

群落分析中的 16S rRNA 及其基因 16S rDNA 优化扩增

周琳, 张杰*

(中国农业科学院植物保护研究所, 植物病虫害生物学国家重点实验室, 北京 100193)

摘要: 本文较为完整地综述了 16S rRNA 在细胞中的地位和作用、转录组成、其全长基因中可变区与保守区的位置、在不同种类中的拷贝数、二级结构, 及其在群落研究过程中 PCR 扩增产生的 16S rRNA 的异常序列与解决方法; 以期为更多的科研工作者提供可靠参考和借鉴, 提高以 16S rRNA 为靶分子的群落研究结果的准确性和真实性。

关键词: 16S rRNA; 可变区; 保守区; 异常序列

中图分类号: Q938 **文献标识码:** A **文章编号:** 0001-6209 (2010) 01-0007-08

随着分子生物学和分子生态学的迅速发展, 使得人们对自然界的研究不再依赖于传统的培养方法, 大量新方法和新技术的产生和广泛应用, 也使人们对自然界的理解和认识也达到了前所未有的广度和深度。其中, 以 16S rRNA 为靶分子进行细菌的系统进化研究^[1]和各种环境内的微生物群落结构组成研究, 包括污水处理系统^[2], 沼气发酵池^[3-4], 植物根部土壤^[5], 口腔^[6], 肠道^[7]等也得以迅速发展和深入。虽然, 16S rRNA 在此类研究中起着极为重要的核心作用, 但目前为止, 未见有关 16S rRNA 全面信息以及其在进行群落分析时在最初的 PCR 扩增中容易产生问题及解决措施的相关报道。本文将有关 16S rRNA 的基础知识以及在 16S rRNA 全长基因的扩增中产生的问题和解决措施进行了总结, 以期为群落研究者提供科学的参考, 以便于进一步提高实验结果的准确性和真实性。

1 16S rRNA 和 16S rDNA

1.1 16S rRNA 在细胞内的地位和作用

核糖体是蛋白质合成的场所, 是极其重要的细

胞器之一。在原核生物中, 核糖体由沉降系数为 50S 的大亚基和沉降系数为 30S 的小亚基组成, 其中的 50S 大亚基由 5S rRNA 和 23S rRNA 与 34 种不同的蛋白质组成; 而 16S rRNA 也作为核糖体的重要组成成分, 与另外 21 种蛋白质组成 30S 的核糖体小亚基^[8]。

1.2 16S rRNA 与 16S rDNA 的关系

16S rDNA 指的是基因组中编码核糖体 16S rRNA 分子的对应的 DNA 序列, 也就是 16S rRNA 的编码基因; 16S rRNA 指的是 16S rDNA 的转录产物。如前所述, 16S rRNA 是核糖体中的一个重要组成部分。一般进行系统进化分析或是对某特定环境进行细菌群落结构分析时, 所分析的对象都是 16S rDNA, 因为 DNA 提取容易, 也比较稳定, 但人们从习惯上还是往往以 16S rRNA 来进行描述。在基因组上, 16S rRNA 基因与 5S rRNA 和 23S rRNA 的各自编码基因组成一个转录单元, 共同转录。16S rRNA 与 23S rRNA 之间为间隔 tRNA, 5S rRNA 后面为收尾 tRNA(图 1)^[8]。

基金项目: 国家“973 项目”(2007CB109203)

* 通信作者。Tel: +86-10-62815921; Fax: +86-10-62812642; E-mail: jzhang@ippcaas.cn

作者简介: (1981-) 女, 河北秦皇岛人, 博士研究生, 主要从事转基因生物生态安全性研究。E-mail: zl382@yahoo.com.cn

收稿日期: 2009-07-09; 修回日期: 2009-09-17

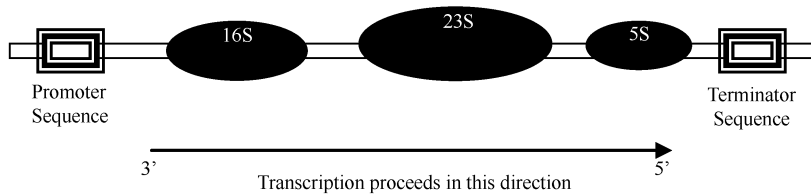


图1 16S rRNA 与 5S rRNA 和 23S rRNA 存在于同一转录单元

Fig.1 16S rRNA, 5S rRNA and 23S rRNA consist of the same transcription unit.

