

鼠疫耶尔森菌 rRNA 基因识别特征研究

张慧娟 魏建春 张恩民 申小娜 梁莹 朱晓宇 徐冬蕾 孙丽娜
张建华 俞东征 海荣

【摘要】 目的 分析鼠疫耶尔森菌(鼠疫菌)的 rRNA 基因识别特征。方法 通过比较基因组学方法,利用 Clustal W 软件,对目前已完成全基因组测序的 9 株鼠疫菌的 rRNA 基因序列进行分析,分别比对 rRNA 基因簇两侧的 2000 bp 序列、基因簇内 16S、23S、5S rRNA 及 16S~23S 基因间隔区序列,以确定鼠疫菌 rRNA 基因的识别特征。结果 菌株 D182038、D106004、Z176003 和 CO92 分别有 6 个 rRNA 基因簇拷贝(6 拷贝菌株),菌株 91001、KIM、Nepal516、Antiqua 和 Pestoides F 分别有 7 个 rRNA 基因簇拷贝(7 拷贝菌株)。按照两侧 2000 bp 序列不同,可将鼠疫菌 rRNA 基因簇分成 13 种类型,并可将 6 拷贝与 7 拷贝菌株进行区分;16S~23S rRNA 基因间隔区含有 4 种不同种类和数量的 tRNA 基因,可将 6 拷贝菌株和 7 拷贝菌株分别进行区分;23S rRNA 基因在不同菌株间及不同 rRNA 拷贝间的碱基突变数方差分析显示,各菌株间 $F=0.548$, $P=0.815>0.05$;各拷贝间 $F=5.228$, $P<0.01$ 。结论 鼠疫菌 rRNA 基因簇两侧序列、间隔区 tRNA 基因和 23S rRNA 基因可作为识别不同菌株和不同 rRNA 基因簇拷贝的指标,将鼠疫菌不同菌株进行分类。

【关键词】 鼠疫耶尔森菌; rRNA 基因; 识别特征

Study on the identification characteristics of rRNA genes on *Yersinia pestis* ZHANG Hui-juan, WEI Jian-chun, ZHANG En-min, SHEN Xiao-na, LIANG Ying, ZHU Xiao-yu, XU Dong-lei, SUN Li-na, ZHANG Jian-hua, YU Dong-zheng, HAI Rong. State Key Laboratory for Infectious Disease Prevention and Control, National Institute of Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Corresponding author: YU Dong-zheng, Email: yudongzheng@icdc.cn; HAI Rong, Email: hairong@icdc.cn

This work was supported by grants from the National Science and Technology Support Projects for the "Tenth Five-Year Plan" of China (No. 2004BA718B07), Major Projects from Department of Science and Technology of China (No. 2008ZX10004-008) and Special Fund from Ministry of Health of China (No. 200802016).

【Abstract】 Objective To study the identification characteristics of rRNA genes on *Yersinia (Y.) pestis*. **Methods** By means of comparative genomics, we compared the rRNA genome sequences of nine completely sequenced strains of *Y. pestis* isolated from China and other countries by Clustal W software. We also compared the 2000 bp sequence adjacent to the rRNA genes, rRNA genes and 16S-23S rRNA spacer region respectively to determine the identification features of rRNA genes for *Y. pestis*. **Results** There were 6 rRNA gene clusters in the strains of D182038, D106004, Z176003 and CO92 respectively (6 copies strain). There were 7 rRNA gene clusters in the strains of 91001, KIM, Nepal516, Antiqua and Pestoides F (7 copies strain). According to the 2000 bp sequence, 13 types of rRNA gene clusters could classify the strains between the 6 copies and 7 copies. There were 4 types of tRNA gene among the 16S-23S rRNA spacer region that could classify the strains among the 6 copies and 7 copies strains respectively. The number of point mutation among the 23S rRNA gene was statistically different in some copies under ANOVA analysis ($F=0.548$, $P=0.815>0.05$ among the strains and $F=5.228$, $P<0.01$ among the copies). **Conclusion** The 2000 bp sequence adjacent to the rRNA genes, tRNA gene and 23S rRNA gene sequence could serve as the identification sign of rRNA genes for classifying the strains of *Y. pestis*.

