

在大肠杆菌和沙门氏菌中整合位点为 tmRNA 的串联基因岛的模式和整合的时序性

宋磊^{①②}, 姜彦竹^①, 张雪洪^{①*}

① 微生物代谢国家重点实验室, 上海交通大学生命科学技术学院, 上海 200240;

② 上海师范大学生命与环境科学学院, 上海 200234

* 联系人, E-mail: xuehzhang@sjtu.edu.cn

2011-03-15 收稿, 2011-06-10 接受

国家自然科学基金(30821005, 30870075)、国家重点基础研究发展计划(2009CB118906)和上海市重点学科建设项目(B203)资助

摘要 tmRNA 是部分 tRNA 和一小段 mRNA 的结合体, 已被证明是作为基因水平转移产物之一的基因岛的整合位点. 我们通过序列比对和比较基因组学的方法确定了散布在肠杆菌科的 13 个属中整合位点为 tmRNA 的 68 个基因岛, 其中 53 个基因岛属于大肠杆菌(*Escherichia coli*)和沙门氏菌(*Salmonella enterica*). 在这 53 个基因岛中, *S. enterica* subsp. *enterica* serovar Agona str. SL483 等 8 个基因组, *E. coli* S88 和 *E. coli* O55:H7 str. CB9615 中发现了 2 个及以上连续的基因岛, 即形成了串联基因岛. 其中, *S. enterica* subsp. *enterica* serovar Typhimurium SL1344 中发现在该位点有 3 个连续的基因岛组成的串联基因岛. 通过在大肠杆菌和沙门氏菌已测序的基因组中分析整合位点为 tmRNA 的可变区, 发现大多数的基因组除了在可变区中包括了基因岛区之外, 还有一段残余的可变区. 确定了串联基因岛进入基因组的时间顺序, 即远离 tmRNA 的基因岛先于靠近 tmRNA 的基因岛整合进入该基因组中. 上述的基因岛中整合酶主要有 3 种类型, 分别是 HP1 整合酶, PhiCTX 整合酶和 P4 整合酶, 以 P4 整合酶所占的比例最多. 在串联基因岛中, 远离 tmRNA 的基因岛中的整合酶是 P4 整合酶, 其次是 PhiCTX 整合酶, 最靠近 tmRNA 的基因岛所用的整合酶是 HP1 整合酶, 由此可以得出 tmRNA 是研究串联基因岛的遗传与进化的重要位点的结论.

关键词

基因岛
tmRNA
多重整合
整合的时序性

自 1990 年首次在大肠杆菌中发现基因岛以来, 实验证明其为原核生物可变区的组成成分^[1,2]. 随着对其研究的逐渐深入, 发现基因岛可以完成各种生物学功能, 如治病性、异源物质降解和抗生素抗性等^[3]. 基因岛有显著的结构特征^[4], 如常以 tRNA 为边界序列, 以其 3'末端在基因岛的 2 个边界上形成正向重复序列, 内部至少含有 1 个整合酶、转座酶或重组酶, 而且正向重复序列是内部整合酶的作用位点, 从而完成该基因岛整合进入相应的基因组中并能从该基因组中环出的动态过程, 并为环出的基因岛转入合适的基因组提供分子基础. 所以基因岛的整合

位点以及由此形成的正向重复序列是确定基因岛和验证其活动性的重要位点.

tRNA 和 tmRNA 作为基因岛的整合位点已经被多次证明^[5-7], 而且认为 tmRNA 比 tRNA 在作用位点上的偏好性更强^[8]. 2004 年, 使用 Islander 法在 106 个已测序的基因组中共找到了 12 个以 tmRNA 为整合位点的基因岛, 比例是 11%. 上述基因岛广泛地分布在变形菌门(*Proteobacteria*), 后壁菌门(*Firmicutes*)和放线菌门(*Actinobacteria*)中^[6]. 我们也在 *Pseudomonas aeruginosa* UCBPP-PA14 中确定了 1 个整合在 tmRNA 位点的基因岛^[9]. tmRNA 同时包括部分 tRNA 和某段

英文版见: Song L, Jiang Y Z, Zhang X H. Chronology and pattern of integration of tandem genomic islands associated with the tmRNA gene in *Escherichia coli* and *Salmonella enterica*. Chinese Sci Bull, 2011, 56, doi: 10.1007/s11434-011-4749-8

mRNA, 其主要的功能是: (1) 释放同有缺陷的 mRNA 结合的核糖体^[10]; (2) 编码蛋白酶水解由缺陷 mRNA 形成的不完整的蛋白质^[10]; (3) 调控某些调节因子的活性, 如 cI 蛋白^[11]; (4) 维持噬菌体的生长^[12]等. 在霍乱弧菌(*Vibrio cholerae*)中发现了 1 个毒力岛 (VPI-1) 可以通过其内部的 P4 整合酶作用于 5'-GCTGGGG-3'(tmRNA 的 3'末端)位点而从该基因组中环出^[13,14]. 在 *E. coli* 的 CP4-57 原噬菌体中也可以通过内部的 P4 整合酶作用于 5'-CCGCCAGC-3'(tmRNA 的 3'末端)位点而从该基因组中环出^[15,16].

整合于 tRNA^{Gly} 基因中的串联基因岛被实验证实存在百日咳博特氏菌(*Bordetella pertussis*)中^[17]. 串联基因岛是研究基因岛的遗传和进化的非常好的材料. 我们在分析已测序的 *S. enterica* 的 tmRNA 位点的可变区时发现, 基因组的该区域中常常含有 2 个或以上的整合酶, 怀疑该区域可能存在串联基因岛, 所以我们首先选取肠菌科的 tmRNA 序列, 通过序列比对和比较基因组学的方法相结合来确定以 tmRNA 序列为整合位点的基因岛. 我们确实在 *E. coli* 和 *S. enterica* 中找到了串联的基因岛, 尤其在 *S. enterica* 中. 我们利用比较基因组学的方法证明了串联基因岛的可靠性, 并通过分析内部整合酶和边界序列特征确定了串联基因岛整合入基因组的时间先后关系, 并对 LT2GI^{tmRNA-1} 的同源性基因岛进行了分析, 为研究基因岛的遗传与进化打开冰山一角.

1 材料与方法

我们从 <http://www.indiana.edu/~tmrna/> 得到所有肠菌科的 tmRNA 序列, 用每个 tmRNA 序列到 NCBI (<http://www.ncbi.nlm.nih.gov/>) 同肠菌科中已测序的基因组作序列比对, 当发现 2 个同 tmRNA 序列相关的正向重复序列并且它们之间的距离在 500 kb 以内, 则 2 个正向重复序列之间的区域被确定为候选的基因岛. 由于该策略中找到的正向重复序列至少在 13 bp 以上, 可能会漏掉很多的基因岛. 我们发现肠菌科中 *E. coli* 和 *S. enterica* 已经有很多已测序的种内基因组, 所以首先定位上述基因组的 tmRNA 所在位置, 然后将该基因组在该位点同所有已测序的其他种内基因组利用 WebACT(<http://www.webact.org/WebACT/generate/>)^[18] 进行比对, 进一步发现候选的基因岛. 将上述方法找到的所有候选基因岛进行 GC 含量和二核苷酸偏向性的分析^[19], 并且确定该区域中是否含

有整合酶、转座酶或重组酶来进一步确定基因岛. 将确定的基因岛中最靠近边界的整合酶、转座酶或重组酶序列从 NCBI 获得, 并通过 CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>)^[20] 注释其功能. 将 LT2GI^{tmRNA-1} 通过 NCBI 序列比对确定其同源的基因岛, 并用 MEGA 软件分析它们之间的进化关系^[21].

2 结果与讨论

2.1 在大肠杆菌和沙门氏菌中 tmRNA 是基因岛的多重整合位点和串联基因岛的存在

tmRNA 是在肠菌科中基因岛的整合热点. 我们利用序列比对和比较基因组学的方法相结合在肠菌科(Enterobacteriaceae)中共发现了 68 个基因岛(表 S1). 在肠菌科中大约 100 个 tmRNA 基因没有成为基因岛的整合位点, 分析发现 NCBI 数据库中在肠菌科测序的基因组分布在 22 个属, 而在 13 个属内都发现了以 tmRNA 为整合位点的基因岛. 在发现的 68 个基因岛中, 有 53 个基因岛属于 *E. coli* 和 *S. enterica*, 其他的 15 个基因岛散落在 *Escherichia*, *Cronobacter*, *Citrobacter*, *Dickeya*, *Enterobacter*, *Klebsiella*, *Pantoea*, *Edwardsiella*, *Shigella*, *Photobacterium*, *Yersinia*, *Sodalis* 中. 在 32 个 *E. coli* 基因组中只有 9 个 tmRNA 基因没有成为基因岛的整合位点, 在 22 个沙门氏菌中只有 3 个 tmRNA 基因没有成为基因岛的整合位点(表 S2 和 S3).

在 32 个已测序 *E. coli* 中, 共发现 23 个整合在 tmRNA 基因的基因岛, 所占的比例是 71.88%(*E. coli* B str. REL606, *E. coli* HS, *E. coli* O111:H- str. 11128, *E. coli* O103:H2 str. 12009, *E. coli* O127:H6 str. E2348/69, *E. coli* SE11, *E. coli* SMS-3-5, *E. coli* SE15 和 *E. coli* 536 中没有确定出该类基因岛)(表 S2). 其中只有 *E. coli* SE15 和 *E. coli* 536 在 tmRNA 的位点没有任何整合过基因岛或残留有可变区的迹象, 其他基因组在该位点都有一段可变区. 在 22 个沙门氏菌基因组中, 共确定了 19 个以 tmRNA 为整合位点的基因岛, 所占的比例是 86.36%(其中只有 *S. enterica* subsp. *enterica* serovars Typhi str. CT18, Gallinarum str. 287/91 和 Enteritidis str. P125109 中没有确定出该类基因岛, 但在该位点都有一段可变区)(表 S3).

tmRNA 是肠菌科中基因岛的多重整合热点. 我们通过比较基因组学分析确定大肠杆菌和沙门氏菌

中在 tmRNA 位点的可变区(表 S2 和 S3). 我们发现除了 *E. coli* SE15 和 *E. coli* 536 在该区域不存在可变区之外, 其余的基因组都存在可变区. *S. enterica* subsp. *enterica* serovar Agona str. SL483, *E. coli* ED1a, *E. coli* S88, *E. coli* APEC O1, *E. coli* UTI89 和 *E. coli* CFT073 的基因组在该位点的可变区完全由 1 个或串联的基因岛构成. 在 32 个 *E. coli* 基因组中, 在该位点的 17 个可变区由基因岛区和残余的可变区组成(表 S2). 在 22 个沙门氏菌基因组中, 在该位点的 18 个可变区由基因岛区和残余的可变区组成(表 S3). 基因岛区可以是 1 个基因岛或是串联基因岛(表 1 和 S1). GC 含量和二核苷酸偏向性分析显示这些残余的可变区也是水平基因转移获得的(表 S2 和 S3). 这些残余的可变区域都是在基因岛区的下游(以 tmRNA 为上游), 由于存在于基因组的时间长, 通过基因组的删除和重排, 从而失去了基因岛的基本结构, 最主要的是整合酶的丢失或假基因化和作为整合酶作用位点的正向重复序列的缺失或突变, 猜测该残余区域是最先来到这个基因组中的. 这样就可以证明这些基因组在 tmRNA 位点上至少发生了 2 次及以上的整合, 并且发现在 *E. coli* 中, 除了 *E. coli* O26:H11 str. 11368 之外, 其余的 *E. coli* 的残余可变区中都含有 tRNA^{Met}, 而 11368GI^{tmRNA} 中含有 2 个 tRNA^{Arg} 和 1 个 tRNA^{Ile}, 可以证明某些 tRNA 序列也是水平基因转移的产物.