1.3 16S rRNA 的可变区与保守区

1.3.1 16S rRNA 可变区与保守区的分布: 16S rRNA 全长基因为 1542 bp, 由 9 个可变区和 10 个保守区组成, 其中保守区反映了生物物种间的亲缘关系, 而可变区则表明物种间的差异。Ashelford 等采用 Pintail 软件对 RDP 数据库中的 4383 个细菌 16S rRNA 的序列进行比对分析, 得到了一级结构中的 16S rRNA 全长基因中 9 个可变区与 10 个保守区的分布示意图^[9]。

Chakravorty 等人也通过对 110 种细菌包括常见的血源性病体、疾控指定特选病原体以及一些环境微生物的 16S rRNA 序列的比对分析, 确定了 9 个可变区的位置和长度(表 1), 并且指出不同的可变区适合不同菌群的鉴定分析^[10]。其中, V1 区适合分离鉴别链球菌(*Streptococcus*), V3 区可以分离鉴别大多数细菌, V6 区则可以用来分离鉴别除肠内细菌(*Enterobacteriaceae*)、埃希氏菌(*Escherichia*)、志贺氏菌(*Shigella*)和沙门氏菌(*Salmonella*)以外的大部分细菌, 而在实际应用中, V8 区往往与 V6 区联合应用。V2、V4、V5、V7、V9 则由于序列变异程度等原因都不太适合用于群落分析和细菌分离。

表 1 16S rRNA 可变区在 16S rRNA 基因上的位置

Table 1 Position of the variable regions in 16S rRNA

Variable regions	Position/bp	Length/bp
V1	69 - 99	31
V2	137 - 242	106
V3	433 - 497	65
V4	576 - 682	107
V5	822 - 879	58
V6	986 - 1043	58
V7	1117 - 1173	57
V8	1243 - 1294	52
V9	1435 - 1465	31

1.3.2 16S rRNA 的二级结构及操纵元的拷贝数:

早在 20 世纪 70 年代, Woese 等就已着手对 16S rRNA 序列的进行深入研究。通过对大量细菌 16S

rRNA 序列的同源性比较, 得出了 16S rRNA 的二级结构图。16S rRNA 具有稳定的二级结构, 并且含有明显的保守区和可变区, 这也为后来的系统进化研究和群落结构分析从分子水平上提供了理论基础。

此外, 不同细菌内含有的 16S rRNA 操纵元拷贝数差别很大, 目前的研究结果显示, 从 1 - 15 个不等(表 2)。本文作者对全基因组已测序的苏云金芽胞杆菌(*Bacillus thuringiensis* serovar *konkukian* str. 97 - 27, NC_005957 和 *Bacillus thuringiensis* str. Al Hakam, NC_008600) 基因组序列分析发现其 16S rRNA 的拷贝数为 14。不同的细菌中的 16S rRNA 基因的拷贝数并不相同, 而且没有任何表型与 16S rRNA 基因的拷贝数相关。一般认为, 16S rRNA 基因高拷贝数是快速生长所必需的。然而在 16S rRNA 操纵元为单拷贝的细菌中, 也观察到了倍增时间很短的现象^[20]。这说明 16S rRNA 基因的拷贝数并不是只在快速生长过程中起作用。Klappenbach 在研究中发现, 16S rRNA 基因的拷贝数与细菌感受周围环境中资源是否可用的反应时间快慢相关。从土壤中分离的细菌涂布营养丰富的培养基上能够快速生长的细菌种类中, 16S rRNA 基因的拷贝数平均为 5.5 个, 而生长缓慢的细菌种类中, 16S rRNA 基因的拷贝数平均仅为 1.4 个。此外, 如果在土壤中加入除草剂 2, 4-Dichlorophenoxyacetic acid(2, 4-二氯苯氧基乙酸, 2, 4-D)后, 土壤中能够降解 2, 4-D 的细菌数量变为优势种群, 该类细菌的 16S rRNA 基因的拷贝数平均为 5.4 个而未作 2, 4-D 处理的土壤中优势细菌种群的 16S rRNA 基因的拷贝数平均为 2.7 个^[21]。这也进一步说明, 16S rRNA 基因高拷贝数的细菌, 对环境中可利用资源的变化能够更快作出反应和适应。由此可见, 16S rRNA 在细菌的生存过程中起着双重作用: 首先是大多数情况下细菌迅速生长所需要的, 其次, 在细菌对外界环境中可利用资源能够做出快速反应中起着重要作用。

表 2 不同细菌中的 16S rRNA 拷贝数

Table 2 Variability of copy numbers of 16S rRNA in different bacteria

The copy numbers of 16S rRNA	Example	Reference
1	<i>Bradyrhizobium japonicum</i>	[11]
2	<i>Helicobacter pylori</i>	[12]
3	<i>Dichelobacter nodosus</i>	[13]
4	<i>Neisseria meningitidis</i>	[14]
6	<i>Lactococcus lactis</i>	[15]
7	<i>Escherichia coli</i>	[16]
8	<i>Vibrio cholerae</i>	[17]
10	<i>Bacillus subtilis</i>	[18]
14	<i>Bacillus thuringiensis</i>	This review
15	<i>Clostridium paradoxum</i>	[19]

