

Microreader™ 23HS Plex ID System 中 23 个常染色体 STR 基因座 在中国北方汉族人群中多态性调查

石妍¹⁾, 江坚²⁾, 陈冲¹⁾, 张京晶³⁾, 贾莉¹⁾, 高知泉⁴⁾, 李惠芬¹⁾, 严江伟⁵⁾, 任贺⁶⁾

(1)北京通达首诚司法鉴定所, 北京 100192; 2)福建南方司法鉴定中心, 福建福州 350000;

3)北京华彦科技有限公司司法鉴定所, 北京 100192; 4)苏州阅微基因技术有限公司,

江苏苏州 215000; 5)山西医科大学法医学院, 山西晋中 030600;

6)北京警察学院刑事科学技术系, 北京 102202)

[摘要] 目的 评价 Microreader™ 23HS Plex ID System 试剂盒中包含的 23 个常染色体 STR 基因座在中国北方汉族人群中位基因频率分布, 获得群体遗传数据, 探究其在法医学中的应用价值。方法 使用 Microreader™ 23HS Plex ID System 试剂盒对中国北方汉族人群 548 例无关样本 DNA 进行检测, 收集分型数据, 计算各基因座的等位基因频率、样本的杂合度(heterozygosity, H)、这些常染色体 STR 基因座的多态信息含量(polymorphism information content, PIC)、个体识别力(power of discrimination, DP)和非父排除率(probability of paternity exclusion, PE)并使用统计软件对各基因座是否符合 Hardy-Weinberg 平衡进行检验; 同时对 Microreader™ 23HS Plex ID System 的累积个体识别能力(CDP)和累积非父排除率(CPE)进行计算。结果 在 548 例无关样本中 23 个 STR 基因座共计检出 260 个等位基因, 等位基因频率为 0.0009 ~ 0.5902, H 为 0.611 ~ 0.885, PIC 为 0.577 ~ 0.864, DP 为 0.815 ~ 0.973(平均 DP 为 0.922), PE 为 0.089 ~ 0.406, CDP=1-3.663 × 10⁻²⁷, CPE=1-2.668 × 10⁻¹⁶, 所有基因座等位基因的分布符合 Hardy-Weinberg 平衡。结论 Microreader™ 23HS Plex ID System 的 23 个基因座在中国北方汉族人群中具有良好的多态性, 在法医学个人识别、群体遗传学研究、亲子鉴定, 特别是复杂亲缘关系鉴定中应用价值较高。

[关键词] 法医遗传学; 遗传多态性; STR 基因座

[中图分类号] D919 **[文献标志码]** A **[文章编号]** 2095-610X(2024)07-0154-06

Polymorphism Investigation of 23 Autosome STR Loci in the Microreader™ 23HS Plex ID System in Han Population in Northern China

SHI Yan¹⁾, JIANG Jian²⁾, CHEN Chong¹⁾, ZHANG Jingjing³⁾, JIA Li¹⁾,

GAO Zhixiao⁴⁾, LI Huifen¹⁾, YAN Jiangwei⁵⁾, REN He⁶⁾

(1) Beijing Tongda Shoucheng Institute of Forensic Science, Beijing 100192; 2) Fujian Nanfang

Judicial Expertise Center, Fuzhou Fujian 350000; 3) Beijing Huayan Judicial Authentication

Institute, Beijing 100192; 4) Suzhou Microread Gene Technology Co., LTD, Suzhou Jiangsu

215000; 5) College of Medicine and Forensics, Shanxi Medical University, Jinzhong

Shanxi 030600; 6) Dept. of Criminal Science and Technology,

Beijing Police College, Beijing 102202, China)

[收稿日期] 2024-03-02

[基金项目] 上海市法医学重点实验室司法部司法鉴定重点实验室开放课题(KF202011)