对 *E. coli* S88 基因组在 tmRNA 位点的可变区进行分析发现, 该区域包括 2 个基因岛, 即 S88GI^{tmRNA-1} 和 S88GI^{tmRNA-2}, S88GI^{tmRNA-2} 同 APECO1GI^{tmRNA} 有极高的相似性, S88GI^{tmRNA-1} 同 ED1aGI^{tmRNA} 有较高的相似性(图 1). APECO1GI^{tmRNA} 和 ED1aGI^{tmRNA} 都整合在 *E. coli* 536 的 tmRNA 基因中, 由此可以证明 *Escherichia coli* S88 在 tmRNA 位点的可变区不是 1 个大的基因岛, 而是 2 个基因岛, 即 S88GI^{tmRNA-1} 和 S88GI^{tmRNA-2} 串联的结果. 在 53 个大肠杆菌和沙门氏菌中找到的基因岛中, *S. enterica* subsp. *enterica* serovar Agona str. SL483, *S. enterica* subsp. *enterica* serovar Newport str. SL254, *S. enterica* subsp. *enterica* serovar Typhimurium LT2, *S. enterica* subsp. *enterica* serovar Paratyphi A str. AKU_12601, *S. enterica* subsp. *enterica* serovar Paratyphi A str. ATCC 9150, *S. enterica* subsp. *enterica* serovar Typhimurium SL1344, *S. enterica* subsp. *enterica* serovar Virchow str. SL491, *S. enterica* subsp. *enterica* serovar Paratyphi B str. SPB7, *E. coli*

S88 和 *E. coli* O55:H7 str. CB9615 中发现串联基因岛的存在(表 1 和图 2). 在 *S. enterica* 已完全测序和基本测序完成的 22 个基因组中, 共发现 8 个基因组中存在串联基因岛, 所占的比例是 36.4%(表 1 和图 2). 在 *E. coli* 已完全和基本测序的 32 个基因组中只发现了 2 个基因组中存在串联基因岛, 所占的比例是 6.25%(表 1 和图 2). 上述的比例数据说明 *S. enterica* 中存在串联基因岛的情况比在 *E. coli* 中更普遍.

2.2 串联基因岛整合的时序性

串联基因岛整合的时序性显示距 tmRNA 这个整合位点越远的基因岛越先整合进入基因组. 在 16 个 *E. coli* 基因组和 11 个 *S. enterica* 基因组中, 在 tmRNA 这个整合位点形成的可变区包括 1 个基因岛和 1 个残余的可变区. 由于残余可变区已经失去了基因岛的基本结构, 所以基因岛比该残余可变区更晚地整合在 tmRNA 位点. 越早整合入该位点的基因岛的结构越不完整, 常常包括缺陷的边界正向重复序列和假基因化或被截短的整合酶. 由于 SL483GI^{tmRNA-2} 和 LT2GI^{tmRNA-2} 中包含假基因化的整合酶和可能的缺陷的边界重复序列, 所以 SL483GI^{tmRNA-1} 比 SL483GI^{tmRNA-2} 后进入基因组, 而 LT2GI^{tmRNA-1} 比 LT2GI^{tmRNA-2} 后进入基因组. 在 *S. enterica* subsp. *enterica* serovar Typhimurium SL1344 中在 tmRNA 位点的可变区包括串联基因岛区(SL1344GI^{tmRNA-1}, SL1344GI^{tmRNA-2} 和 SL1344GI^{tmRNA-3})和一段残余可变区, 残余可变区(2927877~2938549)高同源在于 *S. enterica* subsp. *enterica* serovar Typhimurium str. LT2 中该位点的残余可变区(2905309~2915973). SL1344GI^{tmRNA-3} 是高同源在于 LT2GI^{tmRNA-2}, SL1344GI^{tmRNA-2} 同源在于 *S. enterica* subsp. *enterica* serovar Paratyphi A str. AKU_12601 中的 1 个整合于 tRNA^{Lcu} 基因的基因岛(4456335~4466068), 而 SL1344GI^{tmRNA-1} 同源在于 LT2GI^{tmRNA-1}(图3). 在 *S. enterica* subsp. *enterica* serovar Typhimurium SL1344 残余可变区中有 1 个假基因化的整合酶(SL1344_2747), 没有发现作为整合酶作用位点的边界正向重复序列, 由此可以判定该残余可变区通过 SL1344_2747 最先整合进该基因组中. 随着时间的推移, 这个区域由于边界序列的重排和删除和整合酶的假基因化而失去了活性. 不久之后, SL1344GI^{tmRNA-3} 整合入该位点, 而且由于整合酶

表 1 在大肠杆菌和沙门氏菌中整合在 tmRNA 位点的串联基因岛的特征^{a)}

基因岛	菌株	位置	tmRNA	正向重复序列(<i>attL/attR</i>)	整合酶	通过 CDD(PDB) 注释的整合酶的类型	基因岛的大小 (bp)	基因岛的 GC% (基因岛的 GC%)	δ^* (×1000) 低的区域所占的百分率(%)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率(%)
SL483GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Agona str. SL483	2787019~2820998	2786701..2787063	ATGTAGGAATTCGGACGCGGGTTCAACTC- CCGCCAGCTCCACCAA/ATGTAGGAATTC- GGACGCGGGTTCAACTCCCGCCAGCTCCAC- CAAA	SeAg_B2882 2819809..2820834	HP1 整合酶	33980	51.08 (52.1)	65.518	100
SL483GI ^{tmRNA-2}		2820982~2824289	No	CCGCCAGC/ACGCCAGC	eAg_B2883 2821164..2821745	DNA 断裂- 重联酶	3308	38.24 (52.1)	87.655	91.793
SL254GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254	2846418~2878144	2846103..2846465	AGGATGTAGGAATTCGGACGCGGGTTC- ACTCCCGCCAGCTCCACCAA/ATGTAGGATGT- AGGAATTCGGACGCGGGTTCAACTCCCGC- CAGCTCCACCAAA	SNSL254_A2951 2876970..2877980	DNA 断裂-重联- 酶(HP1 整合酶)	31727	50.52 (52.2)	67.188	100
SL254GI ^{tmRNA-2}		2878119~2888742	No	TTCAACTCCCGCCAGCT/TTCAAGACTCCGCC- AGCT	SNSL254_A2952 (2878326..2879519)	P4 整合酶	10624	47.94 (52.2)	60.835	90.969
LT2GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium LT2	2844265~2878004	2843947..2844309	ATGTAGGAATTCGGACGCGGGTTCAACTC- CCGCCAGCTCCACCAA/ATGTAGGAATTC- GGACGCGGGTTCAACTCCCGCCAGCTCCAC- CAAA	STM2739 2876815..2877840	HP1 整合酶	33740	52.49 (52.2)	67.606	100
LT2GI ^{tmRNA-2}		2877988~2905308	No	CCGCCAGC/ACGCCAGC	STM2740 (2878169..2879233)	P4 整合酶	27321	47.66 (52.2)	55.235	94.35
1260IGI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi A str. AKU_12601	2657604~2691233	2657286..2657648	ATGTAGGAATTCGGACGCGGGTTCAACTC- CCGCCAGCTCCACCAAATAAAACAATGGGT- TACGTGAAAACGTAGCCCTTTTCTT/ATG- TAGGAATTCGGACGCGGGTTCAACTCCCG- CCAGCTCCACCAAATAAAACAAGGGGTTC- GTGAAAACGTAGTCCCTTTTGT	Sspa2423 2690004..2691029	HP1 整合酶	33630	51.60 (52.2)	63.441	99.265
1260IGI ^{tmRNA-2}		2691147~2723458	No	ATGTAGGAATTCGGACGCGGGTTCAACTC- CCGCCAGCTCCACCAA/ATGTAGGAATTCGG- ACCGGGTTCAACTCCCGCCAGCTCCACCA	Sspa2447 2721297..2722310	DNA 断裂-重联- 酶(HP1 整合酶)	32312	49.93 (52.2)	64.371	98.582

(续表 1)

基因岛	菌株	位置	tmRNA	正向重复序列(<i>attL/attR</i>)	整合酶	通过 CDD(PDB) 注释的整合酶的类型	大小 (bp)	基因岛的 GC% (基因组的 GC%)	δ^* (×1000)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率(%)
9150GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi A str. ATCC 9150	2662068~2695697	2661750..2662112	ATGTAGGAATTCGGACGCGGGTTCAACTCCGCCAGCTCCACCAATAAACAATGGGT-TACGTGAAAACGTAGCCCTTTTCTTT/ATGT-AGGAATTCGGACGCGGGTTCAACTCCCGC-CAGCTCCACCAATAAACAAGGGGTACG-TGAAAACGTAGTCCCTTTTGT	Spa2600 2694468..2695493	HP1 整合酶	33630	51.60 (52.2)	63.426	99.265
9150GI ^{tmRNA-2}		2695611~2727922	No	ATGTAGGAATTCGGACGCGGGTTCAACTCCGCCAGCTCCACCA/ATGTAGGAATTCGG-ACGGGGTTCAACTCCCGCAGCTCCACCA	2725761..2726774	DNA 断裂-重联酶(HP1 整合酶)	32312	49.93 (52.2)	64.509	99.291
SL1344GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium SL1344	2855454~2889891	2855136..2855498	ATGTAGGAATTCGGACGCGGGTTCAACTCCGCCAGCTCCACCAAT/ATGTAGGAATTCGGACGCGGGTTCAACTCCCGCAGCTCCA-CCAAAAT	SL1344_2709 2887497..2888522	HP1 整合酶	34438	51.39 (52.2)	206.422	94.444
SL1344GI ^{tmRNA-2}		2889874~2900569	No	CCGCCAGCTCCACCAAA/CCGCCAGCCAC-CAAA	SL1344_2712 2890073..2891269	P4 整合酶	10696	47.57 (52.2)	225.454	93.305
SL1344GI ^{tmRNA-3}		2900554~2927876	No	CCGCCAGC/ACGCCAGC	SL1344_2723 2900734..2901979	P4 整合酶	27323	47.67 (52.2)	226.938	92.857
SL491GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Virchow str. SL491	gcontig. 111270757 serovar str. 3884 2274696~2307156	No	GCTGGCGGAGTTGAACCCCGCTCCGAAATTCCTACTACAT/GCTGGCGGAGTTGAACCCCGCGTCCGAAATTCCTACAT	(2275837..2276850)	DNA 断裂-重联酶(HP1 整合酶)	32461	50.07 (52.2)	65.847	97.872
SL491GI ^{tmRNA-2}		2307070~2342654	2342610..2342972	AACAAAAAGGGGTACGTTTTCACGTAACC-CCTGTTTATTTGGTGGAGCTGGCGGAGT-TGAACCCCGTCCGAAATTCCTACTACAT/AACAAAAAGGGGTACGTTTTCACGTAACCCCTTGTATTTGGTGGAGCTGGCGGAGTTGA-ACCCCGTCCGAAATTCCTACAT	(2307274..2308299)	HP1 整合酶	35585	51.95 (52.2)	71.350	100

(续表 1)