2 16S rRNA 与生物进化系统分类及环境群落结构研究

20 世纪 70 年代末,美国伊利诺斯大学的 Woese 等人对大量微生物和其他生物进行 16S 和 18S rRNA 的核苷酸测序,并比较同源性水平,提出了著名的三域学说,从此奠定了 rRNA 在细菌系统进化中的重要地位,rRNA 甚至被人们称作是细胞中的“活化石”^[22]。

2.1 选取 16S rRNA 进行生物进化和系统分类研究的主要原因

首先,16S rRNA 普遍存在于一切原核生物细胞内,生理功能重要而又恒定,因此可用它比较在进化中的相互关系;其次,编码 16S rRNA 的基因即 16S rDNA 十分稳定,非常保守,在细胞中的含量较高,约占细菌总 rRNA 含量的 80%,而且容易提取;此外,5S rRNA 的全长基因约为 120 bp,所包含的信息量太少,不足以作系统进化分析,而 23S rRNA 的全长基因约为 2.9 kb,分子量明显偏大,并且碱基变异速率要较 16S rRNA 快,对于区分较远亲缘关系的种类不适用。16S rRNA 的全长基因为 1542 bp,分子量相对适中,所含信息量大,易于分析。综合以上因素,16S rRNA 的进化具有良好的时钟性质,在结构与功能上具有高度的保守性,故选取 16S rRNA 进行生物进化和系统分类研究^[22]。

2.2 群落分析中 16S rRNA 基因 16S rDNA 的优化扩增

在以 16S rRNA 为靶标分子的群落结构研究时分为两种策略,一种为从环境的总 DNA 中直接扩增 16S rRNA 的全长基因,再结合其他方法进一步分析,例如常用的克隆建库和测序分析,末端限制片段长度多态性 (terminal restriction fragment length

polymorphism, T-RFLP) 分析、限制片段长度多态性 (restriction fragment length polymorphism, RFLP) 分析和扩增片段长度多态性 (amplified fragment length polymorphism, AFLP) 分析等;另一种为利用 16S rRNA 中某一特定可变区,如荧光原位杂交 (fluorescence *in situ* hybridization, FISH) 分析即根据 16S rRNA 中某一特定区设计寡核苷酸探针,再与环境总 DNA 进行杂交,或者首先扩增 16S rRNA 中某一特定可变区,然后结合变性梯度凝胶电泳 (DGGE),进一步对环境中的群落结构组成进行分析。然而无论是直接的 PCR 扩增 16S rRNA 全长基因,还是 PCR-RFLP、PCR-DGGE 等相结合,在 PCR 扩增中出现的异常序列,包括嵌合基因 (Chimera), 异源双链 (Heteroduplex), 突变 (点突变和缺失突变) 都将直接影响到后续的分析结果,影响试验结果的准确性和真实性。下面将针对上述 PCR 过程中出现的偏差而导致的异常扩增产物及其应对措施做详细介绍。

2.2.1 嵌合基因 (Chimera): Chimera 翻译为凯米拉或奇美拉,源自于希腊神话中,是拥有羊身、狮头和蛇尾的一个杂合怪物。在生物学中,Chimera 翻译为嵌合体,指的是来自不同个体的生物分子、细胞或组织被结合在了一起成为一个整体或生物体。在以 16SrRNA 为靶分子的群落结构分析中,均以混合 DNA 为模板进行 PCR 扩增,极易产生 chimera,但也可以通过优化 PCR 反应参数和条件进行降低和避免。

(1) 嵌合基因形成机理

在 PCR 反应过程中,Chimera 的形成是由于延伸不完全而产生了不完整的扩增产物,这些不完整的扩增产物在下一轮的扩增循环中,与序列相近的其他同源模板退火,继续延伸另一种模板的序列而得到的一种杂合的 DNA 片段(图 2)。

1989 年,Shuldiner 在研究非洲爪蟾 *Xenopus laevis* 非等位的 2 个前胰岛素原基因时,首次发现 chimera。在 PCR 产物中,除了纯合的 I 型和 II 型前胰岛素原基因,还含有 5'端为 I 型前胰岛素原基因而 3'端为 II 型前胰岛素原基因或者 5'端为 II 型前胰岛素原基因和 3'端由 I 型前胰岛素原基因组成的杂合基因^[23]。由此看来,chimera 基因至少有两种来自不同基因的部分序列组成。在后来其它领域的群落研究中,诸如,Choi 等在通过克隆 16S rRNA 部分序列研究牙龈下菌斑样品的群落结构时,也发现存在有 8.6% 的 chimera 序列^[24];Meyerhans 等在

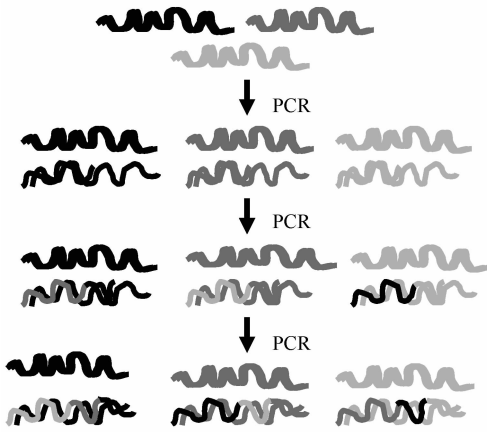


图2 嵌合基因的形成机理示意图

Fig. 2 The formation of chimera.

