

链霉菌看家基因引物的设计与验证

郭银平^{1,2}, 黄 英^{1*}

(¹ 中国科学院微生物研究所 微生物资源前期开发国家重点实验室 北京 100101)

(² 中国科学院研究生院生物系 北京 100049)

摘 要 看家基因的扩增与测序是进行多基因系统进化分析首先需要解决的问题。针对链霉菌这一群高(G+C) mol% 革兰氏阳性细菌, 选定 4 个看家基因: *atpD*、*recA*、*rpoB* 和 *trpB*, 利用 NCBI 数据库中已有的 2 个链霉菌和 3 个分枝杆菌的全基因组序列, 以及另两个链霉菌的 *recA* 基因序列, 通过软件分析设计了各基因的扩增和测序引物, 并优化了扩增反应条件。从所试验的 55 株链霉菌中, 均特异地扩增出了上述 4 个基因的片段, 并成功进行了序列测定, 验证了所设计引物的实用性。所归纳的引物设计方法可用于高(G+C) mol% 革兰氏阳性细菌的其它看家基因, 以促进多基因系统进化研究的开展。

关键词 看家基因; 引物; 链霉菌; 扩增; 测序; 高(G+C) mol%

中图分类号: Q939 文献标识码: A 文章编号: 0001-6209(2007)06-1080-04

链霉菌是产生天然生物活性物质的主要微生物资源, 它们构成了细菌域中最庞大的一个属—链霉菌属。链霉菌分类学已经由表观分类、数值分类、化学分类, 发展到了现今的分子分类。随着以 16S rRNA 基因序列分析为主导的细菌分子分类的发展, 近年来, 国际核酸序列数据库中的 16S rRNA 基因序列信息剧增。然而, 由于 16S rRNA 基因高度保守, 在研究亲缘关系较近的菌株, 特别是诸多链霉菌时, 其间分辨率明显不足^[1]; 另一方面, 16S rRNA 基因只是细菌众多保守基因中的一个, 可能存在种内变异性和基因组内操纵子间的异源性, 因此单独使用该基因来反映菌株间的亲缘关系可能会出现偏差和误导^[2,3]。越来越多的分类学家尝试用其它看家基因作为分子分类的依据, 来弥补或替代 16S rRNA 基因, 如 *gyrB*^[4]、*hsp65*^[5]、*rpoB*^[6]、*trpB*^[1] 等的应用。综合多个看家基因的系统进化分析已经受到关注并逐渐兴起^[7,8]。基于多基因序列分析的需要, 我们首先面对的是如何设计合适的扩增和测序引物, 尤其是针对以链霉菌为代表的高(G+C) mol% 革兰氏阳性细菌, 它们的引物要求较其他类群的细菌更为苛刻。本文以链霉菌为材料, 介绍高(G+C) mol% 革兰氏阳性细菌看家基因引物设计的方法和技巧, 并验证了所设计引物的适用性。

1 材料和方法

1.1 材料

1.1.1 菌株 共 55 株链霉菌(表 1), 其中 49 株为模式菌株。53 株由中国普通微生物菌种保藏中心(CGMCC)提供, 一株由日本微生物保藏中心(JCM)提供, 一株由比利时微生物协

