的 O-抗原基因簇通常位于两个持家基因 *galF* 和 *gnd* 之间。本文根据 *galF* 和 *gnd* 基因设计上下游引物,用长 PCR 方法扩增 O-抗原基因

簇。长 PCR 产物通过 0.8% 琼脂糖凝胶电泳检测,在接近 15kb 的位置出现一条电泳带。为了保证序列的准确性,我们合并 6 管长 PCR 产物构建文库以避免 PCR 反应本身的错配。

根据帕松公式( $P = e^{-m}$ ,  $P$ :未测定碱基的概率; $m$ :测定序列的覆盖率)<sup>[10]</sup>,为了保证序列质量,每个

碱基至少有 3 个以上高质量(大于 90%)的序列覆盖,首先对 96 个插入片段在 1kb 以上的克隆单向测序以达到 90%的碱基测定概率,对余下的 10%的序列挑选克隆进行反向测序。用 Pregap4 和 Gap4 软件拼接和编辑所有序列,从而得到大肠杆菌 O54 的 O-抗原基因簇的全序列,序列全长 14062bp。

2.2 O-抗原基因簇的序列分析

在获得大肠杆菌 O54 O-抗原基因簇的全序列后,用 Orf finder 发现基因,共找到 11 个开放阅读框(orf),都位于 galF 基因和 gnd 基因之间且具有相同的转录方向,都是从 5'到 3'方向。用 Artemis 软件完

成基因注释后得到大肠杆菌 O54 O-抗原基因簇的结构及其(G+C)%含量。大肠杆菌 O54 O-抗原基因簇中的基因从 5'到 3'方向依次为 rmlB、rmlD、rmlA、wzy、orf5、orf6、orf7、orf8、orf9、wzx 和 rmlC,在 wzx 和 rmlC 之间是一段长为 1826bp 的非编码区。这些基因在 O-抗原基因簇中的位置、氨基酸个数和(G+C)%含量见表 2。galF 基因和 gnd 基因分别位于 O-抗原基因簇的 5'末端和 3'末端,它们不属于 O-抗原基因簇,只是用来扩增 O-抗原基因簇。对每个 orf 用 Blast 系列软件与 GenBank 中的基因比较预测出它们的功能(表 2)。

表 2 大肠杆菌 O54 O-抗原基因簇的基因及其功能预测

Table 2 Summary of ORFs in E. coli O54

Gene name	Position of gene	Total number of aa	(G+C)%	Putative function of protein
rmlB	1137 ~ 2198	353	42.7	dTDP-D-glucose 4 β-dehydratase
rmlD	2223 ~ 3122	299	47.7	dTDP-6-deox-L-mannose-dehydrogenase
rmlA	3180 ~ 4058	291	43.5	Glucose-1-phosphatethimidylyltransferase
wzy	4063 ~ 4614	352	28.1	O-antigen polymerase
orf5	5120 ~ 5938	272	31.4	Glycosyl transferase
orf6	5893 ~ 6711	272	29.4	Protein of unknown function
orf7	6708 ~ 7325	205	26.7	Phosphoserine phosphatase
orf8	7322 ~ 8107	261	27.6	Phosphoserine transferase
orf9	8107 ~ 8964	285	31.5	Rhamnosyl transferase
wzx	8961 ~ 10211	346	27.5	O-antigen flippase
rmlC	12039 ~ 12581	180	34.4	dTDP-6-deoxy-D-glucose-3,5-epimerase

2.2.1 单糖合成酶基因:在 O-抗原基因簇中,单糖合成酶基因由于有极高的同源性,很容易通过与已知功能的序列进行比较而鉴定。细菌中的常见糖由于它们为细菌基础代谢底物,其合成酶基因存在于基因组的其它位置,一般不在 O-抗原基因簇里重复出现。而罕见糖如鼠李糖需由 O-抗原基因簇的单糖合成酶基因负责合成。通过进行 Blast 比较发现,orf1 orf2、orf3 和 orf11 都与许多菌的合成鼠李糖前体的 4 个合成酶基因有非常高的相同性,其中 orf1 与鲍氏志贺氏菌的 RmlB 在 353 个氨基酸中有 98%的相同性;orf2 与鲍氏志贺氏菌的 RmlD 在 299 个氨基酸中有 97%的相同性;orf3 与鲍氏志贺氏菌的 RmlA 在 291 个氨基酸中有 97%的相同性;orf11 与鲍氏志贺氏菌的 RmlC 在 171 个氨基酸中有 63%的相同性,75%的相似性,说明它们都有很高的同源性,可以确定 orf1、orf2、orf3、orf11 分别为 dTDP-鼠李糖的 rmlB、rmlD、rmlA、rmlC 基因,因此分别命名为 rmlB、rmlD、rmlA、rmlC。

