

3个STR位点在中国7个民族遗传距离中的研究^{*}

黄辰^{1,2}, 王全丽¹, 宋土生¹, 司履生²,
胡晓岩¹, 倪磊¹, 陈雪玲³, 许健³

(1 西安交通大学 医学院; 2 生命科学与技术学院, 陕西 西安 710061; 3 新疆石河子大学 医学院, 新疆 石河子 832002)

[摘要] 应用PCR技术、质量分数4%变性聚丙烯酰胺凝胶电泳及银染技术, 研究了D16S539, D7S820和D13S713 3个STR位点在新疆锡伯族和哈萨克族中的遗传结构, 结合汉族、黎族、朝鲜族、藏族和维吾尔族相同位点的相关遗传资料, 应用亲缘系数和遗传距离研究了7个民族间的遗传关系, 并根据遗传距离绘制遗传系统树。结果显示, 汉族、朝鲜族、黎族、藏族和锡伯族为一群, 而新疆哈萨克族和维吾尔族为一群。其中, 汉族与朝鲜族遗传距离最近, 其他依次为黎族、藏族和锡伯族。

[关键词] STR位点; 遗传距离; 民族; 遗传关系

[中图分类号] Q343.1⁺2

[文献标识码] A

[文章编号] 1671-9387(2004)01-0081-04

短串联重复序列(short tandem repeats, STR)^[1]是由1~4个核苷酸为核心序列的短串联重复结构组成, 因核心序列重复次数的不同而构成了长度多态性。在人类基因组中, 平均每15~20 kb就存在1个STR位点, 因而人类STR具有分布广、信息量大、高度多态性以及遵循孟德尔共显性遗传等特点, 已成为重要的遗传性标记, 并被广泛应用于人类基因组图谱绘制、遗传连锁分析、法医学中的个体识别和亲权鉴定中。近年来, 国内外学者发现, STR也是民族亲缘关系有效的研究工具^[2~7]。

中国是一个历史悠久的多民族国家, 不同民族间文化、语言和风俗习惯存在明显差异, 比较研究各民族生物学信息和历史资料, 可以为研究民族间演化提供更可靠的理论依据。本研究应用3个STR位点的遗传数据分析了新疆锡伯族、哈萨克族、朝鲜族、汉族、维吾尔族、藏族和黎族7个民族间的亲缘关系, 并以民族间遗传距离为依据建立了7个民族间的遗传系统树, 为探讨民族间亲缘关系研究提供了资料。

1 材料和方法

1.1 标本来源

105份无关正常锡伯族纯系个体抗凝外周血, 来自新疆伊犁地区察布查尔锡伯族自治县; 104份

无关正常哈萨克族纯系个体抗凝外周血, 来自新疆石河子供销学校及沙湾县。

1.2 样品DNA的制备

采用Sino-American Biotechnology公司提供的全血基因组DNA纯化系统(Ready PCRTM Whole Blood Genetic DNA Purification System)提取DNA。

1.3 PCR反应

采用Promega公司提供的Gene Print[®] Silver STRTM III System进行D16S539, D7S820和D13S713 3个位点PCR复合扩增。反应条件参照产品说明书。

1.4 PCR扩增产物的电泳分离和银染

取PCR扩增产物2 μ L与等体积的2 $\times$ STR载样液(Promega)混匀, 于95 $^{\circ}$ C变性120 s后, 立即置于冰盒内, 上样于经预电泳的质量分数4%的聚丙烯酰胺凝胶上, 电压2 000 V, 50~55 $^{\circ}$ C电泳2 h, 然后用银染方法染色。

1.5 统计学处理

根据Hardy-Weinberg平衡定律, 计算各基因型个体数的期望值, 以 χ^2 检验进行Hardy-Weinberg平衡吻合性检验。分别计算出3个STR基因座各自的期望杂合度(heterozygosity, H)、个体识别率(discrimination power, DP)、期望排除概率(probability of exclusion, PE)和多态信息含量

