

用多重 PCR 检测上海地区汉族人群 9 个 STR 基因座的多态性

冯明亮,季芸,陆琼,马俊,稽月华,杨颖

(上海市血液中心,上海 200051)

摘要:利用多重 PCR 和四色荧光(5-FAM, JOE, NED 和 ROX) 自动化检测技术调查上海地区汉族人群 *D3S1358*、*vWA*、*FGA*、*D8S1179*、*D21S11*、*D18S51*、*D5S818*、*D13S317*、*D7S820* 等 9 个 STR 基因座多态性分布并计算该 9 个基因座的基因频率(P_i)、个体鉴别力(DP)、无偏倚期望杂合性(H)、多态性信息含量(PIC)和非父排除概率(PE)。结果显示:9 个 STR 基因座的基因型分布符合 Hardy-Weinberg 平衡,9 个 STR 基因座中 *FGA* 基因座的 DP 值最高为 0.9584,*D8S1179* 的 H 值最高为 0.9403,*D18S51* 的 PIC 值最高为 0.8560,*D18S51* 的 PE 值最高为 0.7391,9 个 STR 基因座累积个体鉴别力(CDP)为 0.9999996,累积非父排除能力(CPE)为 0.99991。9 个 STR 基因座适合作为中国人群的遗传标志,用于人类学、遗传疾病基因连锁分析、法医学亲子鉴定和个体识别等研究领域。

关键词:STR 基因座;上海地区汉族人

中图分类号:Q347

文献标识码:A

文章编号:0253-9772(2002)04-0403-04

The Polymorphism Distributions of Nine STR Loci in Shanghai Han Population

FENG Ming-liang, JI Yun, LU Qiong, MA Jun, JI Yue-hua, YANG Ying

(Shanghai Blood Center, Shanghai 200051, China)

Abstract: By multiplex amplification and four fluorescent technique, the polymorphism distributions of nine STR loci, *D3S1358*, *vWA*, *FGA*, *D8S1179*, *D21S11*, *D18S51*, *D5S818*, *D13S317* and *D7S820* were investigated in Shanghai Han population. Gene frequency (P_i), power of discrimination (DP), polymorphism information content (PIC) expected heterozygosity (H) and probability of paternity exclusion (PE) were calculated. All loci meet Hardy-Weinberg equilibrium. DP of *FGA* locus, H of *D8S1179* locus, PIC of *D18S51* locus and PE of *D18S51* locus are the biggest among nine STR loci. Cumulate DP (CDP) of nine STR loci is 0.9999996, Cumulate PE (CPE) of nine STR loci is 0.99991. Nine STR loci could be used as the genetic markers of Chinese population in the studies of anthropology, linkage analysis of genetic disease genes, individual identification and paternity test in forensic medicine.

Key words: STR locus; Shanghai Han population

短片段串联重复(short tandem repeats, STR), 通常由长度为 2~6 个 bp 的核心序列串联重复而成,广泛存在于人类 DNA 编码区及非编码区,具有高度遗传多态性^[1],遵循孟德尔遗传规律。

本文利用多重 PCR 和四色荧光(5-FAM、

JOE、NED 和 ROX) 自动化检测技术调查上海地区汉族人群 *D3S1358*、*vWA*、*FGA*、*D8S1179*、*D21S11*、*D18S51*、*D5S818*、*D13S317*、*D7S820* 等 9 个 STR 基因座的多态性分布,现将结果报告如下。

1 材料和方法

1.1 检测样本

采集 106 例上海地区汉族人外周静脉血 1ml, 用 ACD-B 保存液抗凝。

1.2 基因组 DNA 抽提

采用快速硫氰酸胍法, 详细操作见文献^[2]。

1.3 PCR 扩增

9 个 STR 基因座通过“Profiler Plus”试剂盒 (ABI 公司) 进行多重 PCR 扩增。PCR 扩增体系为 15 μ l^[3], 包括模板 DNA 6 μ l (0.3~0.75ng), 反应液 6.3 μ l, 金牌 Taq 酶 0.3 μ l, 引物混合液 3.3 μ l。PCR 扩增在 PE9600 自动热循环仪上进行, 循环参数^[4]: 第一步: 95 $^{\circ}$ C 变性 11min, 第二步: 94 $^{\circ}$ C 变性 1min、59 $^{\circ}$ C 复性 1min、72 $^{\circ}$ C 延伸 1min, 循环 28 次, 第三步: 60 $^{\circ}$ C 延伸 45min。

1.4 扩增产物检测^[4]