目前在大肠杆菌、志贺氏菌(Shigella)、沙门氏

菌(Salmonella)和霍乱弧菌(Vibrio cholerae)中共发现有 40 多株菌的 O-抗原中都含有鼠李糖,尽管在不同菌中合成它的 4 个基因排列顺序不同,但是一般都是相邻排列,例如在大肠杆菌、志贺氏菌和沙门氏菌的 O-抗原基因簇中是以 rmlBDAC 的顺序相邻排列<sup>[11~13]</sup>,在霍乱弧菌的 O-抗原基因簇中是以 rml-BADC 的顺序相邻排列。4 个基因在 O-抗原基因簇中分开的情况很少见。只有在霍乱弧菌 O146 和 O149 中,发现 rmlBD 和 rmlAC 之间有两个基因相隔<sup>[14]</sup>,在鲍氏志贺氏菌 9 型中 rmlC 与 rmlBDA 相距 4 个基因<sup>[12]</sup>。在大肠杆菌 O54 中,首次发现 rmlC 基因位于整个 O-抗原基因簇的 3'末端,远离 rmlBDA 基因,这不同于其他细菌。沙门氏菌和霍乱弧菌的研究表明,鼠李糖合成酶基因内的重组在 O-抗原基因簇的横向转移中扮演重要角色,而 rmlC 基因是重组位点<sup>[13,14]</sup>。因此,大肠杆菌 O54 中 rmlC 基因的独特位置暗示它的 O-抗原基因簇是在近期由 rmlC 携带其上游基因通过同源重组介导到 rmlBDA 基因的一端形成的。另外,Blast 比较表明鼠李糖合成酶的

4个基因的序列相似性具有极性,其中5'端序列(*rmlBDA*基因)达到98%以上的相似性,而3'端序列(*rmlC*基因)有75%的相似性,表明*rmlC*基因非常不同,具有更多的可变性和O-抗原特异性。说明这4个基因具有不同的来源,*rmlBDA*基因很早进入这个种中,已进化到与持家基因几乎一样高的相似性,而*rmlC*基因是后来进入的。

**2.2.2 糖基转移酶基因:**在GenBank中寻找保守的功能域,发现*orf5*与PF00535糖基转移酶家族相似,E值为 $2.2 \times e^{-19}$ ,通过进行Blast比较发现,*orf5*与脆弱拟杆菌(*Bacteroides fragilis*)的糖基转移酶有35%的相同性,55%的相似性;也与霍乱弧菌的糖基转移酶有36%的相同性,56%的相似性,因此推测*orf5*也是糖基转移酶基因,命名为*orf5*。*Orf9*与PF00535糖基转移酶家族相似,Blast比较发现*orf9*编码的蛋白与痢疾志贺氏菌的鼠李糖基转移酶I有29%的相同性,48%的相似性;与福氏志贺氏菌的RfbF蛋白有31%的相同性,49%的相似性,而RfbF蛋白是dTDP-鼠李糖的糖基转移酶,所以推测*orf9*也是一个dTDP-鼠李糖的糖基转移酶基因,命名为*orf9*。

**2.2.3 寡糖单位处理酶基因:**在O-抗原基因簇中,寡糖单位处理酶基因包括O-抗原转运酶基因和O-抗原聚合酶基因,它们编码的蛋白是O-抗原中仅有的有典型的8~14个跨膜片段的蛋白,可以用生物信息学方法鉴定<sup>[15,16]</sup>。

