

西藏那曲地区藏族人群 15 个短串联 重复序列位点多态性的研究

李 宁¹, 苏玉虹², 席焕久³, 任 甫³, 朱宝芹², 温有峰³

(1 辽宁医学院生物化学教研室, 锦州 121001;

2 辽宁医学院辽宁省分子细胞生物学与新药开发重点实验室, 锦州 121001;

3 辽宁医学院人类学研究所, 锦州 121001)

摘要: 利用基因扫描技术调查西藏自治区那曲地区藏族人群 D8S1179、D21S11、D7S820、CSF1PO、D3S1358、TH01、D13S317、D16S539、D2S1338、D19S433、VWA、TPOX、D18S51、D5S818 及 FGA 共 15 个短串联重复序列 (STR) 基因座多态性分布, 获得 15 个基因座的群体遗传学数据。结果显示: 15 个 STR 位点在那曲地区藏族人群中具有遗传多态性, 基因型分布符合 Hardy Weinberg 平衡, DP 在 0.758 8—0.960 4 之间, H 在 0.476 2—0.862 0 之间, PIC 在 0.446 4—0.861 5 之间, EPP 在 0.385 0—0.856 0 之间, 累积个体鉴别力为 0.999 999 999, 累积非父排除率为 0.999 999 998。15 个 STR 位点适合作为那曲地区藏族人群的遗传标记用于人类学、疾病连锁分析、法医学亲子鉴定和个体识别等领域的研究。

关键词: 短串联重复序列; 多态性; 藏族; 聚合酶链式反应

中图分类号: Q347 文献标识码: A 文章编号: 1000-3193 (2007) 01-0070-07

短串联重复序列 (short tandem repeat, STR), 通常由长度为 2—6bp 的核心序列串联重复而成, 其广泛存在于人类基因组中^[1], 具有高度的多态性, 并遵循孟德尔遗传定律。STR 由于具有高度多态性、高杂合度、高信息量、检测简便、快捷等特点, 很快被用于遗传连锁图谱的构建、群体遗传学、人类学、民族学、目的基因的筛选、疾病相关基因的定位、法医学个体识别及亲权鉴定^[2,3], 是目前应用最广泛, 研究较深入的遗传标记之一。本文利用多重 PCR 和五色荧光自动化检测技术, 调查西藏那曲地区藏族人群 D8S1179、D21S11、D7S820、CSF1PO、D3S1358、TH01、D13S317、D16S539、D2S1338、D19S433、VWA、TPOX、D18S51、D5S818 及 FGA 共 15 个 STR 位点的多态性分布。从群体水平上阐明那曲地区藏族 STR 遗传结构及其变化规律, 丰富我国民族 STR 遗传资料数据库。

1 材料和方法

1.1 样本

本实验全部血液样本采自西藏那曲地区藏族人群, 共 251 人, 其中男 131 人, 女 120 人。

收稿日期: 2005-06-09; 定稿日期: 2006-09-15

基金项目: 国家自然科学基金 (30270696); 辽宁省教育厅基金 (202172050 和 2004C038)

作者简介: 李宁, 女 (1978), 硕士, 研究方向: 人类基因短串联重复序列多态性的研究。

通讯作者: 席焕久, E-mail: huanjiuxi@sina.com

取上臂静脉血 5ml, 注入装有 EDTA 抗凝剂的试管中, 混匀后置于 4℃ 冰箱内保存。

1.2 基因组 DNA 的提取

采用常规苯酚-氯仿法^[4]。

1.3 PCR 复合扩增及扩增产物的检测

15 个 STR 位点通过 AmpFISTR Identifiler^[5] 试剂盒进行多重 PCR 扩增。PCR 扩增体系为 25μl, PCR 扩增在自动热循环仪上进行, 循环参数为: 第一步: 95℃ 预变性 11min; 第二步: 94℃ 变性 1min, 59℃ 复性 1min, 72℃ 延伸 1min, 共 28 个循环; 第三步: 60℃ 最后延伸 60min。

扩增产物由 ABF3100Avant 型遗传分析仪进行毛细管电泳, 电泳信息由 DataCollection 数据收集软件自动收集, 电泳结束后, 由 GeneMapper 软件^[6] 分析各基因座的基因型。

1.4 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 χ^2 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。按文献报道的方法^[7-10] 计算 15 个 STR 位点的遗传学指标, 包括: 杂合度(H)、多态信息含量(PIC)、个体识别力(DP)、非父排除率(EPP)。运用 χ^2 检验将那曲藏族与美国黑人^[11]、高加索人^[11] 和中国上海汉族^[12] 的 9 个 STR 位点的等位基因频率进行比较。