对2个不同的 HIV-1 的 *tat* 基因进行研究时也检测到了 5.4% 的 chimera 基因^[25]。Brakenhoff 在对人类 γ -晶状体球蛋白基因家族 (γ -crystallin gene family) 相关基因的转录研究时,也得到了 chimera cDNA 克隆^[26]。此外,在其他的领域研究中也分别发现的在 PCR 扩增中存在 chimera 的现象。

(2) 减少 Chimera 形成的措施

在通过 PCR 扩增 16S rRNA 全长基因并结合建库测序分析研究特定环境中的群落结构时,由于 chimera 的干扰,在结果中出现自然界中根本不存在的细菌种类,从而影响结果的准确性,也使试验结果产生偏差。同时在目前公共使用的 16S rRNA 数据库中 (RDP, GenBank, EMBL 等) 含有一定数量的错误序列。在很多研究中也发现了 16S rRNA 的嵌合基因^[27-28]。为此,研究者们投入了大量的精力进行研究,以期最大程度地降低和减少 chimera 的产生^[29-32],并且提高对 chimera 的检测能力^[33-34]等。Hugenholtz 等人使用 partial treeing analysis,对已有的 16S rRNA 数据库进行检测,发现含有 21 个门内种类和 18 个门之间种类形成的 chimera 序列^[35],而 Ashelford 等人以 Pintail 对当时的 RDP 数据库中的已有 16S rRNA 进行分析时,更是进一步发现经保守估计,在 20 个已有 16S rRNA 至少有 1 个就是异常序列即非正常的 16S rRNA,也就是说 5% 的异常基因,其中, chimera 就占了 64.3%^[9]。按此比例计算,在 2009 年 7 月 28 日最新版本的 RDP 10.13 数据库 (<http://rdp.cme.msu.edu>) 所含有的 1,049,433 个 16S rRNA 序列中,52,472 个为异常基因,而 chimera 的数量则为 33,740 个。如此高比例和数目庞大的异常基因直接影响我们研究结果的准确性和真实性以及 RDP 数据库的参考价值,所以,我们

在实际的研究工作中必须采取相应措施最大程度地降低和减少 chimera 的产生:以确保试验结果的准确性和真实性。

(1) 尽量保持 DNA 模板的完整性,防止其降解。Paabo 报道,DNA 在受到诸如紫外线辐射、酶解或超声波破碎等引起的降解后, chimera 产生的频率有所上升,所以在环境总 DNA 提取过程中尽量采取温和方法,操作不要过于猛烈以便保持 DNA 的完整,防止其降解^[36]。Liesack 也报道,当以较低分子量为 4-6 kb 的两种严格嗜压菌的混合 DNA 进行 PCR 扩增时, chimera 的形成数量上升^[28];

(2) 延长延伸时间,最大程度上保证延伸中的引物-模板完成延伸,得到完整的目的链。Wang 等报道,PCR 扩增中的延伸时间从 2 min 增加至 5 min, chimera 的形成百分数略微有所降低^[32]。如果延伸时间增至原来的 6.7 倍则 chimera 的形成频率降低 2.7 倍。

(3) 减少循环数。chimera 形成的数量与 PCR 反应的循环数呈正相关,在以一对同源模板进行扩增,10 个循环和 20 个循环产生的 chimera 百分比分别为 4.8% 和 20.9%^[31]。

(4) 序列分析去除 chimera。

以上是采取各种措施尽量降低或避免在 PCR 过程中产生 Chimera,那么在进入以下一个环节的分析中,我们可以通过一些软件对得到的测序结果进行筛选分析,去除 chimera 序列,再进行聚类或建树分析,这样从 PCR 后进一步保证了结果的准确性和真实性。以下是研究者们常用的软件,方便实用。

Chimera _ Check^[37]、Bellerophon^[38]、Pintail^[9]、Ccode^[39] 充分利用这些软件对测序后的序列进行整理和筛选,尽可能去掉其中的 chimera 序列。

综上所述,在具体的试验研究中,要从 PCR 过程中和测序之后的结果分析双重环节上降低 chimera 的形成数量和有效去除 chimera,这样才能最大程度地保证实验结果的准确性和真实性。

2.2.2 异源双链 (Heteroduplex): 异源双链是指来自不同基因的两条 DNA 单链退火形成的杂合双链。与 chimera 相类似,在以混合 DNA 为模板的 PCR 扩增中容易产生,但也可以避免和降低。

