

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

郭宏杰¹ 冯 露¹ 张 淳¹ 徐建国^{2*} 王 磊^{1*}

(¹南开大学生命科学学院 天津 300071)

(²中国 CDC 传染病所 北京 102206)

摘 要 利用鸟枪法对大肠杆菌 O150 O-抗原基因簇进行测序,序列全长 13551bp,用生物信息学的方法进行序列分析,共发现 11 个基因,分别为鼠李糖合成酶基因(*rmlB*、*rmlD*、*rmlA*、*rmlC*)、糖基转移酶基因(3 个)、O-抗原转运酶基因(*wzx*)和 O-抗原聚合酶基因(*wzy*),另外还有两个基因功能未知。用 PCR 的方法筛选出了针对大肠杆菌 O150 的特异基因,可以用于基因芯片或 PCR 方法对大肠杆菌 O150 的快速检测。另外,通过进化分析发现大肠杆菌 O150 的 O-抗原基因簇中携带有典型的大肠杆菌鼠李糖合成酶基因,并且这些基因参与了 O-抗原基因簇间的重组以形成新的基因簇的过程。

关键词 大肠杆菌, O-抗原, 特异基因, 鼠李糖

中图分类号: Q93 文献标识码: A 文章编号: 0001-6209(2004)01-0034-07

大肠杆菌为革兰氏阴性菌,根据其 O-抗原分为 166 种不同的血清型,一部分 O-抗原特征存在于致病的大肠杆菌中。O-抗原是一类多糖抗原,存在于几乎所有革兰氏阴性菌的表面,由 3 到 6 个单糖组成,具有多样性,成为在 DNA 水平上研究分子进化的极好的材料。O-抗原对细菌的致病性和对环境的耐受力有非常重要的作用,而其多样性的作用主要在于逃逸动物免疫系统和噬菌体的识别。

革兰氏阴性菌 O-抗原的合成过程如下:细菌分步地由糖基转移酶将单糖转移到一个固定在细胞内膜的脂分子上,在内膜的内侧合成寡糖单位。O-抗原的寡糖单位通过转运酶(*Wzx*)而被转移到内膜外侧,然后通过聚合酶(*Wzy*)而聚合成多糖,再被联接到一个糖脂分子上(*LipidA/core*)形成脂多糖分子。编码负责合成 O-抗原的所有酶分子的基因一般都相邻存在,在染色体上形成一个基因簇^[1]。

本研究选择大肠杆菌 O150 为材料,大肠杆菌 O150 在 1972 年首先在患病的动物中被检测到^[2],被证实可以引起牛等牲畜的肠道感染,属于肠产毒性大肠杆菌(*Enterotoxigenic Escherichia coli*, ETEC),其潜在的爆发性流行的危险很大^[3],农业上急需一

个可以快速、准确检测大肠杆菌 O150 的方法。

本实验首先用长 PCR 扩增出大肠杆菌 O150 的 O-抗原基因簇,用鸟枪法进行测序,得到的序列用生物信息学方法进行分析,对 O-抗原基因簇中基因的功能进行了预测,并用 PCR 的方法筛选出了针对大肠杆菌 O150 细菌的特异基因,可用于基因芯片或 PCR 的方法对该菌的快速检测。

另外,本文还对鼠李糖合成酶基因进行了鉴定和进化分析。L-鼠李糖(6-脱氧-L-甘露糖)是许多药物的重要中间体,还可以合成香料或用作食品添加剂^[4],由于鼠李糖的应用前景越来越广,生产鼠李糖的研究备受关注。L-鼠李糖在很多革兰氏阴性菌的 O-抗原中存在,以 dTDP-鼠李糖前体的形式参与 O-抗原的合成,共有 4 个酶按顺序催化从 1-磷酸-葡萄糖合成 dTDP-鼠李糖,分别为 *RmlA*(1-磷酸-葡萄糖胸苷转移酶)、*RmlB*(dTDP-D-葡萄糖-4,6-脱氢酶)、*RmlC*(dTDP-6-脱氧-D-葡萄糖-3,5-变位酶)和 *RmlD*(dTDP-6-脱氧-L-甘露糖脱氢酶)。这 4 个鼠李糖合成酶基因在很多细菌的 O-抗原基因簇中存在,有很高的同源性,通过与已知功能的序列进行比较,可以鉴定大肠杆菌 O150 O-抗原基因簇中的鼠李糖合成

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 86-22-23503151; Fax 86-22-23506281; E-mail: lwang72@vip.sina.com, xujg@public.bta.net.cn