[作者简介] 石妍(1986~), 女, 黑龙江桦南人, 理学学士, 主检法医师, 主要从事法医学物证学相关研究工作。江坚与石妍对本文有同等贡献。

[通信作者] 严江伟, E-mail: yanjw@sxmu.edu.cn

[Abstract] Objective To obtain the distribution of allele frequency and population genetic data of 23 autosomal STR loci contained in the Microreader™ 23HS Plex ID System kit in the Han population of northern China, and explore its application value in forensic medicine. **Methods** The Microreader™ 23HS Plex ID System kit was used to detect the DNA of 548 unrelated samples which from the Han population in northern China. Then, the allele frequencies, heterozygosity(H) in the samples, polymorphism information content(PIC), power of discrimination(DP), and probability of paternity exclusion(PE) were calculated and Hardy-Weinberg equilibrium was tested on each locus based on the genotyping data of 548 samples we collected. The cumulative individual identification ability(CDP) and cumulative non-paternity exclusion rate(CPE) of the Microreader™ 23HS Plex ID System were also calculated. **Results** A total of 260 alleles were detected in 548 unrelated samples from 23 STR loci which conformed to Hardy-Weinberg equilibrium. And the alleles frequencies ranged from 0.0009 to 0.5902, H ranged from 0.611 to 0.885, PIC ranged from 0.577 to 0.864, DP ranged from 0.815 to 0.973 (mean DP was 0.922) and PE ranged from 0.089 to 0.406. At the same time, the CDP of Microreader™ 23HS Plex ID System was $1-3.663 \times 10^{-27}$ and CPE was $1-2.668 \times 10^{-16}$. **Conclusion** The 23 loci of Microreader™ 23HS Plex ID System have a good polymorphism in the north Chinese Han population, and have high application value in population genetics research, forensic personal identification, paternity testing and complex kinship identification.

[Key words] Forensic genetics; Genetic polymorphism; STR loci

短串联重复序列(short tandem repeat, STR)被应用于疾病基因定位研究、法医个体识别及亲权鉴定等多个方面,其核心序列一般为 2~6 bp 串联重复,在人类全基因组序列中均有分布^[1]。2021 年新版的《生物学全同胞关系鉴定技术规范》(SF/T 0117-2021)^[2]中首次引入了状态一致性评分(identity by state score, IBS)概念,规范显示在准确性不低于 99.99%的前提下,检测系统中包含的常染色体 STR 基因座数目达到 55 个时,IBS 与全同胞关系指数 FSI(full sibling index)2 种方法的系统效能分别为 0.9986 和 0.9965。2023 年新版的《生物学半同胞关系鉴定技术规范》(SF/T 0131-2023)^[3]要求 3 种鉴定情形所采用检测系统的系统效能均应大于 0.75,即只有 2 个体配对的半同胞关系鉴定,检测的常染色体 STR 基因座数量应当在 73 个以上;三个体和四个体组合半同胞关系鉴定,因鉴定参考人员数量增加获取的遗传信息更多,需要检测的基因座数量也相应的减少,分别不少于 39 个和 35 个^[4]。2 个标准还同时要求除 19 个必检基因座外,增加检测的基因座不存在连锁不平衡且平均个体识别能力不低于 0.900。因此在遇到复杂亲缘鉴定时,如果没有条件增加参考人员,则应尽可能多的增加检测系统以提高系统效能,降低错判的风险。

Microreader™ 23HS Plex ID System 试剂盒作为常染色体补充试剂盒,检测的基因座包括: D4S2408, D9S2157, D12S391, D20S161, D3S2459, D18S1364, D13S305, D1S2142, D22S1045, D19S400,

D6S1017, D7S1517, D2S1776, D2S1360, D3S1744, D16S3391, D3S1545, D11S4463, D20S85, D1S549, D2S441, D10S2325, D21S2055 及一个 Amelogenin 性别基因,这些 STR 基因座在文献^[5-10]报道中个体识别力 DP 均不小于 0.9。然而对这些 STR 基因座研究结果受限于样本数量和采样范围,其在中国北方汉族人群中的多态性研究仍不充分。因此本研究旨在采用 Microreader™ 23HS Plex ID System 试剂盒对中国北方汉族人群 548 例无关样本进行分析,获得上述基因座的遗传多态性和法医学参数,为 Microreader™ 23HS Plex ID System 试剂盒在个体识别及亲权鉴定,特别是全同胞、半同胞关系鉴定领域提供数据支持。

1 材料与方法

1.1 样本

548 例无关样本 DNA 随机选自北京通达首诚司法鉴定所 DNA 库,在签署知情同意书的基础上,样本主要来源于北京市及附近周边省市的汉族志愿者,符合法医伦理学的要求。本研究通过复旦大学伦理委员会生物医学研究项目伦理批号[复伦研批(FE22025R)号]。