本文用两个不同的软件分析鉴定寡糖单位处理基因以保证结果的准确性。用TMHMM2.0表明*orf4*编码的蛋白中含有10个跨膜片段。用HMMTOP分析也发现同样的10个潜在的跨膜片段,分别位于氨基酸序列中的3~21,27~46,86~104,110~128,156~180,186~205,237~255,261~279,286~307,313~332。从46~86是由40个氨基酸组成的茎环结构(loop),位于第二个和第三个跨膜片段之间。这种大的茎环结构分布在周质区是O-抗原聚合酶(Wzy)的典型特征<sup>[17]</sup>。此外用Blast比较表明*orf4*与许多Wzy蛋白相似,例如,与大肠杆菌O104在342个氨基酸中有27%的相同性及48%相似性,和鲍氏志贺氏菌6型在321个氨基酸中有23%的相同性及48%相似性,和嗜热链球菌在312个氨基酸中有20%的相同性及43%相似性。将这3个菌的Wzy序列用Block maker软件进行分析,共找到4个保守的基序,分别由13、29、35、25个氨基酸组成。将它们的一致序列在数据库中运行3个循环的PSI-

BLAST程序后,发现除了上述3个菌外,还和许多其它菌株的Wzy相似,所以可以确定*orf4*就是wzy基因,命名为wzy。

用TMHMM 2.0分析发现*orf10*编码的蛋白中含有10个跨膜片段。用HMMTOP分析也发现同样的10个潜在的跨膜片段,分别位于氨基酸序列中的16~36,53~73,92~116,154~174,182~202,242~260,294~314,335~355,360~380,385~404。而且Blast比较表明*orf10*与许多Wzx蛋白相似,例如,与热自养甲烷杆菌(*Methanobacterium thermoautotrophicum*)及伴放线菌素放线杆菌(*Actinobacillus actinomycescomitans*)的O-抗原转运酶(Wzx)最相似,分别有53%和50%的相似性。将这3个菌的Wzx序列用Block maker软件进行分析,共找到7个保守的基序,分别为12、13、15、34、35、37、38个氨基酸组成。将它们的一致序列在数据库中运行3个循环PSI-BLAST程序后,发现除了上述两个菌外,还和许多其它菌株的Wzx相似,所以可以确定*orf4*就是wzx基因,命名为wzx。

**2.2.4 合成磷酸丝氨酸侧链的基因:**Blast比较发现,*orf7*与丙酮丁醇梭菌(*Clostridium acetobutylicum*)的磷酸丝氨酸磷酸化家族的酶在194个氨基酸中有24%的相同性和44%的相似性,推测*orf7*也是一个磷酸丝氨酸磷酸化酶,命名为*orf7*。*Orf8*与烷烃双色菌(*Melipona bicolor*)的CDP-2-乙酰甘油-磷酸丝氨酸转移酶有24%的相同性和52%的相似性,推测它是一个磷酸丝氨酸转移酶,命名为*orf7*。*orf7*和*orf8*负责磷酸丝氨酸前体的合成和O-抗原寡糖单位的连接,因此我们推测在大肠杆菌O54的O-抗原中存在丝氨酸。在人体细胞质膜的内小叶上发现有磷脂酰丝氨酸的存在,所以丝氨酸可能是大肠杆菌O54在形成致病菌的过程中,为逃避免疫系统的识别,模仿寄主的表面成分进化而来。通过对已有数据库的搜索,不能确定*orf6*的功能。

**2.2.5 非编码区:**在*orf10*和*rmlC*之间的1826bp的序列中没有大的开放阅读框。在原核细胞的基因组中基因排列很紧密,基因多连续排列,基因之间的间隔区很短。像大肠杆菌O54的O-抗原基因簇中如此长的没有功能的序列十分少见。用Tblastx进行多序列比对发现,这段非编码区与多形拟杆菌(*Bacteroides thetaiotaomicron*)的胞外多糖基因簇中相邻的糖基转移酶(GenBank获得号AA076452)和胞外多糖合成蛋白capA(GenBank获得号AA076453)的DNA序列有46%的相同性,局部的相同性已达到