2 结 果

2.1 15 个 STR 位点的 PCR 扩增和电泳结果

各位点均得到了有效扩增, 电泳图谱显示扩增结果较好, 特异性较高。附图为 1 例样品的电泳峰图。

2.2 等位基因频率、基因型频率分布

通过对 251 例样本的检测, 得到了 15 个 STR 位点的等位基因频率分布资料, 可以看出 15 个 STR 位点的等位基因在西藏那曲地区藏族的分布较均匀, 但也有一些相对的高频、中频和低频等位基因, 构成了 STR 在西藏藏族的的结构特点。经 χ^2 检验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。见表 1。

2.3 统计学分析

根据各位点的等位基因频率和基因型频率分布, 按文献方法计算各位点的 H、DP、PIC、EPP, 15 个 STR 位点 H 在 0.4762—0.8620 之间, PIC 在 0.4464—0.8615 之间, DP 在 0.7588—0.9604 之间, EPP 在 0.3850—0.8560 之间, 累积个体鉴别力为 0.999999999, 累积非父排除率为 0.99999998。表明这些位点具有较高的多态性, 遗传学指标达到了一定标准, 可以用于法医学的个体识别、亲权鉴定和群体遗传学等研究, 结果见表 2。

2.4 各民族间的比较

通过 χ^2 检验将那曲地区藏族与美国黑人、高加索人及中国上海地区汉族的等位基因频率分布进行比较。那曲地区藏族的 9 个 STR 位点等位基因频率(见表 3) 与美国黑人相比均具有显著差异($P < 0.05$), 与高加索人除 CSF1PO 位点外均有显著差异($P < 0.05$), 与上海