同保藏中心(BCCM/LMG)提供。

表 1 55 株实验菌株的名称和保藏号

Table 1 Species names and collection numbers of strains

used in this study			
Species	Collection number	Species	Collection number
<i>S. acrimycini</i>	AS 4.1673 ^T	<i>S. griseoplanus</i>	AS 4.1868 ^T
<i>S. albovinaceus</i>	AS 4.1631 ^T	<i>S. griseus</i> subsp. <i>griseus</i>	AS 4.1419 ^T
<i>S. albovidis</i>	AS 4.1627 ^T	" <i>S. griseus</i> subsp. <i>alpha</i> "	AS 4.1843
<i>S. anulatus</i>	AS 4.1421 ^T	" <i>S. griseus</i> subsp. <i>cretosus</i> "	AS 4.1844
" <i>S. argenteolus</i> "	AS 4.1693	<i>S. griseus</i> subsp. <i>solvifaciens</i>	AS 4.1845 ^T
<i>S. atroolivaceus</i>	AS 4.1405 ^T	<i>S. kanamyceticus</i>	AS 4.1441 ^T
<i>S. aureus</i>	AS 4.1833 ^T	<i>S. laceyi</i>	AS 4.1832 ^T
<i>S. badius</i>	AS 4.1406 ^T	<i>S. luridiscabiei</i>	LMG 21390 ^T
<i>S. bobili</i>	AS 4.1624 ^T	<i>S. mauvecolor</i>	AS 4.1997 ^T
<i>S. californicus</i>	AS 4.570 ^T	<i>S. mediolani</i>	AS 4.1896 ^T
" <i>S. caiscabies</i> "	AS 4.1836	<i>S. microflavus</i>	AS 4.1428 ^T
<i>S. cirratus</i>	AS 4.1679 ^T	<i>S. mutomycini</i>	AS 4.1747 ^T
<i>S. cremeus</i>	AS 4.1625 ^T	<i>S. nojiriensis</i>	AS 4.1897 ^T
<i>S. cyaneofuscatus</i>	AS 4.1612 ^T	" <i>S. ornatus</i> "	AS 4.1321
<i>S. cyaneus</i>	AS 4.1671 ^T	<i>S. peucetius</i>	AS 4.1799 ^T
<i>S. erumpens</i>	AS 4.1626 ^T	<i>S. praecox</i>	AS 4.1782 ^T
<i>S. exfoliatus</i>	AS 4.1407 ^T	<i>S. pulveraceus</i>	AS 4.1928 ^T
<i>S. fimicarius</i>	AS 4.1629 ^T	" <i>S. setonii</i> "	AS 4.1367
<i>S. finlayi</i>	AS 4.1436 ^T	<i>S. sindenensis</i>	AS 4.626 ^T
<i>S. flavidofuscus</i>	AS 4.1617 ^T	<i>S. spiroverticillatus</i>	AS 4.1749 ^T
<i>S. flavogriseus</i>	AS 4.1884 ^T	<i>S. spororaveus</i>	AS 4.1926 ^T
<i>S. floridiae</i>	AS 4.1972 ^T	<i>S. subrutilus</i>	AS 4.1784 ^T
<i>S. fulvorobeus</i>	JCM 9090 ^T	<i>S. tanashiensis</i>	AS 4.1924 ^T
<i>S. galilaeus</i>	AS 4.1320 ^T	<i>S. venezuelae</i>	AS 4.1526 ^T
<i>S. graminofaciens</i>	AS 4.1359 ^T	<i>S. vinaceus</i>	AS 4.1305 ^T
<i>S. griseinus</i>	AS 4.1875 ^T	<i>S. virginiae</i>	AS 4.1530 ^T
<i>S. griseobrunneus</i>	AS 4.1838 ^T	<i>S. yanii</i>	AS 4.1146 ^T
<i>S. griseolus</i>	AS 4.1864 ^T		

基金项目 国家自然科学基金(30670002); 中国科学院知识创新工程重要方向项目(KSCX2-YW-Z-042)

* 通讯作者。Tel: 86-10-64807311; E-mail: huangy@im.ac.cn

作者简介 郭银平(1981-), 女, 河北邯郸人, 硕士研究生, 主要从事链霉菌分子系统学研究。E-mail: guoyinping@yahoo.com.cn

收稿日期 2007-02-15, 接受日期 2007-04-02, 修回日期 2007-07-12

1.1.2 培养基和主要试剂 菌株活化与培养采用 ISP 2 培养基^[9] 28℃培养 7~14d。Taq DNA 聚合酶、dNTP 等 PCR 试剂购自上海申能博彩生物科技有限公司。

1.2 引物设计

1.2.1 基因的保守区和可变区分析 选定 *atpD*、*recA*、*rpoB* 和 *trpB* 4 个看家基因为研究对象。它们均为单拷贝基因,无遗传连锁,在细菌的基础代谢中行使不同的功能。*atpD* 基因编码 F_0F_1 -ATP 合成酶的 β 亚基,*recA* 基因编码重组酶 A,*rpoB* 基因编码 RNA 聚合酶 β 亚基,*trpB* 基因编码色氨酸合成酶 β 亚基。从 NCBI 的 genome database 中下载已有全基因组序列的 *Streptomyces coelicolor* A3(2)、*Streptomyces avermitilis* MA-4680^T、*Mycobacterium tuberculosis* H37Rv^T、*Mycobacterium avium* subsp. *paratuberculosis* K-10 和 *Mycobacterium leprae* TN 的上述 4 个基因的完整核酸序列,另外还找到了 *Streptomyces venezuelae* ISP5230^T 和 *Streptomyces argillaceus* ATCC 12956 的 *recA* 完整基因序列。用 MEGA 3.1^[10] 软件对各基因序列进行比对,根据比对结果确定保守区和可变区,在保守区设计引物。