基因岛	菌株	位置	tmRNA	正向重复序列 (<i>attL/attR</i>)	整合酶	通过 CDD(PDB) 注释的整合酶的类型	大小 (bp)	基因岛的 GC% (基因岛的 GC%)	δ^* ($\times 1000$)	基因组中比基因岛更低的区域所占的百分率(%)
SPB7GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi B str. SPB7	2858026~2868501	2857678..2858039	CCGCCAGCTCCACCAA/C <u>CCGCCAGCCAC</u> CAAA	SPAB_03412 (2858224..2859417)	P4 整合酶	10476	48.78 (52.1)	62.457	90.713
SPB7GI ^{tmRNA-2}		2868476~2876283	No	TATTCAGACTCCGCCAGCCCA/TATTCAGAT-TCCGCCAGCTCA	SPAB_03429 (2868695..2869891)	P4 整合酶	7808	42.17 (52.1)	77.723	92.444
S88GI ^{tmRNA-1}	<i>Escherichia coli</i> S88	2798240~2832381	2797922..2798284	ATGTAGGAATTCGGACGGGGTTCAACTC-CCGCCAGCTCCACCA/ATGTAGGAATTCG-GACGGGGTTCAACTCCCGCCAGCTCCACC-AA	ECS88_2851 (2829972..2830991)	DNA 断裂-重联酶 (HP1 整合酶)	34142	50.31 (50.7)	40.315	89.116
S88GI ^{tmRNA-2}		2832346~2875497	No	TTCGGACGGGGTTCAACTCCCGCCAGCTT-TGGACTGGGGTTCAACTCCCGCCAGC	ECS88_2917 (2872287..2873459)	PhiCTX 整合酶	43152	49.40 (50.7)	45.501	93.966
CB9615GI ^{tmRNA-1}	<i>Escherichia coli</i> O55:H7 str. CB9615	3258465~3313225	3258138..3258500	TTTCGGACGGGGTTCAACTCCCGCCAGCT-CCACCACTTTTTCGGACACGGGTTCAACT-CCCGCCAGCTCCACCACTTT	G2583_3276 (3310496..3311680)	PhiCTX 整合酶	54761	50.98 (50.5)	57.167	98.98
CB9615GI ^{tmRNA-2}		3313207~3337896	No	CCGCCAGCTCCACCA/CTGCCAGCCCA	G2583_3278 (3313387..3314616)	P4 整合酶	24690	36.00 (50.5)	121.065	100

a) 双下画线的碱基是 HP1 整合酶的绑定位点; 虚线的碱基是 PhiCTX 整合酶的绑定位点; 单下画线的碱基是预测的整合酶的酶切位点

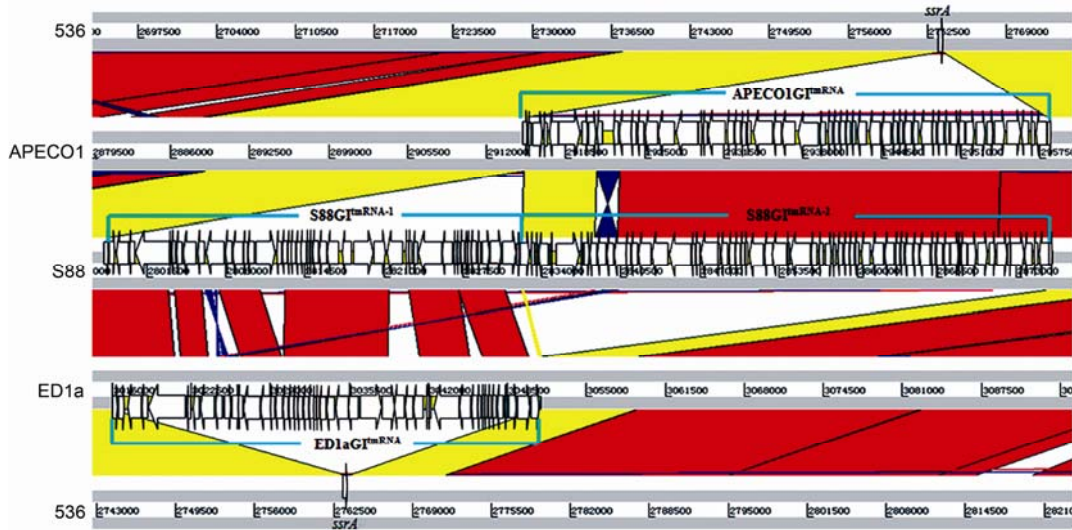


图1 利用 WebACT 在 *Escherichia coli* S88 中 tmRNA 位点鉴定串联基因岛
基因组之间红色和蓝色的区域代表两者的同源区域, 其中蓝色的线代表反向互补区域. 箭头代表基因岛中的基因

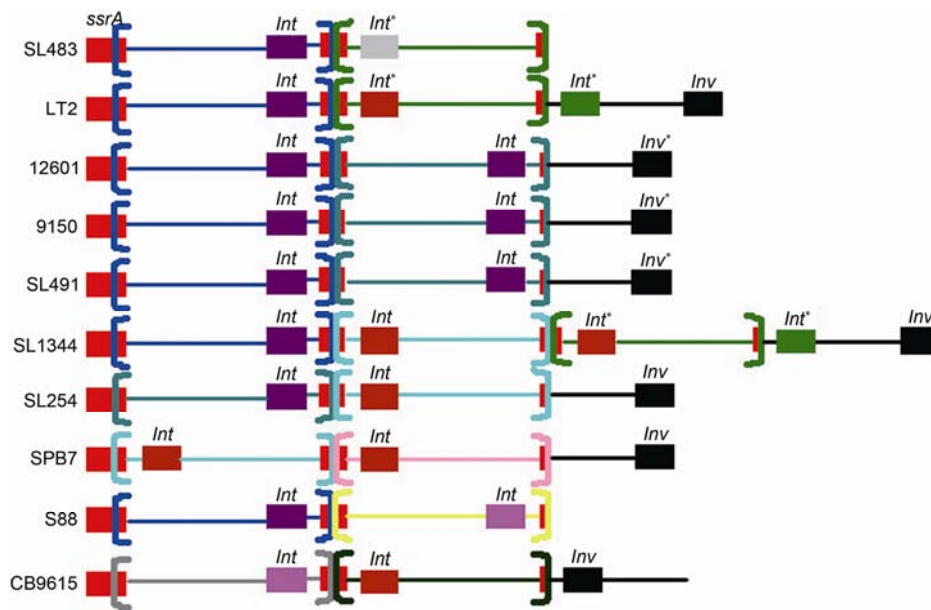


图2 包含有串联基因岛的基因组在 tmRNA 位点的可变区的结构示意图(未按比例)

红色, tmRNA 基因或其片段; 括号代表 att 位点; Int, 整合酶基因; 深紫色代表 HP1 整合酶, 亮紫色代表 PhiCTX 整合酶, 橙色代表 P4 整合酶, 灰色代表 DNA 断裂-重联酶; 黑色, Inv, DNA 转化酶; 括号里的细线代表基因岛区, 相同颜色的括号和细线代表这些区域是相互同源的; 黑色的细线代表残余可变区; 星号标注假基因或截短的整合酶

(SL1344_2723)的假基因化和边界序列(5'-ACGCC-AGC-3')可能的点突变而失去了活动性(图3). SL1344GI^{tmRNA-2}有着完整的边界序列(5'-CCGCC-AGC-3'),同时内部的整合酶是 P4 整合酶. 而 SL1344GI^{tmRNA-1}中的内部整合酶是 HP1 整合酶, 在其边

界序列中该整合酶的绑定位点是 5'-AGGAA-TTT-3'^[22]. 由于已确定的 68 个整合于 tmRNA 位点的基因岛中的整合酶以 P4 整合酶为主, 所以判定含有 P4 整合酶的基因岛应较早地整合入基因组中, 所以 SL1344GI^{tmRNA-2}是第三次整合入该位点的产物. 该

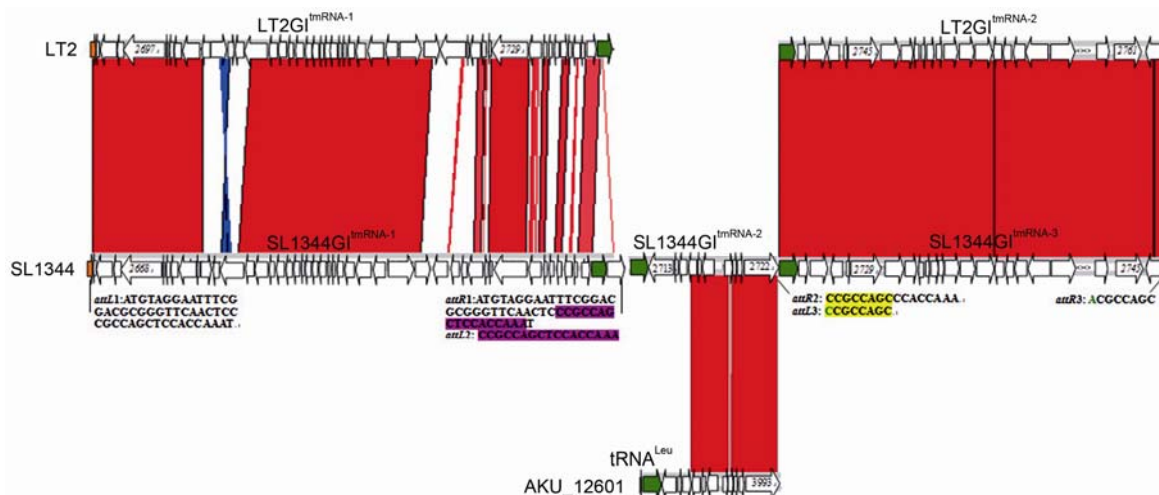


图3 在 *Salmonella enterica* subsp. *enterica* serovar Typhimurium SL1344 中串联基因岛整合时序性的鉴定

基因组之间红色和蓝色的区域代表两者的同源区域, 箭头代表基因岛中的基因, 橙色箭头代表 tmRNA 基因(ssrA); 绿色箭头代表基因岛中的整合酶, 紫色背景下的碱基是 attR1 和 attL2 之间相同的部分; 黄色背景下的碱基是 attR2 和 attL3 之间相同的部分

基因岛含有有功能的整合酶和完整的边界序列, 应具有活性. 由于 SL1344 GI^{tmRNA-1} 含有有功能的 HP1 整合酶和在边界序列中含有完整的该整合酶的绑定序列 (5'-AGGAATTT-3'), 所以 SL1344GI^{tmRNA-1} 是最后 1 个整合入该位点的基因岛. 通过上述分析我们发现, 在 *S. enterica* subsp. *enterica* serovar Typhimurium SL1344 中发现的 3 个基因岛进入基因组的时间先后顺序是 SL1344GI^{tmRNA-3} 先进入, 然后是 SL1344GI^{tmRNA-2}, 最后才是 SL1344GI^{tmRNA-1}. 所以我们可以基本证明对于串联的基因岛来讲, 最先进入基因组的是最远离整合位点的基因岛或可变区, 最后进入基因组的是最靠近整合位点的基因岛.

2.3 基因岛中整合酶的分析

将 68 个基因岛的边界序列和整合酶进行分析(表 S4), 发现整合酶主要被分成 3 种类型: HP1 整合酶, PhiCTX 整合酶和 P4 整合酶. Williams^[23]推测整合在 tRNA 位点或 tmRNA 位点的基因岛中整合酶的作用位点可以在反密码子环, TψC 环和不对称的 3'末端. 在肠菌科中确定的 68 个基因岛中, 其中 17 个基因岛的整合酶是 HP1 整合酶, 预测这类整合酶在边界序列上的绑定位点是 5'-AGGAATTT-3'^[22]. 该整合酶的绑定位在 tmRNA 序列中 tRNA 序列部分的 TψC 环的上游; 10 个基因岛的整合酶是 PhiCTX 整合酶, 这类整合酶的作用位点是 5'-CGGGTTCAACTCCCG-3',

其中绑定位点是 CGGG(CCCG), 酶切位点是在 TψC 环上(TTCAACT). 其余基因岛的整合酶大多是 P4 整合酶, 该类整合酶的作用位点大多是 5'-CCGCCAGC-3', 该位点在 tRNA 序列部分的 TψC 环的下游(表 S4). Wang 等人^[16]对 CP4-57 原噬菌体的删除实验发现, 该噬菌体中的整合酶是一种 P4 整合酶, 该酶的作用位点是 5'-CCGCCAGC-3', 所以可以保证我们找到的内部整合酶是 P4 整合酶的基因岛边界的可靠性. 通过图 2 分析发现, 在串联基因岛中, 最靠近 tmRNA 位点的基因岛中的整合酶常常是 HP1 整合酶, 然后是 PhiCTX 整合酶, 最后是 P4 整合酶. 由此可以预测串联基因岛中整合酶的作用位点常常先是 tRNA 不对称的 3'末端, 然后是 TψC 环, 最后是反密码子环. 也就是说随着基因岛在 tRNA 或 tmRNA 位点的一次又一次地整合, 整合酶的作用位点由 tRNA 或 tmRNA 的 3'端向 5'端延伸.