(1) 异源双链的形成机理

在以混合 DNA 为模板的 PCR 扩增体系中,前期扩增中得到的大量目的基因片段,使反应进行到后期的循环时,引物量相对减少,模板量大量增加。此时,在退火过程中,新旧模板与引物退火的几率

大大降低,模板之间相互退火的几率大大增加,同源性较高的模板之间便相互退火形成异源双链(图2)。

1989年,Nagamine 等人在研究老鼠性别决定基因 mYfin 基因^[40]时发现了 heteroduplex 现象。在用聚丙烯酰胺凝胶电泳检测扩增于小鼠 mYin 基因上的同源序列 *Zfy-1* 和 *Zfy-2* 时发现,电泳图谱上除了 2 条目的条带 *Zfy-1* 和 *Zfy-2* 之外,还含有移动速度较慢的另外 2 条条带,经确认是 *Zfy-1* 和 *Zfy-2* 相互退火形成的异源杂合双链。此外,Ruano 等也以与灵长类动物中一段与 HOX2B 基因相关片段建立了 heteroduplex 的模式研究^[41]。

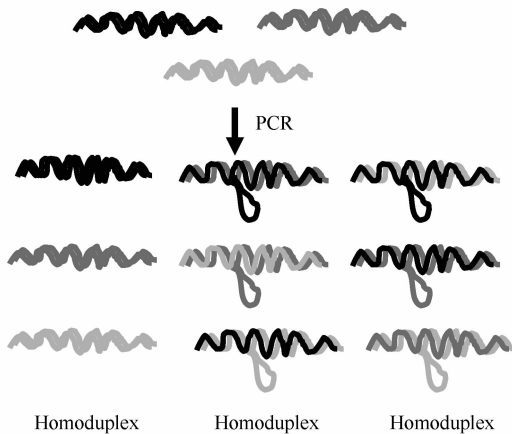


图3 异源双链的形成机理示意图

Fig. 3 The formation of heteroduplex.

(2)降低 heteroduplex 形成的措施

在研究某特定环境中的微生物群落结构时,以环境总 DNA 为模板,进行 16S rRNA 基因的扩增中,同样面临着上述的情况。亲缘关系较近的 16S rRNA 模板之间互相退火形成一个在自然界中根本不存在的异源双链。如果在测序时,分别从 5'端和 3'端进行测序就会得到 2 种不同的结果,在构建系统进化树或聚类分析是,同一个克隆便会处于不同的分枝内,从而影响试验结果的真实性和准确性。那么,在实际的试验过程中可以通过减少 PCR 反应的循环数、减低退火温度、增加反应 buffer 浓度、增加引物用量等措施减少 heteroduplex 的形成。Mark 等研究发现,heteroduplex 的形成很大程度上取决于 PCR 反应的扩增条件。PCR 反应 buffer 浓度提高一倍保持 Mg^{2+} 浓度保持 1.5 mmol/L 不变;引物浓度

由 0.5 $\mu\text{mol/L}$ 上调至 1.0 $\mu\text{mol/L}$;退火温度由 61 $^{\circ}\text{C}$ 降低至 55 $^{\circ}\text{C}$;扩增循环数由 25 降低至 20,在以上的 4 个反应条件中,如果从标准反应条件出发,采用其中的任何一个反应参数,heteroduplex 的形成数量适度下降,而当四种参数同时应用于 PCR 反应体系中时,heteroduplex 的形成数量显著降低^[42]。

此外,Thompson 还报道了一种行之有效的去除 heteroduplex 的措施—reconditioning PCR^[43]。即在 PCR 扩增之后,以 10 倍稀释的首轮 PCR 产物为模板,然后再以相同的 PCR 扩增条件进行 3 个循环的再扩增—reconditioning PCR,heteroduplex 的量可降低到原来的 1/50。此外,首轮 PCR 产物的稀释越高,heteroduplex 的去除效果越好,但在具体实验中也要考虑到产物量和后续的克隆问题,筛选到一个合适的稀释度。

2.2.3 突变 (Mutation): DNA 自身的复制和体外的 PCR 扩增中都存在着一定概率的突变。在群落分析的 PCR 扩增中,影响实验结果的主要包括点突变和缺失突变。

(1)点突变 (point mutation)

在 PCR 扩增反应中存在有一定概率的错配现象,这是 DNA 聚合酶的固有属性。从以往的研究来看,*Taq* 聚合酶的错配率在 $2.6 \times 10^{-5} - 3 \times 10^{-3}$ 之间^[44-45]; *rTth* DNA 聚合酶的错配概率大约为 1.4×10^{-3} ^[46];相对于没有校正活性的 DNA 聚合酶来说,具有 3'-5'外切酶活性的 *Pfu* DNA 聚合酶的错配概率则降低了 10 倍约为 1.6×10^{-6} ^[47]。如此计算下来,对于全长为 1542 bp 的 16S rRNA 基因 16S rDNA 的扩增中,最多会有 5 个碱基的错配,而导致 0.3% 的基因趋异度[▽]。