1.2.2 引物设计 主要采用 Primer Premier 5.0 (PREMIER Biosoft International) 软件,并与手动查找引物相结合。参照选择引物的一般原则,即 (1) 引物 $G+C\% \geq$ 模板 $G+C\%$, 尽量使引物和产物的 T_m 相差在 20℃ 以内,引物间 T_m 相差 5℃~6℃,一般用比引物 T_m 低 5℃ 做最退火温度 (2) 引物与非扩增序列同源性 $\leq 70\%$, 尤其是 3' 端不要有错配 (3) 尽量用 5' 末端稳定 (GC 含量较多), 而 3' 末端不太稳定 (AT 含量较多) 的引物,这种引物的结构可以有效地消除假引发反应 (4) 引物分子内和分子间不形成发夹、二聚体等结构,若有, $\Delta G > -4.5$ kcal/mol 为好,也可放宽至 > -20 kcal/mol; 引物 3' 端不形成任何二级结构,若有, $\Delta G > -2.0$ kcal/mol 为好,也可放宽至 > -10 kcal/mol。

1.2.3 引物评价 同时采用 Oligo 6.0 (Molecular Biology Insights) 和 Primer Premier 5.0 的引物评价功能评价引物的各项指标。

1.2.4 引物的非特异性检查 选择符合上述原则的引物对,用 SPCR 3.0 (<http://moleco.sjtu.edu.cn/moleco/softwares.html>) 进行模拟 PCR, 检查是否有非特异性条带出现。将无非特异性扩增条带的引物再进行 NCBI BLAST 比对,检查引物 3' 端是否在链霉菌基因组的其它位置有匹配情况。3' 端最后 5~7 个核苷酸不应与基因组上非目的扩增位点匹配; 引物与非目的扩增位点匹配上的核苷酸数目应小于引物总核苷酸数的 70%。

选择 BLAST 比对后符合要求的引物进行合成,由上海生工生物工程有限公司完成。

1.3 PCR 扩增和测序

1.3.1 基因组 DNA 模板的制备 从新鲜 ISP 2 平板上刮取适量菌体,用 Moller 的方法^[11] 提取基因组 DNA。

1.3.2 PCR 扩增条件 4 种基因的扩增体系都为 50 μ L: 基因组模板 DNA 1 μ L (50~200 ng), 10 \times PCR buffer 5 μ L, 上下游扩增引物 (20 pmol/L) 各 1 μ L, dNTP mix (10 mmol/L) 1 μ L, MgCl₂ (25 mmol/L) 6 μ L, Taq 酶 2.5U, DMSO 5 μ L, 加双蒸水补足

50 μ L。扩增反应程序: 95℃ 预变性 5 min; 95℃ 30 s, 退火温度 (*atpD*, 63℃; *recA*, 60℃; *rpoB*, 65℃; *trpB*, 66℃) 30 s, 72℃ 1 min, 30 个循环, 72℃ 延伸 10 min。

1.3.3 测序 各基因 PCR 扩增产物经 1% 琼脂糖凝胶电泳检测后,送北京诺赛基因组研究中心有限公司,用所设计测序引物进行序列测定。

2 结果

2.1 设计引物的基因保守区

atpD 的保守区主要有 C1 (199-230)、C2 (649-698)、C3 (766-902)、C4 (1147-1190) 和 C5 (1237-1319), 在 C1 和 C2 区设计上游引物, C4 和 C5 区设计下游引物。对于 *recA*, 由于从 612 位到最后一位碱基, *Streptomyces* 与 *Mycobacterium* 的菌株几乎无共有序列, 所以舍弃 *Mycobacterium* 的序列, 只比较 4 个链霉菌的 *recA* 全基因序列, 得到 *recA* 的保守区 C1 (1-74)、C2 (214-991), 在 C1 设计上游引物, C2 设计下游引物。*rpoB* 的保守区主要有 C1 (247-923)、C2 (1055-1084)、C3 (1144-1203)、C4 (2092-2155) 和 C5 (3004-3368), 同时根据文献^[12], 选择 C2 和 C3 区设计上游引物, C4 区设计下游引物。*trpB* 的保守区主要有 C1 (260-414)、C2 (452-489)、C3 (945-991) 和 C4 (1068-1195), 在 C1 和 C2 区设计上游引物, C3 和 C4 区设计下游引物。以上各区域的位点均以 *Streptomyces coelicolor* A3(2) 相应基因的核酸位置计数。

