

问号钩端螺旋体基因组 DNA 芯片的研制

盛跃颖* 史耀舟** 何平 杨杨 胡宝瑜 赵国屏** 郭晓奎

【摘要】 目的 研制并初步评估问号钩端螺旋体(简称钩体)赖型赖株的基因组 DNA 芯片。方法 利用 Primegens 引物设计软件筛选出问号钩体赖型赖株全基因组中的特异性基因进行引物设计。对成功设计出相应引物的 3 290 个基因用聚合酶链反应方法进行扩增,以纯化后的产物点样制备芯片。并用双色荧光杂交策略对芯片质量进行了初步评估。结果 共获得 3 290 个基因产物用于点样。参考株自身杂交实验结果表明:该芯片有较高的点一致性、信噪比和较低的假阳性率。结论 成功制备了包含问号钩体赖型赖株 3 290 个目的基因的基因组 DNA 芯片,并可用于基于该芯片的问号钩体比较基因组学的研究。

【关键词】 问号钩端螺旋体; 基因组 DNA 芯片

Construction and validation of *Leptospira interrogans* genome DNA microarray

SHENG Yue-ying*, SHI Yao-zhou**, HE Ping, YANG Yang, HU Bao-yu, ZHAO Guo-ping**, GUO Xiao-kui (Department of School of Medicine Microbiology and Parasitology, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China)

[Abstract] Objective To develop and validate a genomic DNA microarray of *L. interrogans* serovar Lai str. 56601. Methods Primers were designed based on specific coding gene sequences of *L. interrogans* serovar Lai str. 56601 using a computer software Primegens. PCR products from these genes were purified and printed onto the glass slides in triplicate. The quality of this microarray is evaluated by the method of two-fluorescence comparative hybridization. Results A total of 3 290 genes were successfully amplified by PCR. Genomic DNAs from an autologous reference strain were labeled with Cy3 or Cy5 fluor and the labeled probes were hybridized against this new microarray. Results showed high consistency of spots, high signal/noise ratio and low false positive rate. Conclusion A microarray was developed containing 3 290 unique gene sequences from *L. interrogans* serovar Lai str. 56601. A research platform based on this microarray can be further established to conduct comparative genomic analysis for *L. interrogans*.

[Key words] *Leptospira interrogans*; genome DNA microarray

2002 年 10 月国家人类南方中心、中国科学院上海生命科学院、上海交通大学医学院等单位完成了我国境内的强毒株——问号钩端螺旋体(简称钩体)赖型赖株(*Leptospira interrogans* serovar Lai str. 56601)的全基因组测序工作^[1]。本研究是在获得其基因组序列及预测开放阅读框架(ORF)初步注释的基础上,制作了包含其大多数特异性 ORF 的基因组 DNA 芯片,为进一步开展基于芯片的钩体比较基因组学研究提供了技术支持。

材料和方法

1. 实验菌株 问号钩体赖型赖株,由中国疾病控制中心微生物研究所提供。

2. 引物设计 采用 Primegens 引物设计软件,以问号钩体赖型赖株全基因组的 4 727 个基因为基础,剔除长度 < 250 bp 的基因后,共成功设计特异性

引物 3 299 对。引物合成是由华大基因上海鼎安科技有限公司完成。加适量水,溶解成 50 μmol/L, -20℃ 可长期贮存。

3. 模板 DNA 的抽提及芯片靶序列的制备 用上海华舜生物工程有限公司的细菌 DNA(小量)抽提试剂盒抽取问号钩体赖型赖株的基因组 DNA,测定浓度后保存于 -20℃ 备用。PCR 扩增:反应总体积为 100 μl。各成分为:10 mmol/L Tris-HCl (pH = 8.0), 50 mmol/L KCl, 2.5 mmol/L MgCl₂, 0.2 mmol/L dNTP, 引物各 10 pmol, 3U Taq DNA 聚合酶(大连 TaKaRa 公司)和 20 ng 模板 DNA。扩增条件为 94℃ 预变性 5 min, 94℃ 30 s, 55℃ 45 s, 72℃ 90 s, 共 35 个循环, 最后 72℃ 5 min。反应结果通过 1% 琼脂糖凝胶电泳进行鉴定。首次未成功扩增出相应产物的基因需改变 PCR 条件进行第 2 次扩增。若仍未获得相应产物,则用 Primer Premier 5.0 重新设计引物。最终成功扩增出 3 290 个基因。