在 0.2ml 离心管中分别加入 2.2 μ l 去离子甲酰胺, 0.4 μ l 加样缓冲液, 0.4 μ l ROX^[500] 内标及 1.5 μ l 扩增产物。95 $^{\circ}$ C 变性 3 分钟, 随即冰浴, 备用。处理过的样品通过 377 型测序仪电泳收集电泳信息, 然后用基因扫描软件 (GeneScan 3.1) 分析扩增片段大小 (bp), 继而用分析软件 (Genotyper 2.1) 分析各基因座的基因型。

1.5 统计分析

1.5.1 基因频率按基因计数法测定。用 χ^2 检验比较期望值与观察值以验证是否复合 Hardy-Weinberg 平衡法则。

1.5.2 个体鉴别力 (power of discrimination, DP; 按 Fisher 法^[5] 计算)

$$DP = 1 - \sum_{i=1}^n P_i^2$$

P_i : 基因型频率

1.5.3 累积鉴别力 (cumulate DP, CDP)

$$C_{DP} = 1 - (1 - D_{P1})(1 - D_{P2}) \cdots (1 - D_{P9})$$

$D_{P1}, D_{P2} \cdots D_{P9}$ 分别表示 9 个 STR 基因座各自的个体鉴别力。

1.5.4 无偏倚期望杂合性 (expected heterozygosity, H ; 按 Nei 法^[6] 计算)

$$H = n \times (1 - \sum_{i=1}^n P_i^2) / (n-1)$$

P_i : 等位基因频率

n : 等位基因数

1.5.5 多态性信息含量 (polymorphism information content, PIC; 按 Botstein 法^[7] 计算)

$$PIC = 1 - \sum_{i=1}^n P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2P_i^2 P_j^2$$

P_i, P_j : 等位基因频率

n : 等位基因数

1.5.6 非父排除概率 (probability of paternity exclusion, PE; 按 Odelberg 法^[8] 计算)

$$P_E = \sum_{i=1}^n P_i(1 - P_i)^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n (P_i P_j)^2 (3P_i + 3P_j - 4)$$

P_i, P_j : 等位基因频率

n : 等位基因数

1.5.7 累积非父排除概率 (cumulate PE, CPE)

$$C_{PE} = 1 - (1 - P_{E1})(1 - P_{E2}) \cdots (1 - P_{E9})$$

$P_{E1}, P_{E2} \cdots P_{E9}$ 分别表示 9 个 STR 基因座各自的非父排除概率。

2 结果

2.1 等位基因频率分布

通过对 106 例样本的检测, 得到 9 个 STR 基因座等位基因频率分布资料, 见表 1。经 χ^2 检验, 结果 $P \geq 0.05$, 表明 9 个 STR 基因座的基因型分布符合 Hardy-Weinberg 平衡法则。

2.2 统计学分析

根据等位基因频率和基因型分布, 按照文献 5~8 方法, 计算出各基因座的个体鉴别力 (DP)、无偏倚期望杂合性 (h)、多态性信息含量 (PIC) 和非父排除概率 (PE), 结果见表 2。9 个 STR 基因座的累积鉴别力和累积非父排除概率分别为 0.9999996 和 0.99991。

2.3 通过 t 检验, 对上海地区汉族和文献 3, 4 提供的中国北方汉族、中国维吾尔族、美国白人和美国黑人群体 9 个 STR 基因座的群体遗传学资料进行分析比较, 结果 P 值均大于 0.05, 证明上述群体之间 9 个 STR 基因座的基因频率总体分布无显著差异。

3 讨论

本研究所选用 D3S1358、vWA、FGA、D8S1179、D21S11、D18S51、D5S818、D13S317、D7S820 等 9 个 STR, 由 2~4 个 bp 的核心序列串联重复而成, 长度在 100~400bp 之间, 该 9 个 STR 基因座的遗传学特征见表 3 所列。