【Key words】 *Yersinia pestis*; rRNA genes; Identification characteristics

DOI: 10.3760/cma.j.issn.0254-6450.2011.04.010

基金项目:国家“十五”科技攻关项目(2004BA718B07); 国家科技部重大专项(2008ZX10004-008); 卫生部专项经费资助项目(200802016)

作者单位:102206 北京, 中国疾病预防控制中心传染病预防控制所 传染病预防控制国家重点实验室

通信作者:俞东征, Email: yudongzheng@icdc.cn; 海荣, Email: hairong@icdc.cn

rRNA 基因在细菌分类鉴定及基因微进化研究中有着重要作用,多株鼠疫耶尔森菌(鼠疫菌)全基因组序列测定的完成为研究不同菌株 rRNA 基因特点提供了资料,通过比较基因组学方法,分析 rRNA 基因在不同来源鼠疫菌之间的差异分布,可以揭示鼠疫菌的 rRNA 基因结构特征。已有研究显示,鼠疫菌中存在不同数目的 rRNA 基因簇拷贝^[1,2],常见 6 拷贝菌株和 7 拷贝菌株,rRNA 基因指纹图谱分析发现中国普遍存在 6 拷贝菌株,并且有明显的疫源地差异^[3]。通过 rRNA 相关基因序列比对可能找到识别某一菌株或某一 rRNA 拷贝的基因特征。本研究选取 rRNA 基因本身及其两侧的 2000 bp 序列进行分析。

材料与方法

1. 菌株:本实验室新近测序的 3 株鼠疫菌玉龙分离菌株 D106004、剑川滇西纵谷生态型菌株 D182038、青藏高原 1 型菌株 Z176003 和目前 GenBank 中已完成全基因组序列测定的 6 株鼠疫菌株(CO92、91001、KIM、Antiqua、Nepal516 和 Pestoides F)。

2. 序列比对及分析:用 Clustal W 软件对上述已完成全基因组测序的 9 株鼠疫菌序列进行多序列比对。分别比对 rRNA 基因簇两侧的 2000 bp 序列、基因簇内部各部分序列,包括 16S、23S、5S rRNA 及 16S~23S、23S~5S 基因间隔区。序列一致率>90% 为同源,以基因序列的同源性和基因间隔区的 tRNA

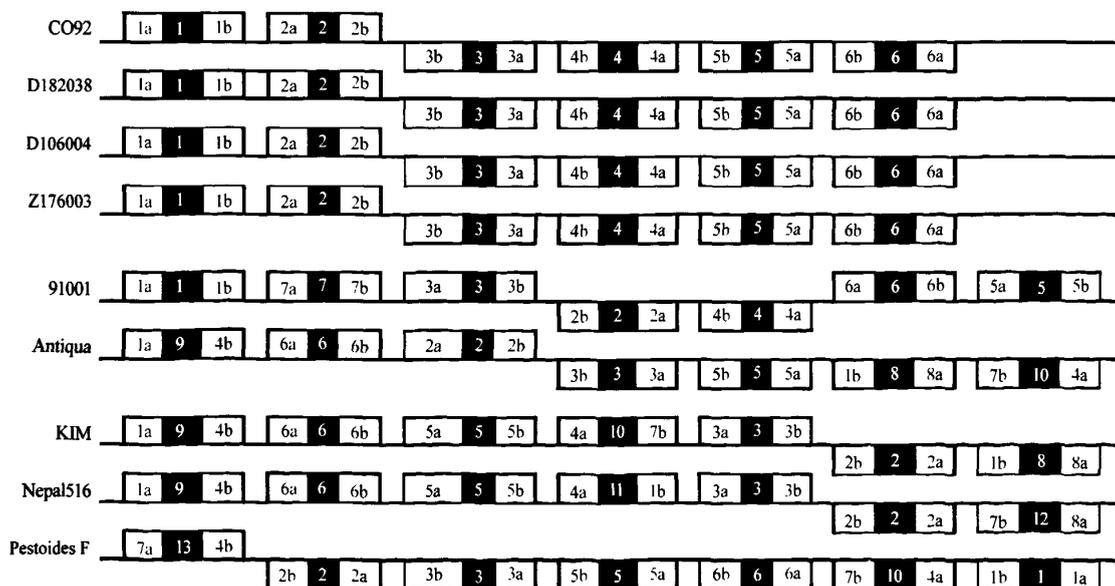
基因类型和数目差异作为识别不同菌株或不同 rRNA 拷贝菌株的指标;对于单碱基突变的基因序列,统计各菌株和各拷贝间的突变数,进行方差分析。