产物的纯化和定量:取成功扩增出的各产物约 300 μl, 经 96 孔 Millipore Multiscreen PCR 纯化板负压抽滤后, 每孔加入 100 μl ddH₂O 溶解纯化产物, 并转入石英板进行定量。60℃ 干燥纯化产物, 每孔加入适量的 50% DMSO, 将产物浓度定量到约 200 ng/μl,

基金项目:国家高技术研究发展计划(863 计划)
(2003AA223031), 国家自然科学基金(30370071)

作者单位:200025 上海, 上海交通大学医学院病原生物学教研室; *上海市疾病预防控制中心; **上海生物芯片公司

通讯作者:郭晓奎, E-mail: xkguo@shsmu.edu.cn

并充分溶解后于 -20℃ 贮存。

4. 芯片矩阵排布与点样 将 96 孔板中的样品转入 384 板, 每孔 10 μl 。现有 3 290 条 PCR 扩增产物, 共组成 9 块 384 板, 对照样品独立组成 1 块 384 板, 包括定位点, 阳性对照(问号钩体赖型赖株 16S rRNA 和 23S rRNA 基因的扩增产物)及阴性对照(50% DMSO, 棉花克隆扩增产物)。芯片点样仪为 Genemachine, 点样针为狭缝针, 片基采用氨基化玻片。所有样品排布于 12 个 18×18 点阵排列的亚矩阵中, 且每个点重复点样 3 次。点样结束后置紫外交联仪中, 以 400 mJ/cm² 能量强度进行交联 20 min。芯片于室温干燥 24 h 后保存于暗室中, 待用。

5. 基因组 DNA 或芯片靶序列 PCR 产物的荧光素标记 标记模板的制备: 经超声波打断的参照株基因组 DNA(片段长度集中于 500 bp ~ 2 kb)在自身比较实验中分别作为参考 DNA 和待测 DNA 或将随机选取的已纯化的 48 个靶序列 PCR 产物(长度 500 bp ~ 2 kb 为佳)作为参考 DNA, 再从其中选取 42 个产物作为待测 DNA。

采用双色荧光标记法, 取 3 μg 模板(参照 DNA 或待测 DNA), 加入 6 μg 随机九聚体和 8 μl 2.5 × reaction buffer(125 mmol/L Tris, 12.5 mmol/L MgCl₂, 25 mmol/L β -巯基乙醇), 于 95℃ 变性 5 min 后, 加入 2 μl 10 × dNTPmix(dATP, dGTP, dTTP 浓度各为 1.2 mmol/L, 0.6 mmol/L dCTP, 10 mmol/L Tris, 1 mmol/L EDTA)、1 μl 的 1 mmol/L Cy5 或 Cy3-dCTP(Amersham)、0.5 μl 的 50 U/ μl Klenow 酶(NEB), 补水至总体积 20 μl , 37℃ 反应 3 h。探针用 QIAquick Nucleotide Removal Kit 纯化, 并转入酶标板定量, 于 60℃ 抽干, -20℃ 可长期贮存。

6. 芯片杂交 取出 Cy3 和 Cy5 标记的探针 50 pmol, 共溶于 9 μl ddH₂O 中, 94℃ 变性 3 min 后, 加入 9 μl 杂交缓冲液和 18 μl 甲酰胺, 充分混匀。将杂交液加于芯片, 盖上盖玻片, 将芯片水平放入加有 PBS 的湿盒, 置于 42℃ 杂交箱中避光杂交 14 ~ 16 h。杂交后芯片于 55℃ 预热的洗液 I(1 × SSC, 0.2 % SDS)中洗涤和洗液 II(0.1 × SSC, 0.2 % SDS)各洗涤 15 min, 转入洗液 III(0.1 × SSC)中室温下洗涤 3 min, 最后于 ddH₂O 中洗涤 1 ~ 2 min 后, 迅速转入离心管中, 1 500 r/min 离心 5 min, 干片后立即扫描或避光保存于干燥箱中。