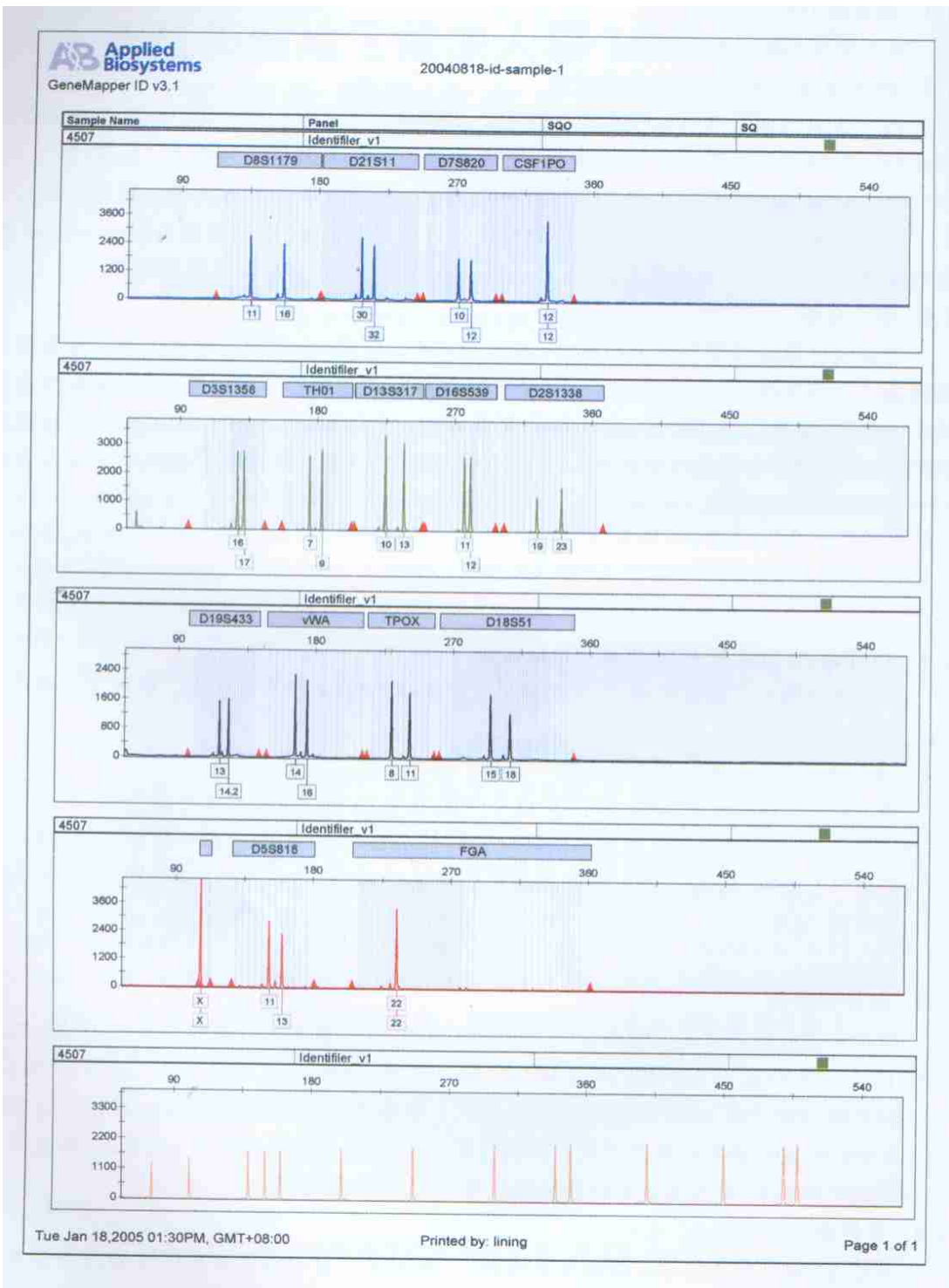


图 1 1 例样品的电泳峰图

The electrophoresis chart of one sample

表 1 那曲地区藏族人群 15 个短串联重复序列(STR)位点的等位基因频率分布
Allele frequency distribution of 15 STR in Naqu Tibetan

D8S1179	A	10	11	12	13	14	15	16	17						
	F	0.078	0.014	0.12	0.259	0.219	0.205	0.083	0.021						
D21S11	A	27	28	28.2	29	29.2	30	30.2	31	31.2	32	32.2	33	33.2	34.2
	F	0.004	0.029	0.04	0.227	0.013	0.187	0.051	0.106	0.062	0.031	0.143	0.004	0.099	0.004
D7S820	A	7	8	9	10	11	12	13							
	F	0.004	0.185	0.113	0.149	0.27	0.249	0.03							
CSF1PO	A	7	8	9	10	11	12	13	14	15					
	F	0.01	0.002	0.029	0.226	0.247	0.386	0.083	0.015	0.002					
D3S1358	A	13	14	15	16	17	18	19							
	F	0.004	0.013	0.274	0.443	0.195	0.069	0.002							
TH01	A	6	7	8	9	9.3	10								
	F	0.081	0.274	0.056	0.511	0.066	0.013								
D13S317	A	8	9	10	11	12	13	14							
	F	0.211	0.097	0.133	0.26	0.184	0.095	0.021							
D16S539	A	8	9	10	11	12	13	14							
	F	0.023	0.208	0.158	0.294	0.223	0.092	0.002							
D2S1338	A	17	18	19	20	21	22	23	24	25	26				
	F	0.042	0.084	0.169	0.139	0.046	0.038	0.266	0.139	0.063	0.013				
D19S433	A	10	12	13	13.2	14	14.2	15	15.2	16	16.2	17			
	F	0.017	0.047	0.231	0.024	0.283	0.134	0.104	0.08	0.019	0.035	0.026			
VWA	A	14	15	16	17	18	19	20							
	F	0.152	0.027	0.275	0.241	0.17	0.13	0.007							
TPOX	A	8	9	10	11	12	13								
	F	0.595	0.133	0.006	0.236	0.028	0.002								
D18S51	A	13	14	15	16	17	18	19	20	21	22				
	F	0.359	0.177	0.122	0.136	0.057	0.031	0.06	0.028	0.012	0.007				
D5S818	A	7	9	10	11	12	13	14	15						
	F	0.009	0.062	0.184	0.394	0.235	0.105	0.004	0.004						
FGA	A	17	18	19	20	21	22	23	24	25	26	27	28	29	
	F	0.009	0.043	0.084	0.032	0.059	0.127	0.247	0.21	0.101	0.06	0.019	0.006	0.006	

注:“A”代表等位基因(Allele);“F”代表等位基因频率(Frequency)

地区汉族除 D7S820 外, 均有显著差异 ($P < 0.05$)。表明不同民族间具有较大的遗传差异, 那曲地区藏族 STR 具有自身的结构分布特点。结果见表 3。

3 讨 论

中华民族含有多个民族, 各个民族由于地理、宗教、风俗等的不同, 使其彼此隔离, 并聚集在相对稳定的区域里, 可能会积累下不同的祖系基因及群体间遗传差异。藏族地处高原地区, 由于不同的地理环境, 高原特殊的气候及特殊的生活习惯, 使其与其他民族形成了较大的差异, 对藏族群体 STR 遗传结构的调