2.4 Enterobacteria phage Fels-2 及其同源基因岛的分析

我们发现找到的 LT2GI^{tmRNA-1} 已被确定为 Enterobacteria phage Fels-2^[24], 通过序列比对发现该基因岛同很多基因组中的某个基因岛同源. 这些基因岛有着相似的噬菌体结构蛋白. 9 个整合在 tmRNA 位点的基因岛已经被确定(表 S1), 其余的 9 个基因岛也被确定(表 S5). 18 个同源的基因岛所在的基因组主

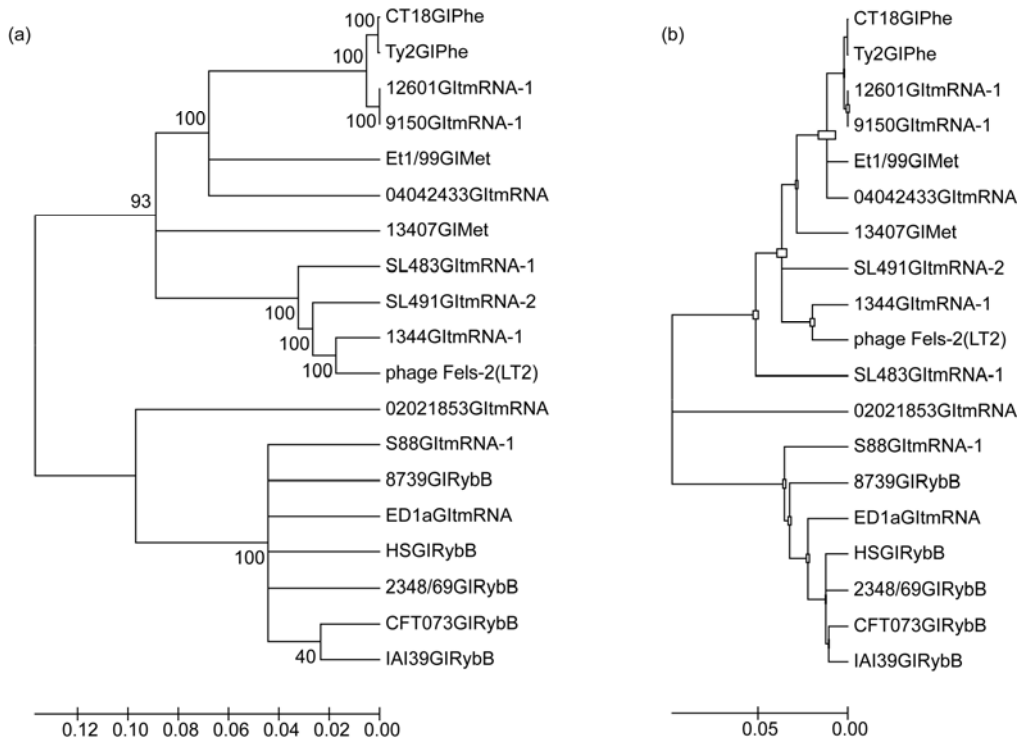


图 4 用 MEGA4 的邻接方法和 MEGA5 的最大似然法建立 *Enterobacteria phage Fels-2* 同源基因岛的系统进化关系。数字显示置信度(500 次重复)的百分率, 刻度显示进化距离。(a) 邻接方法; (b) 最大似然法

要以 *E. coli* 和 *S. enterica* 为主, 但也有 2 个基因岛分布在 *Erwinia tasmaniensis*, *Enterobacter cloacae* 中。tmRNA, RybB 基因, tRNA^{Met} 和 tRNA^{Phe} 基因被确定是这些基因岛的整合位点。将这 19 个基因岛通过 MEGA 软件进行同源性分析(图 4(a),(b)), 发现这 19 个基因岛自动地形成 2 个区域, *Enterobacteria phage Fels-2* 更类似于 *S. enterica* 中发现的基因岛, 这些基因岛的边界以 tmRNA 为主, 在这个区也包含了 *Erwinia tasmaniensis* 和 *Enterobacter cloacae* 中发现的 2 个基因岛, 即 13047GI^{Met} 和 Et1/99GI^{Met}。Enterobacteria phage Fels-2 的同源性离 *E. coli* 中发现的基因岛较远, 这些基因岛的边界以 tmRNA 和 RybB 为主。这是否能说明当基因岛整合进基因组后, 有同该基因组序列趋同性的趋势尚需更多的证明。

3 结论

我们通过序列比对和比较基因组学分析, 在肠菌科中共发现了整合在 tmRNA 位点的分布在 13 个属内的 68 个基因岛, 其中大多存在于 *E. coli* 和 *S. enterica* 中。在 *E. coli* 和 *S. enterica* 中发现了在该位点存在串联基因岛, 远离 tmRNA 的基因岛先进入基因组, 而最靠近 tmRNA 的基因岛最后进入基因组, 并且确定从靠近到远离 tmRNA 的基因岛中使用整合酶的顺序是 HP1 整合酶, 然后是 PhiCTX 整合酶, 最后是 P4 整合酶。分析 *Enterobacteria phage Fels-2* 及其同源的基因岛散落在肠菌科的 19 个基因组中, 以 *E. coli* 和 *S. enterica* 为主, 边界主要以 tmRNA 为主, 以 RybB 为辅。

参考文献

- 1 Hacker J, Bender L, Ott M, et al. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb Athog, 1990, 8: 213-225
- 2 Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. EMBO Rep, 2001, 2: 376-381

- 3 Dobrindt U, Hochhut B, Hentschel U, et al. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*, 2004, 2: 414–424
- 4 Juhas M, van der Meer J R, Gaillard M, et al. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev*, 2009, 33: 376–393
- 5 Boyd E F, Almagro-Moreno S, Parent M A. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol*, 2009, 17: 47–53
- 6 Mantri Y, Williams K P. Islander: A database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Res*, 2004, 32: D55–D58
- 7 Ou H Y, He X, Harrison E M, et al. MobilomeFINDER: Web-based tools for in silico and experimental discovery of bacterial genomic islands. *Nucleic Acids Res*, 2007, 35: W97–W104
- 8 Williams K P. Traffic at the tmRNA gene. *J Bacteriol*, 2003, 185: 1059–1070
- 9 Song L, Zhang X H. Innovation for ascertaining genomic islands in PAO1 and PA14 of *Pseudomonas aeruginosa*. *Chinese Sci Bull*, 2009, 54: 3991–3999
- 10 Muto A, Sato M, Tadaki T, et al. Structure and function of 10Sa RNA: *Trans*-translation system. *Biochimie*, 1996, 78: 985–991
- 11 Retallack D M, Friedman D I. A role for a small stable RNA in modulating the activity of DNA-binding proteins. *Cell*, 1995, 83: 227–235
- 12 Withey J, Friedman D. Analysis of the role of trans-translation in the requirement of tmRNA for lambda daimmP22 growth in *Escherichia coli*. *J Bacteriol*, 1999, 181: 2148–2157
- 13 Karaolis D K, Johnson J A, Bailey C C, et al. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci USA*, 1998, 95: 3134–3139
- 14 Rajanna C, Wang J, Zhang D, et al. The vibrio pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. *J Bacteriol*, 2003, 185: 6893–6901
- 15 Kirby J E, Trempey J E, Gottesman S. Excision of a P4-like cryptic prophage leads to Alp protease expression in *Escherichia coli*. *J Bacteriol*, 1994, 176: 2068–2081
- 16 Wang X, Kim Y, Wood T K. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *ISME J*, 2009, 3: 1164–1179
- 17 Lechner M, Schmitt K, Bauer S, et al. Genomic island excisions in *Bordetella petrii*. *BMC Microbiol*, 2009, 9: 141
- 18 Abbott J C, Aanensen D M, Bentley S D. WebACT: An online genome comparison suite. *Methods Mol Biol*, 2007, 395: 57–74
- 19 van Passel M W, Luyf A C, van Kampen A H, et al. Deltarho-web, an online tool to assess composition similarity of individual nucleic acid sequences. *Bioinformatics*, 2005, 21: 3053–3055
- 20 Marchler-Bauer A, Anderson J B, Chitsaz F, et al. CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res*, 2009, 37: D205–D210
- 21 Tamura K, Dudley J, Nei M, et al. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, 2007, 24: 1596–1599
- 22 Esposito D, Thrower J S, Scocca J J. Protein and DNA requirements of the bacteriophage HP1 recombination system: A model for intasome formation. *Nucleic Acids Res*, 2001, 29: 3955–3964
- 23 Williams K P. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: Sublocation preference of integrase subfamilies. *Nucleic Acids Res*, 2002, 30: 866–875
- 24 Pelludat C, Mirolod S, Hardt W D. The SopEPhi phage integrates into the *ssrA* gene of *Salmonella enterica* serovar Typhimurium A36 and is closely related to the Fels-2 prophage. *J Bacteriol*, 2003, 185: 5182–5191

补充材料

- 表 S1 在肠菌科中整合在 tmRNA 位点的基因岛的特征
- 表 S2 在大肠杆菌中 tmRNA 位点的可变区分析
- 表 S3 在沙门氏菌中 tmRNA 位点的可变区分析
- 表 S4 68 个基因岛的整合酶和边界的分析
- 表 S5 同 *Enterobacteria phage Fels-2* 同源的边界为 *RyxB* 和 tRNA 基因的基因岛的特征

本文的以上补充材料见网络版 csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

表 S1 在肠菌科中整合在 tmRNA 位点的基因岛的特征

基因岛	菌株	位置	整合位置	整合酶或转座酶	大小 (bp)	基因组的 GC(%)	基因岛的 GC(%)	基因组的平均 δ^* (长度是20000)	δ^* (×1000)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
SL483GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	2787019-2820998	2786701..2787063	SeAg_B2882	33980	52.1%	51.08%	28.911	65.518	100%
SL483GI ^{tmRNA-2}	serovar Agona str. SL483	2820991-2824292	No	eAg_B2883	3308	52.1%	38.24%	28.911	87.655	91.793%
SC-B67GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica	2861470-2865465	2861122..2861483	SC2698	3996	52.2%	45.02%	28.897	133.082	99.908%
SL254GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	2846418-2878144	2846103..2846465	SNSL254_A2951	31727	52.2%	50.52%	28.558	67.188	100%
SL254GI ^{tmRNA-2}	serovar Newport str. SL254	2878119-2888742	No	SNSL254_A2952	10624	52.2%	47.94%	28.558	60.835	90.969%
LT2GI ^{tmRNA-1}	<i>Salmonella typhimurium</i> LT2	2844265-2878004	2843947..2844309	STM2739	33740	52.2%	52.49%	28.587	67.606	100%
LT2GI ^{tmRNA-2}		2877988-2905308	No	STM2740	27321	52.2%	47.66%	28.587	55.235	94.35%
D23580GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica	2870953-2898275	2870605..2870966	STMMW_27071	27323	52.2%	47.66%	28.802	55.054	94.944%
SL476GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica	2852495-2879816	2852147..2852509	SeHA_C2911	27322	47.3%	47.68%	29.965	94.262	100%
02021853GI ^{tmRNA} ^A	<i>Salmonella enterica</i> subsp. enterica	2922896-2956228	2922578..2922940	SeD_A3066	33333	52.2%	51.36%	29.295	73.856	100%
12601GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	2657604-2691233	2657286..2657648	Sspa2423	33630	52.2%	51.60%	28.413	63.441	99.265%
12601GI ^{tmRNA-2}	serovar Paratyphi A str. AKU_12601	2691147-2723458	No	Sspa2424	32312	52.2%	49.93%	28.413	64.371	98.582%
9150GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	2662068-2695697	2661750..2662112	Spa2600	33630	52.2%	51.60%	28.734	63.426	99.265%
9150GI ^{tmRNA-2}	serovar Paratyphi A str. ATCC 9150	2695611-2727922	No	Spa2601	32312	52.2%	49.93%	28.734	64.509	99.291%
SL1344GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	2855454-2889891	2855136..2855498	SL1344_2709	34438	52.2%	51.39%	113.106	206.422	94.444%
SL1344GI ^{tmRNA-2}	serovar Typhimurium SL1344	2889874-2900569	No	SL1344_2712	10696	52.2%	47.57%	113.106	225.454	93.305
SL1344GI ^{tmRNA-3}		2900554-2927876	No	SL1344_2723	27323	52.2%	47.67%	113.106	226.938	92.857%
04042433GI ^{tmRNA} ^A	<i>Salmonella enterica</i> subsp. enterica	gcontig_11131159	100071..100433	69953..70978	31529	52.2% ^a	52.68%	28.734	69.934	99.31%
SL491GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica	15067	No	2275837..2276850	32461	52.2% ^a	50.07%	28.734	65.847	97.872%
SL491GI ^{tmRNA-2}	serovar Virchow str. SL491	GA_MM04042433	68587-100115	2274696-2307156	35585	52.2% ^a	51.95%	28.734	71.350	100%
		2307070-2342654	2342610..2342972	2307274..2308299						

