

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¹

(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 GoldeneyeDNA 身份鉴定系统 20A 对 9025 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 0.6161~0.9163, 匹配概率为 0.0128~0.2026, 个体识别率为 0.7974~0.9872, 非父排除率为 0.3108~0.8288, 多态信息含量为 0.5617~0.9136, 累积个体识别率为 0.999999999999999984341, 累积非父排除率为 0.999999989。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 **【文章编号】:** 1004-5619 (2018) 06-0650-06

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 GoldeneyeDNA 身份鉴定系统 20A 对江苏汉族九千余份样本展开 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

¹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

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 10 μL 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 GeneScan3.2 软件进行分型结果分析。

1.3 统计学分析

联合应用 PowerStatsv12 软件和 SPSS19.0 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 Arlequin3.5.2.2 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, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]两两之间的遗传分化系数(F_{st})和 P 值进行计算。采用 MEGA7 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 0.99999999999999984341, CPE 为 0.999999989。

2.2 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。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)。

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。

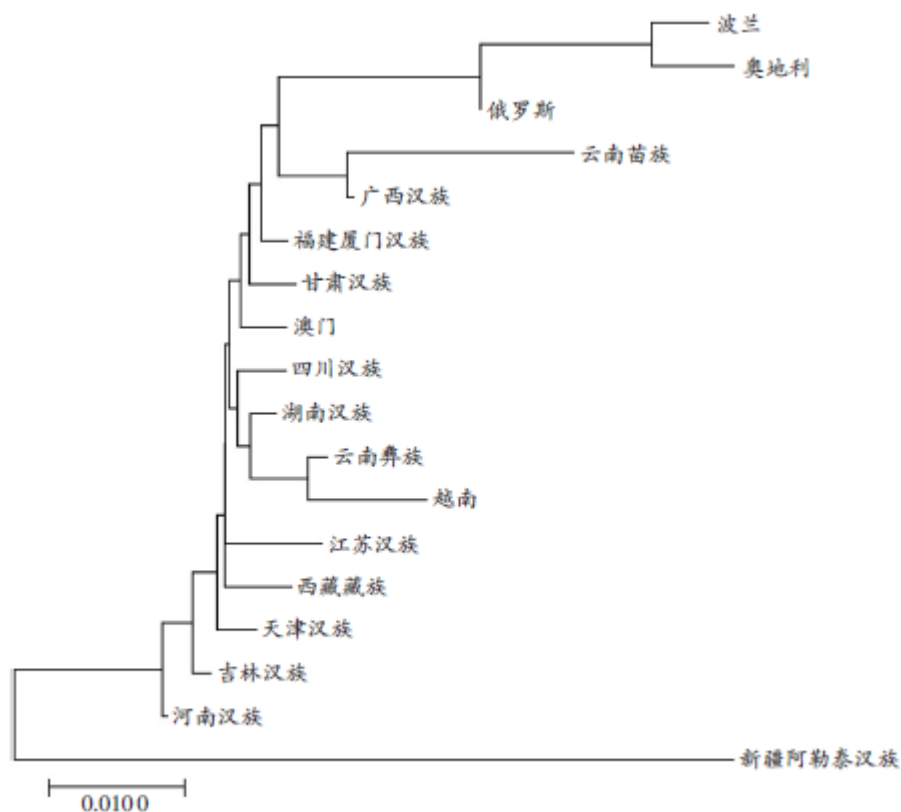


图1 邻接法分析18个人群间的遗传进化距离

3 讨论

本研究采用等位基因频率、 H 、 P_m 、 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 0.6161 (TPOX) 到 0.9163 (PentaE), 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]。

P_m 值从 0.0128 (PentaE) 到 0.2026 (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 0.5617 (TPOX) 到 0.9136 (PentaE), 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 0.7974 (TPOX) 到 0.9872 (PentaE), PE 值从 0.3108 (TPOX) 到 0.8288 (PentaE)。

各基因座之间相互独立, 属于连锁平衡状态, CDP 为 0.99999999999999984341, CPE 为 0.999999989。其中, TPOX (0.3108)、TH01 (0.3424)、D3S1358 (0.4567) 和 CSF1PO (0.4828) 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。

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 5 141 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-DÍZ P, et al. STR data for the AmpFSTR profiler plus loci from Macau (China) [J]. *Forensic Sci Int*, 2001, 123 (1): 74-75.
- [12] PARYS-PROSZEK A, KUPIEC T, WOLAŃSKANOWAK 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 (AmpFSTR 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.