* [收稿日期] 2002-12-20

[基金项目] 陕西省自然科学基金项目(2000K14-G11-8)

[作者简介] 黄辰(1965-), 男, 上海市人, 副教授, 在读博士, 主要从事人类遗传学和肿瘤细胞生物学研究。

(polymorphism information content, PIC)^[8]。民族间相关关系和遗传距离的计算参见 Nei^[9]的方法。

2 结果与分析

2.1 锡伯族和哈萨克族 3 个 STR 位点的遗传数据
105 名新疆锡伯族和 104 名新疆哈萨克族无关个体的 D16S539, D7S820 和 D13S713 3 个 STR 基因座等位片段及其频率的检测结果, 以及朝鲜族、汉

族、维吾尔族、藏族和黎族的相关资料列于表 1。由表 1 可知, 哈萨克族和锡伯族人群 D16S539, D7S820 和 D13S713 3 个 STR 位点基因型频率的观察值与期望值均符合 Hardy-Weinberg 平衡, 说明这两组人群遗传结构较为稳定。法医学数据见表 2, 结果表明, 两个民族 3 个位点的 H 和 DP 均较高, 说明这 3 个 STR 为鉴别能力较高的位点。

表 1 中国 7 个民族 3 个 STR 位点等位基因频率

Table 1 The allele frequency of three STR loci among seven ethnic groups

STR 位点 STR loci	太原汉族 ^[10] Han	朝鲜族 ^[11] Chinese Korean	维吾尔族 ^[12] Uygur	锡伯族 Xibo	哈萨克族 Chinese Kazak	藏族 ^[13] Tibetan	黎族 ^[14] Li
D16S539-6	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.003 9	0.000 0
D16S539-7	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.108 5	0.000 0
D16S539-8	0.018 7	0.015 8	0.015 6	0.022 7	0.024 5	0.050 4	0.000 0
D16S539-9	0.238 3	0.278 9	0.182 3	0.281 7	0.147 1	0.240 3	0.164 3
D16S539-10	0.130 8	0.147 4	0.151 0	0.131 7	0.142 2	0.124 0	0.176 1
D16S539-11	0.252 3	0.268 4	0.281 3	0.168 2	0.220 6	0.259 7	0.323 9
D16S539-12	0.224 3	0.189 5	0.208 3	0.231 7	0.205 9	0.147 3	0.239 4
D16S539-13	0.107 5	0.100 0	0.145 8	0.131 7	0.191 9	0.065 9	0.072 8
D16S539-14	0.028 0	0.000 0	0.015 6	0.031 8	0.063 7	0.000 0	0.000 0
D16S539-15	0.000 0	0.000 0	0.000 0	0.000 0	0.004 9	0.000 0	0.000 0
D7S820-6	0.000 0	0.010 5	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0
D7S820-7	0.009 3	0.010 5	0.016 7	0.004 5	0.009 8	0.000 0	0.000 0
D7S820-8	0.180 6	0.142 1	0.166 7	0.140 9	0.250 0	0.112 4	0.167 5
D7S820-9	0.078 7	0.063 2	0.155 6	0.095 5	0.102 9	0.081 4	0.070 0
D7S820-10	0.129 6	0.226 3	0.227 8	0.245 5	0.230 4	0.174 4	0.135 0
D7S820-11	0.379 6	0.336 8	0.227 8	0.327 3	0.220 4	0.360 5	0.397 0
D7S820-12	0.180 6	0.168 4	0.172 2	0.163 6	0.147 1	0.251 9	0.207 0
D7S820-13	0.032 4	0.042 1	0.027 8	0.013 6	0.039 2	0.019 4	0.022 5
D7S820-14	0.009 3	0.000 0	0.005 6	0.009 1	0.000 0	0.000 0	0.023 5
D13S713-7	0.000 0	0.000 0	0.000 0	0.000 0	0.009 8	0.003 9	0.002 3
D13S713-8	0.291 7	0.278 9	0.218 8	0.204 5	0.235 3	0.255 8	0.354 5
D13S713-9	0.143 5	0.147 4	0.125 0	0.127 3	0.083 3	0.155 1	0.157 3
D13S713-10	0.097 2	0.126 3	0.119 8	0.186 4	0.127 5	0.124 0	0.129 1
D13S713-11	0.236 1	0.205 3	0.197 9	0.268 2	0.201 0	0.271 3	0.244 1
D13S713-12	0.171 3	0.184 2	0.281 3	0.136 4	0.250 0	0.143 4	0.084 5
D13S713-13	0.046 3	0.031 6	0.057 3	0.063 6	0.044 1	0.034 9	0.028 0
D13S713-14	0.013 9	0.026 3	0.000 0	0.009 1	0.004 9	0.011 6	0.000 0
D13S713-15	0.000 0	0.000 0	0.000 0	0.004 5	0.000 0	0.000 0	0.000 0

