



登革病毒cDNA的生物信息学分析及其Oligo探针设计

生物芯片是近年来在生命科学领域中崭露头角的一项新技术, 凭其快速、高效、敏感、经济、平行化、自动化等特点, 在序列分析、基因表达、基因组研究及基因诊断等领域已显示出很好的应用前景[1][2][3][4]。生物芯片的特点决定它必须依赖于生物信息学的手段和方法来进行芯片设计中探针的筛选、芯片分析结果的研究等[5][6]。本研究根据GenBank数据库中的生物信息, 利用美国国家生物技术信息中心(NCBI)中的BLAST免费分析软件找出4型登革病毒的保守序列及各型特异性序列, 然后用生物学软件Oligo6.0设计了48条长度均为70-mer的Oligo探针。这些Oligo探针后期将打印成DNA芯片, 用于登革病毒的基因检测, 并有效地对登革病毒进行分型。

1 材料与方法

1.1 材料

NCBI GenBank中4种血清型登革病毒的cDNA全长序列; NCBI开发的免费生物信息学检索软件BLAST; 生物学软件Oligo6.0。

1.2 方法

1.2.1 BLAST分析 从GenBank数据库获取4型登革病毒的cDNA全长序列, 将各型登革病毒的5'非编码区、3'非编码区及3种结构蛋白(C蛋白、PrM蛋白、E蛋白)和7种非结构蛋白(NS1、NS2a、NS2b、NS3、NS4a、NS4b和NS5蛋白)的cDNA序列分别进行BLAST分析。根据BLAST分析结果获得登革病毒的种属保守序列及各型特异性序列, 以作为设计Oligo探针的备选序列。

1.2.2 Oligo探针的设计 用生物学软件Oligo6.0分别对BLAST检索所得的登革病毒的种属保守序列和型特异性序列逐一进行分析, 设计长度均一的70-mer Oligo探针。为减少非特异性杂交, 保证芯片杂交的敏感度和精确度, 探针设计遵循以下5条基本原则: (1)G+C含量为40%~60%; (2)探针的 T_m 值应接近整个基因组的平均 T_m 值, 上下波动5℃; (3)避免单一碱基连续重复7次以上; (4)探针分子内部互补碱基少于6 bp; (5)每一个探针与非靶基因的相似性必须小于70%, 探针连续同源的碱基数不应超过20个。对备选探针进行筛选时, 各种参数尽量控制在最佳范围, 以保证打印在同一张芯片上的探针在相同的杂交、清洗条件下得到最佳的杂交效果。部分Oligo探针的一些参数值如表1所示。

表 1 Oligo 探针的一些参数值

Tab.1 Some parameters of the oligonucleotide probe

Number	T _m (°C)	GC %	bp of the steadiest duplexes (bp)	ΔG of the steadiest duplexes (kcal/mol)
1	89.3	45.7	3	-1.7
2	90.4	48.6	4	-0.9
3	89.9	47.1	3	-0.3
4	89.9	47.1	3	0.1
5	90.4	48.6	3	-2.6
6	91.0	50.0	4	-0.1
7	90.4	48.6	3	-1.9
8	88.7	44.3	3	1.0
9	89.3	47.1	5	-1.8
10	90.4	48.6	4	-0.5

T_m: melting temperature

2 结果

2.1 登革病毒保守序列及型特异性序列

将各型登革病毒的5'非编码区、3'非编码区及3种结构蛋白和7种非结构蛋白的cDNA序列分别放入GenBank国际数据库系统进行BLAST分析后,获得长度≥70 bp的登革病毒种属保守序列2个及型特异性序列93个。表2所列为登革病毒的2个保守序列及部分型特异性序列。

表 2 BLAST 分析后所得保守序列及部分型特异性序列

Tab.2 Conserved sequences and partial specific sequences acquired by BLAST analysis

Number	Usage	Gene	Location (bp)
1	Genus identification	C	90-170
2	Genus identification	3'UTR	10 619-10 704
3	DEN-1 identification	E	2 200-2 300
4	DEN-2 identification	NS1	1 307-1 487
5	DEN-3 identification	E	2 365-2 440
6	DEN-4 identification	NS1	3 450-3 620

2.2 Oligo探针的碱基序列

遵循探针设计的基本原则，运用Oligo6.0软件从2个登革病毒种属保守序列和93个型特异性序列中挑选出48条70-mer 的Oligo探针，表3所列为部分70-mer 的Oligo探针。