表 2 那曲地区藏族人群 15 个 STR 的 H、PIC、DP、EPP
H、PIC、DP、EPP of 15 STR in Naqu Tibetan population

loci	H	PIC	DP	EPP
D8S1179	0.8145	0.8040	0.9397	0.7846
D21S11	0.8620	0.8615	0.9604	0.8560
D7S820	0.7949	0.7811	0.9251	0.7375
CSF1PO	0.7309	0.7044	0.8879	0.6392
D3S1358	0.6861	0.6414	0.8471	0.5431
TH01	0.6502	0.6448	0.8241	0.5126
D13S317	0.8179	0.8092	0.9412	0.7889
D16S539	0.7867	0.7708	0.9182	0.7178
D2S1338	0.8454	0.8407	0.9540	0.8330
D19S433	0.8259	0.8203	0.9473	0.8142
VWA	0.7972	0.7840	0.9246	0.7894
TPOX	0.5721	0.5595	0.7588	0.3850
D18S51	0.7969	0.7872	0.9291	0.7592
D5S818	0.4762	0.4464	0.8842	0.6342
FGA	0.8515	0.8431	0.9563	0.8403

查及与其他民族的比较必将对藏族群体各方面具有较大的意义。关于藏族群体 STR 的研究报道很少, 位点数也很有限。胡羽^[13] 等对 100 名康巴地区藏族群体的 5 个 STR(D1S549、D3S1375、D12S391、D12S375、D18S865) 基因座进行了遗传多态性调查, 并与成都地区汉族比较, 发现等位基因个数和基因型分布都具有显著差异; 王占海^[14] 等对 136 名青海地区藏族群体的 9 个 STR(D3S1358、VWA、FGA、D8S1179、D21S11、D18S51、D5S818、D13S317、D7S820) 基因座进行了遗传多态性调查, 并与天津地区汉族进行了比较, 结果存在显著性差异。这些报道说明藏族作为一个古老的民族有其自身的遗传学特异性。同时也提示我们在进行同一认定及亲权鉴定时, 应采用本地区的基因频率分布资料。

表 3 那曲地区藏族人群 9 个 STR 等位基因频率与其他人群的比较
Comparison of allele frequency in Naqu Tibetan and other population

位点	那曲藏族 VS 美国黑人	那曲藏族 VS 高加索人	位点	那曲藏族 VS 上海汉族
CSF1PO	$\chi^2 = 85.61, P = 0.000$	$\chi^2 = 8.79, P = 0.456$	D3S1358	$\chi^2 = 19.20, P = 0.004$
D3S1358	$\chi^2 = 54.76, P = 0.000$	$\chi^2 = 80.75, P = 0.000$	FGA	$\chi^2 = 39.91, P = 0.001$
D5S818	$\chi^2 = 110.81, P = 0.000$	$\chi^2 = 61.25, P = 0.000$	D21S11	$\chi^2 = 43.60, P = 0.000$
D7S820	$\chi^2 = 67.23, P = 0.000$	$\chi^2 = 31.79, P = 0.000$	D6S818	$\chi^2 = 16.05, P = 0.025$
D13S317	$\chi^2 = 160.12, P = 0.000$	$\chi^2 = 38.95, P = 0.000$	D7S820	$\chi^2 = 10.64, P = 0.100$
TH01	$\chi^2 = 158.179, P = 0.000$	$\chi^2 = 236.35, P = 0.000$	VWA	$\chi^2 = 16.23, P = 0.013$
TPOX	$\chi^2 = 131.11, P = 0.000$	$\chi^2 = 16.47, P = 0.006$	D8S1179	$\chi^2 = 28.94, P = 0.000$
VWA	$\chi^2 = 92.66, P = 0.000$	$\chi^2 = 36.42, P = 0.000$	D18S51	$\chi^2 = 45.16, P = 0.000$
FGA	$\chi^2 = 55.44, P = 0.000$	$\chi^2 = 152.97, P = 0.000$	D13S317	$\chi^2 = 21.53, P = 0.001$

注: 表中列出了各个位点的 χ^2 值和 P 值。

本研究选取了 251 个西藏藏族无关个体, 进行了 15 个 STR 位点等位基因频率分布调查, 可以说达到了大样本量与多位点, 获得了这些位点在那曲地区藏族群体中的遗传学数据。那曲藏族 9 个 STR 等位基因频率与已报道的美国黑人和高加索人进行比较, 发现那曲

藏族与该两个民族之间均具有显著差异($P < 0.05$); 那曲藏族 9 个 STR 位点与上海汉族相比, 8 个位点有显著差异($P < 0.05$)。那曲藏族由于特殊的地理条件, 长期封闭的生活环境, 与外界交流很少等影响因素从而形成了自身相对独特的遗传特异性。这对我们从基因水平上研究中华民族的起源将有一定的意义; 也进一步说明, 建立不同民族 STR 群体数据库在实际应用中的必要性和重要性。