表 1 9 个 STR 基因座等位基因频率

Table 1 The gene frequency of nine STR locus

<i>D3S1358</i>	<i>N</i>	P_i	<i>vWA</i>	<i>N</i>	P_i	<i>FGA</i>	<i>N</i>	P_i	<i>D8S1179</i>	<i>N</i>	P_i	<i>D21S11</i>	<i>N</i>	P_i
13	1	0.0047	14	49	0.2311	18	2	0.0094	10	24	0.1132	27	1	0.0047
14	14	0.0660	15	4	0.0189	19	12	0.0566	11	19	0.0896	28	7	0.0330
15	59	0.2783	16	38	0.1792	20	8	0.0377	12	19	0.0896	29	58	0.2736
16	71	0.3349	17	52	0.2453	21	25	0.1179	13	50	0.2358	30	68	0.3208
17	51	0.2406	18	43	0.2028	22	33	0.1557	14	36	0.1698	30.2	3	0.0142
18	15	0.0708	19	21	0.0991	22.2	3	0.0142	15	45	0.2123	31	17	0.0802
19	1	0.0047	20	5	0.0236	23	52	0.2453	16	16	0.0755	31.2	15	0.0708
	212	1.0000		212	1.0000	23.2	3	0.0142	17	2	0.0094	32	4	0.0189
						24	34	0.1604	18	1	0.0047	32.2	27	0.1274
						24.2	1	0.0047		212	1.0000	33.2	8	0.0377
						25	19	0.0896				34.2	2	0.0094
						25.2	1	0.0047				35.2	2	0.0094
						26	16	0.0755					212	1.0000
						27	3	0.0142						
						212	212	1.0000						

<i>D18S51</i>	<i>N</i>	P_i	<i>D5S818</i>	<i>N</i>	P_i	<i>D13S17</i>	<i>N</i>	P_i	<i>D7S820</i>	<i>N</i>	P_i
10	1	0.0047	7	7	0.0330	8	56	0.2642	7	2	0.0094
11	1	0.0047	9	11	0.0519	9	40	0.1887	8	34	0.1604
12	11	0.0519	10	40	0.1887	10	27	0.1274	9	10	0.0472
13	46	0.2170	11	66	0.3113	11	51	0.2406	10	31	0.1462
14	41	0.1934	12	51	0.2406	12	28	0.1321	11	72	0.3396
15	30	0.1415	13	37	0.1745	13	8	0.0377	12	57	0.2689
16	29	0.1368		212	1.0000	14	2	0.0094	13	6	0.0283
17	14	0.0660					212	1.0000		212	1.0000
18	10	0.0472									
19	9	0.0425									
20	11	0.0519									
21	5	0.0236									
22	1	0.0047									
23	2	0.0094									
24	1	0.0047									
	212	1.0000									

注: P_i 为等位基因频率 N 为某等位基因的观察个数

表 2 9 个 STR 基因座的 DP, H, PIC 和 PE 值Table 2 The value of DP, H, PIC and PE of nine STR locus

	<i>D3S1358</i>	<i>vWA</i>	<i>FGA</i>	<i>D8S1179</i>	<i>D21S11</i>	<i>D18S51</i>	<i>D5S818</i>	<i>D13S17</i>	<i>D7S820</i>
DP	0.88163	0.92293	0.95835	0.94571	0.92150	0.95710	0.90904	0.92987	0.88768
H	0.86695	0.93616	0.92280	0.94028	0.86328	0.92386	0.93042	0.93518	0.89268
PIC	0.71114	0.79165	0.85329	0.82841	0.77543	0.85600	0.69260	0.79133	0.73861
PE	0.52447	0.63383	0.73498	0.69159	0.62158	0.73909	0.58008	0.63261	0.56534

为了调查 *D3S1358*、*vWA*、*FGA*、*D8S1179*、*D21S11*、*D18S51*、*D5S818*、*D13S17*、*D7S820* 等 9 个 STR 基因座多态性分布的地域和人群的差异, 本文对上海地区汉族群体 9 个 STR 基因座的多态性分布, 与文献 3, 4 提供的群体遗传学资料进行了比较, 结果显示上海地区汉族、中国北方汉族、中国维吾尔族、美国白人和美国黑人群体之间 9 个 STR 基因座的多态性分布无显著差异 ($P > 0.05$)。但是个

别基因频率分布还是有比较明显的不同, 如在汉族群体中 *vWA* - 14, *D18S51* - 13, *D21S11* - 30, *D13S17* - 8, *D7S820* - 12 出现频率较高, 在白人群体中 *FGA* - 20, *D3S1358* - 14, *D18S51* - 13 出现频率较高, 在黑人群体 *D13S17* - 12, *D3S1358* - 14 出现频率较高。另外, 在理论上由 9 个 STR 基因座等位基因构成的基因型数远大于实际观察到的基因型数, 如 *D18S51* 基因座 15 种等位基因应构

