

· 技术与应用 ·

江苏汉族人群 19 个常染色体 STR 基因座基因多态性及遗传距离分析

潘 猛, 崔 鹤, 居晓斌, 刘燕婷, 叶 琴, 陈子庆, 丁小健, 陈 奇, 周 蓉, 顾 民, 周惠英
(南京医科大学第一附属医院 江苏省人民医院司法鉴定所, 江苏 南京 210029)

摘要: 目的 调查 19 个常染色体 STR 基因座在江苏汉族人群中的等位基因分布情况。方法 采用 Goldeneye® DNA 身份鉴定系统 20A 对 9 025 例样本进行检测, 对其中的 19 个常染色体 STR 基因座分型数据进行遗传学分析, 并对其与其他 17 个地区人群间的遗传距离进行分析。结果 19 个常染色体 STR 基因座均符合 Hardy-Weinberg 平衡 ($P>0.05$), 杂合度为 0.616 1~0.916 3, 匹配概率为 0.012 8~0.202 6, 个体识别率为 0.797 4~0.987 2, 非父排除率为 0.310 8~0.828 8, 多态信息含量为 0.561 7~0.913 6, 累积个体识别率为 0.999 999 999 999 998 434 1, 累积非父排除率为 0.999 999 989。江苏汉族人群与天津、湖南和吉林等汉族人群的遗传距离更近, 与新疆阿勒泰汉族人群之间差异具有统计学意义 ($P<0.05$)。结论 获得较客观准确的江苏汉族人群 STR 基因座基因多态性数据和群体遗传学参数, 可为上述 STR 基因座在江苏地区汉族人群的法医学应用提供数据支持。

关键词: 法医遗传学; 多态现象, 遗传; 短串联重复; 遗传距离; 汉族; 江苏

中图分类号: DF795.2 文献标志码: A doi: 10.12116/j.issn.1004-5619.2018.06.016

文章编号: 1004-5619(2018)06-0650-06

Analysis of Genetic Polymorphism and Genetic Distance of 19 Autosomal STR Loci in Jiangsu Han Population

PAN Meng, CUI He, JU Xiao-bin, LIU Yan-ting, YE Qin, CHEN Zi-qing, DING Xiao-jian, CHEN Qi, ZHOU Rong, GU Min, ZHOU Hui-ying

(Forensic Institution of Jiangsu Province Hospital, First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China)

Abstract: Objective To investigate the distribution of alleles in 19 autosomal short tandem repeat (STR) loci in Jiangsu Han population. **Methods** Goldeneye® 20A kit was used to detect 9 025 samples. Genetic analysis was performed on typing data of 19 autosomal STR loci, and genetic distance with other 17 populations was analyzed. **Results** All the 19 autosomal STR loci were consistent with the Hardy-Weinberg equilibrium ($P>0.05$), with the heterozygosity 0.616 1-0.916 3, probability of match 0.012 8-0.202 6, discrimination power 0.797 4-0.987 2, probability of paternity exclusion 0.310 8-0.828 8, and polymorphic information content 0.561 7-0.913 6. The cumulative discrimination power and cumulative probability of exclusion were 0.999 999 999 999 998 434 1 and 0.999 999 989, respectively. The Jiangsu Han population had close genetic distances with the Han population in Tianjin, Hunan and Jilin, and significant difference with Han population in Aletai region in Xinjiang ($P<0.05$). **Conclusion** The STR allele polymorphism data and population genetic parameters of Jiangsu Han population can provide data support for the forensic application of these STR loci in Jiangsu Han population.

Keywords: forensic genetics; polymorphism, genetic; short tandem repeat; genetic distance; Han nationality; Jiangsu

短串联重复(short tandem repeat, STR)序列一般由长度为 2~6bp 的核心序列串联重复而成, 广泛应用

于法医学个体识别和亲缘关系鉴定^[1-2]。本研究采用 Goldeneye® DNA 身份鉴定系统 20A 对江苏汉族九千

基金项目: 江苏省卫计委卫生标准研究课题资助项目(S2017004); 江苏省软科学研究资助项目(BR2018065)