7. 数据读取与分析 芯片扫描采用 Agilent 激光共聚焦扫描仪, 经 Image 5.1 图像分析软件对各信号点进行自动及人工定位, 分析荧光信号强度, 以

列表形式输出原始数据。去除低质量的信号点, 去除标准: 以参照株 DNA-Cy5 为准, 剔除信号中值小于该点背景中值加 2 个标准差的点。继续采用 Microsoft Excel 软件, 计算出扣除背景后的各点信号值、信噪比等参数。采用双色荧光策略时, 采用总值归一法对原始数据进行修正后, 再计算各信号点的待测-Cy3 和参照-Cy5 比值, 并转换成 \log_2 (待测-Cy3/参照-Cy5)。

结 果

1. 问号钩体赖型赖株基因芯片靶序列的设计与制备 基于问号钩体赖型赖株的 3 512 个基因(长度 > 250 bp), 用 Primegens 引物设计软件共成功设计 3 300 对引物。212 个基因未设计出相应的引物, 主要是由于目标基因序列中无特异性区域。经 2 轮 PCR 扩增及重新设计引物后共成功获得 3 290 个产物, 成功率为 99.6%。

2. 芯片的质量评估 用参考株问号钩体赖型赖株的 DNA, 采用单色荧光标记后进行芯片杂交。用读图软件对信号点进行自动定位时可显示点阵的排列、各点的定位是否准确、有无明显的位置偏斜及漏点现象、点的大小形状是否为较均一的圆形, 从而综合评判点样的成功率。杂交图像结果分析显示本实验所研制芯片的点样成功率约为 99%(图 1)。

3. 芯片杂交条件的初步评估 采用双色荧光法, 即用 Cy3 和 Cy5 分别标记问号钩体赖型赖株的基因组 DNA 进行参考株的自身比较实验, 初步评估芯片杂交实验的假阳性率。自身比较实验的杂交结果通过数据读取、归一化处理后, 以选定的通用筛选标准(比较基因组中为基因缺失或基因拷贝数增加的评判标准)进行数据筛选^[2-4]。即认为 \log_2 (参照-Cy3/参照-Cy5) < -1 或 > 1 的基因为假阳性基因, 其占基因总数的百分比为假阳性率。据此统计我们制备的钩体芯片本次杂交实验的假阳性率约为 0.3%(图 2)。

为了进一步检验芯片的杂交实验条件是否适用于问号钩体比较基因组学研究, 选取 48 个制备芯片的靶基因 PCR 产物混合后作为参考标本用 Cy5 标记, 并从中选取 42 个产物作为待测标本用 Cy3。采用与自身比较实验相同的杂交条件进行芯片杂交, 获得较好的杂交结果(图 3)。从杂交图可基本判断实验结果: 每个样点片内重复 3 次, 黄色信号点代表参考和待测标本共有的基因, 共 42 × 3 个; 红色信号点代表相应的基因在待测标本 (下转封三)

(上接第 104 页) 中缺失, 共 6×3 个, 为本次杂交结 果中的阳性基因。

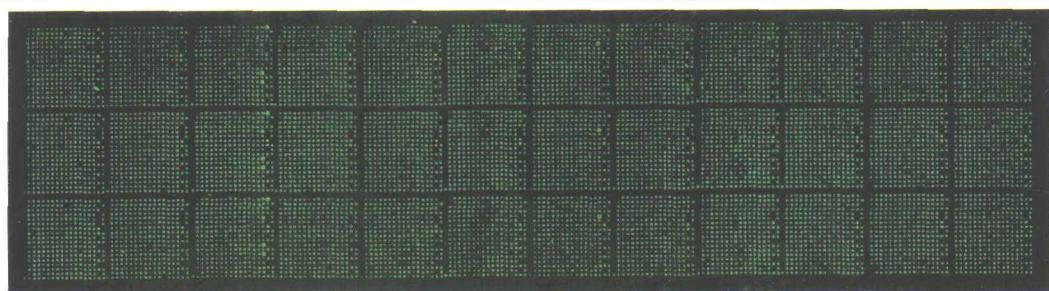


图 1 问号钩体赖型赖株 DNA(参考 DNA)标 Cy3 的单色自身杂交图

Fig 1. Scanning single fluorimetric representation of genomic DNA from *L. interrogans* serovar *Lai* str. 56601 (reference DNA) labeled with Cy3 fluor.