作者简介 郭宏杰(1973-),女,河北省乐亭县人,讲师,博士。主要从事致病菌检测的研究。E-mail: guohj@nankai.edu.cn

其他作者 李 玥

收稿日期 2003-03-31, 修回日期 2003-07-28

酶基因。同时,对这些合成酶基因进行进化上的分析有助于理解细菌 O-抗原基因簇的进化历史。

1 材料和方法

1.1 菌种

大肠杆菌 O150 标准菌株由澳大利亚悉尼大学 Reeves 教授提供,受体菌株(大肠杆菌 DH5 α)由本室保存。

1.2 载体质粒

pGEM-T-Easy 购自 Promega 公司。

1.3 酶和试剂

Expand Long Template PCR 酶购自 Boehringer Mannheim 公司,PCR 纯化试剂盒、T4 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工程技术有限公司。

1.4 基因组 DNA 的提取

采用 Bastin^[5]法提取细菌染色体 DNA。

1.5 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^[6]。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'-TAGTCGCGTGNGC-CTGGATTAAGTTCGC-3')。以大肠杆菌 O150 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。

1.6 O-抗原基因文库的构建

用 *DNase*I 消化 PCR 纯化产物,回收 1~3kb 之间的酶切片段,与 pGEM-T-Easy 载体在 T4 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 DH5 α ,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 孔平板大规模提质粒,用 *EcoR* I 酶切鉴定其中的插入片段的大小,挑选插入片段在 500bp 以上的克隆。

1.7 序列测定和拼接

利用双脱氧终止法,采用 T7-sp6 公用引物,用 ABI377 型 DNA 自动测序仪对克隆中的插入片段进行测序,由上海生物工程有限公司提供服务。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。

1.8 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 orffinder 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 Clustral W 软件做 DNA 序列间的精确比对,用 TMHMM(<http://www.cbs.dtu.dk/services/TMHMM/>)进行蛋白质跨膜片段的分析。

表 1 大肠杆菌 O150 的 O-抗原基因簇中的糖基转移酶基因和寡糖单位处理基因及其中的引物及 PCR 数据

Table 1 PCR testing of the specificity of *E. coli* O150 genes

Gene	Gene function	Base positions of genes	Base positions of forward primers	Base positions of reverse primers	Length of the PCR fragment/bp	Annealing temperature of PCR/°C
<i>wzx</i>	O-antigen flippase	5632 ~ 7134	5706 ~ 5722	6458 ~ 6475	770	50
			5719 ~ 5736	6681 ~ 6698	980	52
			5978 ~ 5995	6789 ~ 6806	829	54
<i>orf8</i>	Glycosyl transferase	8307 ~ 9197	8681 ~ 8699	9169 ~ 9185	505	48
			8418 ~ 8435	8995 ~ 9012	595	48
			8333 ~ 8350	9050 ~ 9067	735	48
<i>orf9</i>	Glycosyl transferase	9184 ~ 10098	9397 ~ 9413	9882 ~ 9899	503	50
			9337 ~ 9354	9899 ~ 9916	580	54
			9489 ~ 9506	10035 ~ 10051	563	54
<i>orf10</i>	Glycosyl transferase	10098 ~ 10994	10302 ~ 10319	10852 ~ 10869	568	50
			10293 ~ 10310	10954 ~ 10971	679	44
			10377 ~ 10394	10973 ~ 10989	613	48
<i>wzy</i>	O-antigen polymerase	11012 ~ 12109	11217 ~ 11234	11985 ~ 12002	786	52
			11225 ~ 11241	12035 ~ 12052	828	53
			11172 ~ 11189	11960 ~ 11977	806	54

1.9 特异基因的筛选

用 PCR 的方法筛选大肠杆菌 O150 的特异基因。针对大肠杆菌 O150 的 O-抗原基因簇中的 *wzx*、*wzy* 和糖基转移酶基因,在基因内部各设计 3 对引物(见表 1),以所有 166 株大肠杆菌和 43 株志贺氏

菌(见表 2)的基因组为模板进行 PCR。PCR 条件: 94℃ 2min, 94℃ 15s, 退火温度因引物的不同而不同(参照表 1),退火时间 50s, 72℃ 2min, 共进行 30 个循环; 72℃ 继续延伸 10min, 反应体系是 25 μ L。