作者简介: 潘猛(1979—), 男, 博士研究生, 副研究员, 主要从事法医物证学、医学免疫学研究; E-mail: 462486743@qq.com

通信作者: 周惠英, 女, 高级实验师, 副主任法医师, 主要从事法医物证学研究; E-mail: 780118873@qq.com

余份样本展开STR基因座群体遗传学调查,旨在总结江苏地区汉族人群19个常用STR基因座的群体遗传学参数等基础数据。

1 材料与方法

1.1 样本

9025份样本来源于本鉴定所2010—2018年进行亲缘关系鉴定的无关个体,其中男性样本5661份,女性样本3364份。采用本机构自主研发的亲权鉴定信息管理系统和司法鉴定科学研究院研发的亲权鉴定信息管理系统,以公民身份证前4位(省市代码)进行严格筛选,确保所有样本均来自江苏省各地级市范围。血样提供者均签署知情同意书。

1.2 DNA提取、扩增及分型

采用Chelex-100法提取模板DNA,或对1mm×1mm血斑免提取进行直接扩增。

扩增体系、扩增条件及分型方法按Goldeneye® DNA身份鉴定系统20A[基点认知技术(北京)有限公司]操作说明书进行。采用10μL反应体系进行扩增,扩增产物在3130基因分析仪(美国AB公司)上进行电泳。采用GeneScan 3.2软件进行分型结果分析。

1.3 统计学分析

联合应用PowerStats v12软件和SPSS 19.0软件对19个常染色体STR基因座进行Hardy-Weinberg平衡检验以及基因频率分析,计算杂合度(heterozygosity,

H)、匹配概率(probability of match, P_m)、个体识别率(discrimination power, DP)、非父排除率(probability of paternity exclusion, PE)、多态信息含量(polymorphic information content, PIC)等群体遗传学参数。应用Arlequin v3.5.2.2软件对江苏汉族人群19个常染色体STR基因座进行连锁不平衡分析,并对其与其他17个不同地区人群(13个国内人群、4个国外人群)^[3-16]两两之间的遗传分化系数(F_{st})和P值进行计算。采用MEGA 7软件进行邻接法进化树绘制。

2 结果

2.1 等位基因频率及群体遗传学参数

江苏汉族人群19个常染色体STR基因座的分布均符合Hardy-Weinberg平衡($P>0.05$),等位基因频率和群体遗传学参数见表1~2。其中,H为0.6161~0.9163, P_m为0.0128~0.2026, DP为0.7974~0.9872, PE为0.3108~0.8288, PIC为0.5617~0.9136。各基因座之间相互独立,属于连锁平衡状态, CDP为0.9999999999999999984341, CPE为0.999999989。

2.2 18个人群间的遗传距离分析

通过对江苏汉族人群和其他17个地区人群两两之间的基因频率差异进行统计学分析,获得不同地区之间遗传距离的 F_{st} 值和P值(表3)。其中,江苏汉族和新疆阿勒泰汉族之间的 F_{st} 值(0.07355)差异具有统计学意义($P<0.05$)。

表1 江苏汉族人群19个STR基因座的等位基因频率

(n=9025)

D19S433		D18S51		D3S1358		Penta E		D8S1179	
等位基因	频率	等位基因	频率	等位基因	频率	等位基因	频率	等位基因	频率
9.2	0.0001	7	0.0001	12	0.0014	5	0.0498	8	0.0019
10	0.0004	8	0.0002	13	0.0013	6	0.0001	9	0.0003
10.2	0.0003	9	0.0002	14	0.0447	7	0.0010	10	0.1023
11	0.0029	10	0.0011	15	0.3564	8	0.0045	11	0.0897
11.2	0.0001	10.2	0.0001	16	0.3274	9	0.0072	12	0.1332
12	0.0392	11	0.0029	17	0.2037	10	0.0436	13	0.2209
12.2	0.0049	12	0.0328	18	0.0597	11	0.1300	14	0.1870
13	0.2967	13	0.1895	19	0.0049	12	0.1089	15	0.1739
13.2	0.0394	13.2	0.0008	20	0.0004	13	0.0508	16	0.0789
14	0.2375	13.3	0.0001	D16S539		14	0.0851	17	0.0109
14.2	0.1088	14	0.2076			14.2	0.0001	18	0.0011
14.3	0.0001	15	0.1840	等位基因	频率	14.4	0.0002	TPOX	
15	0.0702	15.2	0.0001	6	0.0007	15	0.1003		
15.2	0.1481	16	0.1298	7	0.0002	16	0.0830	等位基因	频率
16	0.0114	17	0.0748	8	0.0095	17	0.0808	5	0.0001
16.2	0.0336	17.1	0.0001	9	0.2872	17.4	0.0001	5.2	0.0001
17	0.0015	18	0.0439	10	0.1286	18	0.0818	6	0.0001
17.2	0.0044	19	0.0441	10.2	0.0001	18.4	0.0015	7	0.0004
18.2	0.0003	20	0.0323	11	0.2495	19	0.0618	8	0.5200
		21	0.0250	12	0.2078	19.3	0.0001	9	0.1213
				13	0.1019			10	0.0253