续表 1

基因岛	菌株	位置	整合位置	整合酶或转座酶	大小 (bp)	基因组的 GC(%)	基因岛的 GC(%)	基因组的平均 δ^* (长度是20000)	δ^* (×1000)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
Ty2GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi str. Ty2	2734830-2745481	2734381..2734847	t2647 2735202..2736233	10652	52.1%	49.83%	31.278	62.487	88.641%
62:z4GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. arizonae serovar 62:z4,z23:--	198157-216608	216595..216957	SARI_00233 215199..216470	18452	51.4%	36.27%	31.305	146.636	100%
CVM19633GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Schwarzengrund str. CVM19633	2796896-2819239	2796548..2796910	SeSA_A2892 2797034..2798305	22343	52.2%	39.89%	30.127	109.276	100%
SPB7GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi B str. SPB7	2858026-2868501	2857678..2858039	SPAB_03412 2858224..2859417	10476	52.1%	48.78%	30.489	62.457	90.713%
SPB7GI ^{tmRNA-2}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi B str. SPB7	2868476-2876283	No	SPAB_03429 2868695..2869891	7808	52.1%	42.17%	30.489	77.723	92.444%
SL317GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Newport str. SL317	gcontig_11066329-21440 649769-660392	660739..660367	658992..660185	10624	52.2% ^a	47.96%	28.734	57.636	90.023%
RKS4594GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi C strain RKS4594	2843340-2847336	2842992-2843358	SPC_2805 2843906..2844745 SPC_2804 2843670..2843879	3997	52.2%	45.06%	29.606	133.273	98.925%
14028SGI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Typhimurium str. 14028S	2898213-2925535	2897865..2898226	STM14_3302 2898394..2899458	27323	52.2%	47.66%	28.926	54.934	96.067%
ED1aGI ^{tmRNA}	<i>Escherichia coli</i> ED1a	3016052-3051142	3015734..3016096	ECED1_3109 3049316..3050332	35091	50.7%	51.38%	31.147	41.947	87.162%
S88GI ^{tmRNA-1}	<i>Escherichia coli</i> S88	2798240-2832381	2797922..2798284	ECS88_2851 2829972..2830991	34142	50.7%	50.31%	28.681	40.315	89.116%
S88GI ^{tmRNA-2}	<i>Escherichia coli</i> S88	2832346-2875497	No	ECS88_2917 2872287..2873459	43306	50.7%	49.40%	28.681	45.501	93.966%
APECO1GI ^{tmRNA}	<i>Escherichia coli</i> APEC O1	2915035-2958340	2914707..2915069	APECO1_3866 2954976..2956148	43152	50.5%	49.39%	29.172	46.275	97.436%
UTI89GI ^{tmRNA}	<i>Escherichia coli</i> UTI89	2900009-2943313	2899681..2900043	UTI89_C3009 2939949..2941121	43305	50.6%	49.40%	29.671	46.250	92.241%
CFT073GI ^{tmRNA}	<i>Escherichia coli</i> CFT073	3018510-3067120	3018179..3018541	c3204 3064143..3065315	48611	50.5%	49.16%	31.218	36.787	87.85%
CB9615GI ^{tmRNA-1}	<i>Escherichia coli</i> O55:H7 str. CB9615	3258465-3313225	3258138..3258500	G2583_3276 3310496..3311680	54761	50.5%	50.98%	30.733	57.167	98.98%
CB9615GI ^{tmRNA-2}	<i>Escherichia coli</i> O55:H7 str. CB9615	3313207-3337896	3258138-3258500	G2583_3278 3313387..3314616	24690	50.5%	36.00%	30.733	121.065	100%
11368GI ^{tmRNA}	<i>Escherichia coli</i> O26:H11 str. 11368	3604870-3659140	3604541..3604903	ECO26_3718 3656440..3657615	54271	50.7%	49.95%	32.341	53.074	99.038%

续表 2

基因岛	菌株	位置	整合位置	整合酶或转座酶	大小 (bp)	基因组的 GC(%)	基因岛的 GC(%)	基因组的平均 δ^* (长度是 20000)	δ^* ($\times 1000$)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
IAI39 GI ^{tmRNA}	<i>Escherichia coli</i> IAI39	2933149-2949474	2932781..2933151	ECIAI39_2840 2949057..2949323	16326	50.6%	48.85%	27.965	55.687	92.994%
042GI ^{tmRNA}	<i>Escherichia coli</i> 042	3011732-3038120	3011377..3011746	EC042_2820 3011871..3013142	26389	50.6%	37.67%	30.561	100.890	99.495%
IHE3034GI ^{tmRNA}	<i>Escherichia coli</i> IHE3034	3053522-3096826	3053187..3053557	ECOK1_3021 3093462..3094634	43305	50.7%	49.40%	28.681	45.757	96.552%
DH1GI ^{tmRNA}	<i>Escherichia coli</i> DH1	1105776-1127824	1127808..1128169	EcDH1_1053 1126362..1127603	22049	50.8%	46.15%	25.956	65.168	100%
DH10BGI ^{tmRNA}	<i>Escherichia coli</i> DH10B	2845725-2867773	2845380..2845742	2845946-2847187	22049	50.8%	46.15%	25.988	64.638	99.528%
BW2952GI ^{tmRNA}	<i>Escherichia coli</i> BW2952	2639767-2661820	2639427..2639789	BWG_2379 2639993..2641234	22054	50.8%	46.15%	28.681	64.969	99.517%
55989GI ^{tmRNA}	<i>Escherichia coli</i> 55989	2979078-2994099	2978723..2979171	EC55989_2909 2979298..2980539	15022	50.7%	43.24%	30.295	88.308	98.542%
MG1655GI ^{tmRNA}	<i>Escherichia coli</i> str. K-12 substr. MG1655	2753960-2776008	2753615..2753977	b2622 2754181..2755422	22049	50.8%	46.15%	26.202	65.169	100%
IAI1GI ^{tmRNA}	<i>Escherichia coli</i> IAI1	2822896-2828309	2822541..2822911	ECIAI1_2748 2827827..2828084	5414	50.8%	48.82%	26.477	71.937	93.548%
TW14359GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. TW14359	3534937-3564626	3534580..3534949	ECSP_3593 3554858..3556171	29690	50.5%	43.02%	31.369	68.874	99.462%
EC4115GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. EC4115	3580058-3604399	3579710..3580072	ECH74115_3892 3600133..3601446	24342	50.5%	41.92%	31.697	82.673	100%
SakaiGI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. Sakai	3475950-3500163	3475507..3475962	ECs3512 3497212..3498417	24214	50.5%	41.82%	31.312	81.023	98.678%
EDL933GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 EDL933	3545799-3567453	3545346..3545811	Z3943 3563184..3564509	21655	50.4%	40.40%	31.743	80.816	99.611%
UMN026GI ^{tmRNA}	<i>Escherichia coli</i> UMN026	3048845-3080038	3048490..3048927	ECUMN_2945 3049026..3050273	31194	50.7%	46.13%	29.291%	61.671	96.988%
8739GI ^{tmRNA}	<i>Escherichia coli</i> ATCC 8739	1146183-1157075	1157062..1157423	EcoIC_1063 1155680..1156876	10693	50.9%	50.10%	26.999	43.731	80.69%
E24377AGI ^{tmRNA}	<i>Escherichia coli</i> E24377A	2906274-2944005	2905926..2906288	EcE24377A_2905 2906374..2907684	37732	50.6%	41.70%	28.32	72.742	100%
FAAGI ^{tmRNA}	<i>Escherichia sp.</i> 3_2_53FAA	cont1.29 33305-76575	76541..76903	36634-35462	43271	50.7% ^b	49.39%	28.681	45.830	96.552%
z3032GI ^{tmRNA}	<i>Cronobacter turicensis</i> z3032	3373855-3407526	3373536..3373898	Ctu_32540 3405674..3406693	33672	57.4%	53.84%	27.069	33.393	90.769%
ICC168GI ^{tmRNA}	<i>Citrobacter rodentium</i> ICC168	2730394-2775336	2730065..2730429	ROD_26311 2771293..2772459	44943	54.7%	53.16%	32.046	53.900	94.915%
Ech586GI ^{tmRNA}	<i>Dickeya dadantii</i> Ech586	818286-848199	848157..848519	Dd586_0727 818382..819473	29914	53.6%	53.32%	27.509	57.258	98.137%
35469GI ^{tmRNA}	<i>Escherichia fergusonii</i> ATCC 35469	447996-459401	459319..459756	EFER_0450 458001..459200	11406	49.9%	48.42%	27.781	44.915	85.075%

续表 3

基因岛	菌株	位置	整合位置	整合酶或转座酶	大小 (bp)	基因组的 GC(%)	基因岛的 GC(%)	基因组的平均 δ^* (长度是20000)	δ^* ($\times 1000$)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
638GI ^{tmRNA}	<i>Enterobacter sp.</i> 638	3375108-3386193	3374760..3375119	Ent638_3101 3375316..3376503	11086	53.0%	49.13%	27.581	56.621	93.103%
BAA-894GI ^{tmRNA} _A	<i>Cronobacter sakazakii</i> ATCC BAA-894	589583-601981	602001..602363	ESA_00630 600617..601816	12399	56.8%	46.85%	29.023	71.149	95.184%
78578GI ^{tmRNA}	<i>Klebsiella pneumoniae</i> subsp. pneumoniae MGH 78578	3231639-3259335	3231284..3231655	KPN_02943 3231820..3233067	27697	57.5%	48.04%	27.24	42.981	95.288%
Ech703GI ^{tmRNA}	<i>Dickeya dadantii</i> Ech703	3774856-3780584	3774515..3774877	Dd703_3223 3775059..3776300	5729	55.0%	44.74%	27.509	81.719	96.324%
20103GI ^{tmRNA}	<i>Pantoea ananatis</i> LMG 20103	3268145-3293642	3267797..3268158	PANA_2912 3268343..3269542	25498	53.7%	46.68%	24.27	55.637	98.361%
EIB202GI ^{tmRNA}	<i>Edwardsiella tarda</i> EIB202	2877724-2883742	2877367..2877806	ETA_E_2742 2877923..2879119	6019	59.7%	47.05%	34.01	75.198	88.782%
1012GI ^{tmRNA}	<i>Shigella dysenteriae</i> 1012	gcontig_11126037-61985-118514-124356	124343..124713	120279..120782	5843	51.2%	49.91%	24.431	80.795	96.921%
TTO1GI ^{tmRNA}	<i>Photorhabdus luminescens</i> subsp. laumondii TTO1	4006395-4085215	4006061..4006424	plu3487 4084051..4085147	78821	42.8%	42.42%	31.710	27.032	81.944%
PB1GI ^{tmRNA}	<i>Yersinia pseudotuberculosis</i> PB1/+	1340081-1357049	1356955..1357398	YPTS_1211 1355589..1356830	16969	47.5%	42.95%	26.681	62.185	96.739%
SodalisGI ^{tmRNA}	<i>Sodalis glossinidius</i> str. 'morsitans	3072001-3109919	3071644..3072027	SG1805 3072575..3073552	37919	54.7%	49.63%	28.818	44.935	95.455%

