

·基础研究·

高变异Y染色体短串联重复片段检验体系的建立及对北方汉族群体调查

李运丽^{1,2}, 宋诚诚^{1,2}, 郝世诚^{1,2}, 张金佩^{1,2}, 刘岩^{1,2}, 徐妍^{1,2}, 袁丽^{1,2,3}

(1. 中国政法大学证据科学教育部重点实验室, 北京 100088; 2. 司法文明协同创新中心, 北京 100088; 3. 上海市法医学重点实验室//司法部司法鉴定重点实验室(司法鉴定科学研究院), 上海 200063)

摘要:【目的】建立一套五色荧光复合扩增体系,并调查中国北方汉族遗传多态性和父子家系突变情况。【方法】根据文献筛查出在中国某一群体中突变率为高变异和快速变异的Y-STR基因座,建立一套五色荧光Y-STR复合扩增体系,计算在500个北方汉族无关男性个体中的基因多样性,观察每个样本的Y-STR单倍型,并对中国北方汉族500个父子对进行检测,统计父子间的突变情况。【结果】成功构建了包含21个Y-STR基因座复合扩增检验体系,在北方汉族群体中基因多态性在0.4023~0.9904之间,500个无关个体中未发现有相同的单倍型。在500个父子对中观察到21个Y-STR基因座共发生134次突变,其中突变率最高的基因座是DYF399S1,为 7.40×10^{-2} ;在DYS612、DYS547、DYS627、DYS526b、DYS576、DYS630、DYS449、DYF404S1、DYS390、DYS570和DYS518基因座上突变率 $\geq 1.00 \times 10^{-2}$ 。此外,DYS626、DYS458和DYS439基因座突变率 $\geq 4.00 \times 10^{-3}$ 。有97个父子家系出现1个基因座突变,16个家系同时出现2个基因座突变,仅有1个家系同时出现3个基因座突变,家系突变的概率为0.268。其中127次为一步突变,7次为两步突变,符合逐步突变模式。【结论】高变异Y-STR基因座能够有效区分具有父系亲缘关系的男性个体,可作为联合DNA索引系统(combined DNA index system, CODIS)常染色体和常规低突变Y-STR系统基因座的补充。

关键词: Y染色体;短串联重复序列;复合扩增;单倍型;突变

中图分类号:R441.1

文献标志码:A

文章编号:1672-3554(2021)04-0550-07

DOI: 10.13471/j.cnki.j.sun.yat-sen.univ(med.sci).2021.0410

Establishment of High Mutation Y-Chromosome Short Tandem Repeat Multiplex Amplification System and Investigation on Han Nationality in Northern China

LI Yun-li^{1,2}, SONG Cheng-cheng^{1,2}, HAO Shi-cheng^{1,2}, ZHANG Jin-pei^{1,2}, LIU Yan^{1,2}, XU Yan^{1,2},
YUAN Li^{1,2,3}

[1. Key Laboratory of Evidence Science of Ministry of Education, China University of Political Science and Law, Beijing 100088, China; 2. Collaborative Innovation Center of Judicial Civilization, Beijing 100088, China; 3. Shanghai Key Lab of Forensic Medicine // Key Lab of Forensic Science, Ministry of Justice (Academy of Forensic Science), Shanghai 200063, China]

Correspondence to: YUAN Li; E-mail: yuanliwcy@126.com

Abstract:【Objective】To establish a five-color fluorescence Y-chromosome short tandem repeats (Y-STRs) multiplex amplification system and investigate the genetic polymorphisms and paternal family mutations in the Chinese Northern Han population.【Methods】A total of twenty-one highly and rapidly mutating Y-STR loci were selected according to previ-

收稿日期:2021-03-02

基金项目:中国政法大学“双一流”建设项目(1000-10321021);上海市法医学重点实验室、司法部司法鉴定重点实验室(司法鉴定科学研究院)开放基金(KF202111)