续表 1

<i>D5S818</i>		<i>D18S51</i>		<i>D16S539</i>		<i>Penta E</i>		<i>TPOX</i>	
等位基因	频率	等位基因	频率	等位基因	频率	等位基因	频率	等位基因	频率
6	0.000 1	22	0.016 4	14	0.014 0	19.4	0.000 9	10.3	0.000 1
7	0.016 6	23	0.006 4	15	0.000 8	20	0.047 1	11	0.302 4
8	0.003 1	24	0.004 2	<i>CSFIPO</i>		21	0.028 7	12	0.028 6
9	0.068 4	25	0.002 2	等位基因	频率	22	0.016 8	13	0.001 8
10	0.194 0	26	0.000 6	5	0.000 1	22.3	0.000 1	14	0.000 1
11	0.332 8	27	0.000 5	7	0.001 4	23	0.008 2	<i>TH01</i>	
12	0.234 2	28	0.000 1	8	0.002 4	23.2	0.000 1	等位基因	频率
13	0.138 6	29	0.000 1	9	0.048 6	24	0.004 1	5	0.000 3
14	0.010 4	<i>D6S1043</i>		10	0.239 1	25	0.002 2	6	0.105 3
15	0.001 7	等位基因	频率	10.1	0.000 1	26	0.000 8	7	0.259 0
16	0.000 1	8	0.000 2	11	0.241 6	27	0.000 3	8	0.055 8
<i>D21S11</i>		9	0.001 4	12	0.378 9	<i>D12S391</i>		9	0.517 5
等位基因	频率	10	0.032 0	12.1	0.000 1	等位基因	频率	9.3	0.035 3
25	0.000 1	11	0.097 4	13	0.075 8	14	0.000 2	10	0.025 6
26	0.000 2	12	0.137 6	14	0.011 2	15	0.015 5	11	0.000 8
27	0.002 7	13	0.140 5	15	0.000 8	16	0.007 0	12	0.000 2
28	0.045 4	14	0.141 5	16.1	0.000 1	17	0.093 0	<i>FGA</i>	
28.2	0.007 3	14.3	0.000 1	<i>Penta D</i>		17.3	0.000 4	等位基因	频率
29	0.260 3	15	0.015 1	等位基因	频率	18	0.228 6	13	0.000 2
29.2	0.000 8	16	0.003 4	5	0.000 1	18.2	0.000 1	14	0.000 2
30	0.277 2	17	0.044 0	6	0.003 0	18.3	0.000 1	15	0.000 1
30.2	0.009 9	17.3	0.000 7	7	0.005 8	19	0.219 9	16	0.000 2
30.3	0.004 9	18	0.174 9	8	0.048 7	19.3	0.000 3	16.2	0.000 1
31	0.108 5	18.2	0.000 2	9	0.319 2	20	0.171 4	17	0.002 3
31.1	0.000 2	18.3	0.000 1	10	0.118 1	21	0.106 7	17.2	0.000 1
31.2	0.071 3	19	0.144 8	11	0.147 6	21.3	0.000 1	18	0.022 1
32	0.031 7	19.1	0.000 1	11.2	0.000 1	22	0.079 5	18.2	0.000 1
32.2	0.122 0	19.3	0.000 5	12	0.180 4	23	0.044 6	19	0.047 0
33	0.006 9	20	0.054 0	13	0.126 2	24	0.020 7	20	0.046 6
33.1	0.000 1	20.3	0.000 8	14	0.040 2	25	0.009 5	20.2	0.000 3
33.2	0.044 3	21	0.007 2	15	0.009 0	26	0.002 0	21	0.103 4
34	0.000 7	21.3	0.002 1	16	0.001 3	27	0.000 3	21.2	0.002 3
34.2	0.004 7	22	0.000 3	17	0.000 2	<i>D2S1338</i>		21.3	0.000 1
35	0.000 1	22.3	0.001 3	19	0.000 1	等位基因	频率	22	0.170 9
35.2	0.000 2	23	0.000 1	<i>vWA</i>		13	0.000 1	22.2	0.005 7
36	0.000 2	23.3	0.000 1	等位基因	频率	14	0.000 1	22.3	0.000 1
36.2	0.000 1	<i>D7S820</i>		10	0.000 1	15	0.000 1	23	0.228 0
<i>D13S317</i>		等位基因	频率	12	0.000 1	16	0.011 6	23.1	0.000 1
等位基因	频率	6	0.000 1	13	0.002 3	17	0.066 8	23.2	0.010 3
6	0.000 1	7	0.000 9	14	0.253 4	18	0.114 6	24	0.193 5
7	0.003 2	8	0.141 9	15	0.030 9	19	0.162 5	24.2	0.008 5
8	0.271 9	9	0.058 3	16	0.176 0	20	0.120 3	25	0.099 1
9	0.133 4	9.1	0.002 8	17	0.249 8	21	0.025 2	25.2	0.003 9
10	0.146 8	9.2	0.000 1	18	0.181 4	22	0.046 3	26	0.042 3
11	0.246 3	10	0.158 8	19	0.089 4	23	0.218 9	26.2	0.000 7
12	0.151 5	10.1	0.000 8	20	0.015 3	24	0.156 7	27	0.010 0
13	0.037 2	11	0.348 4	21	0.001 2	25	0.061 3	27.2	0.000 1
14	0.009 2	11.1	0.000 2	22	0.000 1	26	0.011 8	28	0.001 8
15	0.000 3	12	0.243 6	23	0.000 1	27	0.002 6	29	0.000 3
		13	0.038 4			28	0.000 9	30	0.000 3
		14	0.005 4			29	0.000 2		
		15	0.000 3						