表 2 用于筛选特异基因的 166 株大肠杆菌和 43 株志贺氏菌及它们的来源*

Table 2 Bacterial strains and PCR pools used for testing of *E. coli* O150

Pool No.	Strains of which chromosomal DNA included in the pool	Source
1	<i>E. coli</i> type strains for O serotypes O1 ,O2 ,O3 ,O4 ,O10 ,O16 ,O18 ,O39	IMVS ^a
2	<i>E. coli</i> type strains for O serotypes O40 ,O41 ,O48 ,O49 ,O71 ,O73 ,O88 ,O100	IMVS
3	<i>E. coli</i> type strains for O serotypes O102 ,O109 ,O119 ,O120 ,O121 ,O125 ,O126 ,O137	IMVS
4	<i>E. coli</i> type strains for O serotypes O138 ,O139 ,O149 ,O7 ,O5 ,O6 ,O11 ,O12	IMVS
5	<i>E. coli</i> type strains for O serotypes O13 ,O14 ,O15 ,O17 ,O19ab ,O20 ,O21 ,O22	IMVS
6	<i>E. coli</i> type strains for O serotypes O23 ,O24 ,O25 ,O26 ,O27 ,O28 ,O29 ,O30	IMVS
7	<i>E. coli</i> type strains for O serotypes O32 ,O33 ,O34 ,O35 ,O36 ,O37 ,O38 ,O42	IMVS
8	<i>E. coli</i> type strains for O serotypes O43 ,O44 ,O45 ,O46 ,O50 ,O51 ,O52 ,O53	IMVS
9	<i>E. coli</i> type strains for O serotypes O54 ,O55 ,O56 ,O57 ,O58 ,O59 ,O60 ,O61	IMVS
10	<i>E. coli</i> type strains for O serotypes O62 ,O63 ,O64 ,O65 ,O66 ,O68 ,O69 ,O70	IMVS
11	<i>E. coli</i> type strains for O serotypes O74 ,O75 ,O76 ,O77 ,O78 ,O79 ,O80 ,O81	IMVS
12	<i>E. coli</i> type strains for O serotypes O82 ,O83 ,O84 ,O85 ,O86 ,O87 ,O89 ,O90	IMVS
13	<i>E. coli</i> type strains for O serotypes O91 ,O92 ,O95 ,O96 ,O97 ,O98 ,O99 ,O101	IMVS
14	<i>E. coli</i> type strains for O serotypes O112 ,O162 ,O113 ,O114 ,O115 ,O116 ,O117 ,O118	IMVS
15	<i>E. coli</i> type strains for O serotypes O123 ,O165 ,O166 ,O167 ,O168 ,O169 ,O170 ,O171	b
16	<i>E. coli</i> type strains for O serotypes O172 ,O173 ,O127 ,O128 ,O129 ,O130 ,O131 ,O132 ,	c
17	<i>E. coli</i> type strains for O serotypes O133 ,O134 ,O135 ,O136 ,O140 ,O141 ,O142 ,O143	IMVS
18	<i>E. coli</i> type strains for O serotypes O144 ,O145 ,O146 ,O147 ,O148 ,O150 ,O151 ,O152	IMVS
19	<i>E. coli</i> type strains for O serotypes O153 ,O154 ,O155 ,O156 ,O157 ,O158 ,O159 ,	IMVS
20	<i>E. coli</i> type strains for O serotypes O160 ,O161 ,O163 ,O8 ,O9 ,O124 ,O111	IMVS
21	<i>E. coli</i> type strains for O serotypes O103 ,O104 ,O105 ,O106 ,O107 ,O108 ,O110	IMVS
22	<i>Shigella Boydii</i> strains B4 ,B5 ,B6 ,B8 ,B9 ,B11 ,B12 ,B14	d
23	<i>Shigella Boydii</i> strains B1 ,B3 ,B7 ,B8 ,B10 ,B13 ,B15 ,B16 ,B17 ,B18	d
24	<i>Shigella Dysenteriae</i> strains D1 ,D2 ,D3 ,D4 ,D5 ,D6 ,D7 ,D8	d
25	<i>Shigella Dysenteriae</i> strains D9 ,D10 ,D11 ,D12 ,D13	d
26	<i>Shigella Flexneri</i> strains F6a ,F1a ,F1b ,F2a ,F2b ,F3 ,F4a ,F4b ,F5 (v 7) F5 (v 4)	d
27	<i>Shigella Sonnei</i> strains D5 , DR	d

* A total of 27 pools of DNA were made ,each containing 8 to10 strains.

a. Institute of Medical and Veterinary Science ,Anelaide ,Australia ;

b. O123 from IMVS ,others from Statens Serum Institut ,Copenhagen ,Denmark ;

c. O172 and O173 from Statens Serum Institut , Copenhagen ,Denmark , others from IMVS ;

d. Institute of Communicable Disease Prevention and Control , Chinese Center for Disease Prevention and Control.