3. rRNA 基因簇各拷贝的确定:为便于分析和比较,以 CO92 菌株 rRNA 基因簇在全基因组中的位置顺序为参照,将 rRNA 基因簇拷贝依次编号为 1、2、3、4、5、6,每个拷贝两侧的 2000 bp 序列分别用 a、b 表示,靠近 16S rRNA 的 2000 bp 序列命名为 a,靠近 5S rRNA 的 2000 bp 序列命名为 b,如第 1 个拷贝 16S 端即为 1a,5S 端即为 1b。各菌株 rRNA 基因簇两侧 a、b 序列进行比对后,若 a、b 序列一致,则认为是同一个 rRNA 基因簇,即按 CO92 菌株中的 rRNA 编号命名;若出现新的不同的 a 或 b 序列,则依次编号,如 91001 菌株中出现的第 7 拷贝,Antiqua 菌株中出现的第 8、9、10 拷贝等。

结 果

1. rRNA 基因簇两侧序列:按照基因簇两侧的 2000 bp 序列,可将 9 株菌中的 rRNA 基因簇分为 13 种类型,在不同菌株中的位置和方向不同(图 1)。

(1) rRNA 基因簇拷贝在菌株中的位置:菌株 D182038、D106004、Z176003 和 CO92 分别有 6 个 rRNA 基因簇,菌株 91001、KIM、Nepal516、Antiqua 和 Pestoides F 分别有 7 个 rRNA 基因簇。6 拷贝菌株的 rRNA 基因簇位置相同,其他菌株与 CO92 相比,各拷贝在位置上发生了移位或基因反转(与互补链



注:黑色为 rRNA 基因簇;白色为基因簇两侧的 2000 bp 序列;横线下方的基因是由互补链编码

图 1 9 株鼠疫菌的 rRNA 基因簇在全基因组中的位置、方向及两侧 2000 bp 序列比较

的序列同源),并出现了新的 rRNA 基因簇拷贝, KIM、Nepal516 和 Antiqua 中增加了 8a。所有菌株第 2、3、5、6 rRNA 基因簇拷贝两侧 2000 bp 序列在各菌株中保持不变,其余拷贝两侧的 2000 bp 序列由于基因组重排发生了移位,表现为第 1、4、7、8 拷贝的 a 和 b 序列相互重排。

(2)保守序列:a、b 序列两两比对比显示, rRNA 基因两侧有一部分序列是保持不变的。表现为所有 a 序列中紧邻 16S rRNA 基因序列的 237 bp 序列同源(92%),为 rRNA 基因簇的保守序列;b 序列中,2b 和 6b 紧邻 5S rRNA 基因序列的 207 bp 同源(92%),3b 和 7b 紧邻 5S rRNA 基因的 207 bp 同源(92%),1b、3b、5b 和 7b 紧邻 5S rRNA 基因的 75 bp 同源(97%)。

(3)插入序列:将第 1、4、7 拷贝的 a、b 序列与全基因组进行 Blast 比对,结果显示 1b、7b、8a 中均有插入序列存在,而第 4 拷贝周围没有插入序列。表现为 1b 中紧邻 5S rRNA 基因约 430 bp 处有 IS285,长度 1314 bp;7b 中紧邻 5S rRNA 基因约 980 bp 处有 IS1617,长度 611 bp;8a 中紧邻 16S rRNA 基因约 570 bp 处有 IS100,长度 1949 bp。

(4)7 拷贝菌株 a、b 序列特异性:根据 rRNA 基因两侧的 a、b 序列组合独特性进行分析,7 拷贝菌株中,91001 的第 7 拷贝、Nepal516 的第 11、12 拷贝和 Pestoides F 的第 13 拷贝分别是各菌株的惟一 rRNA 基因;而 Antiqua 和 KIM 中的第 8、9 拷贝为该 2 株菌共有。以上序列特征显示了 7 拷贝与 6 拷贝菌株的差异,但不能将 Antiqua 和 KIM 区分开。

2. 基因间隔区内的 tRNA 基因:鼠疫菌 rRNA 基因簇内存在 16S ~ 23S rRNA 和 23S ~ 5S rRNA 基因间隔区,后者较短,在一些菌株中不存在,而 16S ~ 23S rRNA 基因间隔区内夹杂着不同种类和数量的 tRNA 基因,可分为 4 种情况:不含 tRNA 基因(a)、含 1 个 tRNA-Glu(b)、含 1 个 tRNA-Ala(c)、含 tRNA-Ala、Ile 各 1 个(d),见表 1。可以看到,tRNA 基因不随 rRNA 基因簇位置和方向的变化而变化,尽管第 2、3、5、6 拷贝 rRNA 基因簇两侧序列在各菌株中保持不变,其间隔区的 tRNA 基因还是各不相同,说明 tRNA 基因的插入与基因组重排之间没有必然的联系。从不同菌株各拷贝中的 tRNA 基因种类看,6 拷贝的菌株各不相同,KIM 和 Nepal516 相同,其他均不同,提示 tRNA 基因可以较好地区分各菌株。

