

一种以测序为基础的基因分型方法的建立*

刘泽寰, 林蒋海, 符志彦, 潘德京, 徐安龙
(中山大学生命科学学院, 广东 广州 510275)

摘 要: 设计了一种自动化的基因分型方法, 能够直接利用 PCR 产物的 DNA 测序结果作为数据, 通过自动化基因分型程序 (automatic genotyping program) 与一个含有其等位基因序列的数据库进行比较, 来完成高精度的基因分型工作. 用这种方法对 50 个随机挑选的中国南方人的 HLA-DPB1 基因进行了基因分型, 证明该方法是可行的, 并从中发现了 2 个不明确杂合子以及一个新的等位基因序列. 随后对这些样品克隆测序, 发现除了 2 个不明确杂合子得到了进一步的区分外, 其它样品都与前面的分型结果一致, 显示了 96% 的精确性. 通过以上的研究表明, 成功建立了一种高精度的、基于序列的分型 (sequence-based typing) 方法. 既然该方法能成功分型 HLA-DPB1 基因, 预示了该方法同样能应用于其它基因的分型. 该 AGP 程序可以在 <http://genome.zsu.edu.cn> 免费下载.

关键词: 基因分型; 基于序列的分型; 测序; HLA-DPB1

中图分类号: R392.11 **文献标识码:** A **文章编号:** 0529-6579 (2001) 05-0083-05

DNA 基础上的基因分型技术如 PCR-RFLP^[1]、PCR-AFLP^[2]、PCR-SSO^[3]、PCR-SSP^[4]等目前已被广泛应用于日常的基因分型工作中. 这些方法有一个共同的缺点就是只适用于检测已知等位基因上的多态性位点, 而不适用于检测新的等位基因^[5]; PCR^[6]和 SSCP^[7]这两种方法虽然适用于检测新的等位基因, 但是它们却不能够给出进一步的有关新基因序列的具体信息. 然而 SBT 方法却弥补了这两方面的不足, 而且还有助于发现新等位基因. 它是直接根据基因多态性区域的测序结果进行分型的高精度分型方法^[8]. 现今的 SBT 方法主要可以分为两种: 一种是先通过克隆将等位基因分成纯合的 DNA 模板, 然后直接测序定型; 另一种是直接扩增目的基因用于测序, 然后用电脑算出基因型. 前一种方法的优点是可以得到确切无疑的结果, 然而费时费力. 后一种方法的优点在于借助特别的运算完成对杂合子基因型的定型. 杂合子基因型的确定是基因分型中最难解决的问题, 纯合子基因型可以通过简单的序列比较来完成, 而杂合子则要通过复杂得多的运算. 后一种方法的惟一缺点是对某些特定等位基因的组合 (不明确杂合子) 不能给出精确的分型结果^[9], 但是由于不明确杂合子出现的频率不高, 该方法仍然具有实用

价值.

为了使这种 SBT 方法可以服务于我们的日常分型工作, 我们专门设计了 AGP 程序, 该程序通过将测序获得的序列与一个含有相应已知等位基因序列的数据库进行比较而完成对等位基因的自动化指定.

在研究中, 选择了 HLA-DPB1 基因来验证我们建立的 SBT 方法, 这是因为该基因不但在免疫系统中发挥着重要的作用, 而且它的第二外显子呈现高度的序列多样性^[10], 是验证基因分型方法的理想材料.

1 材料和方法

1.1 样 品

随机从中国南方人群中挑选了 50 个健康无关个体作为样品, 并对他们的 HLA-DPB1 基因的第二外显子进行了 PCR 扩增, 扩增产物随后被用于 DNA 测序.

1.2 DNA 抽提和目的片段的扩增

样品的基因组 DNA 采用标准的酚/氯仿抽提、乙醇沉淀法从外周血淋巴细胞中提取. PCR 反应的上游引物为 DPB1 (5'-GAG GAT TAG ATG AGA GTG GCG CCT-3'), 下游引物为 DPB2 (GCC CAA AGC CCT CAC TCA CCT CGG), 它们分别位

* 基金项目: 国家杰出青年基金资助项目 (39725007)

收稿日期: 2001-03-05; 作者简介: 刘泽寰 (1972-), 男, 博士; 通讯联系人: 徐安龙.