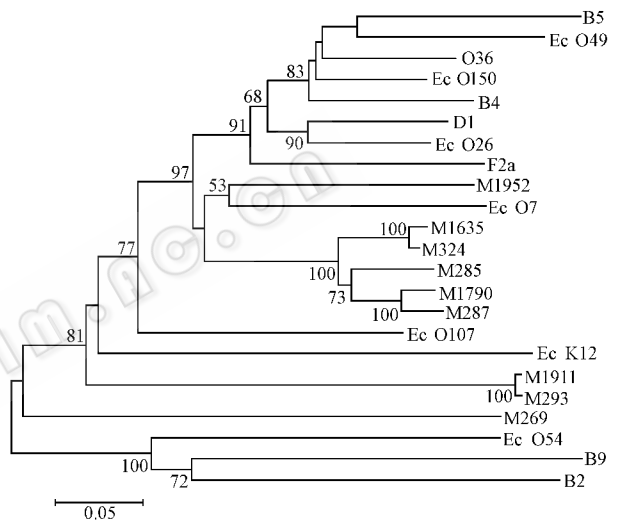
70%。因此我们推测这段非编码区是多形拟杆菌 (*B. thetaiotaomicron*) 中的糖基转移酶基因和胞外多糖合成蛋白 *capA* 基因一同重组进 O-抗原基因簇中,后来在细菌的进化过程中,随着环境的变化不再需要这 2 个基因所编码的蛋白,因此某些碱基突变成终止密码子或发生移码突变从而使序列失去原有的功能。而大肠杆菌 O54 中长的非编码区的存在暗示它的 O-抗原基因簇是在很近的时期才形成。因为进化的时间很短,这段非编码区还没有大段的丢失,基本还保留着两个基因的长度(1826bp)。

**2.2.6 *rmlC* 基因的遗传进化分析** 在大肠杆菌 O54 的 O-抗原基因簇中, *rmlBDA* 的平均(G + C)%含量为 44.54%,而 *rmlC* 的(G + C)%含量仅为 33.85%, (G + C)%含量的不同暗示它们是在不同时间进入到大肠杆菌 O54 的 O-抗原基因簇中,推测 *rmlBDA* 在大肠杆菌 O54 中已存在较长时间,而 *rmlC* 是后来才从低 GC 含量的细菌中重组到大肠杆菌 O54 的 O-抗原基因簇中。进一步与 GenBank 中已知的大肠杆菌、志贺氏菌的 *rmlC* 的(G + C)%含量比较,发现大肠杆菌 O54 中 *rmlC* 的(G + C)%含量最低。一般在 O-抗原基因簇重组到基因组上之后(G + C)%含量将不断改变,向基因组的(G + C)%含量漂移<sup>[18]</sup>,而大肠杆菌 O54 的 *rmlC* 还保持着很低的(G + C)%含量。因此我们推测大肠杆菌 O54 的 O-抗原基因簇是在很近的时期形成的。

另外,稀有密码子的含量也说明基因的外源性。一个基因的稀有密码子含量高要求其相应的同功 tRNA 的含量高,而其相应的同功 tRNA 的含量一般都少,致使相应的蛋白弱表达。而许多稀有密码子富含 A + T,导致基因的(G + C)%含量低。因此用稀有密码子来调控基因的翻译可能是一种选择压力,借此来维持编码区的低(G + C)%含量<sup>[17]</sup>。在大肠杆菌 O54 的 O-抗原基因簇中, *rmlA*、*rmlB*、*rmlD* 的稀有密码子的含量分别为 4.7%、3.4%、4.6%,而 *rmlC* 基因的稀有密码子的含量高达 10%。这也说明 *rmlBDA* 很早进入到 O-抗原基因簇中,而 *rmlC* 基因是在近期才进入大肠杆菌 O54 的 O-抗原基因簇中。

我们也对 *rmlC* 基因做了进化分析。从 GenBank 中提取已知功能的鼠李糖合成酶基因的 *rmlC* 序列,用 Mega2 软件分析、采用 Neighbor-joining 方法得到系统发生树(图 1)。我们发现大肠杆菌 O54 与鲍氏志贺氏菌 9 型都落入一个支中,说明它们的亲缘关系非常近。而且这两个菌的 O-抗原基因簇的共同特征是 *rmlC* 基因都远离 *rmlBDA* 基因。Reeves

等<sup>[19]</sup>在 2000 年发现志贺氏菌是从大肠杆菌进化而来,进化时间在 3 万 5 千年至 27 万年之间。鲍氏志贺氏菌 9 型中 *rmlC* 基因的(G + C)%含量为 38.9%,大肠杆菌 O54 的 *rmlC* 基因的(G + C)%含量比它低 5%,更低(G + C)%含量表明细菌进化的时间更短。此外,研究表明鲍氏志贺氏菌 9 型是在近期新形成的<sup>[12]</sup>,由此我们推测,大肠杆菌 O54 的 O-抗原与鲍氏志贺氏菌 9 型可能起源于共同的祖先,分化的时间非常短,都是通过 *rmlC* 基因介导的重组转移到它们现在所在的细菌中。而大肠杆菌 O54 比鲍氏志贺氏菌 9 型产生的时间更晚,是在很近的时期才形成的。