表 3 Oligo 探针的碱基序列

Tab.3 Probe base sequences of the oligonucleotide probes

Number	Probe base sequence
1	AACAACCAACGGAAAAAGACGGGTCGACCGTCTTTCAATATGCTGAAACGCGCGAGAAACCGCGTGTCAA
2	CTTGGAGGACCATTATGGCTGTGTTGTTTGTGGTCACACTCATTCTTTGTGCAGGACAAGCTGTCTTCA
3	GCCTCAATATCAAACATAACTACGGCAACAAGATGTCCAACGCAAGGAGAGCCTTATCTGAAAGAGGAAC
4	CCACTCAGCTGGCGACCCTAAGGAAGCTATGCATTGAGGGAAAAATTACCAACATAACAACCGACTCAAG
5	CAGGAAAACATGGTAAAGAAGTCAAGATAACACCACAGAGCTCCATCACAGAGGCGGAACCTGACAGGCTA
6	CCAAAAGCGAAACGAGGCACAGCACAAATTATGGAGGTGACAGCCAGGTGGTTATGGGGTTTTCTCTCTA
7	GATGGAGCAGAAAAATGCTGATGACTGGAACATTGGCTGTGTTCCCTCTTCTCACAATGGGACAATTGAC
8	TAGGAGCAGATGTACAGAATACCACCTTCATCATCGACGGCCCAAACACCCCAGAATGCCCTGATAACCA
9	GGTCATTCAAGTTAGAGAAGGAAGTGGCTGAGACCCAGCATGGAACCTTCTAGTGCAGGTTAAATACGA
10	CTCTGAGACACCCAGGATTCACGGTGATAGCCCTTTTTCTAGCACATGCCATAGGAACATCCATCACCCA

近年来登革病毒感染有扩大之势，发病率不断升高[7][8]。登革病毒根据抗原性不同分为4种血清类型，在分类上属黄病毒科黄病毒属。在黄病毒科的70多个成员中，登革病毒是分布最广、致人发病最多、危害最大的一种虫媒病毒。登革病毒感染的诊断，过去主要依靠病毒分离和血清学检测。病毒分离和血清学方法既费时又繁琐，不利于疾病的早期诊断，因此迫切需要建立一种快速、敏感、特异的检测方法。

近几年新兴的DNA芯片技术以其快速、高效、敏感、经济、平行化、自动化等特点，在基因诊断领域内显示了良好的应用前景，正逐步成为一项高通量的现代化诊断技术[9]。本研究利用生物信息学检索软件BLAST及生物学软件Oligo6.0设计特异性高、长度一致、融解温度相近的Oligo探针，为后期打印DNA芯片，用于登革病毒的检测打下了基础。

在DNA芯片制备中，设计或收集特异性和灵敏度高的探针是制备高质量DNA芯片的关键步骤[10]。为确保探针的特异性需行序列比对。序列比对是生物信息学的核心，比对的主要目的在于阐明序列之间的同源关系，目前最常用的序列比对研究方法为BLAST检索。用BLAST分析我们有如下经验：(1)先分别对各基因的cDNA序列进行全长比对；(2)根据比对结果剔除同源性高的区段，对同源性低的区段再行二次比对甚至三次比对，并选择再次比对结果中相似性分数值小于40的区段作为特异性序列。这样经生物信息学软件挑选出来的特异性序列和DDBJ/EMBL/GenBank国际数据库系统中其他现有全部序列的同源性最低，可大大减少非特异性杂交，提高检测的精确度，同时也能有效地提高BLAST检索的效率。

在用生物学软件Oligo6.0对探针进行设计时，选择70-mer长度均一的寡核苷酸作为探针可使杂交条件趋向均一，灵敏度增高。对备选探针进行筛选时，各种参数尽量控制在最佳范围，如Tm值在88.7~91.0℃之间，GC含量为44%~50%，以保证打印在同一张芯片上的探针在相同的杂交、清洗条件下得到最佳的杂交效果。探针最稳定二级结构的配对碱基长度小于6 bp， ΔG 趋于0或为正值，保证探针内部不会形成稳定的二级结构而影响探针与芯片杂交的效率。

利用BLAST系统和生物学软件Oligo6.0设计登革病毒诊断芯片的探针，是一种简便可行、有效的方法。但随着基因组计划的实施，以核酸序列数据库为代表的生物信息数据正以指数形式增加，而对于这些生物信息数据在计算机上的存储检索却远远跟不上这种发展。我们在进行BLAST检索时，就明显感觉到BLAST系统由于数据库的庞大而存在检索速度慢、程序消耗内存大等弱点，因此需要对现有的生物学数据处理工具进行优化和改进。此外还需积极开发更多更好的生物信息学软件[13]，以充分利用现在数量庞大并日益增长的生物信息资源。

参考文献：

- [1] 石嵘, 马文丽, 宋艳斌, 等. 两种限制性标记方法提高基因芯片杂交结果的信噪比[J]. 第一军医大学学报, 2003, 23(2): 124-6.
Shi R, Ma WL, Song YB, et al. Two restriction fluorescence labeling methods for enhancing the signal-to-noise ratio of cDNA microarray hybridization[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2003, 23(2): 124-6.
- [2] De Benedetti VM, Biglia N, Sismondi P, et al. DNA chips: the future of biomarkers [J]. Int J Biol Markers, 2000, 15 (1): 1-92.
- [3] 孙青, 丁彦青, 高雪芹, 等. 肿瘤转移相关基因cDNA芯片的制备与应用[J]. 第一军医大学学报, 2002, 22(12): 1070-5.
Sun Q, Ding YQ, Gao XQ, et al. Development and application of cDNA microarray of tumor metastasis-associated genes[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22 (12): 1070-5.
- [4] 李凌, 马文丽, 毛向明, 等. HIV基因芯片的初步研究[J]. 第一军医大学学报, 2002, 22(8): 724-7.
Li L, Ma WL, Mao XM, et al. Preliminary study of development of gene chips for HIV diagnosis[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(8): 724-7.
- [5] Fortna A, Gardiner K. Genomic sequence analysis tools: a user's guide [J]. Trends Genet, 2001, 17(3): 158-64.