2 结果和分析

2.1 O-抗原基因簇的获得

长 PCR 产物用 0.8% 的琼脂糖凝胶电泳检测,在 10~15kb 的位置出现一条电泳带。为了保证 O-抗原基因簇的序列质量,对大肠杆菌 O150 的基因组共作 6 个长 PCR 反应,然后混合这些产物产生文库,以减小 PCR 反应所带来的误差。由于长 PCR 对模板的完整性要求很高,所以在提取基因组时一定要

注意避免任何可引起基因组 DNA 断裂的操作。

2.2 O-抗原基因簇的测序

本研究采用“鸟枪法”对大肠杆菌 O150 的 O-抗原进行测序,首先对 PCR 纯化产物进行 DNaseI 酶切,提取大小在 1~3kb 之间的片段,与载体进行连接,取 1/10 的连接产物电转化 DH5 α ,共得到 500 个左右的白色克隆。

根据 $P = e^{-m}$ (P :未测定碱基的概率; m :测定序列的覆盖率)^[7],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,我们首先需要对 80 个 500bp 以上的克隆测序以达到 90% 的碱基测定概率。以碱裂解法用 96 孔平板大规模提取质粒,用 *EcoR* I 酶切鉴定其中的插入片段的大小,大于 500bp 插入片段为 50% 以上,说明所构建的大肠杆菌 O150 的 O-抗原的基因文库是成功的。挑选插入片段在 500bp 以上的 80 个克隆首先以 T7 为引物进行单向测序,对于余下的 10% 的序列,再以 *sp6* 为引物测定部分插入片段的反向序列,将所有所得序列用 Pregap4 和 Gap4 软件拼接和编辑,从而得到大肠杆菌 O150 的 O-抗原基因簇的全序列,序列全长 13551bp。

2.3 O-抗原基因簇的序列分析

O-抗原基因簇一般含有 3 种基因:A 单糖合成

基因、B 糖基转移酶基因、C 寡糖单位处理基因。A 类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。C 类基因编码的蛋白由于有典型的 10~14 个跨膜片段,也可以用生物信息学方法鉴定。对于 B 类基因,由于编码此类酶的基因之间的同源性很小(这与其它大多数酶不同)根据与其它已知功能基因的比较,只能判断一个未知基因是否属于这类基因,但不能判断其确切转移哪一个单糖。对大肠杆菌 O150 的 O-抗原基因簇,用 orffinder 发现基因,找到 11 个开放的阅读框,所有开放阅读框的 5'→3' 方向为从 *galF* 基因到 *gnd* 基因。(*galF* 和 *gnd* 不参与 O-抗原的合成,只是用来扩增 O-抗原基因簇^[8,9]),用 blast 系列软件与 GenBank 中的基因比较结果见表 3。

表 3 大肠杆菌 O150 O-抗原基因簇的基因功能预测

Table 3 Summary of ORFs in *E. coli* O150

Gene	Location in sequence	No. of aa	(G + C) % content	Similar proteins (reference)	Identical (%) / Similar (%) (aa overlap)	Putative function
<i>rmlB</i>	1138 ~ 2223	361	43.37	RmlB (<i>Shigella boydii</i>)	98/99(361)	dTDP-D-glucose 4- β -dehydratase
<i>rmlD</i>	2223 ~ 3122	299	47.00	RmlD (<i>Shigella boydii</i>)	95/97(299)	TDP-6-deoxy-L-manno Se-dehydrogenase
<i>rmlA</i>	3180 ~ 4058	292	43.57	RmlA (<i>Shigella boydii</i>)	98/99(292)	Glucose-1-phosphate Thymidyltransferase
<i>rmlC</i>	4063 ~ 4614	183	38.58	RmlC (<i>Shigella boydii</i>)	87/89(172)	dTDP-6-deoxy-D-glucosyl Se-3- β epimerase
<i>orf5</i>	4616 ~ 5599	327	32.72			Unknow
<i>wzx</i>	5632 ~ 7134	500	29.20	Wzx (<i>Bacteroides fragilis</i>)	21/41(419)	O-antigen flippase
<i>orf7</i>	7137 ~ 8282	381	31.76			Unknow
<i>orf8</i>	8307 ~ 9197	296	32.88	Glycosyltransferase (<i>Clostridium acetobutylicum</i>)	27/51(126)	Glycosyltransferase
<i>orf9</i>	9184 ~ 10098	304	31.80	Glycosyltransferase (<i>Shigella flexneri</i> 2a)	47/63(285)	Glycosyltransferase
<i>orf10</i>	10098 ~ 10994	298	32.21	Glycosyltransferase (<i>Shigella flexneri</i> 2a)	48/67(298)	Glycosyltransferase
<i>wzy</i>	11012 ~ 12109	365	30.96	Wzy (<i>Shigella boydii</i>)	25/47(317)	O-antigen polymerase

