

湖南省汉族人群 15 个 STR 基因座多态性分析

欧阳曙明^{1,2}, 殷照初^{1,2}, 江 源^{1,2}, 申 琴^{1,2}, 倪 斌^{*1,2}

(1. 湖南省计划生育研究所 湖南省现代优生技术重点实验室, 中国湖南 长沙 410008 ;

2. 湖南省天衡司法鉴定所, 中国湖南 长沙 410008)

摘 要: 应用美国 AmpFlSTR Identifiler 荧光标记复合扩增试剂盒, 结合 PE9700 型 PCR 仪和美国 ABI 公司 310 型遗传分析仪, 对湖南汉族人群 D8S1179、D21S11、D7S820、CSF1PO、D3S1358、TH01、D13S317、D16S539、D2S1338、D19S433、vWA、TPOX、D18S51、D5S818 和 FGA 共 15 个 STR 基因座进行多态性调查分析. 结果显示 15 个 STR 基因座的基因型分布符合 Hardy-Weinberg 平衡, 其杂合度(H)介于 0.593~0.900, 多态信息含量(PIC)介于 0.54~0.85, 个体识别力(DP)介于 0.780~0.963, 非父排除率(PE)介于 0.282~0.785, 累计个体识别力为 $(1 \sim 1.6 \times 10^{-17}) > 0.999\ 999\ 99$, 累计非父排除率为 0.999 999 5. 证明 15 个 STR 基因座在湖南省汉族人群中具有较高的多态性, 可应用于该地区群体学研究、法医学个体识别和亲权鉴定等.

关键词: 短串联重复序列; 遗传多态性; 法医学

中图分类号: Q347

文献标识码: A

文章编号: 1007-7847(2009)05-0399-04

Genetic Polymorphism Study on 15 STR Loci of Han Population in Hunan

OUYANG Shu-ming^{1,2}, YIN Zhao-chu^{1,2}, JIANG Yuan^{1,2}, SHEN Qin^{1,2}, NI Bin^{*1,2}

(1. Hunan Provincial Key Laboratory of Reproductive Health, Family Planning Institute of Hunan Province, Changsha 410008, Hunan, China; 2. Hunan Tianheng Forensic Institute, Changsha 410008, Hunan, China)

Abstract: The polymorphism distributions of 15 STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA) were investigated in a Hunan Han population using AmpFlSTR identifiler kit, GeneAmp PCR system 9700 and ABI 310 genetic analyzer. Gene frequency, heterozygosity (H), polymorphism information content (PIC), power of discrimination (DP) and probability of paternity exclusion (PE) were calculated, and all loci were tested for Hardy-Weinberg equilibrium. Results indicate that the gene frequency of these 15 STR loci is in Hardy-Weinberg equilibrium. The H is at 0.593~0.900, PIC is at 0.54~0.85, DP is at 0.780~0.963, PE is at 0.282~0.785. Cumulative DP of the 15 STR is $(1 \sim 1.6 \times 10^{-17}) > 0.999\ 999\ 99$, and the cumulative PE is 0.999 999 5. Therefore, the 15 STR loci used in this study are highly polymorphic in Hunan Han population and can be applied to population study, individual identification and paternity testing in forensic science.

Key words: short tandem repeat; genetic polymorphism; forensic science

(Life Science Research, 2009, 13(5): 399~402)

收稿日期: 2008-12-19; 修回日期: 2009-07-10

作者简介: 欧阳曙明(1981-), 男, 湖南湘乡人, 实习研究员, 硕士, 主要从事分子遗传学研究; * 通讯作者: 倪斌(1962-), 男, 湖南桃源人, 湖南省计划生育研究所研究员, 博士, 主要从事遗传学研究, Tel: 0731-4498979, E-mail: ouyang-008@163.com.