于 DPB1 基因第二外显子两端的内含子序列中。所有引物均由上海生工生物工程技术有限公司合成。PCR 反应的组成及条件为: 100 μL 的反应液里包含两种引物各 0.5 $\mu\text{mol/L}$ 、4 种 dNTP 各 0.25 mmol/L 以及 Taq 酶 4 个酶活单位, 于 94 $^{\circ}\text{C}$ 变性 60 s, 56 $^{\circ}\text{C}$ 退火 60 s, 72 $^{\circ}\text{C}$ 延伸 90 s, 共 35 个循环。反应均采用德国 Biometra 公司的温度梯度 PCR 仪。

1.3 测序反应

PCR 反应的产物经由 PCR 纯化试剂盒纯化后, 交由上海基康生物技术公司测序。测序所用的引物为正向的位于第二外显子上游 19 ~ 36 bp 处的 DPBs (5' - GTG GCG CCT CCG CTC ATG - 3')。

1.4 序列分析及基因分型

设计的 AGP 程序用 C++ 语言编写, 其自动基因分型原理如图 1 所示。它是通过将上述 DNA 测序信息与已经建立的包含该基因所有已知等位基因序列的数据库进行循环比较来完成运算的。纯合子的测序结果通过与数据库的简单序列比较就可以找到与之相应的基因型; 对于杂合子的测序结果则需先对数据库中等位基因序列按组合原理进行两两配对成杂合子序列, 然后才与测序结果比较, 由于计算了所有可能出现的杂合概率, 其结果是相当精确的。当然, 当数据库中有超过一对的序列具有相同的杂合结果时, 就会出现前面提到的不明确杂合子的情况, 根据数学统计这种情况出现的比率较小, 本文所选择的 HLA-DPB1 基因理论上出现不明确杂合子的机率大约为 5%。

该程序的另一优点是有助于预见新等位基因的存在。当运算结果显示数据库中没有与测序结果相匹配的数据时, 该测序结果必定含有新序列。

本文所用的 HLA-DPB1 基因数据库是我们根据 WHO 国际命名委员会最近更新发布的 93 种 DPB1 等位基因第二外显子序列整理建立的^[10]。

1.5 分型结果的验证

由 AGP 程序所计算出的 DPB1 等位基因分型结果, 均以单克隆测序检验。克隆中所用的载体为 pUC19, 测序引物采用 M13 系列载体通用的测序引物, 均购自北京华美生物技术公司。

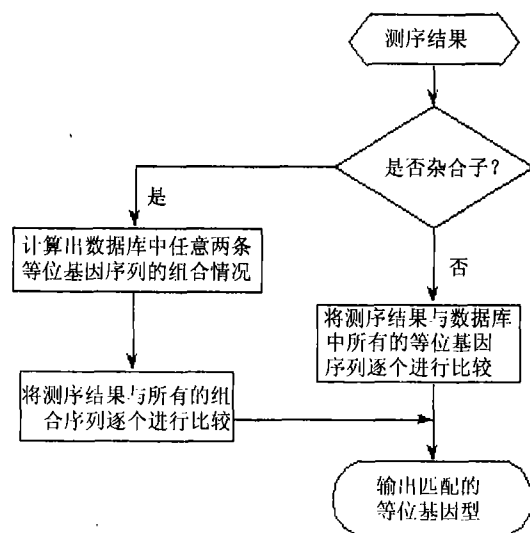


图 1 自动分型程序的原理

Fig. 1 Diagram of automatic genotyping program

2 结果

2.1 PCR 测序结果

50 个样品经过 PCR 扩增及序列测定后, 其中 24 个为 DPB1 纯合子, 26 个为杂合子。它们的分型结果将在下面阐述。

2.2 基因分型及结果验证

图 2 是样品 S06 的自动测序图谱, 我们在此把它作为示例来说明我们建立的 SBT 方法是如何运作的。图上黑点的位置即是所谓的杂合位点。可以看出在同一核苷酸位置上具有两种不同的碱基峰, 说明是由两种等位基因序列叠加的结果。

AGP 程序则将这个杂合序列与数据库中所有可能出现的两两杂合情况进行比较, 从而定出该样品的基因型。

表 1 列出了包括该样品在内的所有 50 个样品的分型结果, 其中样品 S16 和 S22 为不明确杂合子, 无法计算出精确的分型结果, 而样品 S42 由于没有与数据库匹配的运算结果, 推测含有新的等位基因序列。表 1 还列出了对分型结果进行验证的克隆测序结果, 从两者的数据来看, 除了不明确杂合子外, 分型结果是完全一致的, 不明确杂合子和新等位基因序列只能通过克隆测序解决。