表 S2 在大肠杆菌中 tmRNA 位点的可变区分析

菌株	<i>smpB</i>	<i>SsrA</i> (tmRNA)	可变区	基因岛区和残余可变区	基因组的 GC %	残余可变区的 GC%	残余可变区的 δ^* ($\times 1000$)	基因组中比残余可变区的 δ^* 值更低的区域所占的百分率
<i>Escherichia coli</i> ED1a	3015037..3015519	3015727..3016113	3016098..3051096	ED1aGI ^{tmRNA}				
<i>Escherichia coli</i> S88	2797225..2797707	2797915..2798301	2798286..2875468	S88GI ^{tmRNA-1} S88GI ^{tmRNA-2}				
<i>Escherichia coli</i> APEC O1	2914010..2914492	2914700..2915070	2915223..2958157	APECO1GI ^{tmRNA}				
<i>Escherichia coli</i> UTI89	2898984..2899466	2899674..2900044	2900197..2943130	UTI89GI ^{tmRNA}				
<i>Escherichia coli</i> CFT073	3017482..3017964	3018172..3018542	3018695..3066942	CFT073GI ^{tmRNA}				
<i>Escherichia coli</i> IHE3034	3052497..3052979	3053187..3053557	3053710..3096644	IHE3034GI ^{tmRNA}				
<i>Escherichia coli</i> O55:H7 str. CB9615	3257441..3257923	3258138-3258500	3258501..3348776	CB9615GI ^{tmRNA-1} CB9615GI ^{tmRNA-2} 3337897..3348776	50.5%	42.58%	69.720	94.141%

续表 1

菌株	<i>smpB</i>	<i>SsrA</i> (tmRNA)	可变区	基因岛区和 残余可变区	基因组的 GC %	残余可变区的 GC%	残余可 变区的 δ^* ($\times 1000$)	基因组中比残 余可变区的 δ^* 值更低的区域 所占的百分率
<i>Escherichia coli</i> O26:H11 str. 11368	3603844.. 3604326	3604541.. 3604903	3604917.. 3667152	11368GI ^{tmRNA} 3659141..3667152	50.7%	39.42%	83.347	94.225%
<i>Escherichia coli</i> IA139	2932091.. 2932573	2932781.. 2933151	2933303.. 2954887	IA139GI ^{tmRNA} 2949475.. 2954887	50.6%	36.47%	96.771	97.785%
<i>Escherichia coli</i> 042	3010687.. 3011169	3011377.. 3011746	3011747.. 3049019	042GI ^{tmRNA} 3038121.. 3049019	50.6%	39.02%	86.376	99.286%
<i>Escherichia coli</i> DH1	1128384.. 1128866	1127808.. 1128169	1094977.. 1127805	DH1GI ^{tmRNA} 1094977..1105775	50.8%	42.31%	79.234	98.364%
<i>Escherichia coli</i> str. K-12 substr. DH10B	2844683.. 2845165	2845380.. 2845742	2845744.. 2878572	DH10BGI ^{tmRNA} 2867774..2878572	50.8%	42.30%	78.460	98.152%
<i>Escherichia coli</i> BW2952	2638730.. 2639212	2639427.. 2639789	2639791.. 2672619	BW2952GI ^{tmRNA} 2661821-2672619	50.8%	42.31%	78.927	98.582%
<i>Escherichia coli</i> 55989	2978033- 2978515	2978723.. 2979171	2979094.. 2999401	55989GI ^{tmRNA} 2994100-2999401	50.7%	35.97%	88.829	94.856%
<i>Escherichia coli</i> str. K-12 substr. MG1655	2752918.. 2753400	2753615.. 2753977	2753979.. 2786807	MG1655GI ^{tmRNA} 2776009.. 2786807	50.8%	42.31%	79.131	99.301%
<i>Escherichia coli</i> IA11	2821851.. 2822333	2822541.. 2822911	2823062.. 2833621	IA11GI ^{tmRNA} 2828310..2833621	50.8%	35.84%	102.858	98.529%
<i>Escherichia coli</i> O157:H7 str. TW14359	3533890.. 3534372	3534580.. 3534949	3534950.. 3569933	TW14359GI ^{tmRNA} 3564627.. 3569933	50.5%	36.25%	99.175	97.022%
<i>Escherichia coli</i> O157:H7 str. EC4115	3579013.. 3579495	3579710.. 3580072	3580073.. 3615208	EC4115GI ^{tmRNA} 3604400..3615208	50.5%	42.62%	67.124	67.124
<i>Escherichia coli</i> O157:H7 str. Sakai	3474905.. 3475387	3475507.. 3475962	3475965.. 3510972	SakaiGI ^{tmRNA} 3500164.. 3510972	50.5%	42.62%	67.392	92.126%
<i>Escherichia coli</i> O157:H7 EDL933	3544754.. 3545236	3545346.. 3545811	3545814.. 3578262	EDL933GI ^{tmRNA} 3567454.. 3578262	50.4%	42.58%	67.262	91.781%
<i>Escherichia coli</i> UMN026	3047800.. 3048282	3048490.. 3048927	3048865.. 3092139	UMN026GI ^{tmRNA} 3080039.. 3092139	50.7%	43.44%	61.601	92.774%
<i>Escherichia coli</i> ATCC 8739	1158120.. 1157638	1157062.. 1157423	1135384.. 1157060	8739GI ^{tmRNA} 1135384.. 1146182	50.9%	42.36%	79.905	98.178%
<i>Escherichia coli</i> E24377A	2905229.. 2905711	2905926.. 2906288	2906290.. 2949309	E24377AGI ^{tmRNA} 2944006.. 2949309	50.6%	35.94%	100.304	97.228%
<i>Escherichia coli</i> B str. REL606	2676191.. 2676673	2676881.. 2677254	2677252.. 2683466	No	50.8%	38.78%	103.558	98.925%
<i>Escherichia coli</i> HS	2795173.. 2795655	2795869.. 2796231	2796384.. 2806230	No	50.8%	40.08%	73.164	96.815%
<i>Escherichia coli</i> O111:H- str. 11128	3315480.. 3315962	3316177.. 3316539	3316540.. 3352326	No	50.6%	50.89%	53.973	95.333%
<i>Escherichia coli</i> O103:H2 str. 12009	3243520.. 3244002	3244217.. 3244579	3244580.. 3251308	No	50.7%	36.83%	106.017	98.022%
<i>Escherichia coli</i> O127:H6 str. E2348/69	2998326.. 2998808	2999022.. 2999384	2999386.. 3014716	No	50.6%	41.15%	80.018	96.904%
<i>Escherichia coli</i> SE11	3013728.. 3014210	3014418.. 3014787	3014788.. 3022802	No	50.8%	39.36%	89.002	98.851%
<i>Escherichia coli</i> SMS-3-5	2850460.. 2850942	2851150.. 2851520	2851671.. 2862401	No	50.5%	42.76%	69.447	94.28%
<i>Escherichia coli</i> SE15	No	2679357.. 2679731	No	No				
<i>Escherichia coli</i> 536	2762649.. 2763131	2763339.. 2763709	No	No				

表 S3 在沙门氏菌中 tmRNA 位点的可变区分析

菌株	<i>smpB</i>	<i>SsrA</i> (tmRNA)	可变区	基因岛区和残余可 变区	基因组 的 GC %	残余可变区 的 GC%	残余可变 区的 δ^* ($\times 1000$)	基因组中比 残余可变区 的 δ^* 值更低的 区域所占 的百分率
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Agona str. SL483	2769236.. 2769718	2786701.. 2787062	2787103.. 2824445	SL483GI ^{tmRNA-1} SL483GI ^{tmRNA-2}				
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Choleraesuis str. SC-B67	2843616.. 2844143	2861122.. 2861483	2861493.. 2870655	SC-B67GI ^{tmRNA} 2865466..2870655	52.2%	45.38%	71.953	90.066%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL254	2828948.. 2829430	2846103.. 2846465	2846529.. 2892611	SL254GI ^{tmRNA-1} SL254GI ^{tmRNA-2} 2888743..2892611	52.2%	45.83%	61.283	77.466%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. LT2	2826484.. 2826966	2843947.. 2844308	2844349.. 2915973	LT2GI ^{tmRNA-1} LT2GI ^{tmRNA-2} 2905309..2915973	52.2%	43.76%	76.857	96.044%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. D23580	2853141.. 2853623	2870605.. 2870966	2870970.. 2908958	D23580GI ^{tmRNA} 2898276..2908958	52.2%	43.78%	75.029	94.518%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Heidelberg str. SL476	2834681- 2835163	2852147.. 2852509	2852512.. 2890479	SL476GI ^{tmRNA} 2879817..2890479	52.1%	43.71%	74.968	93.668%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Dublin str. CT_02021853	2905328- 2905855	2922578.. 2922939	2922980.. 2960670	02021853GI ^{tmRNA} 2956229..2960670	52.5%	45.14%	93.663	94.495%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi A str. AKU_12601	2639817.. 2640299	2657286.. 2657647	2657688.. 2727331	12601GI ^{tmRNA-1} 12601GI ^{tmRNA-2} 2723459..2727331	52.2%	45.77%	63.539	80.034%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi A str. ATCC 9150	2644281.. 2644763	2661750.. 2662111	2662152.. 2731795	9150GI ^{tmRNA-1} 9150GI ^{tmRNA-2} 2727923..2731795	52.2%	45.78%	63.401	81.234%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium SL1344	2837672.. 2838154	2855136.. 2855498	2855538.. 2938549	SL1344GI ^{tmRNA-1} SL1344GI ^{tmRNA-2} SL1344GI ^{tmRNA-3} 2927877-2938549	52.2%	43.76%	232.121	93.305%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Javiana str. GA_MM04042433	117417.. 117944	100440.. 100054	54936.. 100031	04042433GI ^{tmRNA} 54936..68586	52.8%	38.67%	81.253	100%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Virchow str. SL491	No	2342979.. 2342593	2270818.. 2342570	SL491GI ^{tmRNA-1} SL491GI ^{tmRNA-2} 2270818..2274695	52.8%	45.80%	91.190	84.444%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. Ty2	2717618.. 2718100	2734487.. 2734848	2734850.. 2754903	Ty2GI ^{tmRNA} 2745482..2754903	52.1%	52.86%	90.569	94.488%
<i>Salmonella enterica</i> subsp. <i>arizonae</i> serovar 62:z4:z23:-	233968.. 234495	216595.. 216957	192601.. 216594	62:z4GI ^{tmRNA} 192601..198156	51.4%	50.56%	74.729	86.594%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Schwarzengrund str. CVM19633	2779083- 2779856	2796548.. 2796910	2796911.. 2823095	CVM19633GI ^{tmRNA} 2819240..2823095	52.2%	45.62%	57.975	73.055%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi B str. SPB7	2840163.. 2840690	2857678.. 2858039	2858045.. 2880114	SPB7GI ^{tmRNA-1} SPB7GI ^{tmRNA-2} 2876284..2880114	52.1%	45.86%	52.260	62.461%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Newport str. SL317	677715.. 678242	660739.. 660367	645903.. 660366	SL317GI ^{tmRNA} 645903..649708	52.8%	45.80%	88.723	86.957%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Paratyphi C strain RKS4594	2825527.. 2826009	2842992- 2843358	2843363.. 2851187	RKS4594 GI ^{tmRNA} 2847337..2851187	52.2%	45.37%	91.168	95.624%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. 14028S	2880401.. 2880883	2897865.. 2898226	2898230.. 2936208	14028SGI ^{tmRNA} 2925536..2936208	52.2%	43.77%	75.658	95.175%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Enteritidis str. P125109	2778594- 2779076	2796059- 2796420	2796423.. 2800903	No	52.2%	45.12%	94.017	95.12%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Gallinarum str. 287/91	2769465.. 2769947	2786912.. 2787273	2787273.. 2791756	No	52.2%	45.07%	93.842	94.894%
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	2742393- 2742857	2759262.. 2759623	2759664.. 2769046	No	52.1%	52.93%	92.164	95.508%

