

云南汉族人群 15 个常染色体 STR 基因座的遗传多态性与 OL 等位基因的研究

余建华¹⁾, 高静²⁾, 向超杰¹⁾, 李安¹⁾, 史斌¹⁾, 许冰莹²⁾

(1) 昆明市公安局刑事科学技术研究所, 云南昆明, 650011) 2) 昆明医科大学法医学院, 云南昆明 650500)

[摘要] **目的** 调查云南地区汉族人群 15 个常染色体 STR 基因座的遗传多态性, 获得法医学参数以及 off-ladder 等位基因频率. **方法** 用 PowerPlex[®] 16 system 试剂盒检测 20204 个云南地区汉族群体样本, 计算 15 个常染色体 STR 基因座的等位基因频率、法医学参数以及各 STR 基因座的 OL 等位基因的频率. **结果** 所检测的 15 个常染色体 STR 基因座等位基因频率范围为 0.000 025 至 0.522 743, 在云南地区汉族人群的累积匹配概率为 3.13×10^{-18} , 累积非父排除概率为 99.998 44%. 经 Bonferroni 校正, 除基因座 CSF1PO 外其余 14 个基因座均符合 Hardy Weinberg 平衡. 在 15 个 STR 基因座中共有 14 个 STR 基因座检测到 OL 基因, 其中有 5 个 STR 基因座至今没有文献报道. **结论** 基于对大样本的常染色体 STR 基因座的调查, 对建立法医基础 DNA 数据库以及发现常染色体 STR 基因座的 OL 等位基因有重要意义, 能够更好地应用于法医学实践.

[关键词] PowerPlex[®] 16 system; STR; 等位基因频率; OL 等位基因; 云南汉族

[中图分类号] R89 **[文献标识码]** A **[文章编号]** 2095 - 610X (2014) 07 - 0027 - 05

Genetic Variability and Off-ladder Allele of 15 Autosomal STR Loci in Yunnan Han Population of China

YU Jian-hua¹⁾, GAO Jing²⁾, XIANG Chao-jie¹⁾, LI An¹⁾, SHI Bin¹⁾, XU Bing-ying²⁾

(1) Institute of Criminal Science and Technology of Kunming Police Bureau, Kunming Yunnan 650011; 2) School of Forensic Medicine, Kunming Medical University, Kunming Yunnan 650500, China)

[Abstract] **Objective** To get genetic polymorphism of 15 STR loci in Yunnan Han population with PowerPlex[®] 16 System Kit, and obtain the allele frequencies, the frequency of off-ladder allele and forensic parameters. **Methods** PowerPlex[®] 16 system kit was used to detect 20204 samples of Yunnan Han population. The allele frequencies of 15 STR loci and OL alleles and other forensic parameters were calculated. **Results** The allele frequencies of 15 STR loci were 0.000 025 to 0.522743. The combined matching probability was 3.13×10^{-18} and the combined power of exclusion was 99.998 44%. After Bonferroni correction, departure from expectation of HWE was observed in only CSF1PO. The off-ladder alleles were discovered in 14 gene loci, of which 5 loci had not been reported in the literature. **Conclusions** Investigation of gene frequency in large samples is good to establish DNA database, and find OL alleles in STR loci. It has great significance for individual identification and paternity testing.

[Key words] PowerPlex[®] 16 system; STR; Allele frequencies; Off-ladder allele; Yunnan Han population

[基金项目] 云南省应用基础研究基金资助项目 (2010CD226, 2011CA022, 2011FZ113)

[作者简介] 余建华 (1979~), 男, 云南昆明市人, 医学学士, 主检法医师, 主要从事法医 DNA 分析. 高静和余建华对本文有同等贡献.

[通讯作者] 许冰莹. E-mail: bingying_xu@126.com

STR 即短串联重复序列, 是一类广泛存在于真核生物基因组中的 DNA 重复序列. 目前, PowerPlex[®] 16 system 试剂盒广泛应用于法医物证实验室, 用于进行个人识别和亲子鉴定. 本文采用 PowerPlex[®] 16 system 试剂盒复合扩增 20 204 名云南地区汉族人群无关个体血样的 15 个常染色体 STR 基因座, 用 ABI 3130 型全自动遗传分析仪对扩增产物进行检测, 获得云南地区汉族人群的群体遗传学数据, 建立云南汉族群体基础 DNA 数据库, 从而应用于法医研究及实践中.