**图 1 用 Mega2 软件产生的 *rmlC* 的进化树**  
Fig.1 Phylogenetic trees for the *rmlC* genes generated by Mega2  
Sequences used included those from eight *E. coli* strains (EcO54, Patent No. 200310107159.5), *E. coli* O49(EcO49, unpublished data in our laboratory), *E. coli* O36(EcO36, unpublished data in our laboratory), *E. coli* O7(EcO7, AAC63615), *E. coli* O107(EcO107, Patent No. 03109588.7), *E. coli* O26(EcO26, AAN60457), *E. coli* K12(Eck12, P37745), *S. boydii* 5(Ecbd5, AAL27325), *S. boydii* 9(Ecbd9, AAL27354), *S. boydii* 2(Ecbd2, Patent No. 200310107158.0), *S. dysenteriae* 1(D1, unpublished data in our laboratory), *S. flexneri* 2a(F2a, P37780), nine *S. enterica* strains: M1952(AAG09516), M1635(AAG09501), M324(AAG09540), M1790(AAG09505), M287(AAG09529), M1911(AAG09512), M293(AAG09533), M285(AAG09525), M269(AAG09521). The scale bar indicates the 0.05 evolutionary distance unit.

**2.2.7 *galF* 基因和 *gnd* 基因的进化分析** 在细菌中,由于选择压力小,不易发生变异,所以持家基因一般都很稳定,更保守,不容易发生重组。因此用持家基因建立的系统发生树能更好的说明不同种之间的进化关系<sup>[20]</sup>。*galF* 和 *gnd* 基因是持家基因,在我们得到的 O-抗原基因簇中从 1 到 765bp 编码 GalF

(从第 44 个 aa 到最后一个氨基酸);O-抗原基因簇 3'端的 1339bp 编码 Gnd(从第 1 个 aa 到第 446 个 aa)。我们把这些序列和其他菌的 *galF* 和 *gnd* 基因进行了比较,用 Mega2 软件建立了系统发生树(图 2)。 *galF* 和 *gnd* 基因的进化树很清楚的表明大

肠杆菌属和沙门氏菌属形成两个分离的组,而志贺氏菌都落在大肠杆菌属中。这说明志贺氏菌属与大肠杆菌属的亲缘关系更近,揭示志贺氏菌属与大肠杆菌属在进化上属于同一菌属。

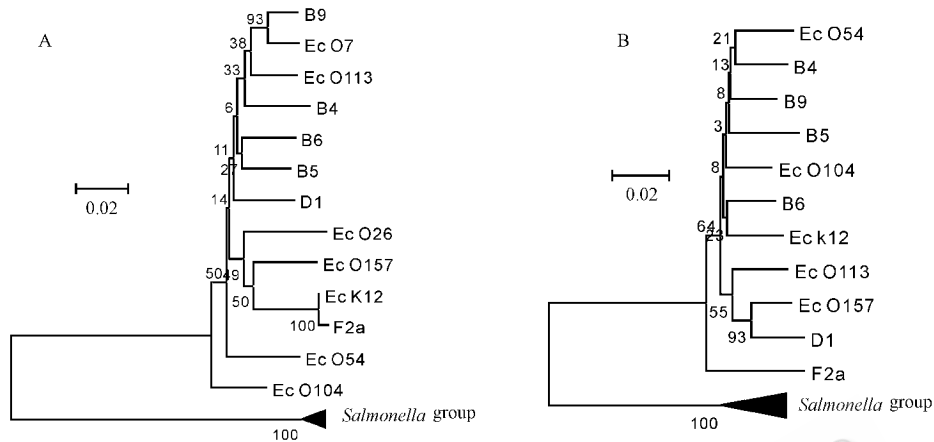


图 2 用 Mega2 软件产生的 *galF* 基因的进化树(A)和 *gnd* 基因的进化树(B)

Fig.2 Phylogenetic trees for the *galF* (A) and *gnd* (B) genes generated by Mega2 (neighbor-joining method)