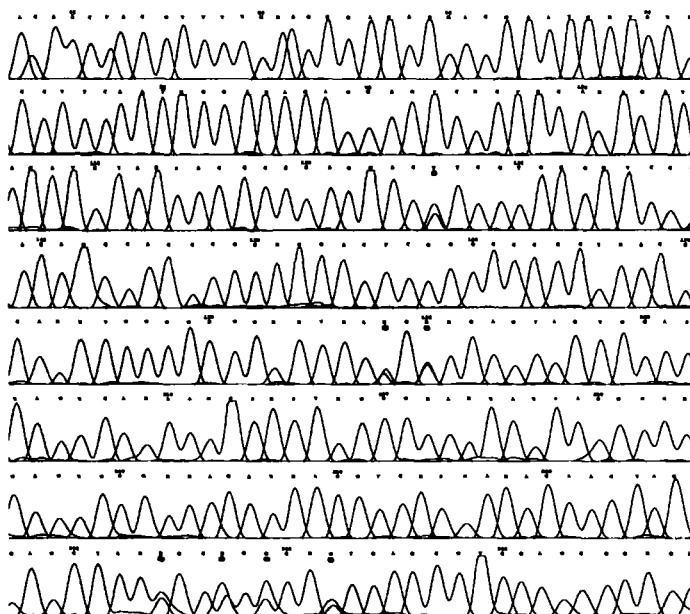


图2 样品 S06 的 DNA 自动测序图谱

Fig.2 Sequencing result of S06 by automated sequencer

3 讨论

本文叙述了一种由自动分型程序 AGP 辅助的 SBT 分型方法。该方法的重大优势在于基因分型的精确度高。由于它是利用 DNA 测序结果作为数据来源的，因此序列上每一个核苷酸都将一视同仁的被用于分型，不管它是在保守位置还是高变位置。这造成了两方面的优势：第一，在新等位基因的寻找方面，由于绝大多数的基因分型方法都只参考序列上高变区的核苷酸构成，从而遗漏了在序列保守区发生突变的新等位基因，本方法可以弥补这一缺陷；第二，在已知等位基因的鉴定上，其它的分型方法有可能造成误判，最典型的例子就是在本文的 DPB1 分型中，发现了两个个体含有 DPB1 * 3801 这一等位基因，而这一基因型在以往的研究中，从未在中国人群中被发现，这可以通过 DPB1 * 0501 和 DPB1 * 3801 的序列比较来说明，两者只在第二外显子的第 49 位核苷酸上存在 G 与 C 的不同，而第 49 位核苷酸在第二外显子上并不属于高变位置，因此其它的分型方法往往忽略这一位置，从而错把 DPB1 * 3801 当成了 DPB1 * 0501 来判断。该分型方法的另一个优势在于等位基因型的识别面大。就以 DPB1 基因作为例子，一般的基因分型方法诸如 SSP，在设计的时候由于引物的数量和位置、操作的复杂度和耗时等因素限制，通常只能识别 93 种 DPB1 等位基因中常见的 30 种，而本

方法则对它们全体均具有识别能力。最好的例子就是在本文的 DPB1 分型中，发现了 DPB1 * 1501、DPB1 * 20011、DPB1 * 2201、DPB1 * 2801 以及 DPB1 * 3801 等好几种以往从未报导在中国人群中出现的基因型。

由于每种基因的新等位基因型都在不断更新，大部分的基因分型方法都不能很好的适应这一转变，只能通过不断的增加限制性内切核酸酶、特异性寡核苷酸探针以及引物的种类和数量来弥补这一切。然而，这样做的最终结果是造成这些方法变得越来越费时、费力、得不偿失^[8]。本文所建立的分型方法，由于使用了可以存储的数据库文件，所以具有更新功能，只须把新公布的等位基因序列添加到程序的数据库中即可，剩下的工作交给电脑就可以了。当然，美中不足的是这一方法仍然存在对不明确杂合子无法精确分型的缺陷。在本实验中，这一不明确杂合子所占的比例为 4%，与理论预期值 5% 基本符合。然而，由于这一比例在统计中并不足以影响整体的分型结果，因此本研究参照绝大多数同类的分型方法不做进一步的弥补。除了上述原因外，还因为本文研究的是一种通用的基因分型方法，对于不同基因的不明确杂合子，不可能一一给出解决方案。在此，我们只能给出解决这一问题的基本策略：可以通过使用组特异性 PCR 引物来达到在测序之前将不明确杂合子分组归类的目的；也可以在测序之后，加一步特异性引物扩增或特异