成 120 种基因型,但本文仅发现其中 40 种,可见基因型分布并不均一。

DP 、 H 、 PE 和 P_{IC} 是衡量一个遗传标记物在人类学、遗传疾病基因连锁分析、法医学等研究领域应用价值的指标。Gill 等将 $DP \geq 0.9$, $H \geq 0.7$, $P_{IC} \geq 0.7$ 的基因座作为高鉴别能力的遗传标记物^[9],从表 2 可见,9 个 STR 中, vWA 、 FGA 、 $D8S1179$ 、 $D21S11$ 、 $D18S51$ 、 $D13S317$ 基因座 $DP > 0.9$, $H > 0.9$, $P_{IC} > 0.77$, 表明它们属于高鉴别力、高杂合度、高信息量基因座,极有应用价值。另外, $D3S1358$ 、 $D5S818$ 属于中高度多态形基因座,也可作为法医学个体识别^[10]的重要证据。

用多重 PCR 和四色荧光自动化检测技术分析 9 个微卫星基因座的基因型,分辨率和准确率高,可以分辨 1 个 bp 的长度,采用分子质量内标及标准品(ladder),避免了因电泳引起的误差,具有很好的重复性。由于 STR 基因座片段较短,大小分布于 100~400bp 之间,等位基因之间片段长度差异较小,且多个 STR 基因座可以同时复合扩增,分型简便、灵敏。

综上所述,9 个 STR 基因座遗传多态性丰富,遵循孟德尔遗传规律,适合作为中国人群的遗传标志^[11],用于人类学研究、遗传疾病基因连锁分析、法医学亲子鉴定和个体识别等研究领域。

表 3 9 个 STR 基因座遗传学特征

Table 3 Genetics characteristic of nine STR locus

基因座 Locus designation	染色体 Chromosome location	重复序列 Common sequence motif	长度范围(bp) Size range
<i>D3S1358</i>	3p	TCTA (TCTG) ₁₋₃ (TCTA) _n	114-142
<i>vWA</i>	12p12-pter	TCTA (TCTG) ₃₋₄ (TCTA) _n	157-197
<i>FGA</i>	4q28	(TTTC) ₃ TTTT TTCT (CTTT) _n CTCC (TTCC) ₂	219-267
<i>D8S1179</i>	8	(TCTR) _n ^c	128-168
<i>D21S11</i>	21	(TCTA) _n (TCTG) _n [(TCTA) ₃ (TCTA) ₃ TCA (TCTA) ₂ TCCA TA] (TCTA) _n	189-243
<i>D18S51</i>	18q21.3	(AGAA) _n	273-341
<i>D5S818</i>	5q21-31	(AGAT) _n	135-171
<i>D13S317</i>	13q22-31	(GATA) _n	206-234
<i>D7S820</i>	7q11.21-22	(GATA) _n	258-294

R:可以是 A 或 G

参考文献(References):

- [1] Schuler G D, Boguski, Stewart, *et al.* A gene map of the human genome[J]. *Science*, 1996, 274: 540~546.
- [2] 冯明亮, 季芸, 马俊, 等. 反向 PCR-SSOP 技术行 HLA-AB 分型与临床应用[J]. *中国实验血液学杂志*, 2001, 9(4): 359~362.
- [3] 杜志淳, 李莉, 林源. “中国罪犯 DNA 数据库”模式库—13 个 STR 基因座基因在中国人人群中的分布[J]. *法医学杂志*, 2000, 2: 1~5.
- [4] PERKIN ELMER Corporation. AmpliSTR Profiler plus™ PCR Amplification kit User's Manual [R]. 1998, 3: 1~41; 13: 1~9.
- [5] Fish R A. Standard calculations for evaluating a blood group system[J]. *Heredity*, 1951, 5: 95~102.
- [6] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individual[J]. *Genetics*, 1978, 89: 583.
- [7] Botstein P. Construction of a genetic linkage map in man using restriction length polymorphism[J]. *Am J Hum Genetic*, 1980, 32: 314.
- [8] Odelberg SJ. Repetitive DNA: molecular structure, polymorphism and forensic application. In DNA and other polymorphism in forensic science [M]. 1st ed. Chicago: Year Book Medica Publishers Inc, 1990, 26.
- [9] Gill P. A new method of STR interpretation using inferential logic—development of a Criminalintelligence database[J]. *Int J Leg Med*, 1996, 109: 14.
- [10] Hammond H A, Jin L, Zhong Y, *et al.* Evaluation of 13 short tandem repeat loci for use in personal identification applications [J]. *Am J hum Genet* 1994, 55: 175~189.
- [11] 黄代新, 张林, 吴梅筠, 等. 短串联重复序列 *D7S2201* 基因座的群体遗传学研究[J]. *遗传*, 2001, 23(2): 107~110.