- [6] 李衍达, 孙之荣. 生物信息学: 基因和蛋白质分析的实用指南[M]北京: 清华大学出版社, 2000. 70-84.
- [7] Gubler DJ. Dengue and dengue hemorrhagic fever[J]. Clin Microbiol, 1998, 11(4): 480-96.
- [8] Jelinek T. Dengue fever in international travelers[J]. Clin Infect Dis, 2000, 31(1): 144-52.
- [9] Bavykin SG, Akowski JP, Zakhariev VM, et al. Portable system for microbial sample preparation and oligonucleotide microarray analysis[J]. Appl Environ Microbiol, 2001, 67(2): 922-8.
- [10] Kenichi K, Akira S. Probe design for DNA chips[J]. Gen Inf, 1999, 10(2): 225-6.
- [11] 毛向明, 马文丽, 姜立, 等. 应用RD-PCR方法制备K562细胞表达谱芯片基因探针[J]. 第一军医大学学报, 2002, 22(6): 548-50.
- Mao XM, Ma WL, Jiang L, et al. Application of restriction display PCR in preparing the probes of the gene chip for investigating gene expression profile of K562 cells[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(6): 548-50.
- [12] 张宝, 马文丽, 石嵘, 等. 探针的纯化与否对基因芯片重复利用的影响[J]. 第一军医大学学报, 2002, 22(3): 208-10.
- Zhang B, Ma WL, Shi R, et al. Effect of probe purification on the reutilization of gene chip[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(3): 208-10.
- [13] Higgins DG, Bleasby AJ. Improved software for multiple sequence alignment[M]. Cabios, 1991, 89(6): 189-92.

参考文献:

- [1] 石嵘, 马文丽, 宋艳斌, 等. 两种限制性标记方法提高基因芯片杂交结果的信噪比[J]. 第一军医大学学报, 2003, 23(2): 124-6.
- Shi R, Ma WL, Song YB, et al. Two restriction fluorescence labeling methods for enhancing the signal-to-noise ratio of cDNA microarray hybridization[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2003, 23(2): 124-6.
- [2] De Benedetti VM, Biglia N, Sismondi P, et al. DNA chips: the future of biomarkers [J]. Int J Biol Markers, 2000, 15 (1): 1-92.
- [3] 孙青, 丁彦青, 高雪芹, 等. 肿瘤转移相关基因cDNA芯片的制备与应用[J]. 第一军医大学学报, 2002, 22(12): 1070-5.
- Sun Q, Ding YQ, Gao XQ, et al. Development and application of cDNA microarray of tumor metastasis-associated genes[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(12): 1070-5.
- [4] 李凌, 马文丽, 毛向明, 等. HIV基因芯片的初步研究[J]. 第一军医大学学报, 2002, 22(8): 724-7.
- Li L, Ma WL, Mao XM, et al. Preliminary study of development of gene chips for HIV diagnosis[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(8): 724-7.
- [5] Fortna A, Gardiner K. Genomic sequence analysis tools: a user's guide [J]. Trends Genet, 2001, 17(3): 158-64.
- [6] 李衍达, 孙之荣. 生物信息学: 基因和蛋白质分析的实用指南[M]北京: 清华大学出版社, 2000. 70-84.
- [7] Gubler DJ. Dengue and dengue hemorrhagic fever[J]. Clin Microbiol, 1998, 11(4):

- [8] Jelinek T. Dengue fever in international travelers[J]. Clin Infect Dis, 2000, 31(1): 144-52.
- [9] Bavykin SG, Akowski JP, Zakhariev VM, et al. Portable system for microbial sample preparation and oligonucleotide microarray analysis[J]. Appl Environ Microbiol, 2001, 67(2): 922-8.
- [10] Kenichi K, Akira S. Probe design for DNA chips[J]. Gen Inf, 1999, 10(2): 225-6.
- [11] 毛向明, 马文丽, 姜立, 等. 应用RD-PCR方法制备K562细胞表达谱芯片基因探针[J]. 第一军医大学学报, 2002, 22(6): 548-50.
- Mao XM, Ma WL, Jiang L, et al. Application of restriction display PCR in preparing the probes of the gene chip for investigating gene expression profile of K562 cells[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(6): 548-50.
- [12] 张宝, 马文丽, 石嵘, 等. 探针的纯化与否对基因芯片重复利用的影响[J]. 第一军医大学学报, 2002, 22(3): 208-10.
- Zhang B, Ma WL, Shi R, et al. Effect of probe purification on the reutilization of gene chip[J]. J First Mil Med Univ/Di Yi Jun Yi Da Xue Xue Bao, 2002, 22(3): 208-10.
- [13] Higgins DG, Bleasby AJ. Improved software for multiple sequence alignment[M]. Cabios, 1991, 89(6): 189-92.