表2 江苏汉族人群19个STR基因座的群体遗传学参数

(n=9 025)

基因座	H	Pm	DP	PE	PIC	P值 ¹⁾
D19S433	0.811 6	0.058 4	0.941 6	0.620 8	0.791 9	0.529 9
D5S818	0.768 9	0.086 1	0.913 9	0.542 9	0.738 5	0.195 8
D21S11	0.814 7	0.055 7	0.944 3	0.626 6	0.796 5	0.792 2
D18S51	0.853 8	0.037 1	0.962 9	0.702 3	0.840 7	0.696 7
D6S1043	0.870 5	0.029 7	0.970 3	0.735 6	0.860 8	0.451 0
D3S1358	0.717 9	0.129 3	0.870 7	0.456 7	0.668 9	0.518 4
D13S317	0.801 5	0.069 1	0.930 9	0.601 9	0.771 6	0.573 1
D7S820	0.773 6	0.087 1	0.912 9	0.550 9	0.737 4	0.640 4
D16S539	0.788 8	0.081 5	0.918 5	0.578 7	0.750 3	0.663 5
CSFIPO	0.733 9	0.116 0	0.884 0	0.482 8	0.689 1	0.457 9
Penta D	0.808 3	0.060 5	0.939 5	0.614 6	0.786 1	0.650 4
vWA	0.797 4	0.069 9	0.930 1	0.594 4	0.770 2	0.243 7
D8S1179	0.846 4	0.044 0	0.956 0	0.688 0	0.825 4	0.077 4
TPOX	0.616 1	0.202 6	0.797 4	0.310 8	0.561 7	0.531 0
Penta E	0.916 3	0.012 8	0.987 2	0.828 8	0.913 6	0.232 9
TH01	0.640 5	0.174 3	0.825 7	0.342 4	0.597 6	0.598 6
D12S391	0.837 0	0.045 3	0.954 7	0.669 3	0.821 1	0.569 7
D2S1338	0.866 5	0.034 7	0.965 3	0.727 6	0.847 5	0.309 8
FGA	0.848 8	0.038 6	0.961 4	0.692 6	0.836 8	0.479 5

注:1)Hardy-Weinberg平衡检验

2.3 遗传进化树分析

进化树用邻接法推断,本次比对数据按0.0100比例绘制。进化树结果表明:江苏汉族人群和国内其他汉族人群遗传距离较近,而与新疆阿勒泰汉族、云南苗族、俄罗斯、波兰和奥地利人群距离较远。