本研究不仅为进一步研究中华民族 STR 遗传结构, 建立我国不同民族 STR 数据库奠定了基础, 同时对于保护中华民族遗传资源具有科学意义和历史意义, 而且将为遗传制图、基因分离、疾病连锁分析、法医学个体识别和亲权鉴定等领域的理论与应用研究提供基础数据。

参考文献:

- [1] Collette D, Sabine F, Celle F, *et al.* A comprehensive genetic map of the human genome based on 5264 microsatellite[J]. *Nature*, 1996, 380(14): 152- 154.
- [2] Amar A, Brautbar C, Motro U, *et al.* Genetic variation of three tetrameric tandem repeats in four district Israel ethnic groups[J]. *Journal of Forensic science*, 1999, 44(5): 983- 986.
- [3] Schafer AJ, Hawkins JR. DNA variation and the future of human genetics[J]. *Nature Biotechnology*, 1998, 16(1): 33- 39.
- [4] J. 萨姆布鲁克, D. W. 拉塞尔著, 黄培堂等译. 分子克隆实验指南(第三版) [M]. 科学出版社, 2002: 463- 469.
- [5] AmpFISTR Identifier PCR Amplification Kit user's manual, Applied Biosystem, 2003.
- [6] GeneMapper ID v3. 1 user's manual, Applied Biosystem, 2003.
- [7] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individual[J]. *Genetics*, 1978, 89: 583- 584.
- [8] Boston P. Construction of a genetic linkage map in man using restriction length polymorphism[J]. *Am J Hum Genetic*, 1980, 32: 314- 315.
- [9] Fish RA. Standard calculations for evaluating a blood group system[J]. *Heredity*, 1951, 5: 95- 102.
- [10] Odelberg SJ. Repetitive DNA: Molecular structure, polymorphism and forensic application[A]. In *DNA and Other Polymorphism in Forensic Science* [M]. 1st ed. Chicago: Year Book Medical Publishers Inc, 1990: 26- 27.
- [11] Werrett DJ. The national DNA database[J]. *J Forensic Sci Int*, 1997, 88(4): 40- 42.
- [12] 冯明亮, 季芸, 陆琼, 马俊, 稽月华, 杨颖. 用多重 PCR 检测上海地区汉族人群 9 个 STR 基因座多态性[J]. *遗传*, 2002, 24(4): 403- 406.
- [13] 胡羽, 廖森, 周斌, 等. 中国康巴地区藏族群体 5 个 STR 基因座的遗传多态性研究[J]. *四川大学学报*, 2004, 35(1): 21- 24.
- [14] 王占海, 郭燕霞, 刘开会, 等. 青海地区藏族人群 9 个 STR 基因座的遗传多态性[J]. *中国法医学杂志*, 2004, 19(2): 100- 101.

Genetic Polymorphisms of 15 STR Loci in Naqu Tibetan Population

LI Ning¹, SU Yu-hong², XI Huan-jiu³, REN Fu³, ZHU Bao-qin², WEN You-feng³

(1. Teaching and Research Section of Biochemistry, Liaoning Medical College, Jizhou 121000;

2. Key Laboratory of Molecular Cell Biology and New Drug Development of Liaoning Province,

Liaoning Medical College, Jizhou 121000; 3. Anthropology Laboratory, Liaoning Medical College, Jizhou 121000)

Abstract: Using multiplex amplification and five fluorescent techniques (6FAM, VIC, NED, PET and LIZ), 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 the Naqu Tibetan population. Gene frequency, power of discrimination (DP), heterozygosity (H), polymorphism information content (PIC) and probability of paternity exclusion (EPP) were determined. In addition, Hardy-Weinberg equilibrium of the frequencies were also tested for all loci. Our results indicate the following:

1) Allele frequencies in the 15 STR loci meet the Hardy-Weinberg equilibrium.

2) The power of discrimination of the 15 STR loci is 0.7588—0.9604, heterozygosity of the 15 STR loci is 0.4762—0.8620, the polymorphism information content of the 15 STR loci is 0.4464—0.8615, and the probability of paternity exclusion of the 15 STR loci is 0.3850—0.8560.

3) The cumulative power of discrimination of the 15 STR is 0.999999999 and the cumulative probability of paternity exclusion is 0.999999998. On the basis of these results, the 15 STR loci could be used as the genetic markers for the Naqu Tibetan population in studies of anthropology, linkage analysis of genetic disease, individual identification and paternity test in forensic medicine.

Key words: Short tandem repeat (STR); Polymorphism; Tibetan