Genomic DNA from the reference strain was labeled with Cy3 fluor and the labeled probes were hybridized with the new *L. interrogans* microarray. Results show that the final microarray has high consistency of spots.

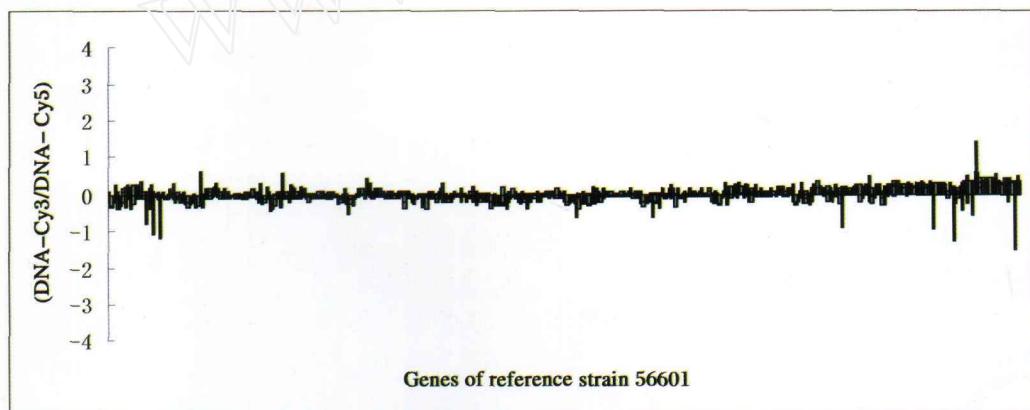


图 2 参考 DNA + 参考 DNA 的杂交统计结果

Fig 2. Statistic result of the microarray comparison of reference DNA and reference DNA

Microarray hybridization result comparing reference DNA to reference DNA was displayed by the Microsoft Excel software.

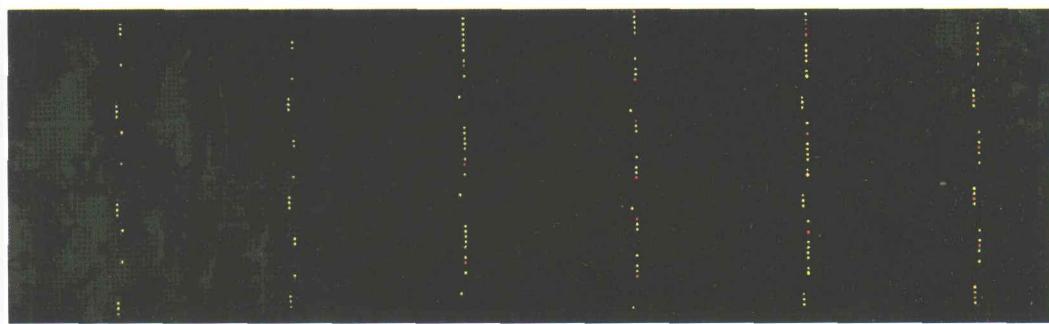


图 3 PCR 产物自身杂交图(待测 PCR 产物标 Cy3 + 参考 PCR 产物标 Cy5)

Fig 3. Scanning fluorimetric representation of the microarray comparison of test PCR products (labeled with Cy3) and reference PCR products (labeled with Cy5)

Forty-two yellow fluorescent spots indicate equivalent hybridization by test and reference PCR products; six red fluorescent spots indicate unopposed hybridization by reference sample, suggesting deletion of relevant products in test sample.