1.2 方法

按照 Microreader™ 23HS Plex ID System 试剂盒(苏州阅微基因技术有限公司)说明书进行 PCR 复合扩增,扩增程序为:预变性 95 ℃, 5 min;退火延伸 94 ℃, 20 s, 60 ℃, 90 s, 重复 28 个循

环；后延伸 60 ℃，20 min；4 ℃ 保存备用。扩增体系：Microreader™ 2.5 × Master Mix I，4 μL；Microreader™ 23HS Plex 5 × Primer Mix，2 μL；DNA，0.5–2 ng；补充 Nuclease-Free Water 至 10 μL。

扩增产物通过美国 AB 公司的 ABI 3500XL 遗传分析仪进行电泳检测，检测结果使用 GeneMapper ID-X v1.5 软件进行分析。

1.3 数据分析

用 Excel 2003 完成数据库的建立，使用 Count 函数统计每个基因座的等位基因频数并计算等位基因频率；使用 Cervus 3.0.7 软件计算每个基因座的 H、PIC、DP、PE [11]，同时使用卡方检验进行 Hardy-Weinberg 平衡测试。根据文献计算 CDP=

$1 - (1 - DP_1)(1 - DP_2) \cdots (1 - DP_k) = 1 - \prod (1 - DP_k)$ 和 $CPE = 1 - (1 - PE_1)(1 - PE_2) \cdots (1 - PE_k) = 1 - \prod (1 - PE_k)$ [12]。公式中的 k 表示第 k 个遗传标记的 DP 值和 PE 值，CDP 越大表示体系的个体识别能力越强，CPE 越大表示体系的鉴别能力越强。

2 结果

2.1 中国北方汉族人群等位基因频率

548 例无关样本的 23 个常染色体基因座共检出 260 个等位基因，单个基因座等位基因分型数量介于 6 个 (D4S2408) ~ 23 个 (D10S2325)，等位基因的频率分布在 0.0009 ~ 0.5902 之间，见表 1。

表 1 23 个常染色体 STR 基因座等位基因频率分布 (n = 548)(1)

Tab. 1 Allele frequency distribution of 23 autosomal STR loci (n = 548)(1)

D10S2325		D11S4463		D13S305		D16S3391		D19S400	
基因型	频率	基因型	频率	基因型	频率	基因型	频率	基因型	频率
6	0.0138	9	0.0018	12	0.0037	8	0.0009	7	0.0009
7	0.3226	11	0.0009	13	0.0101	9	0.0419	8	0.1564
8	0.0194	12	0.0583	14	0.0221	10	0.1883	9	0.0505
9	0.0652	13	0.2054	15	0.0419	11	0.0563	10	0.0496
10	0.0991	14	0.3382	16	0.0948	12	0.151	11	0.3507
11	0.1321	14.2	0.0009	17	0.1575	13	0.2852	12	0.1673
12	0.1438	15	0.2613	17.2	0.0009	14	0.2111	13	0.0878
13	0.0991	16	0.1132	18	0.2503	15	0.0534	14	0.069
14	0.0555	17	0.0221	18.2	0.0009	16	0.0037	15	0.0525
15	0.0222	18	0.0027	18.3	0.0009	D18S1364		16	0.0092
16	0.011	D12S391		19	0.2275	基因型	频率	17	0.0009
17	0.0055	基因型	频率	20	0.1371	11	0.0009	D18S2142	
18	0.0009	15	0.0129	20.2	0.0009	12	0.0353	基因型	频率
19	0.0009	16	0.0101	21	0.0429	13	0.1906	12	0.3479
19.3	0.0009	17	0.1162	22	0.0092	14	0.1906	13	0.1193
20	0.0027	18	0.2491	23	0.0009	15	0.2031	14	0.041
21	0.0037	19	0.2193	D1S549		16	0.1985	15	0.2077
22	0.0009	20	0.1499	基因型	频率	17	0.0391	16	0.1861
23	0.0009	21	0.0979	9	0.0009	18	0.0838	17	0.0958
25	0.0009	22	0.0729	11	0.0651	19	0.0534	18	0.0101
26	0.0009	23	0.0438	12	0.1531	20	0.0101	D2S1360	
27	0.0009	24	0.0175	13	0.1009	D22S1045		基因型	频率
31	0.0009	25	0.0101	14	0.2663	基因型	频率	15	0.0009
D20S161		26	0.0046	15	0.3232	11	0.2406	19	0.0055
基因型	频率	D20S85		16	0.0709	12	0.0037	20	0.0037
15	0.0602	基因型	频率	16.2	0.0009	13	0.0082	21	0.175
16	0.2263	10	0.4301	17	0.0138	14	0.0119	22	0.5902
17	0.378	11	0.0101	D21S2055		15	0.2454	23	0.1235
18	0.1319	12	0.0486	基因型	频率	16	0.2903	24	0.0306
19	0.1245	13	0.0888	9.1	0.0989	17	0.1783	25	0.0259
20	0.0612	14	0.1542	10.1	0.0082	18	0.0175	26	0.0184
21	0.0156	15	0.164	11.1	0.0027	19	0.0064	27	0.0166
22	0.0009	16	0.0898	12.1	0.0018	21	0.0009	28	0.0092