2.3.1 单糖合成基因:组成 O-抗原的糖通常为中性糖、氨基糖,有时为罕见糖如 6-脱氧己糖、3,6-双脱氧己糖等,所有的单糖均需以核苷酸二磷酸(NDP)-单糖前体的形式参与合成 O-抗原。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中,一般不在 O-抗原基因簇里重复出现。大肠杆菌 O150 的 O-抗原的结构未知,但通过进行 blast 比较, *orf1*, *orf2*, *orf3* 和

orf4 分别与合成鼠李糖前体的 dTDP-鼠李糖的 4 个合成酶基因有非常高的同源性,其中, *orf1* 与鲍氏志贺氏菌的 RmlB 在 361 个氨基酸中有 98% 的相同性,说明它们有极高的同源性,可以肯定 *orf1* 基因为 *rmlB* 基因,命名为 *rmlB*; *orf2* 与鲍氏志贺氏菌的 RmlD 在 299 个氨基酸中有 95% 的相同性,说明它们有极高的同源性,可以肯定 *orf2* 基因为 *rmlD* 基因,命名为 *rmlD*; *orf3* 与鲍氏志贺氏菌的 RmlA 在 292

个氨基酸中有 98% 的相同性,说明它们有极高的同源性,可以肯定 *orf3* 基因为 *rmlA* 基因,命名 *rmlA*; *orf4* 与鲍氏志贺氏菌的 RmlC 在 172 个氨基酸中有 87% 的相同性,说明它们有极高的同源性,可以肯定 *orf4* 基因为 *rmlC* 基因,命名为 *rmlC*; 大肠杆菌 O150 的结构未知,但从以上结果可以肯定大肠杆菌 O150 的 O-抗原结构中有鼠李糖存在。在已知序列的含有鼠李糖的 O-抗原基因簇中,4 个合成酶基因一般以 *rmlB*、*rmlD*、*rmlA*、*rmlC* 的顺序在 O-抗原基因簇中呈一个簇排列,在大肠杆菌 O150 中也为同样的顺序。

2.3.2 糖基转移酶基因: *orf8* 通过进行保守区域比较,与糖基转移酶家族 2 相似(pfam00535, 3e-07),进行 blast 比较,与丙酮丁醇梭菌的糖基转移酶有 27% 的相同性,51% 的相似性,说明它们之间有高度的同源性,可以推测 *orf8* 也为糖基转移酶基因,暂命名为 *orf8*。*orf9* 与福氏志贺氏菌 2a 的 *rfbF* 基因在 285 个氨基酸中有 47% 的相同性,63% 的相似性,*rfbF* 基因编码 dTDP-鼠李糖的糖基转移酶基因,所以推测 *orf9* 是一个糖基转移酶基因,暂命名为 *orf9*。*orf10* 与福氏志贺氏菌 2a 的 *rfbG* 基因在 289 个氨基酸中有 48% 的相同性,67% 的相似性,*rfbG* 基因编码 dTDP-鼠李糖的糖基转移酶基因,所以推测 *orf10* 是一个糖基转移酶基因,暂命名为 *orf10*。

2.3.3 寡糖单位处理基因:通过 HMMTOP 分析,*orf6* 和 *orf11* 是仅有的两个编码蛋白具有 10 个以上跨膜片段的基因。*orf6* 编码的蛋白通过 HMMTOP 分析发现含有 14 个潜在跨膜片段,通过 Eisenberg 等^[10]的算法也发现 *orf6* 有 14 个潜在的穿膜区,而且与许多 Wzx 蛋白相似,例如,和热球菌的 Wzx 在 419 个氨基酸中有 21% 的相同性,41% 的相似性,所以可以推测 *orf6* 很可能是 *wzx* 基因,命名为 *wzx*。*orf11* 编码的蛋白通过 HMMTOP 分析发现含有 10 个潜在跨膜片段,它的内膜的拓扑结构具有众所周知的 O-抗原聚合酶(Wzy)的典型特征,此外,它与鲍氏志贺氏菌编码 O-抗原聚合酶的 Wzy 在 317 个氨基酸中有 25% 的相同性,47% 的相似性,说明它们之间有高度的同源性,所以命名 *orf11* 为 *wzy* 基因。