(2)缺失突变 (deletion mutation)

在 PCR 反应中,如果模板具有稳定的二级结构会导致扩增效率降低或扩增产物中含有缺失突变^[48-50]。而众所周知,16S rRNA 具有稳定的二级结构^[51]。为了减少和避免由于 16S rRNA 本身二级结构而引起的问题,Chou 等提出在 PCR 反应中加入 DNA 单链结合蛋白或是使用 Processivity 优于 *Taq* 酶的 DNA 聚合酶^[52]。减少 point mutation 形成的措施包括降低模板浓度和减少循环数。Qiu 等人报道在 PCR 反应中,每种模板的浓度由 10 ng 降至 100 pg 时,突变的概率由 17.3% 降低至 10.5%。相应

[▽]但作者在实际工作中发现,*Pfu* 等 DNA 聚合酶虽然保真性能力较强,但是扩增能力相对较差。以同一环境总 DNA 为模板,*Pfu* 和 *Taq* DNA 聚合酶分别扩增得到的特定片段(250bp)进行变性梯度凝胶电泳,*Taq* 酶扩增产物的电泳图谱较 *Pfu* 扩增电泳图谱要相对丰富,所以,要根据实际工作需要慎重选择相应的 DNA 聚合酶。

地降低模板浓度和减少循环数能够降低点突变的概率^[53]。此外,循环数的降低也能够一定程度上减少突变的产生。

3 展望

16S rRNA 基因分析是目前认识和探索细菌分类、种系发生以及特定环境群落结构和物种多样性的极其重要的窗口,这种不依赖于培养的核酸检测技术使得对自然环境中的微生物种类和数量的发现和研究的长足快速的发展。如果不能有效降低或消除 PCR 过程中产生的异常序列,就会对细菌的种系发生构建错误的系统进化树以及对特定环境中物种多样性产生错误认识。

近年来,越来越多的研究领域以 16S rRNA 为对象,对某种特定的环境系统进行研究和分析并用以解决实际问题。严兴等通过对工业装置和实验室装置中好氧池生物膜 16S rDNA 克隆文库的比较分析得知,工业装置好氧池中微生物群落中缺少适当的硝化菌群是造成其硝化能力低下的重要原因,而提高菌群中 *Nitrosomonas* 属和 *Nitrospira* 属的多度是解决工业装置硝化能力低下的关键^[54]。刘彬彬等也通过构建 16S rDNA 克隆文库,对实验室规模的用于降解喹啉的厌氧反应器的种子污泥和运行稳定的厌氧生物膜反应器的微生物区系组成进行了对比分析。结果显示,在驯化过程中, γ -*Proteobacteria* 亚纲与 *Desulfobacter postgatei* 种的微生物显著增加,这种动态变化表明这些细菌可能是在厌氧条件下对喹啉的降解起关键作用的微生物^[55]。在诸如上述应用中,16S rDNA 扩增的精确与否在实际问题的诊断中尤为关键,也决定着解决实际问题的中所采取措施的正确与否。

微生物群落结构和多样性是微生物生态学和环境科学研究的难点和热点,对于开发微生物资源,阐明微生物群落与其生境的关系、揭示群落结构与功能的联系从而指导微生物群落功能的定向调控具有重要意义^[56]。自从以 16S rRNA 基因作为群落结构研究的靶标分子,人类对自然界中微生物的认识逐渐走向全面,并不断深入,有益微生物不断得到合理、有效的利用。尽管如此,我们对自然界存在的微生物多样性的认识还远远不够,在今后相当长的时间内以 16S rRNA 基因为基础的群落结构研究方法仍然是我们揭示和明确环境微生物群落结构的主要手段,而且也将继续在进一步开发微生物资源为我们的生产和生活所用,包括目前废水和废