表 1 23 个常染色体 STR 基因座等位基因频率分布 ($n = 548$)(2)
Tab. 1 Allele frequency distribution of 23 autosomal STR loci ($n = 548$)(2)

D10S2325		D11S4463		D13S305		D16S3391		D19S400	
基因型	频率	基因型	频率	基因型	频率	基因型	频率	基因型	频率
D2S1776		17	0.0147	13.1	0.0009	D3S1545		29	0.0027
基因型	频率	18	0.0037	15	0.0018	基因型	频率	30	0.0018
8	0.0009	D2S441		16	0.0027	7	0.0903	D3S1744	
9	0.1596	基因型	频率	17	0.0259	8	0.0009	基因型	频率
10	0.0621	9	0.0009	18	0.1884	9	0.0018	13	0.0184
11	0.2813	9.1	0.041	19	0.2275	10	0.0182	14	0.1131
11.1	0.0009	10	0.2454	20	0.0259	11	0.229	15	0.0999
12	0.3883	10.1	0.0018	21	0.0231	12	0.3422	16	0.1111
13	0.0818	11	0.33	22	0.0306	12.3	0.0009	17	0.3045
14	0.0296	11.3	0.0362	23	0.0259	13	0.2308	18	0.1827
15	0.0064	12	0.1772	24	0.0203	13.3	0.0018	19	0.1019
D3S2459		13	0.0259	25	0.041	14	0.0785	20	0.0467
基因型	频率	14	0.1298	26	0.0798	15	0.0046	21	0.0073
8	0.024	14.1	0.0009	27	0.1193	16	0.0009	D7S1517	
9	0.0027	15	0.0147	28	0.0563	D9S2157		基因型	频率
10	0.0101	16	0.0018	29	0.0166	基因型	频率	15	0.0018
11	0.185	D4S2408		30	0.0082	7	0.024	16	0.0009
12	0.275	基因型	频率	D6S1017		9	0.0037	17	0.0175
13	0.1884	8	0.224	基因型	频率	11	0.0203	18	0.011
14	0.1895	9	0.3033	8	0.2298	12	0.0838	19	0.075
15	0.0999	10	0.3046	9	0.0046	13	0.2564	20	0.1342
16	0.0156	11	0.1424	10	0.3606	14	0.2088	21	0.1438
17	0.0027	12	0.0278	11	0.0296	15	0.2955	22	0.1165
18	0.0009	13	0.0009	12	0.2851	16	0.0968	23	0.1545
				13	0.0709	17	0.0138	24	0.1103
				14	0.0055	18	0.0018	25	0.1632
								25.2	0.0009
								26	0.0344
								27	0.0166
								28	0.0092
								29	0.0009

2.2 中国北方汉族人群中群体遗传学参数

23 个常染色体 STR 基因座的杂合度 H 为 0.611(D2S1360) ~ 0.885(D21S2055), 多态信息指数 PIC 为 0.577(D2S1360) ~ 0.864(D21S2055), 个体识别力 DP 为 0.815(D2S1360) ~ 0.973(D21S2055), 平均 DP 为 0.922, 非父排除率 PE 为 0.577(D2S13-60) ~ 0.864(D21S2055), $CDP=1-3.663 \times 10^{-27}$, $CPE=1-2.668 \times 10^{-16}$ 。 χ^2 检验表明所有基因座均未显著偏离 Hardy-Weinberg 平衡 ($P > 0.05$), 见表 2。