2.3.4 未知功能基因:*orf5* 和 *orf7* 通过与已知序列比较,同源性比较低,无法推测其可能的功能。

2.3.5 O-抗原基因簇的(G + C)%:大肠杆菌 O150 的 O-抗原基因簇中(G + C)% 平均为 35.58%,而基因组中其它位置(G + C)% 一般为 50%。目前普遍认为 GC 差异说明 DNA 片段的外源性,并随着在所

在种属内的时间的增长而进化到与其它片段相似的 GC 比,这一特征说明 O-抗原起源于低 GC 含量的细菌中,近期才通过横向转移进入他们现在所在的细菌中^[11]。

2.3.6 鼠李糖合成酶基因的进化分析:从 GenBank 中提取已知功能的鼠李糖合成酶基因的序列,用 ClustalW 进行分析(见图 1)。由图可以看出,大肠杆菌 O150 具有典型的大肠杆菌鼠李糖合成酶基因的特点,并且大肠杆菌和志贺氏菌的亲缘关系非常近,尤其在 *rmlB* 和 *rmlD* 中,二者的差异低于 5%。志贺氏菌属细菌共有 33 种不同的 O 抗原,Reeves 教授等^[12]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,其它的研究工作包括多位点酶切电泳(Multilocus enzyme electrophoresis, MLEE)、核糖体分型(Ribotyping)和对 8 个持家基因进行的分析都表明志贺氏菌属细菌属于大肠杆菌属^[13]。我们对鼠李糖合成酶基因的进化分析也支持了上述论断。另外,选择了 4 株沙门氏菌 O-抗原基因簇中的鼠李糖合成酶基因进行了研究,*rmlB* 中大肠杆菌与沙门氏菌的差异平均为 20%,二者用持家基因分析得出的 15% 的差异略高一些。这两个属的细菌在 14000 万年前分化,沙门氏菌有 47 种 O 抗原,其中 3 种与大肠杆菌相同。说明它们分别在过去的 14000 万年中形成了大约 164 种和 43 种新的抗原。另外,通过对很多大肠杆菌和沙门氏菌 O-抗原基因簇进行序列分析和比较,发现不同个体的基因簇之间的重组和转座子引起的重组为其进化的主要途径,本研究通过对鼠李糖合成酶基因的进化分析也可以看出,如果这 4 个合成酶基因是从同一祖先进化而来,没有发生过重组,那么 4 个基因的进化树应该有相似的拓扑结构,但发现其 5' 端基因(包括 *rmlB*、*rmlD* 和 *rmlA*)保守性比较强,具有种属特异性,而 3' 端基因(*rmlC*)具有更大的多变性,具有 O-抗原的特异性,尤其在大肠杆菌中(包括志贺氏菌)更为明显。同时,对 *rmlB*、*rmlD*、*rmlA* 和 *rmlC* 基因进行(G + C)% 的分析,平均值分别为 43%、48%、45% 和 35%,*rmlC* 基因比其它 3 个基因的(G + C)% 明显偏低,也说明这 4 个基因具有不同的起源,可以推测 *rmlB*、*rmlD* 和 *rmlA* 基因在大肠杆菌中已经存在很长时间了,而 *rmlC* 基因以及其后的 O-抗原特异性基因是在鼠李糖合成酶基因介导的重组中转移到它现在所在的细菌中而形成新的抗原基因簇。