短串联重复序列(STR)是人类基因组中广泛存在的一类遗传标志,具有高度多态性和遗传稳定性^[1]. STR 基因座的等位基因数目多,杂合度高,个体识别力强,同时扩增片段短,检测灵敏度高,是目前群体遗传学,法医学研究及应用领域十分理想的 DNA 遗传标记^[2]. 本研究应用美国 AmpFISTR Identifiler 试剂盒复合扩增技术结合毛细管电泳技术,对湖南汉族人群的 15 个 STR 基因座的遗传多态性进行了调查,为群体学研究,法医学应用及 DNA 数据库建设提供基础数据.

1 材料和方法

1.1 样本及其 DNA 提取

自 2007 年 11 月至 2008 年 6 月收集本实验室个案累计血样共 270 例,所有个体均来源于湖南省汉族,且三代之类均居住于湖南,用 Chelex-100 法抽提 DNA^[3].

1.2 PCR 复合扩增及扩增产物的检测

15 个 STR 基因座采用美国 AmpFISTR Identifiler 试剂盒进行复合 PCR 扩增. PCR 体系及程序设定均按照试剂盒说明书进行,然后于

美国 PE 公司 9700 型 PCR 仪进行扩增. 扩增产物在美国 ABI310 型遗传分析仪上进行毛细管电泳,电泳信息由 Data Collection 数据软件自动收集,电泳结束后,由 GeneMapper3.1 软件分析各基因座的基因型.

1.3 统计分析

按基因计数法计算等位基因频率、基因型频率. 应用卡方检验比较基因型的观察值和期望值之间是否符合 Hardy-Weinberg 平衡. 运用 PowerStatsV1.2 软件计算 15 个 STR 基因座的遗传学指标^[4],包括:杂合度(H)、多态信息含量(PIC)、个体识别力(DP)、非父排除率(PE).

运用 χ^2 检验将湖南汉族与北京人群^[5]、四川人群^[6]、陕西汉族人群^[7]、东北汉族人群^[8]、厄瓜多尔黑人^[9]、美洲印第安人^[9]和日本人^[10]的多个 STR 位点的等位基因频率进行比较.

2 结果

2.1 PCR 扩增及电泳结果

15 个 STR 位点均得到了有效扩增,电泳图谱显示扩增结果较好,特异性高.

表 1 湖南汉族人群 15 个 STR 基因座等位基因频率
Table 1 The allele frequencies of 15 STR loci of Han population from Hunan

Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
5				0.002											
6			0.002			0.107									
7				0.007		0.267	0.006							0.026	
8			0.165			0.050	0.296	0.006				0.524		0.004	
9			0.057	0.039		0.485	0.163	0.291				0.111		0.061	
9.2			0.002												
9.3						0.037									
10	0.130		0.126	0.250		0.054	0.119	0.141				0.017	0.002	0.231	
10.2				0.006											
11	0.107		0.350	0.224			0.228	0.243		0.002		0.322	0.009	0.317	
12	0.120		0.246	0.376	0.004		0.163	0.228		0.028		0.024	0.043	0.224	
12.2										0.004		0.002			
13	0.180		0.041	0.087	0.004		0.022	0.081		0.278	0.004		0.194	0.122	
13.2										0.039					
14	0.215		0.011	0.009	0.046		0.004	0.011		0.241	0.252		0.185	0.013	
14.2										0.107					
15	0.170				0.344					0.081	0.024		0.189	0.002	
15.2										0.163					
16	0.069				0.319				0.013	0.017	0.183		0.135		
16.2										0.033					
17	0.007				0.215				0.069		0.219		0.076		0.004
17.2										0.006					0.002
18	0.002				0.059				0.120		0.181		0.048		0.002
18.2									0.002	0.002		0.002	0.002		0.019
19					0.009				0.170		0.124		0.033		0.069
20									0.120		0.013		0.031		0.056
21									0.030				0.024		0.100
22									0.057				0.024		0.219
22.2															0.004
23									0.196				0.004		0.191
23.2															0.019
24									0.157						0.159
24.2															0.006
25									0.059						0.096
25.2															0.007
26									0.006						0.033
26.2															0.002
27		0.002													0.015
28		0.048													
28.2		0.004													
29		0.287								0.002					0.002
30		0.244													
30.2		0.004													
31		0.010													
31.2		0.072													
32		0.041													
32.2		0.146													
33		0.009													
33.2		0.035													
34		0.002													
34.2		0.006													