作者简介:李运丽,硕士生,研究方向:法医DNA鉴定,E-mail:470322773@qq.com;袁丽,通信作者,教授,主任法医师,研究方向:法医学鉴定,E-mail:yuanliwcy@126.com

ous studies to establish a five-color fluorescence Y-STR multiplex amplification system. Genetic diversity and mutation rate were analyzed in 500 unrelated male individuals or 500 paternal (father and son) pairs separately in the Chinese Northern Han population.【Results】A multiplex amplification system containing 21 Y-STR loci was successfully established. The genetic diversity of each Y-STR ranged from 0.402 3 to 0.990 4, and no identical haplotypes were found in the 500 unrelated individuals. The mutation rate was 0.268, and one hundred and thirty-four mutations were observed in the 21 Y-STR loci among the 500 paternal pairs, of which the highest mutation rate was 7.40×10^{-2} at DYF399S1. The mutation rate of the loci, including DYS612, DYS547, DYS627, DYS526b, DYS576, DYS630, DYS449, DYF404S1, DYS390, DYS570 and DYS518, was $\geq 1.00 \times 10^{-2}$. Further, the mutation rate of DYS626, DYS458 and DYS439 loci was $\geq 4.00 \times 10^{-3}$. A total of ninety-seven father and son pairs were one locus mutation, sixteen pairs had two loci mutations, and only one pair had three loci mutations at the same time. A total of one hundred and twenty-seven cases were one-step mutations and seven cases were two-step mutations, which were consistent with the stepwise mutation pattern.【Conclusion】The rapidly mutating Y-STRs could effectively distinguish male individuals with paternal kinship and could be used as a complement to the Combined DNA Index System (CODIS) autosomal STRs and general Y-STRs.

Key words: Y chromosome; short tandem repeats; multiplex amplification; haplotype; mutation

[J SUN Yat-sen Univ (Med Sci), 2021, 42(4): 550-556]

Y染色体上的短串联重复片段(short tandem repeat, STR)是父系遗传关系鉴定和强奸案混合斑中男性个体识别的有效手段,是常染色体遗传标记的重要补充^[1-2]。随着男性家系检测的增多和Y-STR数据库的建立,发现现有的Y-STR试剂由于低突变率常常不能有效区分不同男性家系^[3]。快速变异的Y-STR (rapidly mutating Y-STR, 突变率 $\geq 1.00 \times 10^{-2}$)^[4-5]有助于缩小侦查范围、锁定犯罪嫌疑人,是法医DNA领域新的研究、检测方向。然而,在欧美白人中被定义为快速变异的Y-STR基因座有的在中国人群中突变率并不高,并且个别基因座融入复合扩增难度大^[6]。为了提升Y-STR鉴别男性家系的能力,本研究筛查了既往研究报告中在我国群体突变率高的21个Y-STR基因座,建立五色荧光复合扩增体系,并对北方汉族群体进行调查,为高突变的Y-STR在父子间突变情况提供基础性资料。

1 材料和方法

1.1 样本

北方汉族健康男性无关个体志愿者500人,每人及其儿子组成父子对500对,经知情同意采集外周血,保存在FTA卡上,父子关系均经“母-子-父”或“父-子”亲子鉴定,亲权指数均超过10 000。

1.2 方法

1.2.1 基因座的筛选 根据本实验室以往研究结果^[6],从国外13个RM Y-STR中选取在中国群体中突变率高的DYF399S1、DYF404S1、DYS449、

DYS518、DYS526b、DYS547、DYS570、DYS576、DYS612、DYS626和DYS627 11个基因座,另外,根据文献筛选在中国民族群体突变率 $\geq 4.0 \times 10^{-3}$ 具有高变异的9个Y-STR基因座^[7-10],包括DYS390、DYS439、DYS456、DYS458、DYS460、DYS481、DYS549、DYS552和DYS630基因座。另外,本体系也纳入了常用的DYS391基因座,其在中国南方汉族人群^[7]突变率为 3.3×10^{-3} 。除了DYS481和DYS612系三核苷酸重复序列外,其他均为四核苷酸重复序列(表1)。