The left fig is Phylogenetic tree for the *galF* gene, the right fig is Phylogenetic trees for the *gnd* gene. Sequences used included those from *E. coli* O54( *EcO54*, Patent No. 200310107159.5 ), *E. coli* O49( *EcO49*, unpublished data in our laboratory ), *E. coli* O36( *EcO36*, unpublished data in our laboratory ), *E. coli* O7( *EcO7*, GenBank accession No. AAC63611 and AAC27540 ), *E. coli* O150( *EcO150*, Patent No. 03109586.0 ), *E. coli* O107( *EcO107*, Patent No. 03109588.7 ), *E. coli* O26( *EcO26*, GenBank accession No. AAN60453 and AAN60466 ), *E. coli* K12( *Eck12*, GenBank accession No. AAC75103 and AAC75090 ), *S. boydii* 5( *Ecbd5*, GenBank accession No. AAL27321 and AAL27335 ), *S. boydii* 9( *Ecbd9*, GenBank accession No. AAL27346 and AAL27356 ), *S. boydii* 2( *Ecbd2*, Patent No. 200310107158.0 ), *S. dysenteriae* 1( *D1*, unpublished data in our laboratory ), *S. flexneri* 2( *F2a*, GenBank accession No. NP\_837664 and NP\_837649 ), *S. enterica*( GenBank accession No. AAB49385 and AAB49393 ). The scale bar indicates the 0.02 evolutionary distance unit.

2.3 大肠杆菌 O54 特异基因的鉴定

我们将大肠杆菌 O54 的 *wzx*、*wzy* 基因通过 Blast 在 GenBank 已有的数据中搜索,预测到它们的特异性高,因此大肠杆菌 O54 的 *wzx*、*wzy* 基因可用于对该菌的检测。我们根据它们设计 PCR 引物,在每个基因内设计两对引物,每对引物分布在相应基因内的不同地方以确保其检测的准确性。以含有大肠杆菌 O54 的第 9 组菌作为正对照,对所有 166 种大肠杆菌 43 株志贺氏菌进行 PCR 反应。每对引物除了在第 9 组中做 PCR 后得到了预期大小的正确的一条带外,在其他组包括去掉大肠杆菌 O54 的第 27 组中都没有扩增到任何 PCR 产物带,所以 *wzx*、*wzy* 基因对大肠杆菌 O54 是高度特异的。这两个基因内的 DNA 片段可用于快速准确地检测人体和环境中的大肠杆菌 O54。

由于一些多糖特殊的存在于致病菌中,以表面多糖为目标的血清学反应自上世纪 30 年代一直被用于对细菌的分型和血清学鉴定。但传统的血清学方法耗时长,灵敏度低,且血清不全,容易漏检。本

研究从合成 O-抗原的基因簇中筛选特异 DNA 分子,弥补了传统的血清学方法的不足。在这些基因内的特异的 DNA 片段对用现代的分子生物学方法(基因芯片技术或 PCR 方法)快速鉴定和控制致病菌有重要意义。

参 考 文 献

[ 1 ] Whitfield C. Biosynthesis of lipopolysaccharide O antigens. *Trends in Microbiology*, 1995, 3: 178 – 185.

[ 2 ] Reeves P R. Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology*, 1996, 4: 495 – 503.

[ 3 ] Curd H, Liu D, Reeves P R. Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. *J Bacteriol*, 1998, 180( 4 ):1002 – 1007.

[ 4 ] Sugiyama T, Kato N, Kato Y, et al. Evolutionary relationship among *rfb* gene clusters synthesizing mannose homopolymer as O-specific polysaccharides in *Escherichia coli* and *Klebsiella*. *Gene*, 1997, 198: 111 – 113.

[ 5 ] Garabal J I, Gonzalez E A, Vazquez F, et al. Serogroups of *Escherichia coli* isolated from piglets in Spain. *Vet Microbiol*, 1996, 48 ( 1 – 2 ): 113 – 123.