2.2 统计学分析

通过对 270 份样本的检测,得到了 15 个 STR 基因座的等位基因频率分布资料(表 1)。其中 D8S1179 等位基因数为 9 个,分布范围为 10~18;D21S11 等位基因数为 14 个,分布范围为 27~34.2;D7S820 等位基因数为 9 个,分布范围为 6~14;CSF1PO 等位基因数为 9 个,分布范围为 6~14;D3S1358 等位基因数为 8 个,分布范围为 12~19;TH01 等位基因数为 6 个,分布范围为 6~10;D13S317 等位基因数为 8 个,分布范围为 7~14;D16S539 等位基因数为 7 个,分布范围为 7~14;D2S1338 等位基因数为 12 个,分布范围为 16~29;D19S433 等位基因数为 13 个,分布范围为 11~18.2;vWA 等位基因数为 8 个,分布范围为 13~20;TPOX 等位基因数为 6 个,分布范围为 8~13;D18S51 等位基因数为 15 个,分布范围为 10~23;D5S818 等位基因数为 9 个,分布范围为 7~15;FGA 等位基因数为 18 个,分布范围为 17~29。15 个 STR 基因座中,D21S11、CSF1PO、TH01、D19S433 与 FGA 属于复杂重复序列,其他 10 个 STR 基因座属于简单重复序列。经卡方检验分析比较基因型的观察值和期望值,结果 $P > 0.05$,表明各 STR 基因座的基因型分布

符合 Hardy-Weinberg 平衡。

根据各位点的等位基因频率和基因型频率分布,运用 PowerStatsV12 软件计算得出 15 个 STR 基因座的 H、PIC、DP、PE (表 2)。其中 H 介于 0.593~0.900,PIC 介于 0.54~0.85,DP 介于 0.780~0.963,PE 介于 0.282~0.785。

表 2 湖南汉族人群 15 个 STR 基因座多态性统计学指标
Table 2 Statistical indices of 15 STR loci of Han population from Hunan

Loci	Heterozygosity	PIC	DP	PE
D8S1179	0.878	0.830	0.953	0.750
D21S11	0.800	0.790	0.943	0.599
D7S820	0.733	0.730	0.915	0.482
CSF1PO	0.704	0.690	0.887	0.434
D3S1358	0.726	0.680	0.879	0.469
TH01	0.663	0.630	0.845	0.373
D13S317	0.800	0.760	0.922	0.599
D16S539	0.781	0.740	0.913	0.565
D2S1338	0.900	0.850	0.963	0.795
D19S433	0.848	0.790	0.940	0.691
vWA	0.826	0.780	0.930	0.648
TPOX	0.593	0.540	0.780	0.282
D18S51	0.893	0.850	0.960	0.780
D5S818	0.800	0.740	0.911	0.599
FGA	0.889	0.850	0.959	0.773

表 3 湖南汉族人群 STR 等位基因频率同其他人群之间的比较

Table 3 Comparison of allele frequency in Han population from Hunan and other population

Loci	Beijing	Sichuan	Han in Shanxi	Han in Northeast China	Ecuadorian Black	Amerindian	Japanese
D3S1358	0.693	0.249	0.761	0.728	0.893	0.003	0.814
vWA	0.233	0.499	0.504	0.287	0.000	0.000	0.013
FGA	0.113	0.672	0.095	0.167	0.000	0.000	0.002
D8S1179	0.037	0.089	0.041	0.000	0.000	0.000	
D21S11	0.066	0.203	0.360	0.375	0.000	0.000	
D18S51	0.769	0.670	0.007	0.308	0.000	0.000	
D5S818	0.382	0.910	0.167	0.430	0.000	0.000	
D13S317	0.198	0.560	0.015	0.054	0.000	0.000	
D7S820	0.412	0.167	0.450	0.002	0.000	0.000	
TH01							0.000
TPOX							0.046
CSF1PO							0.406

注:黑体加粗数据 $P < 0.05$,具有显著性差异。

Notes: Numbers in bold type indicate those loci that depict significant differences between Han population from Hunan and other population.