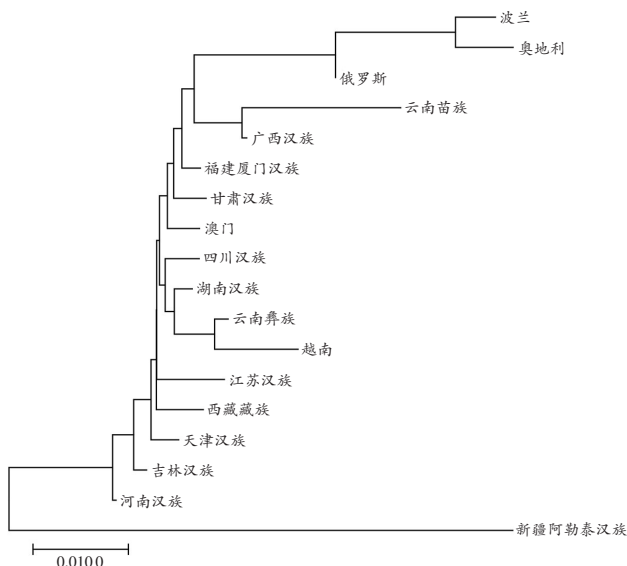


图1 邻接法分析18个人群间的遗传进化距离

3 讨论

本研究采用等位基因频率, H、Pm、DP、PE和PIC等群体遗传学参数对江苏汉族人群19个常染色体STR基因座的遗传多态性进行分析。Hardy-Weinberg平衡检验P值均大于0.05,说明所收集的群体数据基因

型分布的观察值和预期值之间差异无统计学意义,样本收集有效。H值范围从0.616 1(*TPOX*)到0.916 3(*Penta E*),表明体系内所有基因座的两个等位基因在随机抽样情况下区分为两个不同等位基因(杂合子)的可能性较高^[17]。Pm值从0.012 8(*Penta E*)到0.202 6(*TPOX*),表明随机抽取个体在进行个体识别时匹配可能性较低^[17]。PIC值从0.561 7(*TPOX*)到0.913 6(*Penta E*),当该值大于0.5时STR标记物可提供高度多态性信息^[18]。DP和PE值反映遗传标记系统在个体识别和亲权鉴定中的能力,通常当DP大于0.9、PE大于0.5能表明基因座具有高度多态性^[19],本研究江苏汉族人群在19个STR基因座的DP值从0.797 4(*TPOX*)到0.987 2(*Penta E*),PE值从0.310 8(*TPOX*)到0.828 8(*Penta E*)。各基因座之间相互独立,属于连锁平衡状态,CDP为0.999 999 999 999 998 434 1, CPE为0.999 999 989。其中,*TPOX*(0.310 8)、*TH01*(0.342 4)、*D3S1358*(0.456 7)和*CSFIPO*(0.482 8)4个基因座PE值小于0.5,说明在江苏汉族人群中上述4个基因座多态性欠佳,可通过大样本分析获得更符合江苏人群特征的基因座予以替代。以上统计数据表明,本研究选择的STR基因座在江苏汉族人群中具有较高的应用可行性。

对18个不同人群间的 F_{st} 和P值进行比较,结果表明,江苏汉族和新疆阿勒泰汉族人群之间的 F_{st} 值差异具有统计学意义($P<0.05$),和其他人群间的遗传距离差异无统计学意义。新疆阿勒泰汉族人群与大

部分汉族人群间差异均有统计学意义。江苏汉族人群与天津汉族、湖南汉族、吉林汉族、福建厦门汉族、四川汉族、澳门(民族不详)、甘肃汉族、河南汉族、西藏藏族、云南彝族、广西汉族、越南人群的遗传距离依次增加,如以江苏汉族和天津汉族人群的遗传距离(0.01097)作为参考,江苏汉族与云南苗族、俄罗斯、波兰和奥地利等人群有超过2倍以上的遗传距离,用进化树可直观展示上述差异。遗传距离统计分析提示,人群的基因遗传差异与实际的物理距离成正相关。

目前,江苏地区各鉴定机构采用不同来源的基因频率作为累积亲权指数值计算的依据。由于这些频率可能会存在样本量相对较少、地区代表性低等问题,对区域性亲权鉴定的标准化和规范化造成影响。通过本次9025例大样本数据的分析总结,可得到较以往更科学、可靠的江苏地区19个常染色体STR基因座的群体遗传学数据,将会为江苏地区法医物证专业的有序化、规范化发展提供必要的基础数据支撑。同时,希望通过本研究成果的发表,面向江苏省法医物证专业进行基因频率推广,达到在江苏省内统一基因频率的目标。

(感谢中山大学孙宏钰教授对本文的悉心指导!)