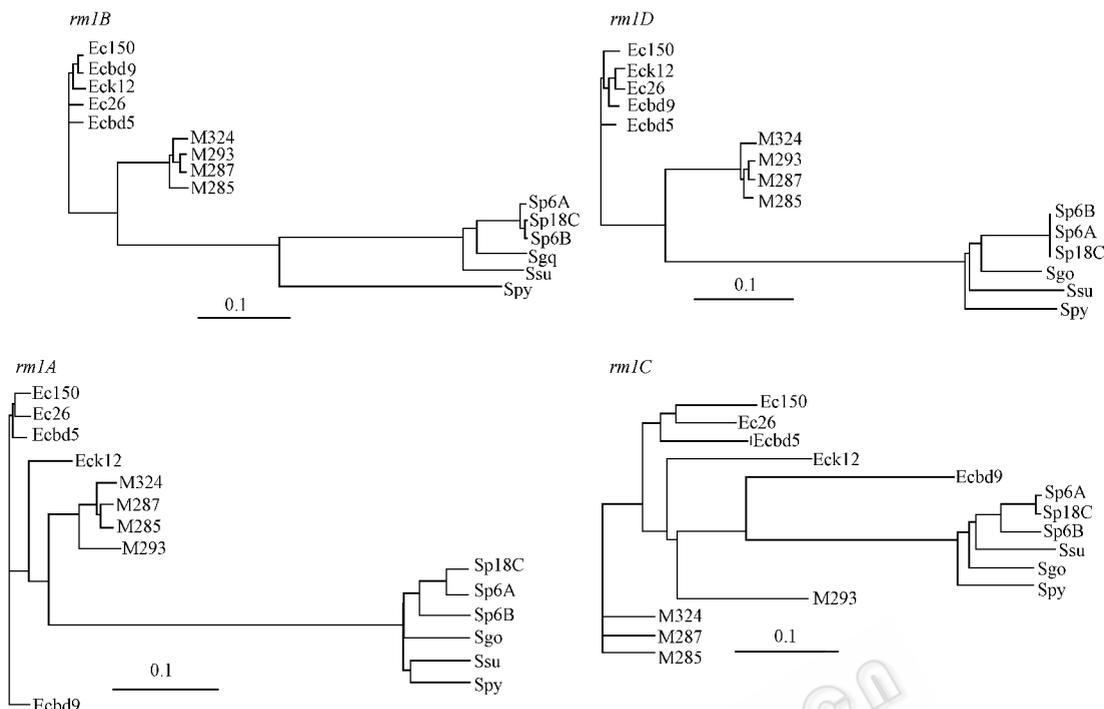


图 1 用 Clustal W 产生的鼠李糖合成酶基因 (*rmlB*, *rmlD*, *rmlA*, 和 *rmlC*) 的进化树

Fig. 1 Phylogenetic trees for the *rmlB*, *rmlD*, *rmlA*, and *rmlC* genes generated by Clustal W (neighbor-joining method)

Sequences used included those from three *E. coli* strains [*E. coli* O150 (*Ec*150), *E. coli* O26 (*Ec*26)], *E. coli* K12 (*Eck*12)], two *S. boydii* strains (*Ec*bd5 and *Ec*bd9)], four *S. enterica* strains (*M*324, *M*293, *M*287, *M*285) and six *Streptococcus* strains [*S. pneumoniae* 6A, 6B, and 18C (*Sp*6A, *Sp*6B, *Sp*18C)]; *S. gordonii* (*Sgo*), *S. suis* (*Ssu*), and *S. pyogenes* (*Sp*y)]. The scale bar indicates the 0.1 evolutionary distance unit.

2.4 大肠杆菌 O150 特异基因的鉴定

最近研究发现, O-抗原基因簇中的 B 和 C 类基因为每株菌的特异基因, 可用于对该菌的检测^[14]。根据这一理论, 设计针对 B 和 C 组基因的 PCR 引物, 以含有大肠杆菌 O150 的一组菌作为正对照, 对所有 166 种大肠杆菌 A3 株志贺氏菌进行 PCR 反应。不能产生正反应产物的基因将极有可能是大肠杆菌 O150 的特异基因。

本研究针对大肠杆菌 O150 的 O-抗原基因簇中的 *wzx*、*wzy*、*orf8*、*orf9*、*orf10* 基因设计引物, 在每个基因内, 各设计了 3 对引物, 每对引物分布在相应基因内的不同地方以确保其特异性。每对引物除了在第 18 组中做 PCR 后得到了预期大小的一条带外, 在其他组中都没有扩增到任何大小正确的带, 所以 *wzx*、*wzy*、*orf8*、*orf9*、*orf10* 基因对大肠杆菌 O150 是高度特异的。可用于快速准确地检测人体和环境中的大肠杆菌 O150。

3 讨论

大肠杆菌是人类和其它脊椎动物中营共生生活

的肠道寄生菌, 但在身体衰弱或免疫力低下的宿主中, 原本正常的非致病性大肠杆菌也会引起宿主的感染^[15]。另外还有很多大肠杆菌的致病性极强, 致病性大肠杆菌按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 (ETEC)、肠出血性大肠杆菌 (Enterohaemorrhagic *E. coli*, EHEC) 和肠聚集性大肠杆菌 (Enteraggative *E. coli*, EaggEC) 等。在发展中国家 5 岁以下的儿童中, 估计每年腹泻人数 15 亿, 死亡人群约 300 万, 这些地区的儿童社区研究中, ETEC 是最常见的肠道致病菌。由于致病大肠杆菌的主要携带体为牛、猪等牲畜, 我国人口众多, 各地区发展不平衡, 一些地区的卫生条件较差, 人畜交叉感染机会很多, 这就为大肠杆菌大规模爆发流行提供了基础。另外, 一些农业国家 (如澳大利亚和加拿大) 也急需快速鉴定手段以保证他们的出口牲畜及产品不含致病菌。一些大的食品公司也急需快速鉴定方法以确保食品的质量。

