
* 专论与综述 *

原核生物的基因组短重复序列

覃筱婷* 周俊初 李阜棣

(华中农业大学农业微生物重点实验室 武汉 430070)

众所周知, 真核生物的基因组比原核生物的大得多, 也复杂得多, 已知在真核生物基因组中, 有大量的重复序列存在。原核生物的基因组较为简单: 一般只有一个环状双链的染色体, 其大小亦不超过 10^6 bp, 如此小的基因组要保证原核生物的生长繁殖及其它功能, 其非编码区的重复序列必须维持最低值。在细菌的基因组中, DNA 序列多以单拷贝的形式存在, 但这并不意味着原核生物基因组中没有多拷贝基因。例如, rRNA 基因、tRNA 基因和插入因子 (IS) 等, 它们对于原核生物维持快速而大量的基因表达和对基因表达进行精细的调节有着重要的作用。近年来的许多研究表明, 原核生物基因组中还存在着另一类短重复序列。它们与上述的 rRNA 基因等多拷贝序列不同, 大小不超过 200 bp, 反向重复, 可以转录成 RNA 但不编码蛋白。这类短重复序列广泛分布在原核生物特别是细菌的基因组中, 一般位于基因之间的区域^[1]。

目前, 已报道的重复序列有 REP 序列 [Repetitive Extragenic Palindrome 基因外重复的回文结构或 PU (Palindrome Units, 回文单元)], ERIC 序列 (Enterobacteria Repetitive Intergenic Consensus 肠杆菌基因间的重复的一致序列), Ng-REP 序列, Dr-REP 序列, Mp-REP 序列, STRR 序列 (Short Tandemly Repeated Repetitive 短的串联的重复序列), 本文将对此作一简要介绍。

1 REP 序列

该序列最早是由 Gilson^[2] 在 *E. coli* 及 Higgins^[3] 在 *Salmonella typhimurium* 中发现的, 是一段 38 bp 的反向重复序列。它能组成一个稳定的茎环结构, 存在于基因组中, 并在序列中心区域形成一 5 bp 可变环。在此序列的对称臂上, 如果一侧臂的某一碱基发生变化, 则在其另一臂上也相应地发生变化, 这种变

化是维持其回文结构所必需的^[1-4]。

REP 序列一般位于多顺反子的操纵子 (Operon) 中顺反子之间的区域, 也可位于 3' 端可转录的非编码区。估测在 *E. coli* 中, REP 序列约有 500 ~ 1000 拷贝, 占整个基因组的 1%^[7]。Gilson^[5-6] 的研究还发现, REP 序列在 *E. coli* 基因组中常串联成簇 (cluster), 簇中还可包括其它类型的重复因子, 他将这一特殊的 REP 序列排列方式称为“细菌中零散分布的镶嵌因子 BIMEs (bacterial interspersed mosaic element)”, 估计在 *E. coli* 约有 200 个左右的 BIMEs, 平均六个基因有一个。

用 DNA-RNA 杂交和 PCR 技术进行的研究表明: 与 REP 序列同源的短重复序列在绝大部分细菌中都存在, 它们在不同的菌株、种和属间具有高度的保守性^[1-7,8], 目前已证实基因组中有 REP 序列存在的细菌有: *Agrobacterium* spp.; *Bordetella pertussis*; *Brucella* spp.; *Cyanobacteria* spp.; *Deinococcus radiodurans*; *Escherichia coli*; *Halobacterium* spp.; *Leptospira interrogans*; *Mycobacterium tuberculosis*; *Mycoplasma pneumoniae*; *Myxococcus xanthus*; *Neisseria gonorrhoeae*; *Neisseria meningitidis*; *Pneumocystis carinii*; *Rhizobium* spp.; *Rhodocyclidium raneliae*; *Salmonella typhimurium*; *Spiroplasma* spp.。

有关 REP 序列的功能目前尚不很清楚, 由于它可以形成稳定的茎环结构。Higgins^[7] 推测其功能可能包括:

- (1) 转录后可以稳定上游 mRNA 的结构;
- (2) 稳定的上游 mRNA 才具有翻译活性;
- (3) 通过调整上游 mRNA 的稳定性来调整基因的表达;

* 目前地址: 中国农科院植保所 北京 100094

1995-09-12 收稿

(4) 在多顺反子的操纵子中, REP 序列负责不同基因表达水平的调控;

(5) 影响染色体的结构并与染色体的重组有关。

所有上述功能均与 REP 序列能以茎环形式存在于基因组有关。由于 REP 序列常位于 3' 末端, 曾有人认为它可能是转录的终止子。事实上, 人们却未能在 REP 序列中发现典型的终止密码子; 将它克隆在启动子与结构基因之间, 也没有发现基因表达水平的下降, 以上说明它并不是终止子。REP 序列在操纵子内各基因间的定位是随机的, 因此可以推测它们只是在数量水平上影响操纵子内各基因的表达^[7]。