参考文献:

- [1] ZHANG S, NIU Y, BIAN Y, et al. Sequence investigation of 34 forensic autosomal STRs with massively parallel sequencing[J]. *Sci Rep*, 2018, 8(1): 6810.
- [2] ZHANG S, BIAN Y, TIAN H, et al. Development and validation of a new STR 25-plex typing system[J]. *Forensic Sci Int Genet*, 2015, 17:61-69.
- [3] ZHANG X, HU L, DU L, et al. Genetic analysis of 20 autosomal STR loci in the Miao ethnic group from Yunnan Province, Southwest China[J]. *Forensic Sci Int Genet*, 2017, 28:e28-e29.
- [4] LIU Q L, CHEN Y F, HUANG X L, et al. Population data and mutation rates of 19 STR loci in seven provinces from China based on Goldeneye™ DNA ID System 20A[J]. *Int J Legal Med*, 2017, 131(3):653-656.
- [5] SUN M, ZHANG X, WU D, et al. Population genetic data of 15 STR loci in Gansu Han population from China[J]. *Int J Legal Med*, 2015, 129(4): 735-736.
- [6] XU X M, FENG Z, YAO J. Population data of 15 autosomal STR loci in Chinese Han population from Jilin Province, Northeast China[J]. *Forensic Sci Int Genet*, 2017, 27:187-188.
- [7] LU Y, SONG P, HUANG J C, et al. Genetic polymorphisms of 20 autosomal STR loci in 5141 individuals from the Han population of Xiamen, Southeast China[J]. *Forensic Sci Int Genet*, 2017, 29:e31-e32.
- [8] CHEN P, WANG B, HE G, et al. Population genetic analyses of 20 autosomal STR loci in Chinese Han population from Tianjin in Northern China[J]. *Forensic Sci Int Genet*, 2017, 27:184-185.
- [9] KANG L L, ZHAO J, LIU K, et al. Allele frequencies of 15 STR loci of Tibetan lived in Tibet Lassa[J]. *Forensic Sci Int*, 2007, 168(2-3):236-240.
- [10] ZHANG X, HU L, DU L, et al. Population data for 20 autosomal STR loci in the Yi ethnic minority from Yunnan Province, Southwest China[J]. *Forensic Sci Int Genet*, 2017, 28:e43-e44.
- [11] GUSMÃO L, PRATA M J, SÁNCHEZ-DIZ P, et al. STR data for the AmpFℓSTR profiler plus loci from Macau (China)[J]. *Forensic Sci Int*, 2001, 123(1): 74-75.
- [12] PARYS-PROSZEK A, KUPIEC T, WOLAŃSKA-NOWAK P, et al. Genetic variation of 15 autosomal STR loci in a population sample from Poland[J]. *Leg Med (Tokyo)*, 2010, 12(5):246-248.
- [13] STEPANOV V A, MELNIKOV A V, LASH-ZAVADA A Y, et al. Genetic variability of 15 autosomal STR loci in Russian populations[J]. *Leg Med (Tokyo)*, 2010, 12(5):256-258.
- [14] NEUHUBER F, RADACHER M, MEISNER N, et al. Nine STR markers plus amelogenin (AmpFℓSTR Profiler Plus): A forensic study in an Austrian population[J]. *Int J Legal Med*, 1999, 113(1):60-62.
- [15] SHIMADA I, BRINKMANN B, TUYEN N Q, et al. Allele frequency data for 16 STR loci in the Vietnamese population[J]. *Int J Legal Med*, 2002, 116(4):246-248.
- [16] LI X, LI L, WANG Q, et al. Population genetic analysis of the Globalfiler STR loci in 3 032 individuals from the Altay Han population of Xinjiang in Northwest China[J]. *Int J Legal Med*, 2018, 132(1): 141-143.
- [17] 侯一平. 法医物证学[M]. 4版. 北京:人民卫生出版社, 2016.
- [18] 赵勤松,任峥,张红玲,等. 贵州汉族19个STR基因座多态性及法医学应用[J]. *法医学杂志*, 2017, 33(4): 388-392.
- [19] GILL P, URQUHART A, MILLICAN E, et al. A new method of STR interpretation using inferential logic--development of a criminal intelligence database[J]. *Int J Legal Med*, 1996, 109(1):14-22.

(收稿日期:2018-03-22)

(本文编辑:张素华)