表1 分型结果以及克隆测序验证结果的比较

Tab.1 Typing results verified by cloning and sequencing

样品	分型结果	分型结果的克隆测序验证
S01	DPB1 * 0501	DPB1 * 0501
S02	DPB1 * 0501/DPB1 * 1301	DPB1 * 0501/DPB1 * 1301
S03	DPB1 * 0501	DPB1 * 0501
S04	DPB1 * 0501	DPB1 * 0501
S05	DPB1 * 0501/DPB1 * 1301	DPB1 * 0501/DPB1 * 1301
S06	DPB1 * 0501/DPB1 * 0402	DPB1 * 0501/DPB1 * 0402
S07	DPB1 * 1301/DPB1 * 3801	DPB1 * 1301/DPB1 * 3801
S08	DPB1 * 0301/DPB1 * 2801	DPB1 * 0301/DPB1 * 2801
S09	DPB1 * 0501/DPB1 * 2801	DPB1 * 0501/DPB1 * 2801
S10	DPB1 * 0501	DPB1 * 0501
S11	DPB1 * 0501/DPB1 * 1301	DPB1 * 0501/DPB1 * 1301
S12	DPB1 * 0501	DPB1 * 0501
S13	DPB1 * 1401	DPB1 * 1401
S14	DPB1 * 0501/DPB1 * 01011	DPB1 * 0501/DPB1 * 01011
S15	DPB1 * 0501	DPB1 * 0501
S16	DPB1 * 02012/DPB1 * 1501 or DPB1 * 0402/DPB1 * 2201	DPB1 * 0402/DPB1 * 2201
S17	DPB1 * 2101/DPB1 * 2801	DPB1 * 2101/DPB1 * 2801
S18	DPB1 * 0501	DPB1 * 0501
S19	DPB1 * 0202/DPB1 * 1301	DPB1 * 0202/DPB1 * 1301
S20	DPB1 * 0401	DPB1 * 0401
S21	DPB1 * 2101	DPB1 * 2101
S22	DPB1 * 1001/DPB1 * 20011 or DPB1 * 1701/DPB1 * 2501	DPB1 * 1001/DPB1 * 20011
S23	DPB1 * 0301	DPB1 * 0301
S24	DPB1 * 0501	DPB1 * 0501
S25	DPB1 * 0501/DPB1 * 0402	DPB1 * 0501/DPB1 * 0402
S26	DPB1 * 0401	DPB1 * 0401
S27	DPB1 * 0202/DPB1 * 0401	DPB1 * 0202/DPB1 * 0401
S28	DPB1 * 02012/DPB1 * 0402	DPB1 * 02012/DPB1 * 0402
S29	DPB1 * 0402	DPB1 * 0402
S30	DPB1 * 0501/DPB1 * 0402	DPB1 * 0501/DPB1 * 0402
S31	DPB1 * 0501	DPB1 * 0501
S32	DPB1 * 0402	DPB1 * 0402
S33	DPB1 * 0202/DPB1 * 0402	DPB1 * 0202/DPB1 * 0402
S34	DPB1 * 0402/DPB1 * 2801	DPB1 * 0402/DPB1 * 2801
S35	DPB1 * 0202/DPB1 * 0402	DPB1 * 0202/DPB1 * 0402
S36	DPB1 * 0501	DPB1 * 0501
S37	DPB1 * 02012/DPB1 * 0402	DPB1 * 02012/DPB1 * 0402
S38	DPB1 * 0501	DPB1 * 0501
S39	DPB1 * 0501	DPB1 * 0501
S40	DPB1 * 0501	DPB1 * 0501
S41	DPB1 * 0501/DPB1 * 3801	DPB1 * 0501/DPB1 * 3801
S42	Unable to find matching allele	DPB1 * 0501/new
S43	DPB1 * 2101	DPB1 * 2101
S44	DPB1 * 0501/DPB1 * 0402	DPB1 * 0501/DPB1 * 0402
S45	DPB1 * 0402	DPB1 * 0402
S46	DPB1 * 0501	DPB1 * 0501
S47	DPB1 * 0501/DPB1 * 20011	DPB1 * 0501/DPB1 * 20011
S48	DPB1 * 1301	DPB1 * 1301
S49	DPB1 * 0501/DPB1 * 1401	DPB1 * 0501/DPB1 * 1401
S50	DPB1 * 02012/DPB1 * 0402	DPB1 * 02012/DPB1 * 0402