2.3 各人群之间 STR 等位基因频率间的比较

通过 χ^2 检验将湖南汉族人群同其他人群的等位基因频率分布进行比较(表 3)。国内人群之间的比较发现,湖南汉族人群的 9 个 STR 位点的等位基因频率同四川人相比均无显著性差异

($P > 0.05$),同北京人相比仅 D8S1179 位点有显著性差异 ($P < 0.05$),同东北汉族相比 D8S1179、D7S820 位点有显著性差异 ($P < 0.05$),同陕西汉族相比则 D8S1179、D18S51、D13S317 位点有显著性差异($P < 0.05$)。同国外人群相比,湖南汉族

人群的 9 个等位基因频率同美洲印第安人相比均有显著性差异($P < 0.05$),同厄瓜多尔黑人相比则除 D3S1358 位点无显著性差异($P > 0.05$)外,其余均有显著性差异,而同日本人群相比的 6 个 STR 位点等位基因频率的比较中,有 vWA 位点、FGA 位点、TH01 位点和 TPOX 位点有显著性差异($P < 0.05$)。

3 讨论

STR 具有种类多、分布广、高度多态、易于检测等特点,同时由于 STR 基因座在群体之间和群体内部的异质性水平较高,因此在不同群体以及亲缘关系相近的群体之间的等位基因和基因型频率分布有明显的地理差异,地理因素在遗传学多样性形成中的作用比语言因素要更显著^[11]。所以在将 STR 应用于人类进化研究和法医学个体识别、亲权鉴定时,必须先对群体进行多态性调查,以获得其频率分布的数据。

本研究采用复合扩增并荧光检测技术,调查了 D8S1179、D21S11、D7S820、CSF1PO、D3S1358、TH01、D13S317、D16S539、D2S1338、D19S433、vWA、TPOX、D18S51、D5S818 和 FGA 共 15 个 STR 基因座在湖南省 270 名汉族无关个体中的多态性分布情况,得到了反映该地区群体遗传特征的等位基因频率数据。对这些基础数据的统计分析表明,所得数据完全符合 Hardy-Weinberg 平衡定律($P > 0.05$),证明了此次调查资料在群体遗传学上的可靠性和科学性。

在 STR 研究中,遗传标记的多态性及应用价值一般可用杂合度(H)、多态信息含量(PIC)、个体识别力(DP)和非父排除率(PE)来衡量。杂合度能客观地反映样本的遗传变异水平。杂合度越大,表明群体内遗传差异越大,多态性则越高。PIC 直接反映某个遗传标记所包含或所能提供的遗传信息的总量。PIC > 0.5 时,该遗传标记具有高度的可提供信息性。DP 和 PE 则反映了遗传标记在法医学个体识别及亲权鉴定中的能力。一般认为 DP > 0.8, PE > 0.5 时,属于高度多态性遗传标记,具有较高的应用价值^[12]。

根据 15 个 STR 基因座多态性指标数据,可推断大多数基因座具有较高多态性,其中 D2S1338 各项多态性指标均为最高,具有较好的一致性。而 TPOX、TH01、CSF1PO、D3S1358、D7S820 等 5 个 STR 基因座 PE 均小于 0.5,其中

TPOX 的 DP 值小于 0.8,说明这 5 个 STR 基因座在湖南汉族群体中的多态性不是很高。因此应用于法医学实践时,应尽量避免使用这几个 STR 基因座,而应联合其他多态性较高的 STR 位点。计算 15 个 STR 基因座累计 PE 为 0.999 999 5,累计 DP 为 $(1 \sim 1.6 \times 10^{-17}) > 0.999 999 99$,说明这 15 个 STR 基因座联合应用适合于湖南汉族的多态性研究,在人类遗传学、法医学等领域有重要的应用价值。