2.2 引物

结合软件评价和非特异性检查结果, 选定以下引物对 (表 2) 进行各基因片段的扩增和测序。

2.3 PCR 扩增和序列测定结果

用表 2 中的扩增引物, 所有 55 株链霉菌的 4 个基因片段都得到了成功扩增, 片段大小在 800~1000 bp 之间。

用表 2 中的测序引物成功测定了所有 55 株链霉菌的 4 个基因片段的序列, 且正反向测序结果吻合。将所测各基因序列分别用 MEGA 3.1^[10] 软件进行比对后, 截取包含可变区的 500 bp 左右的片段, 提交 GenBank 进行注册。注册号为: *atpD*, EF031272 ~ EF031326; *recA*, EF055014 ~ EF055068; *rpoB*, EF055069 ~ EF055123; *trpB*, EF055124 ~ EF055178。

3 讨论

由于链霉菌的基因组为高 (G+C) mol%, 解链温度高, 扩增引物对的 T_m 值也应该比较高, 这点不同于一般引物的设计要求。在 PCR 扩增引物设计过程中, 用软件评价引物的各项参数很重要, 如: 引物自身和引物对之间二级结构的自由能, 引物引发扩增的效率, 引物 (G+C) mol% 等。文献和网络文章中报道的引物设计一般原则仅能用作参考, 不能完全按照那些原则来设计引物。为了避免非特异扩增条带的出现和提高引物 (G+C) mol% 以达到模板 DNA 的 (G+C) mol%, 我们使用了较长的扩增引物 (30~40mer 左右), 而不是一般原则上的 20~25mer; 引物二级结构的自由能 (ΔG) 可以比原则上的 (-4.5 kcal/mol) 更小, 如本研究中的经验值 $\Delta G > -20$ kcal/mol, 3' 端二级结构 $\Delta G > -10$ kcal/mol。通过提高 PCR 的退火温度完全可以把这些二级结构打开, 同时

表 2 扩增和测序引物

Table 2 Primers for amplification and sequencing

Gene	Function	Primer sequence (5'→3')	Position*
atpD	Amplification	atpDPF, GTCGGCGACTTCACCAAGGGCAAGGTGTCAACACC	283-318
		atpDPR, GTGAACTGCTTGGCGACGTGGGTGTCTGGGACAGGAA	1243-1280
	Sequencing	atpDF, ACCAAGGGCAAGGTGTCAAA	295-314
		atpDR, GCCGGGTAGATGCCCTTCTC	1027-1046
recA	Amplification	recAPF, CCGCRCTCGCACAGATTGAACGSCAATTC	35-63
		recAPR, GCSAGGTCGGGGTTGTCTTSAGGAAGTTGCG	916-947
	Sequencing	recAF, ACAGATTGAACGGCAATTCG	45-64
		recAR, ACCTTGTCTTGACCACCTT	733-752
rpoB	Amplification	rpoBPF, GAGCGCATGACCACCCAGGACGTCGAGGC	1162-1190
		rpoBPR, CCTCGTAGTTGTGACCTCCACGGCATGA	2126-2155
	Sequencing	rpoBF1, TTCATGGACCAGAACAACC	1273-1291
		rpoBR1, CGTAGTTGTGACCCTCCC	2135-2152
trpB	Amplification	trpBPF, GCGCGAGGACCTGAACCCACCCGGCTCACACAAGATCAACA	267-308
		trpBPR, TCGATGGCCGGGATGATGCCCTCGGTGCCGCGACAGCAGGC	1049-1088
	Sequencing	trpBF, GGCTCACACAAGATCAACAA	289-308
		trpBR, TCGATGGCCGGGATGATGCC	1069-1088

* Nucleotide positions were numbered according to corresponding genes of *Streptomyces coelicolor* A3(2).

保证扩增的特异性。测序引物的(G+C)mol%则无需考虑模板(G+C)mol%高低问题,只需使测序引物的(G+C)mol%满足 T_m 值略高于50℃~55℃(测序时常用的退火温度)即可。

在探索最适扩增反应条件时,退火温度的控制很关键。本研究中的trpB基因在65℃的退火温度下扩增效率很低;而将退火温度提高1℃,其它条件不变时,实验中所有模板DNA都得到了特异性扩增。这是因为trpB基因的扩增引物较长(trpBPF 42mer, trpBPR 40mer),引物分子内部或分子之间容易形成稳定的发夹和二聚体等二级结构,退火温度较低时这些二级结构不能完全打开,从而影响引物与模板的结合,无法有效引发扩增反应。因此,对于较长的引物,适当提高退火温度反而有利于引物与模板结合,提高扩增的效率和特异性。此外,对于高(G+C)mol%的基因组模板,在扩增反应体系中加入10%体积的DMSO可使扩增效果更好。