弃物的处理、电子污染等亟待解决的问题中继续发挥重要作用。

参考文献

- [1] Woese CR. Bacterial evolution. *Microbiological Reviews*, 1987, 51(2): 221-271.
- [2] Lydmark P, Almstrand R, Samuelsson K, et al. Effects of environmental conditions on the nitrifying population dynamics in a pilot wastewater treatment plant. *Environmental Microbiology*, 2007, 9(9): 2220-2233.
- [3] Malin C, Illmer P. Ability of DNA content and DGGE analysis to reflect the performance condition of an anaerobic biowaste fermenter. *Microbiological Research*, 2008, 163(5): 503-511.
- [4] Kim TW, Lee JH, Kim SE, et al. Analysis of microbial communities in doenjang, a Korean fermented soybean paste, using nested PCR-denaturing gradient gel electrophoresis. *International Journal of Food Microbiology*, 2009, 131(2-3): 265-271.
- [5] Kim MC, Ahn JH, Shin HC, et al. Molecular analysis of bacterial community structures in paddy soils for environmental risk assessment with two varieties of genetically modified rice, Iksan 483 and Milyang 204. *Journal of Microbiology and Biotechnology*, 2008, 18(2): 207-218.
- [6] Siqueira JF Jr, Rôças IN, Debelian GJ, et al. Profiling of root canal bacterial communities associated with chronic apical periodontitis from Brazilian and Norwegian subjects. *Journal of Endodontics*, 2008, 34(12): 1457-1461.
- [7] Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science*, 2005, 308: 5728.
- [8] 孙乃恩, 孙东旭, 朱德煦. 分子遗传学. 第一版. 南京: 南京大学出版社, 1997.
- [9] Ashelford KE, Chuzhanova NA, Fry JC, et al. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, 2005, 71(12): 7724-7736.
- [10] Chakravorty S, Helb D, Burday M, et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, 2007(69): 330-339.
- [11] Kuëndig C, Beck C, Hennecke H, et al. A single rRNA gene region in *Bradyrhizobium japonicum*. *Journal of Bacteriology*, 1995, 177, 5151-5154.
- [12] Tomb JF, White O, Kerivage AR, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 1997, 388: 539-547.

- [13] La Fontaine S, Rood JI. Organization of ribosomal RNA genes from the footrot pathogen *Dichelobacter nodosus*. *Microbiology*, 1996, 142, 889-899.
- [14] Klappenbach JA, Saxman PR, Cole JR, et al. rncdb, the ribosomal RNA operon copy Number database. *Nucleic Acids Research*, 2001, 29(1): 181-184.
- [15] Tulloch DL, Finch LR, Hillier AJ, et al. Physical map of the chromosome of *Lactococcus lactis* subsp. *lactis* DL11 and localization of six putative rRNA operons. *Journal of Bacteriology*, 1991, 173: 2768-2775.
- [16] Kiss A, Sain B, Venetianer P. The number of rRNA genes in *Escherichia coli*. *FEBS Letters*, 1977, 79: 77-79.
- [17] Heidelberg JF, Eisen JA, Nelson WC, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*, 2000, 406(6795): 477-483.
- [18] Stewart GC, Wilson FE, Bott KF. Detailed physical mapping of the ribosomal RNA genes of *Bacillus subtilis*. *Gene*, 1982, 19: 153-162.
- [19] Rainey FA, Ward-Rainey NL, Janssen PH, et al. *Clostridium paradoxum* DSM 7308^T contains multiple 16S rRNA genes with heterogeneous intervening sequences. *Microbiology*, 1996, 142: 2087-2095.
- [20] Zillig W. The order *Thermococcales*. New York: Springer-Verlag, 1992.
- [21] Klappenbach JA, Dumba JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, 2000, 66(4): 1328-1333.
- [22] 周德庆. 微生物学教程. 第二版. 北京: 高等教育出版社, 2002.
- [23] Shuldiner AR, Nirula A, Roth J. Hybrid DNA artifact from PCR of closely related target sequences. *Nucleic Acids Research*, 1989, 17(11): 4409.
- [24] Choi BK, Paster BJ, Dewhirst, FE, et al. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. *Infection and Immunity*, 1994, 62: 1889-1895.
- [25] Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. *Nucleic Acids Research*, 1990, 18: 1687-1691.
- [26] Brakenhoff RH, Schoenmakers JG, Lubsen NH. Chimeric cDNA clones: a novel PCR artifact. *Nucleic Acids Research*, 1991, 19(8): 1949.
- [27] Koczynski ED, Bateson MM, Ward DM. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Applied and Environmental Microbiology*, 1994, 60: 746-748.
- [28] Liesack W, Weyland H, Stackebrandt E. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed culture of strict barophilic bacteria. *Microbial Ecology*, 1991, 21: 191-198.
- [29] Qiu XY, Wu LY, Huang HS, et al. Evaluation of PCR-generated chimeras; mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, 2001, 67: 880-887.
- [30] Speksnijder A, Kowalchuk, GA, et al. (2001). Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Applied and Environmental Microbiology*, 2001, 67: 469-472.
- [31] Wang, GC, Wang Y. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiology*, 1996, 142: 1107-1114.
- [32] Wang GC, Wang Y. Frequency of formation of chimeric molecules is a consequence of PCR co-amplification of 16S rRNA genes from mixed bacterial genomes. *Applied and Environmental Microbiology*, 1997, 63: 4645-4650.
- [33] Komatsoulis G, Waterman M. A new computational method for detection of chimeric 16S rRNA artifacts generated by PCR amplification from mixed bacterial populations. *Applied and Environmental Microbiology*, 1997, 63: 2338-2346.
- [34] Robinson-Cox JF, Bateson MM, Ward DM. Evaluation of nearest-neighbor methods for detection of chimeric small subunit rRNA sequences. *Applied and Environmental Microbiology*, 1995, 61: 1240-1245.
- [35] Hugenholtz P, Huber T. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *International Journal of Systematic and Evolutionary Microbiology*, 2003, 53: 289-293.
- [36] Paabo S, Irwin DM, Wilson AC. DNA damage promotes jumping between templates during enzymatic amplification. *The Journal of Biological Chemistry*, 1990, 265: 4718-4721.
- [37] Maidak BL, Cole JR, Lilburn TG, et al. (2001). The RDP-II (Ribosomal Database Project). *Nucleic Acids Research*, 2001, 29: 173-174.
- [38] Huber T, Faulkner G, Hugenholtz P. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics*, 2004, 20(14): 2317-2319.
- [39] Gonzalez JM, Zimmermann J, Saiz-Jimenez C. Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics*, 2005, 21(3): 333-337.
- [40] Nagamine CM, Chan K, Lau YF. A PCR artifact: generation of heteroduplexes. *The American Journal of Human Genetics*, 1989, 45: 337-339.
- [41] Ruano G, Kidd KK. Modeling of heteroduplex formation during PCR from mixtures of DNA templates. *Genome Research*, 1992, 2: 112-116.