性寡核苷酸探针杂交，从而进一步将它们分开。具体的情况还要看操作者权衡利弊来定。

参考文献：

- [1] OTAL M, SEKI T, NOMURA N, et al. Modified PCR-RFLP method for HLA-DPB1 and DQA1 genotyping[J]. Tissue Antigens, 1991, 38: 60 – 71.
- [2] BAXTER-LOWE L A. Nucleic acid-based methods for HLA typing: the wave of the future[J]. Am Soc Hist Immunogenetics, 1992, 16: 6.
- [3] TIERCY J M, JEANNET M, MACH B. A new approach for the analysis of HLA class II polymorphism: HLA oligotyping[J]. Blood Reviews, 1990, 4: 9 – 15.
- [4] SALAZAR M, DEULOFEUT R, YUNIS J J, et al. A fast PCR-SSP method for HLA-DQ genetic typing[J]. Tissue Antigens, 1993, 41: 102 – 106.
- [5] KNIPPER A J, HINNEY A, SCHUCH B, et al. Selection of unrelated bone marrow donors by PCR-SSP typing and subsequent non-radioactive sequence-based typing for HLA-DRB1/3/4/5, DQB1, and DPB1 alleles[J]. Tissue Antigens 1994, 44: 275 – 284.
- [6] CLAY T M, BIDWELL J L, HOWARD M R, et al. PCR – fingerprinting for selection of HLA matched unrelated marrow donors[J]. The Lancet 1991, 337: 1049 – 1052.
- [7] CARRINGTON M, MILLER T, WHITE M, et al. Typing of HLA-DQA1 and DQB1 using single-strand conformation polymorphism[J]. Hum Immunol, 1992, 33: 208 – 212.
- [8] SANTAMARIA P, BOYCE-JACINO M T, LINDSTROM A, et al. HLA class II typing: Direct sequencing of DRB, DQB, and DQA genes[J]. Hum Immunol, 1992, 33: 69 – 81.
- [9] VERSLUIS L F, ROZEMULLER E, TONKS S, et al. High-resolution HLA-DPB typing based upon computerized analysis of data obtained by fluorescent sequencing of the amplified polymorphic exon 2[J]. Hum Immunol, 1993, 38: 277 – 283.
- [10] ROBINSON J, WALLER M J, PARHAM P, et al. IMGT/HLA database-a sequence database for the human major histocompatibility complex [J]. Nucleic Acids Res, 2001, 29(1): 210.

A Sequence-based Typing Method for HLA-DPB1 Genotyping

LIU Ze-huan, LIN Jiang-hai, FU Zhi-yan, PAN De-jing, XU An-long

(School of Life Sciences, Zhongshan University, Guangzhou 510275, China)

Abstract: An automatic sequence-based typing (SBT) method is established for genotyping. It is a high-resolution genotyping method by utilizing the automatic genotyping program (AGP) to compare the DNA sequence data obtained by direct sequencing of PCR products with a database containing known allelic sequences. It offers SBT the ability to distinguish heterozygous individuals as well as identifying homozygous individuals. To validate this typing system, 50 unrelated southern Chinese individuals were typed in the HLA-DPB1 locus as an example. Two ambiguous heterozygotes and one novel allelic sequence were found in these individuals. The typing results were confirmed by cloning and subsequent sequencing of the cloned DNA without any discordance except two ambiguous heterozygotes, indicated a 96% concordance. Thus, an efficient and accurate SBT method was established by using AGP program. It was proved applicable in genotyping of HLA-DPB1 locus, and can be easily applied and extended to other gene.

Keywords: genotyping; sequence-based typing; sequencing; HLA-DPB1