这4个基因的扩增引物共有的优点是(1)特异性高,扩增反应中很少出现非特异性条带;(2)普遍适用性好,能有效扩增所有链霉菌实验菌株。另外,这4个基因的正反向测序引物也都能稳定、特异地引发测序反应,得到的序列长度均较长,包括了需要的可变区。

本研究针对链霉菌属这一高(G+C)mol%革兰氏阳性菌代表菌属,设计了4个看家基因的扩增和测序引物,并获得了很好的扩增和测序效果。所设计的引物可用于更多链霉菌甚至其它放线菌相应基因的扩增与测序,所归纳的引物设计方法也可用于其它看家基因,以促进多基因系统进化研究的开展。此外,我们所扩增测序的基因区段涵盖了较多的可变区,基于这些区段的序列分析可望得到分类分辨率较高的链霉菌多基因系统发育图。

参 考 文 献

[1] Egan S, Wiener P, Kallifidas D, et al. Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie van Leeuwenhoek*, 2001, **79**: 127-133.

[2] Cilia V, Lafay B, Christen R. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol*, 1996, **13**: 451-461.

[3] Anderson I AS, Wellington EMH. The taxonomy of *Streptomyces* and related genera. *Int J Syst Evol Microbiol*, 2001, **51**: 797-814.

[4] Hatano K, Nishii T, Kasai H. Taxonomic re-evaluation of whorl-forming *Streptomyces* (formerly *Streptoverticillium*) species by using phenotypes, DNA-DNA hybridization and sequences of gyrB, and proposal of *Streptomyces luteireticuli* (ex Katoh and Arai 1957) corrig., sp. nov., nom. rev. *Int J Syst Evol Microbiol*, 2003, **53**: 1519-1529.

[5] Rodriguez-Nava V, Couble A, Devulder G, et al. Use of PCR-restriction enzyme pattern analysis and sequencing database for hsp65 gene-based identification of *Nocardia* species. *J Clin Microbiol*, 2006, **44**: 536-546.

[6] Kim BJ, Kim CJ, Chun J, et al. Phylogenetic analysis of the genera *Streptomyces* and *Kitasatospora* based on partial RNA polymerase beta-subunit gene (rpoB) sequences. *Int J Syst Evol Microbiol*, 2004, **54**: 593-598.

[7] Stackebrandt E, Frederiksen W, Garrity GM, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol*, 2002, **52**: 1043-1047.

[8] Devulder G, Perouse de Montclos M, Flandrois JP. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *Int J Syst Evol Microbiol*, 2005, **55**: 293-302.

[9] Shirling EB, Gottlieb D. Methods for characterization of *Streptomyces* species. *Int J Syst Bacteriol*, 1966, **16**: 313-340.

[10] Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, 2004, **5**: 150-163.

[11] Moller EM, Bahnweg G, Sandermann H, et al. A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucl Acids Res*, 1992, **20**: 6115-6116.

[12] Kim BJ, Lee SH, Lyu MA, *et al.* Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase

gene (*rpoB*). *J Clin Microbiol*, 1999, **37**: 1714–1720.

Design and validation of primers for housekeeping genes of streptomycetes

GUO Yin-ping^{1,2}, HUANG Ying^{1*}

(¹ State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China)

(² Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract :Multigene-based phylogenetic analyses are becoming more widely used in molecular taxonomy because of its lab-to-lab portability and reproducibility. The first step that needs to be settled for this approach is to amplify and sequence several housekeeping genes. Four housekeeping genes *atpD*, *recA*, *rpoB* and *trpB* were chosen for streptomycetes, which are representatives of high G + C mol% Gram-positive bacteria. Primer pairs for amplification and sequencing of the gene fragments were designed according to the known genome sequences of two streptomycetes and three mycobacteria, as well as the available *recA* gene sequences of another two streptomycetes in NCBI database, by using software packages Primer premier 5.0, Oligo 6.0 and SPCR 3.0, and NCBI BLAST program. Performance of the primer sets were validated by specific amplification of all gene fragments from the 55 streptomycete tested strains under optimized PCR reaction conditions, and by successful sequencing of the amplification products. It is concluded that the primer sets designed here are effective, and the primer design procedure and guidelines are valuable for other housekeeping genes of high G + C mol% bacteria.

Keywords : housekeeping gene ; primer ; streptomycetes ; amplification ; sequencing ; high G + C mol%

Foundation item : National Natural Science Foundation of China (30670002) ; Knowledge Innovation Project of Chinese Academy of Sciences (KSCX2-YW-Z-042)

* Corresponding author. Tel : 86-10-64807311 ; E-mail : huangy@im.ac.cn

Received : 15 February 2007 / Accepted : 2 April 2007 / Revised : 12 July 2007