表 2 中国新疆锡伯族和哈萨克族 3 个 STR 位点法医学数据

Table 2 The forensic data of three STR loci in Chinese Kazak and Xibo

位点 Loci	锡伯族 Xibo					哈萨克族 Chinese Kazak				
	等位片段 A llele fragment	H	DP	PE	PIC	等位片段 A llele fragment	H	DP	PE	PIC
D7S820	8	0.8875	0.9137	0.6488	0.7430	7	0.9356	0.9312	0.6276	0.7723
D13S713	8	0.9286	0.9368	0.6365	0.7862	8	0.9304	0.9398	0.6501	0.7885
D16S539	8	0.9362	0.9319	0.6100	0.7738	8	0.9439	0.9458	0.6713	0.8019

2.2 7 个民族间亲缘关系以及遗传树

根据表 1 的数据进行计算, 7 个民族间的相关关系和遗传距离分别列于表 3 和表 4。结果表明, 太原汉族和朝鲜族之间的亲缘关系最为密切, 其次为维吾尔族和哈萨克族。运用非加权平均法对 7 个民

族间的遗传距离进行聚类分析, 获得遗传系统树(图 1)。由图 1 可知, 7 个民族可以分为两个遗传群, 太原汉族、朝鲜族、黎族、藏族和锡伯族为一群, 而维吾尔族和哈萨克族为一群。

表 3 中国 7 个民族间的相关关系

Table 3 Similarity coefficient and genetic distances among the seven ethnic groups

民族 Ethnic group	太原汉族 Han	朝鲜族 Chinese Korean	维吾尔族 Uygur	锡伯族 Xibo	哈萨克族 Chinese Kazak	藏族 Tibetan	黎族 Li
太原汉族 Han							
朝鲜族 Chinese Korean	0.984						
维吾尔族 Uygur	0.947	0.961					
锡伯族 Xibo	0.963	0.973	0.942				
哈萨克族 Chinese Kazak	0.939	0.942	0.980	0.936			
藏族 Tibetan	0.970	0.970	0.925	0.953	0.898		
黎族 Li	0.979	0.962	0.914	0.929	0.901	0.959	

表 4 中国 7 个民族间的遗传距离

Table 4 Genetic distances among the seven ethnic groups

民族 Ethnic group	太原汉族 Han	朝鲜族 Chinese Korean	维吾尔族 Uygur	锡伯族 Xibo	哈萨克族 Chinese Kazak	藏族 Tibetan	黎族 Li
太原汉族 Han							
朝鲜族 Chinese Korean	0.016						
维吾尔族 Uygur	0.054	0.040					
锡伯族 Xibo	0.038	0.027	0.060				
哈萨克族 Chinese Kazak	0.063	0.060	0.020	0.066			
藏族 Tibetan	0.030	0.030	0.078	0.048	0.108		
黎族 Li	0.021	0.039	0.090	0.074	0.104	0.042	