表 S4 68 个基因岛的整合酶和边界的分析

基因岛	菌株	attL/attR	整合酶或转座酶	CDD (PDB)
SL483GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Agona str. SL483	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAA/ATGT <u>AGGAATTT</u> CGGACGCG GGTTCAACTCCCGCCAGCTCCAC- CAAA	SeAg_B2882 2819809..2820834	HP1 整合酶
SL254GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Newport str. SL254	AGGATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAAT/AGGATGT <u>AGGAATTT</u> CGGA CGCGGGTTCAACTCCCGCCAGCTCC ACCAAAT	SNSL254_A2951 2876970..2877980	DNA 断裂-重联酶(HP1 整合酶)
LT2GI ^{tmRNA-1}	<i>Salmonella typhimurium</i> LT2	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAA/ATGT <u>AGGAATTT</u> CGGACGCG GGTTCAACTCCCGCCAGCTCCAC- CAAA	STM2739 2876815..2877840	HP1 整合酶
02021853GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Dublin str. CT_02021853	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATAAAAACAAGGGGTTAC- GTGAAAAC- GTAGCCCTTTTT/ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATATAACAAGGGGTTAC- GTGAAAACGTAGCCCTTTTT	SeD_A3066 2955002..2956027	HP1 整合酶
12601GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi A str. AKU_12601	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATAAAAACAATGGGTTAC- GTGAAAAC- GTAGCCCTTTTTCTT/ATGT <u>AGGAA</u> <u>TTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATAAAAACAAGGGGTTAC- GTGAAAACGTAGTCCCTTTTTGTT	Sspa2423 2690004..2691029	HP1 整合酶
12601GI ^{tmRNA-2}		ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCA/ATGT <u>AGGAATTT</u> CGGACGCGGG TTCAACTCCCGCCAGCTCCACCA	Sspa2424 2691562..2692320 SSPA2447 2721297..2722310	DNA 断裂-重联酶 (HP1 整合酶) DNA 断裂-重联酶 (HP1 整合酶)
9150GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi A str. ATCC 9150	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATAAAAACAATGGGTTAC- GTGAAAAC- GTAGCCCTTTTTCTT/ATGT <u>AGGAA</u> <u>TTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAATAAAAACAAGGGGTTAC- GTGAAAACGTAGTCCCTTTTTGTT	Spa2600 2694468..2695493	HP1 整合酶
9150GI ^{tmRNA-2}		ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCA/ATGT <u>AGGAATTT</u> CGGACGCGGG TTCAACTCCCGCCAGCTCCACCA	Spa2601 2696026..2696784 2725761-2726465	DNA 断裂-重联酶 (HP1 整合酶) DNA 断裂-重联酶 (HP1 整合酶)
04042433GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Javiana str. GA_MM04042433	TGGTGGAGCTGGCGGGAGTT- GAACCCGCGTCCG <u>AAATTCCT</u> TAC- TAC- AT/TGGTGGAGCTGGCGGGAGTTGA ACCCGCGTCCG <u>AAATTCCT</u> TACAT	69953..70978	HP1 整合酶

续表 1

基因岛	菌株	attL/attR	整合酶或转座酶	CDD (PDB)
SL491GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Virchow str. SL491	GCTGGCGGGAGTT- GAACCCGCGTCCG <u>AAATTCCTAC</u> - <u>TAC</u> - AT/GCTGGCGGGAGTTGAACCCGCGT CCG <u>AAATTCCTAC</u> AT	2275837..2276850 2276146.. 2276850	DNA 断裂-重联酶 (HP1 整合酶) DNA 断裂-重联酶 (HP1 整合酶)
SL491GI ^{tmRNA-2}		AACAAAAAGGGGCTACGTTTTAC- GTAACCCCTTGTTTTATTGGTG- GAGCTGGCGGGAGTT- GAACCCGCGTCCG <u>AAATTCCTAC</u> - <u>TAC</u> - AT/AACAAAAAGGGGCTACGTTTTCA CGTAACCCCTTGTTTTATTGGTG- GAGCTGGCGGGAGTT- GAACCCGCGTCCG <u>AAATTCCTAC</u> AT	2307274..2308299	HP1 整合酶
SL1344GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Typhimurium SL1344	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAAT/ATGT <u>AGGAATTT</u> CGGACGCG GGTTCAACTCCCGCCAGCTCCAC- CAAAT	SL1344_2709 2887497..2888522	HP1 整合酶
ED1aGI ^{tmRNA}	<i>Escherichia coli</i> ED1a	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAA/ATGT <u>AGGAATTT</u> CGGACGCGG GTTCAACTCCCGCCAGCTCCACCAA	ECED1_3109 3049316..3050332	HP1 整合酶
S88GI ^{tmRNA-1}	<i>Escherichia coli</i> S88	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAA/ATGT <u>AGGAATTT</u> CGGACGCGG GTTCAACTCCCGCCAGCTCCACCAA	ECS88_2851 2829972..2830991	DNA 断裂-重联酶 (HP1 整合酶)
z3032GI ^{tmRNA}	<i>Cronobacter turicensis</i> z3032	ATGT <u>AGGAATTT</u> CGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAAA/ATGT <u>AGGAATTT</u> CGGACGCG GGTTCAACTCCCGCCAGCTCCAC- CAAA	Ctu_32540 3405674..3406693	HP1 整合酶
Ech586GI ^{tmRNA}	<i>Dickeya dadantii</i> Ech586	AAGTGGTGGAGCTGGGGGGAGTT- GAACCCCGTCCG <u>AAATTCCTC</u> - <u>TACA</u> /AATTGGTGGAGCTGGGGGA GTTGAACCCCGTCCG <u>AAATTCCTC</u> - <u>TACA</u>	Dd586_0727 818382..819473	HP1 整合酶
morsitansGI ^{tmRNA}	<i>Sodalis glossinidius</i> str. 'morsitans'	<u>CGGGTTCAAATCCCGCCAGCTCCAC</u> CA/ <u>CGGGTTCAAATCCCGCCAGCTCC</u> ACCA	SG1805 3072575..3073552	转座酶
S88GI ^{tmRNA-2}	<i>Escherichia coli</i> S88	TTCGGAC- <u>GCGGGTTCAACTCCCGCCAGC/TTTG</u> GACT <u>GCGGGTTCAACTCCCGCCAGC</u>	ECS88_2917 2872287..2873459	PhiCTX 整合酶
APECO1GI ^{tmRNA}	<i>Escherichia coli</i> APEC O1	TTCGGAC- <u>GCGGGTTCAACTCCCGCCAGC/TTTG</u> GACT <u>GCGGGTTCAACTCCCGCCAGC</u>	APECO1_3866 2954976..2956148	PhiCTX 整合酶
UTI89GI ^{tmRNA}	<i>Escherichia coli</i> UTI89	TTCGGAC- <u>GCGGGTTCAACTCCCGCCAGC/TTTG</u> GACT <u>GCGGGTTCAACTCCCGCCAGC</u>	UTI89_C3009 2939949..2941121	PhiCTX 整合酶
CFT073GI ^{tmRNA}	<i>Escherichia coli</i> CFT073	GGAC- <u>GCGGGTTCAACTCCCGCCAGCTCCA</u> CCAATCAT/GGACA <u>GCGGGTTCAACTC</u> <u>CCCGCCAGCTCCACCAATCAT</u>	c3204 3064143..3065315	PhiCTX 整合酶
FAAGI ^{tmRNA}	<i>Escherichia sp.</i> 3_2_53FAA	GCTGGCGGGAGTTGAACCCGAG- TCCAAA/GCTGG <u>GCGGGAGTTGAACCC</u> <u>GCGTCCGAA</u>	36634..35462	PhiCTX 整合酶

续表 2

基因岛	菌株	attL/attR	整合酶或转座酶	CDD (PDB)
CB9615GI ^{tmRNA-1}	<i>Escherichia coli</i> O55:H7 str. CB9615	TTTCGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCAC- TTT/TTTCGGACACGGGGTTCAACTCC CGCCAGCTCCACCACTTT	G2583_3276 3310496..3311680	PhiCTX 整合酶
11368GI ^{tmRNA}	<i>Escherichia coli</i> O26:H11 str. 11368	TCGGAC- GCGGGTTCAACTCCCGCCAGCTCCA C/TCGGGCACGGGGTTCAACTCCCGCC AGCTCCAC	ECO26_3718 3656440..3657615	PhiCTX 整合酶
IHE3034GI ^{tmRNA}	<i>Escherichia coli</i> IHE3034	GGAC- GCGGGTTCAACTCCCGCCAG/GGACT CGGGTTCAACTCCCGCCAG	ECOK1_3021 3093462..3094634	PhiCTX 整合酶
ICC168GI ^{tmRNA}	<i>Citrobacter rodentium</i> ICC168	TTTCGGAC- GCGGGTTCAACTCCCGCCAGCTCCA CCA/TTTCGGACCGCTGGTTCAACTCC CGCCAGCTCCACCA	ROD_26311 2771293..2772459	PhiCTX 整合酶
TTO1GI ^{tmRNA}	<i>Photobacterium luminescens</i> subsp. laumondii TTO1	AC- GCGGGTTCAACTCCCGCCAGCTCCA CCA/ACGCGGGTTCAACTCCCGCCAG CTCCACCA	plu3487 4084051..4085147	PhiCTX 整合酶
SL1344GI ^{tmRNA-2}	<i>Salmonella enterica</i> subsp. enterica serovar Typhimurium SL1344	CCGCCAGCTCCAC- CAAA/CCGCCAGCCACCAAAA	SL1344_2712 2890073..2900377	P4 整合酶
Ty2GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi str. Ty2	AACTCCGCCAGCTCCAC- CA/AACTCCGCCAGCCACCA	t2647 2735202..2736233	P4 整合酶
DH1GI ^{tmRNA}	<i>Escherichia coli</i> DH1	TTGGTGGGCTGGCG- GAG/TTGGTGGAGCTGGCGGAG	EcDH1_1053 1126362..1127603	P4 整合酶
DH10BGI ^{tmRNA}	<i>Escherichia coli</i> DH10B	CTCCGCCAGCTCCAC- CAA/CTCCGCCAGCCACCAA	2845946..2847187	P4 整合酶
BW2952GI ^{tmRNA}	<i>Escherichia coli</i> BW2952	TTCAACTCCGCCAGCTCCAC- CAA/TTCAACTCCGCCAGCCACCA A	BWG_2379 2639993..2641234	P4 整合酶
MG1655GI ^{tmRNA}	<i>Escherichia coli</i> str. K-12 substr. MG1655	CTCCGCCAGCTCCAC- CAA/CTCCGCCAGCCACCAA	b2622 2754181..2755422	P4 整合酶
E24377AGI ^{tmRNA}	<i>Escherichia coli</i> E24377A	CCGCCAGCTCCACAAA- TA/CCGCCAGCTCAACAATA	EcE24377A_2905 2906374..2907684	P4 整合酶
SL254GI ^{tmRNA-2}	<i>Salmonella enterica</i> subsp. enterica serovar Newport str. SL254	TTCAACTCCGCCAGCT/TTCAACT CCGCCAGCT	SNSL254_A2952 2878326..2879519	P4 整合酶
SPB7GI ^{tmRNA-1}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi B str. SPB7	CCGCCAGCTCCAC- CAAA/CCGCCAGCCACCAA	SPAB_03412 2858224..2859417	P4 整合酶
SPB7GI ^{tmRNA-2}		TATTCAGACTCCGCCAGCCCA/TATT CAGATTCCGCCAGCTCA	SPAB_03429 2868695..2869891	P4 整合酶
SL317GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Newport str. SL317	AGCTGGCGGAG- TCTGA/AGCTGGCGGGAGTTGA	658992..660185	P4 整合酶
35469GI ^{tmRNA}	<i>Escherichia fergusonii</i> ATCC 35469	TGGTGGGCTGGCGG/TGGTGGAGCTG GCGG	EFER_0450 458001..459200	P4 整合酶
BAA-894GI ^{tmRNA}	<i>Cronobacter sakazakii</i> ATCC BAA-894	TTT- GGTGGGCTGGCGG/TTTGGTGGAGCT GGCGG	ESA_00630 600617..601816	P4 整合酶
78578GI ^{tmRNA}	<i>Klebsiella pneumoniae</i> subsp. pneumoniae MGH 78578	CCGCCAGC/CCGCCAGC	KPN_02943 3231820..3233067	P4 整合酶