3 讨论

目前, 亲权鉴定虽然有检测 SNP 位点等其他手段, DNA STR 分型技术仍然占据主流地位, 是国内外进行亲权鉴定的主要检测手段^[13]。为了提升比对效率, 方便数据库建立, 目前广泛应用于法庭科学的常染色体 STR 检测试剂盒的一般都包含 CODIS 基因座。而 STR 基因座因其特殊的序列

表 2 23 个 STR 基因座群体遗传学参数($n = 548$)Tab. 2 Population genetic parameters of 23 STR loci ($n = 548$)

基因座	等位基因数	H	PIC	DP	PE	P
D10S2325	23	0.837	0.815	0.955	0.860	0.6301
D11S4463	10	0.768	0.722	0.905	0.724	0.173
D12S391	12	0.845	0.820	0.955	0.855	0.6883
D13S305	16	0.832	0.806	0.949	0.838	0.2321
D16S3391	9	0.792	0.775	0.934	0.796	0.8283
D18S1364	10	0.845	0.816	0.953	0.846	0.7616
D19S400	11	0.794	0.776	0.937	0.813	0.4041
D1S2142	7	0.792	0.749	0.921	0.764	0.9114
D1S549	9	0.772	0.748	0.920	0.764	0.3281
D20S161	8	0.763	0.734	0.913	0.752	0.5123
D20S85	9	0.754	0.723	0.910	0.752	0.2897
D21S2055	21	0.885	0.864	0.973	0.911	0.5406
D22S1045	10	0.772	0.728	0.906	0.722	0.4319
D2S1360	13	0.611	0.577	0.815	0.594	0.8615
D2S1776	9	0.755	0.704	0.895	0.709	0.1836
D2S441	12	0.790	0.752	0.922	0.769	0.6277
D3S1545	12	0.774	0.726	0.907	0.730	0.2043
D3S1744	9	0.797	0.797	0.945	0.830	0.4183
D3S2459	11	0.796	0.777	0.934	0.794	0.8106
D4S2408	6	0.750	0.702	0.891	0.686	0.3683
D6S1017	7	0.703	0.671	0.872	0.651	0.3375
D7S1517	16	0.859	0.862	0.971	0.901	0.1698
D9S2157	10	0.796	0.755	0.923	0.770	0.6738

H: 杂合度; PIC: 多态信息含量; DP: 个体识别力; PE: 非父排除率; P: Hardy-Weinberg 不平衡精确检验的概率值, $P > 0.05$ 表明该基因座符合 Hardy-Weinberg 平衡。

结构在遗传过程中易产生变异, 法医领域广泛使用的 STR 基因座突变率一般在 0.2% 左右, 个别基因座突变率高达 0.5%^[14]。由于 STR 基因座遗传过程中突变的存在, 在亲权鉴定时检测到不符合孟德尔遗传定律的基因座数量小于等于 2 个, 一般需要增加检测 STR 基因座的数量, 而不是立即作出排除结论^[15]。在全同胞关系鉴定、半同胞关系鉴定或其它缺失亲本的复杂亲缘关系鉴定中, 需要检测更多数量的基因座, 对检测基因座的性能要求更高^[4, 16]。

本研究中中国北方汉族人群的 23 个常染色体 STR 基因座除 D2S1360 和 D6S1017 外, PIC 值均大于 0.7, DP 值均大于 0.89, H 值均大于 0.75。23 个 STR 基因座 PE 值均大于 0.50, DP 值均大于 0.8, $CDP=1-3.663 \times 10^{-27}$, $CPE=1-2.668 \times 10^{-16}$ 。通常当 $DP \geq 0.8$ 、 $PE \geq 0.5$ 时说明标记物属于具有高度多态性的遗传标记, 其应用价值较高^[17]。同时, 23 个 STR 基因座均符合哈迪-