1 材料与方 法

1.1 样本及 DNA 提取

20 204 例云南地区汉族无关个体样本采集于云南省内不同地区. 采用 Chelex-100 法^[1]提取样本 DNA.

1.2 PCR 扩增及检测

PCR 扩增按照 PowerPlex[®] 16 System Kit (Promega, 美国) 说明书进行. 25 μ L 扩增体系包括 2.5 μ L PCR 复合扩增引物, 2.5 μ L 缓冲液, 0.8 μ L Taq DNA 聚合酶 (5 U/ μ L), 和 2.5 μ L 模板 DNA. 采用 GeneAmp_PCR System 9700 (Applied

Biosystems, 美国) 进行 PCR 扩增.

PCR 产物采用 ABI 3130 遗传分析仪 (Applied Biosystems, 美国) 进行电泳, GeneMapper ID v. 3.2 (Applied Biosystems, 美国) 读取样本分型结果.

1.3 统计学处理

15 个常染色体 STR 基因座的基因频率及相关统计学参数由李慧敏博士采用软件 Matlab (MathWorks, 美国) 依据法医物证学统编教材^[2]公式进行计算. 收集已发表的基因频率数据^[3-13], 采用 dispan 软件计算遗传距离, 采用 MEGA v4 软件构建系统进化树^[14].

2 结果

15 个 STR 基因座的等位基因频率、杂合度期望值 (He)、杂合度观察值 (Ho)、个人识别率 (DP)、非父排除率 (PE)、Hardy-Weinberg 平衡 (P) 及新发现的 OL 基因见表 1, 累积匹配概率为 3.13×10^{-18} , 累积非父排除概率为 99.99844%.

与已报道的国内云南汉族、云南布朗族、云南拉祜族、闽南汉族、河南汉族、台湾汉族、云南白族和青海土族群体数据进行比较获得系统进化树. 结果显示, 对于同样地处中国的云南汉族、云南布

表 1 15 个 STR 基因座的基因频率及相关统计学参数 (1)

Tab. 1 Allelic frequency and other parameters of the 15 STR loci (1)

	CSF1PO	D13S317	D16S539	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	Penta D	Penta E	TH01	TPOX	
3.2											0.000 025				
4		0.000 198 Δ											0.000 099		
5			0.000 396								0.000 125	0.047 837	0.000 198		
6	0.000 049	0.000 594	0.000 099					0.000 124			0.002 671		0.096 07		
6.3													0.000 025		
7	0.005 147	0.001 213	0.000 074	0.000 025			0.023 461	0.001 93	0.000 049		0.016 603	0.001 782	0.272 892	0.000 52	
7.3													0.000 025		
8	0.001 609	0.293 407	0.007 919	0.000 099			0.002 302	0.146 654	0.001 485		0.046 687	0.006 707	0.057 736	0.522 743	
9	0.042 912	0.135 666	0.264 057	0.000 272			0.066 62	0.065 433	0.000 495		0.339 617	0.007 622	0.496 461	0.123 515	
9.3													0.041 601		
10	0.235 547	0.1370 03	0.120 05	0.000 916			0.198 525	0.158 162	0.124 926		0.126 08	0.041 502	0.034 399	0.023 213	0.000 025
11	0.245 818	0.239 185	0.267 2	0.003 861		0.000074	0.320 432	0.357 256	0.094189		0.147 126	0.171 748	0.000 445	0.302 712	0.000 049
12	0.382 226	0.148 213	0.218 769	0.035 686		0.001015	0.238 987	0.228 767	0.123 119		0.155 715	0.111 76	0.000 025	0.025 589	0.000 124
13	0.074 738	0.037 493	0.105 944	0.193 774		0.001361	0.139 255	0.036 899	0.209 538		0.117 067	0.054 544		0.001 262	0.001 831
13.2				0.000 025											
13.3													0.000 025		
14	0.010 691	0.006 88	0.014 576	0.206 048		0.040487	0.009082	0.004727	0.178 232		0.037 949	0.088 077		0.000 396	0.247 773
15	0.001 163	0.000 148	0.000 916	0.180 806		0.33632	0.001287	0.000049	0.175 089	0.000 025	0.008 738	0.092 828		0.000 049	0.029 771
15.2															0.000 025 Δ
16	0.000 074			0.130 395		0.33528	0.000049		0.076 569	0.000 619	0.001 323	0.080 603			0.175 485
17	0.000 025			0.07496		0.214537			0.013 537	0.001 336	0.000 175	0.068 724			0.238 195
17.2										0.000 074					
18				0.045 931		0.064542			0.002 598	0.032 568		0.072 436			0.193 675