同时我们选取国内和国外部份人群的 STR 等位基因频率与检测得到的湖南省汉族人群数据进行比较,发现了湖南汉族人群的 STR 位点群体学数据与四川人群的最为接近,统计学分析其中 9 个 STR 位点均无显著性差异($P > 0.05$),与其他国内人群的比较中也只发现少数位点有显著性差异。而与部分国外人群的比较中,发现检测的 STR 位点等位基因分布大部分具有显著性差异,尤以美洲印第安人最为明显,所有对比的 9 个 STR 位点均有显著性差异($P < 0.01$),厄瓜多尔黑人则有 8 个位点具有显著性差异($P < 0.01$),而与日本人群相比的 6 个 STR 位点中,则有 4 个具有显著性差异($P < 0.05$)。这也在一定程度上反应了 STR 数据具有一定的地域性和群体性,相比国外人群,湖南汉族人群与国内主要人群的 STR 数据更为接近,但是也存在少数位点具有显著性差异,所以在以后的研究中,我们应该继续扩大样本采集的数量和范围,为建立我国不同地区的 STR 数据库奠定基础。

参考文献(References):

- [1] 侯一平. 法医物证学[M]. 北京:人民卫生出版社(HOU Yi-ping. Forensic Genetics[M]. Beijing: People Health Press), 2004. 76-77.
- [2] SCHAFER A J, HAWKINS J R. DNA variation and the future of human genetics[J]. Nature Biotechnology, 1998, 16 (1): 33-39.
- [3] WALSH P S, METZGER D A, HIGUCHI R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material[J]. Biotechniques, 1991, 10(4): 506-513.
- [4] TEREBA A. Tools for analysis of population statistics[J]. Profiles DNA, 1999, 11(3): 14-16.
- [5] FUNG W K, YE J, HU L, et al. Allele frequency for nine STR loci in Beijing Chinese[J]. Forensic Science International, 2001, (121): 207-209.
- [6] ZHANG Hai-jun, LI Ying-bi, HOU Yi-ping, et al. Analysis of 15 STR loci in Chinese population from Sichuan in West China[J]. Forensic Science International, 2007, (171): 222-225.

(下转第 421 页)

3 讨论

本研究以基因枪轰击的方式将构建好的重组表达载体导入洋葱表皮细胞,发现 *MuMADS1-GFP* 融合蛋白在洋葱细胞内表达,且在细胞核中表达. 表明 *MuMADS1* 基因编码的蛋白作为一种转录因子应在细胞核内直接或间接参与一些功能基因的启动表达. 另一方面,也可说明携带 GFP 的植物表达载体构建成功且在植物细胞确实能够正确表达出带有绿色荧光的融合蛋白.

GFP 在真核或原核细胞内表达后,可在蓝光或紫光的激发下产生明亮的绿色荧光,其融合表达载体已广泛用于基因编码产物的定位或基因表达的报告基因^[12]. 与其他报告基因相比,GFP 没有本底表达,无需底物、辅因子等物质,只需应用荧光显微镜可方便地检测到目的基因的表达,筛选出阳性克隆,为日后转基因植物的检测提供方便. 虽然通过用 GFP 作为标记物来观察 *MuMADS1* 基因表达蛋白在细胞中的分布和定位,可为我们进一步研究该基因的功能提供线索和思路,然而,对于该基因更精确的亚细胞定位乃至对基因具体功能的阐述则需要将构建的植物表达载体转入植株内并且永久稳定表达后再进行各个方面的检测研究.

参考文献 (References):