讨 论

基因芯片技术是近年来兴起的一种高通量的检测方法,已被广泛地用于微生物的基因表达、比较基因组学等方面的研究^[5,6]。其技术原理与核酸分子杂交相似,但技术特点是采用微量点样技术或光刻原位合成法将大量靶基因片段(cDNA, PCR 产物或寡核苷酸)包被在同一个固相载体上,即可在 1 次实验中进行成千上万个样品核酸与载体上固定靶标之间的分子杂交反应。

本研究所研制的问号钩体基因组 DNA 芯片是属于 PCR 产物芯片,即芯片上的靶序列是根据特异性的基因序列设计相应的引物进行 PCR 扩增而获得。可见,制备芯片的第一步是针对每个特异基因设计 1 对引物。我们选用了 1 个适用于芯片大批量引物设计工作的引物设计软件——Primegens^[7]。该软件的特点是快速、自动化,能在 3~4 h 内自动设计约 3 000 对特异性的引物。在引物设计时,考虑到小基因的 PCR 扩增效率与产物的纯化得率都较低,且杂交实验的信号不强等原因,因此未对长度<250 bp 的基因设计引物。

在完成芯片制备后,最重要的是通过参考株自身杂交实验对芯片质量进行初步评估,并对标记探针的制备、杂交温度及洗片条件进行了摸索与优化。

通过参考株自身杂交实验首先是对芯片点样情况进行全面地评估。由于芯片点样结果会受到点样室的温、湿度,洁净程度,样品的性状及点样仪的参数设定等多种因素的影响,因此一方面要严格控制点样室的各种环境指标,另一方面需利用少量样品进行预点实验,以保证大批量点样时获得高品质的芯片。

其次,采用双色荧光杂交法^[8](用 Cy3 和 Cy5 分别标记问号钩体赖型株的基因组 DNA)进行参考株自身比较实验,可进一步全面评估芯片质量。首先可通过分析芯片上设置的阴性质控点的信号值,监控芯片制备和杂交过程。数据分析显示在参考株自身比较实验中所有阴性对照的信号值在 100 以下,说明阴性对照的选择及芯片实验流程基本符合标准。另一方面通过数据处理与分析可综合评价芯片实验的假阳性率。理论上自身比较实验的杂交结果经数据读取、归一化处理分析后,获得的所有基因

的 $\log_2(Cy3/Cy5)$ 应为 0,但由于实验中系统误差和随机误差的存在,必有部分基因的该值会偏离 0。以通用的 cutoff 值筛选假阳性基因,计算假阳性率可初步评估芯片实验的可靠性。

再次,为了进一步检验芯片的杂交实验条件是否适用于问号钩体的比较基因组学研究,我们设计了芯片靶基因自身杂交实验。将选好的靶基因 PCR 产物混合后作为标记反应的模板,并使待测 DNA 相对于参考 DNA 缺失了部分已知的基因。通过该实验来检验所选取的芯片杂交条件和制备的芯片是否能特异性地检出已知的缺失基因。

在完成芯片制备及相应评估实验后,以本研究所初步建立的相应杂交体系、数据分析方案,可进一步开展基于芯片的比较基因组学研究。即通过比较问号钩体同种不同菌株(如有毒株与无毒株)或遗传学上与其相近的不同菌种在基因组上的差异与相似性,深入研究菌株间或菌种间基因组的多样性,可较全面地探究问号钩体血清学分型的分子基础、毒力相关基因的分布及钩体基因组进化的基本规律等。

参 考 文 献

- Zhao GP, Guo XK, Ren SX, et al. Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature*, 2003, 422: 888~892
- Bjorkholm B, Gordon JI, Engstrand L, et al. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infect Immun*, 2001, 69: 7832~7838
- Israel DA, Falkow S, Peek RM, et al. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J Clin Investig*, 2001, 107: 611~620
- Israel DA, Salama N, Peek RM, et al. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci USA*, 2001, 98: 14625~14630
- Luechini S, Thompson A, Hinton JCD. Microarray for microbiologists. *Microbiology*, 2001, 147: 1403~1414
- Fitzgerald JR, Musser JM. Evolutionary genomics of pathogenic bacteria. *Trends Microbiol*, 2001, 9: 547~553
- Xu D, Li G, Xu Y, et al. Primegens: robust and efficient design of gene-specific probes for microarray analysis. *Bioinformatics*, 2002, 18: 1432~1437
- Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res*, 1996, 6: 639~645

(收稿日期:2006-05-08)

ISSN 1673-6184 CN 31-1966/R 邮发代号:4-341 定价:12.00 元