表 1 不同 rRNA 基因簇的鼠疫菌中 tRNA 种类及数量

rRNA 基因簇	菌 株								
	CO92	D182038	D106004	Z176003	91001	KIM	Nepal516	Antiqua	Pestoides F
1	a	c	b	b	a	b	b	b	b
2	b	b	b	b	b	b	b	b	b
3	c	c	b	b	c	c	c	c	b
4	c	b	c	c	c	a	a	a	d
5	c	c	b	c	c	c	c	c	c
6	b	c	b	b	b	b	b	b	b
7					c	b	b	c	b

注:以 CO92 rRNA 基因 5S 端为参考进行比较

3. 23S rRNA 基因:rRNA 基因各部分中 16S 和 5S rRNA 基因变异较小,23S rRNA 突变较多,表现为单碱基突变,共 27 个突变位点。突变位置多发生在碱基的第 1171 ~ 1879 bp 之间,突变类型共有 14 种,以 G→A 置换及 C→T 置换突变居多。27 个突变位点中,一些突变仅限于少数拷贝,将突变少于 10 个拷贝的位点排除外,则有 7 个位点发生了单碱基突变,这 7 个位点分别是处于核苷酸 1171、1172、1178、1826、1879、2799、2807 处,为 23S rRNA 的多突变区(表 2、3)。

表 2 鼠疫菌中 23S rRNA 基因序列单碱基突变数量及位置

顺序	碱基位置	突变碱基	主体碱基	突变数量
1	1171	G	A	21
2	1172	T	C	21
3	1178	C	T	23
4	1826	A	G	19
5	1879	G	C	13
6	2799	A	G	17
7	2807	T	C	18

表 3 不同鼠疫菌株中不同拷贝的 23S rRNA 碱基突变数

菌株	拷 贝						
	1	2	3	4	5	6	7
CO92	0	3	0	0	6	3	
D182038	1	0	1	3	2	3	
D106004	0	3	3	3	2	1	
Z176003	0	1	2	4	4	6	
91001	2	4	0	3	3	6	0
KIM	1	1	1	1	4	6	3
Antiqua	0	0	3	0	6	3	3
Nepal516	1	3	1	1	7	3	6
Pestoides F	0	0	0	3	0	3	3

注:同表 1

对不同菌株间及不同 rRNA 基因簇拷贝间突变数进行方差分析,结果显示各菌株间 $F=0.548, P=0.815 > 0.05$,各菌株间突变数差异无统计学意义;各拷贝间 $F=5.228, P < 0.01$,各拷贝的突变数差异有统计学意义,利用 LSD 法进行两两比较,结果显示,

第 1 拷贝与 5、6、7 拷贝间差异均有统计学意义,第 2 拷贝与 5、6 拷贝间差异有统计学意义,第 3 拷贝与 5、6 拷贝间差异有统计学意义,第 4 拷贝与 5、6 拷贝间差异有统计学意义,其他各拷贝间差异无统计学意义。第 1、2、3、4 拷贝突变较少,5、6 拷贝突变最多。

讨 论

已有研究显示,绝大多数鼠疫菌 rRNA 基因以基因簇的形式存在,保守性较强。一般种内不同菌株间同源性 >99.5%;而 23S rRNA 基因和 16S ~ 23S rRNA 基因间隔区内的 tRNA 基因变异性较大,在许多研究中作为细菌属内种间或种内分类的依据^[4]。

基于鼠疫菌中 rRNA 基因的保守性,本研究用 rRNA 基因簇两侧的序列作为分析对象,试图从另一个方向出发,找到不同菌株间及不同拷贝菌株间 rRNA 基因的识别标志。同一株菌中每个 rRNA 基因簇两侧的序列是不同的,以此序列作为该 rRNA 基因簇的识别特征,可以将 6 拷贝和 7 拷贝菌株区分开。按照基因簇两侧的序列不同,将已测序菌株的 rRNA 基因簇分为 13 种,其中 6 拷贝菌株中均含有第 1 ~ 6 个拷贝,而在 7 拷贝菌株中存在第 7 ~ 13 个拷贝,这些 rRNA 基因簇两侧的序列可以作为识别特征区分 6 拷贝和 7 拷贝菌株。同时,还可区分 7 拷贝菌株间除了 Antiqua 和 KIM 外的其他菌株。另外,本研究中发现的 rRNA 基因簇 16S 侧存在的保守序列和部分 5S 侧的保守序列也可作为不同 rRNA 基因簇识别的参考指标。