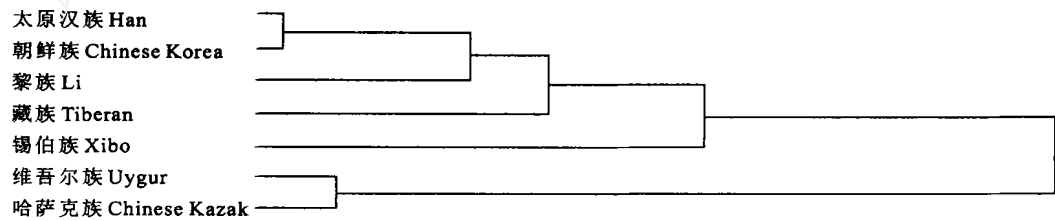


图 1 中国 7 个民族遗传系统树

Fig 1 Phylogenetic trees among 7 ethnic groups

3 讨 论

人类 STR 广泛分布于基因组的非编码区域, 其变异多不受自然选择的影响, 因而与其他基因多态性相比, 能够更完整地反映群体的遗传信息。此外, STR 位点高杂合度的特点也决定了 STR 在人类学研究中具有较高的准确性和敏感性。

本研究中的 7 个民族在人类学上可分为两类: 黄种人(蒙古人种)和白种人。前者包括汉族、朝鲜族、黎族、藏族和锡伯族, 后者包括哈萨克族和维吾尔族。陆舜华等^[15]对中国 11 个北方民族肤纹进行聚类分析, 结果表明汉族与朝鲜族遗传距离较近, 其次为藏族, 而哈萨克族和维吾尔族为另一聚类群。本研究结果与之一致, 汉族与朝鲜族遗传距离最近, 其次分别为黎族、藏族和锡伯族, 而哈萨克族和维吾尔族也为另一聚类群。

黎族居于中国海南岛, 地理位置相对隔离。本研究结果证明, 黎族遗传结构与汉族较为接近, 这与 ABO 血型研究结果一致^[16]。从历史角度考察, 黎族与白越族具有较深的渊源, 在春秋战国时期, 楚国灭了越国后, 迫使白越族南迁, 并和南方土著融合, 其中一支形成现在的黎族。因此, 黎族的形成经历了复杂的民族间的交融。

藏族在遗传结构上也与汉族具有很近的遗传距离。考古研究证明, 藏族可能起源于横断山脉的朵康(现甘青南部、川西高原和西藏东南部), 可能是罗刹女种系与崇拜猕猴的民族(羌族)发生融合, 形成藏族, 并在以后形成六支^[17]。而汉族为炎黄子孙, 炎帝所在部落也为羌族。另外, 在唐朝(公元 641)时, 文成公主进藏, 在汉藏之间发生联姻, 因此藏族在形成过程中与汉族间关系较为密切。本研究结果也与历史分析一致。

锡伯族起源于辽宁,在 1764 年部分锡伯族官兵赴新疆伊犁地区屯垦戍边,并定居于察布查尔锡伯族自治县,在体征上与汉族相似,同属蒙古人种,因此与汉族属于同一聚类群。

新疆哈萨克族在形成过程中融合了包括汉族、俄罗斯等 10 余个民族血统,并在 15 世纪稳定下来。

而维吾尔族形成于漠北回鹘汗国时期(公元 744~840 年),有 10 多个部落融合形成,并在喀喇汗王朝(公元 840~1212 年)与突厥语族融合,形成伊斯兰化。因此,哈萨克族和维吾尔族在亲缘关系上与汉族相对较远。

[参考文献]