- [1] 吴乃虎,刁丰秋. 植物转录因子与发育[J]. 科学通报 (WU Nai-hu, DIAO Feng-qiu. Transcription factors and development of plants[J]. Chinese Science Bulletin), 1998, 43(20): 2133-2139.
- [2] 胡丽芳,金志强,徐碧玉. MADS-box 基因对花的发育及开花早晚的影响[J]. 生命科学研究 (HU Li-fang, JIN Zhi-qiang, XU Bi-yu. The effect of MADS-box gene to the development of flower and flowering-time [J]. Life Science Research), 2004, 8(4): 7-12.
- [3] HUANG H, TUDOR M, WEISS C A, et al. The Arabidopsis MADS-box gene *AGL3* is widely expressed and encodes a sequence-specific DNA-binding protein[J]. Plant Mol Biol, 1995, 28(3): 549-567.
- [4] IMMINK R G, FERRARIO S, BUSSCHER L J, et al. Analysis of the petunia MADS-box transcription factor family [J]. Mol Genet Genomics, 2003, 268(5): 598-606.
- [5] MOORE S, VREBALOV J, PAYTON P, et al. Use of genomics tools to isolate key ripening genes analyse fruit maturation in tomato[J]. Exp Bot, 2002, 53(377): 2023-2030.
- [6] VREBALOV J, RUEZINSKY D, PADMANABHAN V, et al. MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (*rin*) locus[J]. Science 2002, 296(5566): 343-346.
- [7] INABA A, LIU X, YOKOTANI, et al. Differential feedback regulation of ethylene biosynthesis in pulp and peel tissues of banana fruit[J]. J Exp Bot, 2007, 58(5): 1047-1057.
- [8] XU B Y, SU W, LIU J H, et al. Differentially expressed cDNAs at the early stage of banana ripening identified by suppression subtractive hybridization and cDNA microarray[J]. Planta, 2007, 226: 529-539.
- [9] 韦带莲,胡丽芳,苏伟,等. 香蕉 *MuMADS1* 基因的生物信息学分析[J]. 生物技术通讯(WEI Dai-lian, HU Li-fang, SU Wei, et al. Bioinformatics analysis of *MuMADS1* from banana[J]. Biotechnology Bulletin), 2006(Suppl), Z1: 310-311.
- [10] LIU J H, XU B Y, HU L F, et al. Involvement of a banana MADS-box transcription factor gene in ethylene-induced fruit ripening[J]. Plant Cell Rep, 2009, 28(1): 103-111.
- [11] 萨姆布鲁克 J, 拉塞尔 D W. 黄培堂, 王嘉玺, 朱厚基, 等(译). 分子克隆实验指南(第三版)[M]. 北京: 科学出版社 (SAMBROOK J, RUSSELL D W. HUANG Pei-tang, WANG Jia-xi, ZHU Hou-ji, et al (Translate). Molecular Cloning: A Laboratory Manual, 3rd ed[M]. Beijing: Science Press), 2002. 26-27, 96-99.
- [12] WAHLFORS J, LOIMAS S, PASANEN T, et al. Green fluorescent protein (GFP) fusion constructs in gene therapy research[J]. Histochem Cell Biol, 2001, 115 (1): 59-65.

(上接第 402 页)

- [7] WANG Zhen-yuan, YUAN Rong-jun, WANG Fang, et al. Genetic polymorphisms of 15 STR loci in Han population from Shanxi (NW China)[J]. Forensic Science International, 2005, (147): 89-91.
- [8] WANG Wei, JIA Hui-ling, WANG Qin, et al. STR polymorphism of "forensic loci" in the northern Han Chinese population[J]. J Hum Genet, 2003, (48): 337-341.
- [9] FABRICIO G A, DORA S Q, BEGONA M J. Genetic analysis of the Amerindian Kichwas and Afroamerican descendents populations from Ecuador characterized by 15 STR-PCR polymorphisms[J]. Forensic Science International, 2006, (160): 231-235.
- [10] HARA M, KIDO A, SAITO K, et al. Population data for nine STR loci FGA, vWA, D3S1358, CSF1PO, TPOX, TH01, D7S820, D13S317 and D5S818 in Japanese [J]. International Congress Series, 2003, (1239): 231-234.
- [11] ROSSER Z H, ZERJAL T, HURLES M E. Y-chromosomal diversity in Europe is clinical and influenced primarily by geography, rather than by language[J]. Am J Hum Genet, 2000, 67(6): 1526-1543.
- [12] NEI M, TAKEZAKI N. The root of phylogenetic tree of human population[J]. Mol Bio Evol, 1996, 13 (1): 170-176.