温伯格平衡($P > 0.05$), 与规范中要求的 19 个必检基因座不存在连锁不平衡(D12S391 包含在必检基因座中), 且平均个体识别力为 0.922, 联合苏州阅微基因技术有限公司的 Microreader™ 23sp ID System 与 PowerPlex® 21 System 系统(美国 Promega 公司)使用, 可使检测基因座数目达到 60 个, 增加 AGCU 21+1 STR 荧光检测系统(无锡中德美联生物技术有限公司), 基因座数目达到 73 个, 能够满足全同胞与半同胞规范中对补充基因座的要求。

本文统计分析的数据结果, 是对目前常染色体 STR 检测系统的有效补充, 大大增加了可用的常染色体 STR 基因座数量, 显著提高了检测体系的效能, 提供了一些北方汉族人群常染色体遗传学数据, 对丰富法医学 STR 分型数据库也具有一定的实际意义, 为 Microreader™ 23HS Plex ID System 应用于全同胞关系鉴定和半同胞关系鉴定提供了可靠的数据支持。

[参考文献]

- [1] 苏永东, 万洪静, 陈冲, 等. 云南地区汉族群体 21 个常染色体短串联重复序列基因座遗传多样性及群体遗传学关系 [J]. 海南医学, 2023, 34(19): 2749-2754.
- [2] 李成涛, 刘希玲, 孙宏钰, 等. SF/T 0117-2021, 生物学全同胞关系鉴定技术规范 [S]. 中华人民共和国司法部.
- [3] 李成涛, 孙宏钰, 张素华, 等. SF/T 0131-2023, 生物学半同胞关系鉴定技术规范 [S]. 中华人民共和国司法部.
- [4] 谭政, 马冠车, 付丽红, 等. 生物学半同胞关系鉴定策略 [J]. 法医学杂志, 2023, 39(3): 262-270.
- [5] 宋天福, 朱洋, 郑冰洁. 浙江汉族人群新常染色体 STR 基因座的遗传多态性分析 [J]. 百科论坛电子杂志, 2021(22): 2806-2808.
- [6] Shi M, Yu X, Bai R, et al. Genetic polymorphism of 14 non-CODIS STR loci for forensic use in southeast China population [J]. *Forensic Science International*, 2008, 174(1): 77-80.
- [7] 吴微微, 刘冰, 郝宏蕾, 等. 中国 28 个省/区汉族人群 41 个 STR 基因座多态性数据分析 [J]. 中国法医学杂志, 2016, 31(1): 27-32.
- [8] 吴谨, 李英碧, 侯一平, 等. 成都汉族群体七个短串联重复序列基因座的遗传多态性和法医学应用研究 [J]. 中华医学遗传学杂志, 2006, 23(2): 230-233.
- [9] 白雪, 丛斌, 李淑瑾, 等. 分子克隆技术制备 miniSTR 基因座等位基因分型标准物 [J]. 法医学杂志, 2009, 25(2): 106-108.
- [10] 曹枝, 滕少康, 黄世宁, 等. 广西壮族 D6S477 等四个 STR 基因座的遗传多态性 [J]. 解剖科学进展, 2007, 13(2): 134-137.
- [11] 梁淑桢, 陈俊超. 广东省韶关地区汉族人群 20 个 STR 基因座的遗传多态性 [J]. *广东医科大学学报*, 2020, 38(02): 153-156+159.
- [12] 侯一平. 法医 DNA 分析的科学证据意义 [J]. 中国法医学杂志, 2001, 16(2): 118-120.
- [13] 申琴, 杨红旗, 袁朝晖, 等. DNA 鉴定技术在法医学物证学中的现状和展望 [J]. 生物技术世界, 2015, 10: 252-253.
- [14] 蔡颖, 周广彪, 赵书民, 等. 中国人群亲权鉴定常用 STR 基因座平均突变率的估计 [J]. 中国司法鉴定, 2010, 52(05): 56-59.
- [15] 伍新尧, 童大跃, 朱运良, 等. 用 STR 分型技术作亲权鉴定时判断标准的研究 [J]. 中山大学学报(医学科学版), 2010, 31(01): 1-6.
- [16] 李海霞, 李建金, 李燃, 等. 两个已知全同胞参与的同胞鉴定分析 [J]. 中国法医学杂志, 2019, 34(02): 159-164.
- [17] Budowle B, Sajantila A. Short tandem repeats-how microsatellites became the currency of forensic genetics [J]. *Nat Rev Genet*, 2024, 25(7): 450-451.