续表 3

基因岛	菌株	<i>attL/attR</i>	整合酶或转座酶	CDD (PDB)
638GI ^{tmRNA}	<i>Enterobacter sp.</i> 638	<u>CCCCCAGCTCCA/CCCCCAGCCCA</u>	Ent638_3101 3375316..3376503	P4 整合酶
20103GI ^{tmRNA}	<i>Pantoea ananatis</i> LMG 20103	<u>CCGCCAGCTCCAC-</u> <u>CA/CCGCCAGCCACCA</u>	PANA_2912 3268343..3269542	P4 整合酶
62:z4GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. arizonae serovar 62:z4,z23:--	<u>TGGTGGGCTGGCGT/TGGTGGAGCTG</u> <u>GCGG</u>	SARI_00233 215199..216470 SARI_00216 200549..201604	P4 整合酶 DNA 断裂-重联酶
CVM19633GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Schwarzengrund str. CVM19633	<u>CCGCCAGCTCCAC-</u> <u>CA/ACGCCAGCCACCA</u>	SeSA_A2892 2797034..2798305	P4 整合酶
SL483GI ^{tmRNA-2}	<i>Salmonella enterica</i> subsp. enterica serovar Agona str. SL483	<u>CCGCCAGC/ACGCCAGC</u>	SeAg_B2883 2821164..2821745	DNA 断裂-重联酶
14028SGI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi- murium str. 14028S	<u>CCGCCAGC/ACGCCAGC</u>	STM14_3302 2898394..2899458	P4 整合酶
SL1344GI ^{tmRNA-3}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi- murium SL1344	<u>CCGCCAGC/ACGCCAGC</u>	SL1344_2712 2890073..2891269	P4 整合酶
LT2GI ^{tmRNA-2}	<i>Salmonella typhimurium</i> LT2	<u>CCGCCAGC/ACGCCAGC</u>	STM2740 2878169..2879233	P4 整合酶
D23580GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi- murium str. D23580	<u>CCGCCAGC/ACGCCAGC</u>	STMMW_27071 2871134..2872379	P4 整合酶
SL476GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Heidelberg str. SL476	<u>CCGCCAGC/ACGCCAGC</u>	SeHA_C2911 2852676..2853740	P4 整合酶
UMN026GI ^{tmRNA}	<i>Escherichia coli</i> UMN026	<u>CCGCCAGC/ACGCCAGC</u>	ECUMN_2945 3049026..3050273	P4 整合酶
042GI ^{tmRNA}	<i>Escherichia coli</i> 042	<u>CCGCCAGC/ACGCCAGC</u>	EC042_2820 3011871..3013142	P4 整合酶
EC4115GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. EC4115	<u>CCGCCAGC/ACGCCAGC</u>	ECH74115_3892 3600133..3601446	噬菌体整合酶
SakaiGI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. Sakai	<u>CCGCCAGC/ACGCCAGC</u>	ECs3492 3481763..3482260 ECs3493 3482457..3482804 ECs3511 3495897..3497210 ECs3512 3497212..3498417	Rve 整合酶 Rve 整合酶 噬菌体整合酶 DNA 断裂-重联酶
EDL933GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 EDL933	<u>CCGCCAGC/ACGCCAGC</u>	Z3943 3563184..3564509	噬菌体整合酶
PB1GI ^{tmRNA}	<i>Yersinia pseudotuberculosis</i> PB1+	<u>TGGTGCTGGGG-</u> <u>GA/TGGAAGCTGGGGGA</u>	YPTS_1211 1355589..1356830	P4 整合酶
Ech703GI ^{tmRNA}	<i>Dickeya dadantii</i> Ech703	<u>TTCAACTCCCGCCAGCTCCAC-</u> <u>CA/TTCAACTTCCCCAGCCACCA</u>	Dd703_3223 3775059..3776300	P4 整合酶
8739GI ^{tmRNA}	<i>Escherichia coli</i> ATCC 8739	<u>TGGTGGGCTGG-</u> <u>CAG/TGGTGGAGCTGGCGG</u>	EcolC_1063 1155680..1156876	P4 整合酶
55989GI ^{tmRNA}	<i>Escherichia coli</i> 55989	<u>CCGCCAGCTCCAC-</u> <u>CAAA/CTGCCAGCCACCAAA</u>	EC55989_2909 2979298..2980539	P4 整合酶

续表 4

基因岛	菌株	attL/attR	整合酶或转座酶	CDD (PDB)
CB9615GI ^{tmRNA-2}	<i>Escherichia coli</i> O55:H7 str. CB9615	<u>CCGCCAGG</u> TCCAC- CA/ <u>CTGCCAGC</u> CCACCA	G2583_3278	P4 整合酶
			3313387..3314616	
			G2583_3284	DNA 断裂-重联酶
			3320630..3320818	
			G2583_3295	噬菌体整合酶
			3333629..3334954	
EIB202GI ^{tmRNA}	<i>Edwardsiella tarda</i> EIB202	<u>CCGCCAGG</u> TCCAC- CA/ <u>CCGCCAGG</u> CCACCA	ETA_E_2742	P4 整合酶
			2877923..2879119	
RKS4594GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Paratyphi C strain RKS4594	<u>CCGCCAGCTCCACCAA</u> A- TA/ <u>CAGCCAGCTCCACCAA</u> ATA	SPC_2804	转座酶
			2843670..2843879	
			SPC_2805	转座酶
			2843906..2844745	
IAI1GI ^{tmRNA}	<i>Escherichia coli</i> IAI1	<u>CCGCCAGCTC</u> / <u>CCGCCAGCTC</u>	ECIAI1_2746	Rve 整合酶
			2825616..2826449	
			ECIAI1_2748	丝氨酸类重组酶
			2827827..2828084	
TW14359GI ^{tmRNA}	<i>Escherichia coli</i> O157:H7 str. TW14359	<u>CCGCCAGT</u> TCCAC- CA/ <u>CTGCCAGC</u> CCACCA	ECSP_3593	噬菌体整合酶
			3554858..3556171	
SC-B67GI ^{tmRNA}	<i>Salmonella enterica</i> subsp. enterica serovar Cholerae-suis str. SC-B67	<u>CCGCCAGCTCCAC</u> - CAAAT/ <u>CAGCCAGCTCCACCAA</u> AT	SC2698	转座酶
			2862036..2862875	
1012GI ^{tmRNA}	<i>Shigella dysenteriae</i> 1012	<u>TTGGTGGGCTGGC</u> / <u>TTGGTGGAGCTG</u> <u>GC</u>	120279..120782	转座酶
IAI39GI ^{tmRNA}	<i>Escherichia coli</i> IAI39	CAA/CAA	ECIAI39_2840	丝氨酸类重组酶
			2949057..2949323	

双下划线的碱基是 HP1 整合酶的绑定位点; 虚线的碱基是 PhiCTX 整合酶的绑定位点; 单下划线的碱基是预测的整合酶的酶切位点.

表 S5 同 Enterobacteria phage Fels-2 同源的边界为 RyeB 和 tRNA 基因的基因岛的特征

基因岛	菌株	位置	整合位点	大小 (bp)	整合酶或转座酶	基因组的 GC%	基因岛的 GC%	基因组的平均 δ^* (长度是 20000)	δ^* ($\times 1000$)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
Et1/99GI ^{Met}	<i>Erwinia tasmaniensis</i> Et1/99	626788-657198 ETA_05520-ETA_t023	tRNA ^{Met} (657153..657229)	31411	ETA_05530 628263..629291	53.7%	53.59%	25.659	25.980	75.591%
E2348/69GI ^{RybB}	<i>Escherichia coli</i> O127:H6 str. E2348/69	852346-886514 E2348C_0798- E2348C_ncR09	RybB (886501..886579)	34169	E2348C_0798 852446..853498	50.6%	50.32%	28.58	41.343	91.724%
CFT073GI ^{RybB}	<i>Escherichia coli</i> CFT073	908857-942364 c0932-c0979	RybB (942351..942429)	33508	c0932 909332..910363	50.5%	50.31%	31.218	42.156	88.462%
CT18GI ^{Phe}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi str. CT18	4409517-4543123 STYt072_ STYt073	tRNA ^{Phe} (4409511..4409574)	133607	STY4680 4541654..4542913	52.1%	49.68%	31.141	86.855	100%

续表 1

基因岛	菌株	位置	整合位点	大小 (bp)	整合酶或转座酶	基因组的 GC%	基因岛的 GC%	基因组的平均 δ^* (长度是20000)	δ^* ($\times 1000$)	基因组中比基因岛的 δ^* 值更低的区域所占的百分率
Ty2GI ^{Phe}	<i>Salmonella enterica</i> subsp. enterica serovar Typhi str. Ty2	4394167-4526025 <u>t4226</u> -t4372	t4372 tRNA ^{Phe} (4525975.. 4526050)	131859	t4371 4524556.. 4525815 t4369 4522608.. 4522886	52.1	49.75%	31.278	87.436	100%
HSGI ^{RybB}	<i>Escherichia coli</i> HS	916800-948289 <u>EcHS_A0909</u> -EcHS_A4738	EcHS_A4738 RybB (948276.. 948354)	31490	EcHS_ A0909 917273.. 918304	50.8%	52.00%	26.758	48.994	97.279%
13047GI ^{Met}	<i>Enterobacter cloacae</i> subsp. cloacae ATCC 13047	4519470-4551289 ECL_tRNA069-E <u>CL_04443</u>	tRNA ^{Met} (4519441.. 4519516)	31820	ECL_04443 (4550149.. 4551234)	54.8%	53.82%	26.438	40.560	93.413%
8739GI ^{RybB}	<i>Escherichia coli</i> ATCC 8739	3023189-3058101 <u>EcolC_2750</u> <u>EcolC_2795</u>	RybB (3023124.. 3023202)	34913	EcolC_2757 3030569.. 3031135	50.9%	50.54%	26.999	43.020	96.296%
IAI39GI ^{No}	<i>Escherichia coli</i> IAI39	1882151-1905453 ECIAI39_1790- ECIAI39_1822	No	23303	ECIAI39_ 1820 1904043.. 1904933	50.6%	54.84%	27.965	60.191	97.727%