总之, REP 序列是遍布于细菌整个基因组的、在许多细菌中都发现存在的短的重复序列。

2 ERIC 序列

在细菌的基因组中, 特别是肠杆菌群 (Enterobacteria) 中, 普遍存在的短重复序列称为 ERIC 序列。它是一段长为 126 bp 的反向重复序列^[2, 9]。其特点与 REP 序列相似, 也定位在基因组内可转录的非编码区或与转录有关的区域。ERIC 序列的中心是一对双重反向重复序列, 曾有人推测该序列所转录的 mRNA 可以形成茎环结构, 但迄今尚未确证。已知若在该序列中心的对称臂上有一个碱基改变, 则在其对称臂上相对应的碱基也相应地发生改变^[10]。

Hulton^[11] 比较了 7 株 *E. coli* 及 7 株 *Salmonella typhimurium* 中的 ERIC 序列后, 发现 ERIC 序列在基因组内的不同位点、种内的不同菌株间乃至种间都具有高度的保守性。已知不同菌株间的 ERIC 序列在染色体上的定位不是固定的, 在菌株间相对应的染色体位置上, ERIC 序列是否存在, 不会引起该位点附近序列的改变。

ERIC 序列既可以以完整的序列形式存在于基因组中, 也可以以部份序列存在。已知: *S. typhimurium* cysB 位点的 ERIC 序列的 3' 端是缺失的; *Klebsiella pneumoniae* fim 位点只包含 ERIC 序列靠近 3' 末端的一半; *Yersinia pseudotuberculosis* pel 位点只包含有 ERIC 序列的 5' 端和 3' 端, 中间序列被别的序列替换, 其长度正好与 ERIC 序列中间的长度相符^[11]。

与 REP 序列相似, ERIC 序列的功能也尚未完全了解。只知 ERIC 序列都是可转录的, 但它们与编码区的序列没有固定的关系。在一个多顺反子的操纵子

中, 迄今还没有发现该序列的存在与操纵子功能之间有相关性^[11]。

REP 序列与 ERIC 序列是普遍存在于原核生物中的短重复序列, 它们之间有许多相似性^[11] (如均定位在可转录的非编码区、反向重复等)。根据这两种序列的长短与碱基排列的差异基本上可以排除它们在进化上的相关性; 根据它们在染色体上定位的不同又可以排除它们在功能上的相似性。现在人们通常认为它们都是细菌基因组中两种独立的 DNA 序列, 都是“自私”(Selfish) DNA, 通过基因转化在细菌基因组中普遍传播并遗传下来^[11]。

3 其它短重复序列

除了上述的两种短重复序列外, 近年来在原核生物基因组内, 又陆续报道了一些与这两种序列完全不同的短重复序列。

3.1 Ng-REP 序列: *Neisseria gonorrhoeae* 中, Correia^[9] 通过杂交方法发现了一段长为 26 bp 的短重复序列。比较同一基因组中 4 个不同位点该重复序列后发现: 26 bp 的序列中有 5 个碱基是可变的, 21 个碱基是固定的。用 21 bp 的寡核苷酸链探针与其它菌株杂交, 发现在 *Neisseria meningitidis* 中杂交也呈阳性。用根据此序列设计的引物进行 PCR 扩增, 结果比用 RAPDs 方法表现出更强的菌株与种的专一性。因此, 有人推测, Ng-REP 序列可能是另一类广泛分布的短重复因子^[12]。

3.2 Dr-REP 序列: James^[13] 等用一个克隆的 DNA 片段做探针与 *Deinococcus radiodurans* 的总 DNA 文库进行杂交, 得到一个相当复杂的杂交模型。这一结果表明在该菌的基因组中, 存在着多拷贝该克隆片段或该克隆片段的部份序列, 用该探针在 *D. radiodurans* SARK 菌株中也鉴定出一段重复 Dr-REP 序列。该序列的长度是可变的, 约为 150 ~ 192 bp, 其中心也是一对双重反向重复的序列, 用 SARK 菌株的克隆片段与 *E. coli* 的基因文库杂交, 大约有 3% 的克隆表现为阳性, 说明 Dr-REP 序列在 *E. coli* 中普遍存在。

3.3 Mx-REP 序列: Gilson 和 Fujita^[17] 用探针杂交法从 *Mycobacterium xanthans* 中也鉴定了一个 87 bp 的反向重复序列, 该序列有一个反向重复中心^[12]。有意思的是 Mx-REP 序列常位于 rpoD 基因的下游, 该位点在 *E. coli* 中则恰好是 REP 序列分布的位点。