表 1 15 个 STR 基因座的基因频率及相关统计学参数 (2)
Tab. 1 Allelic frequency and other parameters of the 15 STR loci (2)

	CSF1PO	D18S317	D16S539	D18S51	D21S11	D3S1358	DS8818	D7S820	D8S1179	FGA	Penta D	Penta E	TH01	TPOX	vWA
18.2										0.000 099					
19				0.043 704		0.00584			0.000198△	0.053 603	0.000 075	0.054 222			0.096 12
19.2										0.000 049					
20				0.030 885		0.00047				0.059 741	0.000 025△	0.046 699			0.015 096
20.2										0.000 421					
21				0.021 778		0.000074			0.110 721			0.025 416			0.001 757
21.2										0.004 504					
22				0.017 868						0.170 214		0.014 205			0.000 074
22.2										0.006 434					
23				0.006 954						0.208 746		0.007 424			
23.2										0.009 874					
24				0.004 306						0.169 422		0.004 133			
24.2										0.010 889					
25				0.001 138						0.097 802		0.001 732△			
25.2										0.004 801					
26				0.000346	0.000 099					0.042 12					
26.2										0.001 906					
27				0.000198	0.002 054					0.010 344					
27.2					0.000 223					0.000 247					
28				0.000025	0.044 323					0.002 722					
28.2					0.007 894					0.000 074					
29					0.259 28					0.000 544					
29.2					0.001 683										
30					0.272 471					0.000 074					
30.2					0.018 437					0.000 025					
31					0.100 797										
31.2					0.073 03										
32					0.027 445										
32.2					0.136 013										
33					0.003 514										
33.2					0.046 723										
34					0.000 445										
34.2					0.005 098										
35					0.000 049										
35.2					0.000 396										
36					0.000 025										
H _c	0.733 287	0.799 228	0.788 185	0.860 921	0.822 404	0.725 442	0.779 371	0.770 853	0.851 305	0.870 626	0.999 997	0.915 612	0.666 182	0.621 08	0.806 357
H ₀	0.718 076	0.791 922	0.783 558	0.853 247	0.818 204	0.722 926	0.770 293	0.769 452	0.843 15	0.863 245	0.802 017	0.899 822	0.665 759	0.614 779	0.804 94
PD	0.883 633	0.929 003	0.920 251	0.963 773	0.944 989	0.873 725	0.916 1	0.911 761	0.958 265	0.968 506	0.939 045	0.985 897	0.839 284	0.793 749	0.931 921
PE	0.467 598	0.597 863	0.576 357	0.576 257	0.558 615	0.458 878	0.552 89	0.543 321	0.649 435	0.554 671	0.607 263	0.232 488	0.341 91	0.354 081	0.584 153
P	0.002 414	0.102 209	0.315 316	0.004 104	0.354 981	0.746 819	0.015 422	0.117 11	0.170 754	0.396 44	0.355 137	0.011 876	0.277 857	0.280 223	0.222 794

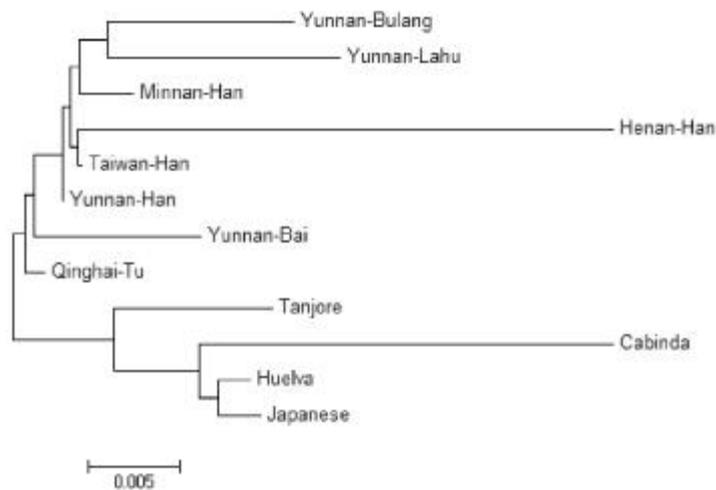


图 1 12 个民族 9 个 STR 基因座的 NJ 遗传距离系统进化树
Fig. 1 NJ-phylogenetic tree of 9 STRs in 12 populations