目前传统的大肠杆菌鉴定方法仍采取血清学鉴定,但这个方法的缺点在于需要培养单菌落,耗时长、灵敏度低且容易漏检。很多科学家都在探索用分子生物学的方法进行这项工作,但都没有发现合适的 DNA 分子。本研究从合成表面多糖抗原的基因簇中筛选特异 DNA 片段用于基于基因芯片或 PCR 方法对大肠杆菌的快速检测,既继承了传统,又弥补了传统的血清学方法的不足。

参 考 文 献

- [1] Whitfield C. Biosynthesis of lipopolysaccharide O antigens. *Trends in Microbiology*, 1995, **3**: 178 - 185.
- [2] Furowicz A J, Orskov F. Two new *Escherichia coli* O antigens, O150 and O157, and one new K antigen, K92, in strains isolated from veterinary diseases. *Acta Pathol Microbiol Scand Microbiol Immunol*, 1972, **80**(3): 441 - 444.
- [3] Danbara H, Komase K, Arita H, et al. Molecular analysis of enterotoxin plasmids of enterotoxigenic *Escherichia coli* of 14 different O serotypes. *Infect Immun*, 1988, **56**(6): 1513 - 1517.
- [4] 浦跃武, 刘卫斌, 陈志明. 鼠李糖的研究现状及其应用. *食品工业科技*, 2002, **2**: 84 - 85.
- [5] 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.
- [6] Reeves P R. Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology*, 1996, **4**: 495 - 503.
- [7] 徐建国. 分子医学细菌学. 北京: 科学出版社, 2000.
- [8] Marolda C L, Valvano M A. The GalF protein of *Escherichia coli* is not a UDP-glucose pyrophosphorylase but interacts with the GalU protein possibly to regulate cellular levels of UDP-glucose. *Mol Microbiol*, 1996, **22**: 827 - 840.
- [9] Stevenson G, Neal B, Liu D, et al. Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J Bacteriol*, 1994, **176**(13): 4144 - 4156.
- [10] Eisenberg D, Schwarz E. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol*, 1984, **179**: 125 - 142.
- [11] Lan R, Reeves P R. Gene transfer is a major factor in bacterial evolution. *Molecular Biology and Evolution*, 1996, **13**: 47 - 55.
- [12] Reeves P R. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *PNAS*, 2000, **97**: 10567 - 10572.
- [13] Lan R, Reeves P R. *Escherichia coli* in disguise: molecular origins of *Shigella*. *Molecular and Infection*, 2002, **4**: 1125 - 1132.
- [14] 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.
- [15] Nataro J P, Kaper J B. Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, 1998, **11**: 142 - 201.

Sequence of *Escherichia coli* O150 O-antigen Gene Cluster and Identification of Rhamnose Pathway Genes

GUO Hong-Jie¹ FENG Lu¹ ZHANG Chun¹ XU Jian-Guo^{2*} WANG Lei^{1*}

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

(² Institute of Communicable Disease Prevention and Control, Chinese Center for Disease Prevention and Control, Priority Laboratory for Molecular Medical Bacteriology, Ministry of Health, P O BOX 5, Changping, Beijing 102206, China)

Abstract: The *Escherichia coli* O150 O-antigen gene cluster was sequenced. It contains the genes for biosynthesis of nucleotide sugars dTDP-rhamnose, as well as genes for O unit flippase(*wzx*), O-antigen polymerase(*wzy*) and potential transferase genes. By polymerase chain reaction against representative stains for the 166 *E. coli* and 43 *Shigella* O serogroups, we identified 5 genes that are highly specific to *E. coli* O150. This work provides the basis for a sensitive test for the rapid detection of *E. coli* O150. Phylogenetic trees for the *rmlB*, *rmlD*, *rmlA*, and *rmlC* genes was generated, and we found that these genes are typical *E. coli* genes and may have been involved in recombination events between O-antigen gene clusters.

Key words: *Escherichia coli*, O-antigen, Specific gene, Rhamnose

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

* Corresponding author. Tel 86-22-23503151 Fax 86-22-23506281 E-mail: jwang72@vip.sina.com, xujg@public.bta.net.cn

Received date: 03-31-2003