3.4 Mp-REP 序列: Wenzel^[14] 研究 *Mycoplasma pneumoniae* 后发现该菌株基因组中, 至少有 5 种长度为 1.1 ~ 2.2 kb, 拷贝数为 8 ~ 10 的重复序列。除了上述的 5 种重复序列之外, 还发现两种更短重复序列, 即 300 bp 的 RepMpl 和 400 bp 的 SDCI 序列, 经测定该菌的基因组长度仅为 840 kb, 但它的重复序列含量却高达 65%。

3.5 STRR 序列: *Cyanobacterium cacoethrix* 的基因组内, 有 3 种 7 个核苷酸为一组的短重复序列, 称为 STRR 序列。通过探针杂交试验研究后发现该序列在 *Cyanobacterium* 其它种也有存在^[8]。

从上面的介绍可知: 重复序列不仅在真核生物中广泛分布, 在原核生物特别是在细菌中也广泛分布着一些短重复序列。虽然这些短重复序列的准确功能还不完全确定, 它们为何广泛分布在基因组内? 在基因组内的不同位点, 不同的基因组间, 它们的相似值是多少? 所有这些问题目前都还不能回答, 有待人们去进一步研究。对于它们的传播及繁殖, 一种普遍推测的机制是: 它们作为一种“自私”DNA, 通过基因转化在基因组间扩散的。虽然人们对这些短重复序列了解还不深入, 但短重复序列在原核生物及真核生物中都广泛分布的事实表明: 短重复序列在基因组的结构及生物进化方面可能起着某种重要作用。

4 应用

短重复序列目前已广泛应用于分子生物学及生物的指纹图谱(fingerprint)研究。由于 REP 及 ERIC 序列在原核生物基因组中的随机及广泛分布, 人们已根据此两类序列设计出相应的引物, 用供试生物的基因组为模板, 进行 PCR 扩增。只要引物能与基因组内的重复序列结合, 并且结合的位置和方向都处于 Taq 酶所能涉及的范围之内, 就可以扩增出具有种乃至菌株的特异性指纹图谱, 此技术也称为分子标记技术^[11-15]。

短重复序列的还可用于分析插入序列(如转座子)引起的插入突变。根据 REP 序列设计出引物对的一半, Tn5 专一性序列设计出引物对的另一半, 选择好适宜的引物扩增方向, 进行 PCR 扩增就可以得到一段包括有插入序列与 REP 序列附近序列的扩增 DNA 产

物。用该扩增片段为探针, 与供试基因组的基因文库杂交, 则可以快速鉴定出 Tn5 插入染色体的位点。目前该技术已用于分析细菌的基因图谱。

现在的研究表明, 原核生物短重复序列的发现为微生物的鉴定及基因分析提供了快速简便的新方法, 具有深远的意义。

参考文献

- [1] James R L, Weinstock G M. J Bacteriol, 1992, 174: 4525 ~ 4529.
- [2] Gilson E, Clement M, Brutiag D et al. EMBO J, 1984, 3: 1417 ~ 1421.
- [3] Higgins C F, Ames G F L, Barnes W M et al. Nature, 1982, 298: 760 ~ 762.
- [4] Stern M J, Ames G F L, Smith N H et al. Cell, 1984, 37: 1015 ~ 1026.
- [5] Gilson E, Hofneng M. Nucleic Acids Res, 1990, 18: 3941 ~ 3952.
- [6] Gilson E, Bachelier S, Perrin S et al. Res Microbiol, 1990, 141: 1103.
- [7] Higgins C F, McLaren R S, Newbury S F. Gene, 1988, 72: 3 ~ 14.
- [8] Versalovic J, Koeuth T, Lupski J R. Nucleic Acids Res, 1991, 19: 6823.
- [9] Correia F F, Inouye S, Inouye M. J Bacteriol, 1986, 167: 1009 ~ 1015.
- [10] Shalples G J, Lloyd R G. Nucleic Acids Res, 1990, 18: 6503 ~ 6508.
- [11] Hulton C S, Higgins C F, Shalp P M. Mol Microbiol, 1991, 5: 825 ~ 834.
- [12] Giseile N, Lindrostrom K. Syst Appl Microbiol, 1994, 17: 265 ~ 273.
- [13] James V, Maria S F. Methods in Mol Cellular Biology, 1994, 5: 25.
- [14] Wenzel R, Herrmann R. Nucleic Acids Res, 1988, 16: 5275 ~ 5279.
- [15] Lennon E, Gutman P D. J Bacteriol, 1991, 173: 2137 ~ 2140.
- [16] Subramanian P S. PCR Methods Application, 1992, 1: 187 ~ 194.
- [17] Gilson E, Perrin D, Clement J M. et al. FEBS Lett, 1986, 206: 323 ~ 328.
- [18] Welsh J, McClelland M. Nucleic Acids Res, 1991, 19: 5275 ~ 5279.