朗族、云南拉祜族、闽南汉族、河南汉族、台湾汉族、云南白族和青海土族同处于一个分支，西班牙的 HUELVA、安哥拉的 Cabinda、日本、印度的 Tanjore 属另一个分支，见图 1。

3 讨论

结果显示，在调查的 15 个常染色体 STR 基因座中，除 4 个 STR 基因座（CSF1PO, D18S51, D5S818 和 Penta E）外，其余 11 个 STR 基因座均达到 Hardy-Weinberg 平衡。经 Bonferroni 校正后，除基因座 CSF1PO 外，其余 STR 基因座均达到 Hardy-Weinberg 平衡。究其原因，可能与特殊地理位置、人口特征和频繁的基因交流有关，这种情况在相关文献里有类似地报道^[15-17]。云南地处中国西南边界，除汉族外，有 25 个少数民族，是我国拥有最多少数民族的省份，考虑云南汉族人群可能与少数民族人群有着频繁的基因交流；另外不同的抽样方法和较小的样本量均可能导致抽样误差^[18]从而影响人口数据统计结果，需要做进一步的调查以明确误差原因。

系统进化树显示（图 1），对于同样地处中国的云南汉族、云南布朗族、云南拉祜族、闽南汉族、河南汉族、台湾汉族、云南白族和青海土族同处于一个种族，西班牙的 Huelva、安哥拉的 Cabinda、日本、印度的 Tanjore 属于另一个分支。四个不同地域的汉族在系统进化树表现为非常接近，提示不同地域之间的汉族的关系比不同民族之间的关系更近。云南汉族与台湾汉族遗传关系最近，然而，与河南汉族遗传关系最远。与其他国家相比，云南汉族与 Huelva 的遗传关系最近，而与 Cabinda 的遗传关系最远。

在笔者的研究中，云南汉族有 14 个基因座检测到 OL 基因 (off-ladder)，根据 PowerPlex[®] 16 System Kit 说明书和常染色体 STR 基因座 DNA 数据库^[19]，我们检测到的部分 OL 基因在相关文献中已经报道^[15, 19-22]，但在表 1 中用 "△" 标记的 5 个基因座则至今没有文献报道过。每个 OL 等位基因片段都经 DNA 测序进行确认。

所谓 Off-ladder 是指在 STR 分型图谱中，常会遇到部分样本的少量片段峰未被程序化数字命名，在分型图谱中被标识为 off-ladder。出现 off-ladder 峰有两种情况：一种情况是 off-ladder 大部分是由于电泳过程中的漂移作用使它们被软件测量的片段大小稍稍偏离相应的等位基因，不能被软件自

动命名。对这种 off-ladder 峰通过分析前后已自动化正确命名的片段相对与 ladder 等位基因的漂移偏离情况，再对 off-ladder 峰进行漂移校正，可以人工修改为数字化命名。另一种情况是当分型标准品的基因座等位基因未包含检测样本的等位基因时，相当于样本中出现了一个新的等位基因，它能够被软件识别片段大小和荧光标记种类，但在与 ladder 对比时，找不到对应的 ladder 等位基因，这有两种情况：一种情况是出现在某个基因座 ladder 最大和最小等位基因范围内，是由于等位基因片段出现个别碱基的缺失或插入引起的基因微变异所致。另一种情况是超过这个范围而位于两个邻近基因座 ladder 之间，是由于 STR 基因座滑动突变形成的稀有等位基因。本文统计的 off-ladder 等位基因就是属于这种情况。

现如今，常染色体 STR 基因座已经广泛应用于法医学实践中，有必要收集和调查云南各民族人群的基因型和等位基因频率，法医学应用参数（H、Dp、PE、PIC）等信息，建立云南各民族人群常染色体 STR 基因座基础 DNA 数据库，用于进行父权指数、匹配概率、似然率等生物统计学计算，从而评估遗传标记提供的证据强度和系统效能，以保证法庭科学 DNA 数据库的应用和运行的科学性。

目前，笔者收集和调查了云南地区各个民族的样本，建立了法庭科学基础 DNA 数据库并且应用于法医学实践。巨大的样本量的调查可以更好地发现 STR 基因座的一些稀有等位基因，而正确辨别 OL 等位基因，在法医学的个人识别和亲子鉴定实践中，具有重要的应用价值和实际意义。

[参考文献]