- [1] Yuan B, Vaske D, Weber J D, et al. Improved set of short-tandem-repeat polymorphisms for screening the human genome[J]. *Am J Hum Genet*, 1997, 60: 459- 460
- [2] Etler D A. Recent developments in the study of human biology in China: a review[J]. *Hum Biol*, 1992, 64: 567- 585
- [3] Chu J Y, Huang W W, Kuang S Q, et al. Genetic relationship of populations in China[J]. *Proc Natl Acad Sci USA*, 1998, 95: 11763- 11768
- [4] Rolf B, Horst B, Eigel A, et al. Microsatellite profiles reveal an unexpected genetic relationship between Asian populations[J]. *Hum Genet*, 1998, 102: 647- 652
- [5] Zhao T, Lee T D. Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation[J]. *Hum Genet*, 1989, 83: 101- 110
- [6] Xie X, Wang X, He X, et al. Population study of three STR loci- HUM TH01, TPOX and CSF1PO in Dong-Xiang, Hui and Han from northwest China[J]. *Forensic Sci Int*, 2001, 119: 258- 259
- [7] Lin Z, Ohshima T, Gao S, et al. Genetic variation and relationships at five STR loci in five distinct ethnic groups in China[J]. *Forensic Sci Int*, 2000, 112: 179- 189
- [8] Lereu M V, Phillips C P, Carracedo A, et al. Investigation of the STR locus HUM TH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in patterning investigation[J]. *Forensic Sci Int*, 1994, 66: 41- 52
- [9] Nei M. *Molecular Evolutionary Genetics*[M]. New York: Columbia University Press, 1987.
- [10] 梁景青, 赵芳芳, 王向伟, 等. 太原地区汉族人群 3 个 STR 基因座的多态性分析[J]. *中国法医学杂志*, 2001, 16(1): 43- 44
- [11] 卜晓波, 宋洁, 王英. 中国牡丹江地区朝鲜族人群 3 个 STR 基因座遗传多态性[J]. *中国法医学杂志*, 2001, 16(2): 97- 98
- [12] 袁俐, 袁育康, 冯继东, 等. 新疆维吾尔族四个 STR 位点遗传多态性分析[J]. *生命科学研究*, 2001, 5(4): 291- 301
- [13] Xie X D, Wang X L, He X L, et al. Tibetan population data on the multiplex short tandem repeat loci D16S539, D7S820, and D13S713[J]. *遗传学报*, 2001, 28(6): 487- 492
- [14] 易升生, 刘雅诚, 霍振义. 中国黎族人群 13 个 STR 基因座的多态性调查[J]. *中国法医学杂志*, 2001, 16(4): 224- 225
- [15] 陆舜华, 郑斌, 张火敏. 内蒙地区蒙古、汉、回、朝鲜族肤纹特征比较研究[J]. *人类学报*, 1995, 14(3): 240- 246
- [16] 李晓军, 蔡巧青, 陈月甘. 海南汉族及黎族人群 ABO 血型基因频率的调查分析[J]. *海南医学院学报*, 1995, 1(2): 55- 58
- [17] 石硕. 一个隐含藏族起源真相的文本——对藏族始祖传说中“猕猴”与“罗刹女”含义的释读[J]. *中国社会科学*, 2000, (4): 167- 177.

The studies of three STR loci on the genetic distances among seven ethnic groups in China

HUANG Chen^{1,2}, WANG Quan-li¹, SONG Tu-sheng¹, SILu-sheng²,
HU Xiao-yan¹, NILei¹, CHEN Xue-ling³, XU Jian³

(1 Medical College; 2 Department of Genetics and Cytobiology, Jiaotong University, Xi'an 710061, China;

3 Department of Medical Immunology, Medical College of Shihezi University, Shihezi, Xinjiang 832002, China)

Abstract: Three short tandem repeat (STR) loci of D7S820, D13S713 and D16S539 were investigated in two ethnic groups living in China (104 unrelated Kazak and 105 unrelated Xibo). The DNA was amplified by PCR. The PCR products were analyzed by 4% denaturing PAGE and detected by silver stain. The data of the three STR on other five ethnic groups (Han, Chinese Korea, Li, Uyghur and Tibetan) were collected. The genetic similarity coefficient and genetic distances among the seven ethnic groups were calculated, and Phylogenetic trees based on the genetic distances were established. The result showed the seven ethnic groups were divided into two clusters. The Han, Chinese Korea, Li, Tibetan and Xibo were involved in an identical cluster, and Kazak and Uyghur in another.

Key words: short tandem repeat; genetic distance; ethnic groups; genetic coefficient