- [42] Mark AJ, Straus N. Effect of PCR Conditions on the formation of heteroduplex and single - stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. *Genome Research*, 1993, 3: 186-194.
- [43] Thomposon JR, Marcelino LA, Polz MF. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Research*, 2002, 30(9): 2083-2088.
- [44] Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods and Applications*, 1991, 1: 17-24.
- [45] Ford JE, McHeyzer-Williams MG, Lieber MR. Chimeric molecules created by gene amplification interfere with the analysis of somatic hypermutation of murine immunoglobulin genes. *Gene*, 1994, 142: 279-283.
- [46] Stewart ACM, Gravitt PE, Cheng S, et al. Generation of entire human papillomavirus genomes by long PCR: frequency of errors produced during amplification. *PCR Methods and Applications*, 1995, 5: 79-88.
- [47] Lundberg KS, Shoemaker DD, Adams MWW, et al. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*, 1991, 108: 1-6.
- [48] Cariello NF, Thilly WG, Swenberg JA, et al. Deletion mutagenesis during polymerase chain reaction: dependence on DNA polymerase. *Gene*, 1991, 99: 105-108.
- [49] Pallansch L, Beswick H, Talian J, et al. Use of an RNA folding algorithm to choose regions for amplification by the polymerase chain reaction. *Analytical Biochemistry*, 1990, 185: 57-62.
- [50] Holstege FCP, van der Vliet PC, Timmers HTM. Instability of human TATA-binding protein CAG triplet repeats during amplification by PCR. *Biochimica et Biophysica Acta*, 1994, 1219: 157-159.
- [51] Gutell RR, Larsen N, Woese CR. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiological Reviews*, 1994, 58: 10-26.
- [52] Chou Q. Minimizing deletion mutagenesis artifact during *Taq* DNA polymerase. *Nucleic Acids Research*, 1992, 20(16): 4371.
- [53] Qiu XY, Wu LY, Huang HS, et al. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology*, 2001, 67(2): 880-887.
- [54] 严兴, 徐正茂, 冯晓西, 等. 焦化废水工业处理装置和实验室装置硝化菌群的分子比较. *微生物学报 (Acta Microbiologica Sinica)*, 2007, 47(2): 301-306.
- [55] 刘彬彬, 张峰, 冯晓西, 等. 对降解喹啉的厌氧生物反应器中重要功能菌群的鉴定. *生态学报 (Acta Ecologica Sinica)*, 2006, 26(5): 1390-1395.
- [56] 车玉伶, 王慧, 胡洪营等. 微生物群落结构和多样性解析技术研究进展. *生态环境 (Ecology and Environmental Sciences)*, 2005, 14(1): 127-133.

16S rRNA gene in community structure analysis and the optimized amplification – A review

Lin Zhou, Jie Zhang*

(State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: The purpose of this review was to offer reference for the study of community analysis in which 16S rRNA was taken as target molecule. We summarized the information of 16S rRNA including its function in cells, transcription organization, the position of variable regions and conserved regions, copies numbers of 16S rRNA gene, secondary structure, some artifacts generated during PCR amplification and the solution to avoid the artifacts.

Keywords: 16S rRNA; variable region; conserved region; artifacts

(本文责编:王晋芳)

Supported by the Key Project of Chinese National Programs for Fundamental Research and Development (2007CB109203)

* Corresponding author. Tel: +86-10-62815921; Fax: +86-10-62812642; E-mail: jzhang@ippcaas.cn

Received: 9 July 2009 / Revised: 17 September 2009