© 中国科学院微生物研究所期刊联合编辑部 <http://journals.im.ac.cn>

- [ 6 ] De Rycke J , Milon A , Oswald E , *et al.* Necrotoxic *Escherichia coli* ( NTEC ) : two emerging categories of human and animal pathogens . *Vet Res* , 1999 , **30** ( 2 - 3 ) : 221 - 233 .
- [ 7 ] Toth I , Oswald E , Mainil J G , *et al.* Characterization of intestinal cnf1 + *Escherichia coli* from weaned pigs . *Int J Med Microbiol* , 2000 , **290** ( 6 ) : 539 - 542 .
- [ 8 ] Bastin D A , Reeves P R . Sequence and analysis of the O antigen gene ( *rfb* ) cluster of *Escherichia coli* O111 . *Gene* , 1995 , **164** : 17 - 23 .
- [ 9 ] Feng L , Sof ' ya N Senchenkova , Yang J h , *et al.* Structural and genetic characterization of the *Shigella boydii* type 13 O antigen . *J Bacteriol* , 2004 , **186** ( 2 ) : 383 - 392 .
- [ 10 ] 徐建国 . 分子医学细菌学 . 北京 : 科学出版社 , 2000 .
- [ 11 ] Jocelyne M S , Wang L , Reeves P R . Sequence of the *Escherichia coli* O26 O antigen gene cluster and identification of O26 sepecific genes . *Gene* , 2002 , **297** ( 1 - 2 ) : 123 - 127 .
- [ 12 ] Wang L , Qu W J , Reeves P R . Sequence analysis of four *Shigella boydii* O-antigen loci : implication for *Escherichia coli* and *Shigella* relationships . *Infection and Immunity* , 2001 , **69** : 6923 - 6930 .
- [ 13 ] Li Q , Reeves P R . Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica* . *Microbiology* , 2000 , **146** : 2291 - 2307 .
- [ 14 ] Li Q , Matthew Hobbs , Reeves P R . The variation of dTDP-L-rhamnose pathway genes in *Vibrio cholerae* . *Microbiology* , 2003 , **149** : 2463 - 2474 .
- [ 15 ] Wang L , Reeves P R . Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes . *Infection and Immunity* , 1998 , **66** : 3545 - 3551 .
- [ 16 ] Wang L , Curd H , Qu W J , *et al.* Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-Specific genes . *Journal of Clinical Microbiology* , 1998 , **36** ( 11 ) : 3182 - 3187 .
- [ 17 ] Marolda R , Mavris M , Fallarino A , *et al.* Characterization of the *rfe* region of *Shigella flexneri* . *J Bacteriol* , 1994 , **176** ( 3 ) : 733 - 747 .
- [ 18 ] Reeves P R , Wang L . Genomic organization of LPS-specific loci . *Curr Top Microbiol Immunol* , 2002 , **264** ( 1 ) : 109 - 135 .
- [ 19 ] Pupo G M , Lan R , Reeves P R . Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics . *Proc Natl Acad Sci* , 2000 , **97** : 10567 - 10572 .
- [ 20 ] Lan R , Reeves P R . *Escherichia coli* in disguise : molecular origins of *Shigella* . *Microbes Infect* , 2002 , **4** ( 11 ) : 1125 - 1132 .

## Sequence and Evolutional Analysis of *Escherichia coli* O54 O-antigen Gene Cluster

YANG Jing-Hua YU Hao REN Yi WANG Lei FENG Lu \*

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

**Abstract** : The *Escherichia coli* O54 O-antigen gene cluster was sequenced . It was found to contain genes responsible for the biosynthesis of nucleotide sugar dTDP-rhamnose ( *rmlBDA* and *rmlC* ) , genes encoding O unit flippase ( *wzx* ) and O-antigen polymerase ( *wzy* ) and putative transferase . Unlike most of the cases in which *rml* genes are located together , *rmlC* was apart from other *rml* genes . Further analysis indicated that *rmlC* gene was recently introduced into the O-antigen gene cluster of *Escherichia coli* O54 . It was also found that *E. coli* O54 and *Shigella boydii* 9 are closely related . By polymerase chain reaction for the 166 *E. coli* and 43 *Shigella* O serogroups , *wzx* and *wzy* as specific genes for *E. coli* O54 were identified . This work provides the basis for a sensitive test for the rapid detection of *E. coli* O54 .

**Key words** : *Escherichia coli* O54 , O-antigen gene cluster , *rmlC* gene , Evolutionary , Specific gene

Foundation item : Chinese National Programs for High Technology Research and Development ( 2002AA2Z2051 ) ; Chinese National Natural Science Foundation ( 30270029 )

\* Corresponding author . Tel 86-22-66229592 ; Fax 86-22-66229596 ; E-mail : lfeng@tedamail.nankai.edu.cn

Received date : 12-04-2003