- [1] WALSH P S, PETZGER D A, HIGUCHI R. Chelex-100 as medium for simple extraction of DNA for PCR-based typing from forensic material[J]. *Biotechniques*, 1991, 10 (4):506 - 513.
- [2] 侯一平. 法医物证学[M]. 第3版. 北京:人民卫生出版社, 2010:140 - 141.
- [3] SBELEZA, CALVES, FREIS, et al. 17 STR data AmpFI-STR Identifier and Powerplex 16 System from Cabinda (Angola) [J]. *Forensic Science International*, 2004, 141 (2-3):193 - 196.
- [4] SHUQIN HUANG, TUNLIANG Z H U, XINTANG S H E N, et al. Genetic variation analysis of 15 autosomal STR loci of AmpFISTR Sinofiler PCR Amplification Kit in Henan (central China) Han population [J]. *Legal*

- Medicine, 2010, 12(3):160 – 161.
- [5] SITALAXIMI T, MTECH, RTRIVEDI, et al. Genotype Profile for thirteen tetranucleotide repeat loci and two pentanucleotide repeat loci in four endogamous tamil population groups of India [J]. *J Forensic Sci*, 2002, 47(5):1 – 6.
- [6] CLOTILDE COUDRAY, ROSARIO CALDERON, EVELYNE GUITARD, et al. Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain) [J]. *Forensic Science International*, 2007, 168(2-3):e21 – e24.
- [7] MHASHIYADA, YITAKURA, TNAGASHIMA, et al. Polymorphism of 17 STRs by multiplex analysis in Japanese population [J]. *Forensic Science International*, 2003, 133(8):250 – 253.
- [8] BOFENG ZHU, JIANGWEI YAN, CHUNMEI SHEN, et al. Population genetic analysis of 15 STR loci of Chinese Tu ethnic minority group [J]. *Forensic Science International*, 2008, 174(2-3):255 – 258.
- [9] CHIH-WEI WANG, DING-PING CHEN, CHI-YUAN CHEN, et al. STR data for the AmpFISTR SGM Plus and Profiler loci from Taiwan [J]. *Forensic Science International*, 2003, 138(1-3):119 – 122.
- [10] 吴德清. 闽南地区汉族人群常染色体15个STR基因座遗传多态性的研究[D]. 汕头:汕头大学, 2006:16 – 17.
- [11] 赖江华, 张保华, 朱波峰, 等. 云南白族STR遗传多态性研究[J]. *西安医科大学学报*, 2002, 23(3):242 – 245.
- [12] 刘建兴, 秦洁, 景强, 等. 云南布朗族群体15个STR基因座遗传多态性研究[J]. *昆明医学院学报*, 2007, 28(2):1 – 5.
- [13] 赖跃, 王传海, 秦洁, 等. 云南拉祜族群体15个STR基因座遗传多态性研究[J]. *昆明医学院学报*, 2007, 28(3):5 – 9.
- [14] TAMURA K, DUDLEY J, NEI M, et al. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0 [J]. *Mol Biol Evol*, 2007, 24(8):1 596 – 1 599.
- [15] VADIM A, STEPANOV, ALEXANDER V, et al. Genetic variability of 15 autosomal STR loci in Russian populations [J]. *Legal Medicine*, 2010, 12(5):256 – 258.
- [16] RAK S A, ZALAN A, SZABADOS G. Population genetic data on 15 STR loci in the Hungarian population [J]. *Forensic Sci Int Genet*, 2011, 5(5):543 – 547.
- [17] CHANG HO SHIN, PHILIP JANG, KYEONG MAN HONG, et al. Allele frequencies of 10 STR loci in Koreans [J]. *Forensic Sci Int*, 2004, 140(1):133 – 135.
- [18] 费传宝. 统计误差分析 [J]. *读与写杂志*, 2009, 6(10):87 – 88.
- [19] 刘秋玲, 梁艳芳, 吕德坚, 等. Powerplex? 16TM System 在中国人群中OL等位基因的研究 [J]. *中山大学学报 (医学科学版)*, 2009, 30(3S):29 – 32.
- [20] 陆惠玲, 台运春, 刘超, 等. Powerplex TM 16 体系OL等位基因序列分析及命名探讨 [J]. *法医学杂志*, 2006, 22(3):186 – 189.
- [21] MIZUNO N, SEKIGUCHI K, SATO H, et al. Variant alleles on the Penta E locus in the PowerPlex 16 kit [J]. *J Forensic Sci*, 2003, 48(2):358 – 361.
- [22] 张红玲, 王杰, 黄映康, 等. 贵州地区汉族Penta D和Penta E 基因座遗传多态性 [J]. *中国法医学杂志*, 2009, 24(6):410.

(2014 – 02 – 25 收稿)