rRNA 基因簇内各部分序列两两比对结果显示,不同菌株间基因组表现为单碱基突变差异,16S ~ 23S rRNA 基因间隔区因所含 tRNA 基因类型和数目不同而不同。因 16S 和 5S rRNA 基因在研究的 9 株菌株间同源性 >99.5%,不能提供可以区分和识别各菌株和不同 rRNA 基因拷贝的信息,因此重点分析了 23S rRNA 基因和间隔区内的 tRNA 基因。23S rRNA 基因在各菌株间的碱基突变位点较多,方差分析显示,不同菌株间差异无统计学意义,但在不同的 rRNA 基因簇拷贝间的突变数差异有统计学意义,第 5、6 拷贝的突变较多,即该拷贝在不同菌株中差异较大。

16S ~ 23S rRNA 基因间隔区内的 tRNA 基因在不同研究菌株和不同 rRNA 基因簇拷贝中类型和数目差异较大,以此为标识,可将 rRNA 基因簇分为 4 种不同的类型。在 6 拷贝菌株中分布差异明显,7 拷贝菌株中的 KIM 和 Antiqua 株所含 tRNA 基因类型

不同,可以进行区分,因此间隔区的 tRNA 基因可以作为 rRNA 基因簇的识别特征。

2009 年 Liang 等^[5]对 8 株已测序鼠疫菌的基因组重排现象进行了分析,显示插入序列在基因组重排中起到重要作用。本研究中,不同菌株间 rRNA 位置和方向的差异也反映了基因组重排的过程,7 拷贝菌株比 6 拷贝菌株多出的 rRNA 拷贝周围均存在插入序列,该 rRNA 拷贝很可能在基因组重排时丢失,变成 6 拷贝的菌株。同时,tRNA 基因并不随着 rRNA 基因簇在基因组中的移位而移位,可见 tRNA 的变化不是由于基因组重排导致的,而是独立进行的。

综上所述,本研究比较分析了已测序鼠疫菌 rRNA 基因簇本身及两侧的序列,rRNA 基因两侧序列可以作为识别鼠疫菌 rRNA 基因簇的 1 个重要指标;两侧较短的保守序列、间隔区的 tRNA 基因和 23S rRNA 基因序列的不同可以作为补充指标,鉴别不同菌株和不同 rRNA 基因簇拷贝,将不同菌株间的鼠疫菌加以分类。本研究方法和结果丰富了鼠疫菌 rRNA 基因的遗传特征,从分子流行病学角度为鼠疫疫情的溯源及预防控制工作提供一定的理论依据。全基因组重排与 rRNA 基因簇、tRNA 基因重排的关系分析,为鼠疫菌的基因进化研究提供一定的参考。在中国存在 18 种不同生态型的鼠疫疫源地,有些生态型菌株与已测序菌株理论上接近,但没有实验验证或全序列测定,本研究中得出的结论能否代表中国鼠疫菌 rRNA 基因簇的类型,还需要更多的实验进行研究。

(本研究的序列分析工作得到北京华大基因研究院王琪的大力帮助,谨此感谢)

参 考 文 献

- [1] Parkhill J, Wren BW, Thomson NR, et al. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*, 2001, 413 (6855):523-527.
- [2] Deng W, Burland V, Plunkett G 3rd, et al. Genome sequence of *Yersinia pestis* KIM. *J Bacteriol*, 2002, 184(16):4601-4611.
- [3] Wei JC, Yu DZ, Hai R, et al. The geographical distribution of ribotypes of *Yersinia pestis* in China. *Chin J Epidemiol*, 2003, 24 (11):1027-1030. (in Chinese)
魏建春,俞东征,海荣,等.中国鼠疫菌核糖体型地理分布. *中华流行病学杂志*, 2003, 24(11):1027-1030.
- [4] Shen DX, Feng ZC. Review about the rRNA cluster on the classify of bacteria. *Med Recapitulate*, 2004, 10(2):69-71. (in Chinese)
沈德新,封志纯.核糖体核糖核酸基因簇在细菌系统分类中的研究进展. *医学综述*, 2004, 10(2):69-71.
- [5] Liang Y, Hou X, Wang Y, et al. Genome rearrangements of completely sequenced strains of *Yersinia pestis*. *J Clin Microbiol*, 2010, 48(5):1619-1623.

(收稿日期:2010-12-24)

(本文编辑:万玉立)