1.2.2 引物的设计与合成 在GeneBank中查询所选取的高变异Y-STR基因座的基因序列信息,利用Primer 3软件(<http://bioinfo.ut.ee/primer3/>)设计引物,并利用UCSC(<http://genome.ucsc.edu/>)中的BLAT软件检验引物的特异性。根据各基因座核心序列情况设计其在体系中的位置,每对引物的前导链5'端分别标记6-FAM、HEX、TAMRA、ROX荧光染料,内标选用橙色的ILS550,委托上海生工生物技术有限公司合成引物。

1.2.3 复合扩增体系的建立与引物优化 PCR体系中使用2×MasterMix(北京金马晟和科技有限公司)缓冲液,内含Taq酶。在离心管中配置10 μL直扩体系,包含:2×MasterMix 4 μL,引物1~5 μL,用去离子水补足至10 μL。用0.5 mm的打孔器对打孔1片加入各个离心管中。PCR热循环参数设置为:95℃ 11 min、94℃ 10 s、59℃ 1 min、72℃ 30 s, 27个循环,60℃ 60 min,4℃保持,在9700扩增仪(AB公司)上进行PCR反应。在PCR扩增体系分

表 1 筛选的 Y-STR 基因座及其基本信息
Table 1 Selected Y-STR loci and their basic information

Locus	Core repeat motif	Position
DYS390	((TCTG)8(TCTA)9-14(TCTG)1(TCTA)4	Yq11.221
DYS391	(TCTA)6-14	Yq11.21
DYF399S1	(GAAA)3N7-8(GAAA)10-23	Yq11.23
DYF404S1	(TTTC)10-20N42(TTTC)3	Yq11.23
DYS439	(GATA)3N32(GATA)5-19	Yq11.21
DYS449	(TTCT)13-19N22(TTCT)3N12 (TTCT)13-19	Yp11.2
DYS456	(AGAT)13-18	Yp11.2
DYS458	(GAAA)11-24	Yp11.2
DYS460	(TAGA)8-13	Yq11.222
DYS481	(CTT)22-32	Yp11.2
DYS518	(AAAG)3(GAAG)1(AAAG)14-22 (GGAG)1 (AAAG)4N6 (AAAG)11-19N27(AAGG)4	Yq11.221
DYS526b	(CCCT)3N20(CTTT)11-17(CCTT)6-10N113 (CCTT)10-17	Yp11.2
DYS547	(CCTT)9-13T(CTTC)4-5N56(TTTC)10-22N10 (CCTT)4(TCTC)1(TTTC)9-16N14 (TTTC)3	Yq11.221
DYS549	(GATA)9-15	Yq11.222
DYS552	(TCTA)3(TCTG)1(TCTA)11-16N40 (TCTA)7-12	Yq11.21
DYS570	(TTTC)14-24	Yp11.2
DYS576	(AAAG)13-22	Yp11.2
DYS612	(CCT)5(CTT)1(TCT)4(CCT)1 (TCT)19-31	Yq11.21
DYS626	(GAAA)14-23N24 (GAAA)3N6 (GAAA)5(AAA)1(GAAA)2-3 (GAAG)1(GAAA)3	Yq11.223
DYS627	(AGAA)3N16(AGAG)3 (AAAG)12-24N82 (AAGG)3	Yp11.2
DYS630	(AAAG)4(AGAG)3N18 (AAAG)12-21	Yq11.222

The bold type loci were from 13 RM Y-STR loci.

别设置各个基因座的引物浓度,进行引物浓度优化,使基因座之间的等位基因峰值达到平衡。

1.2.4 扩增片段的分离与检测 PCR 扩增产物采用 ABI 3130 型基因分析仪进行毛细管电泳分离。PCR 扩增产物 1 μL 、ILS 500 内标 0.15 μL 、Hi-Di™ Formamide 10 μL 混匀,加入自动进样盘,编制进样表,采用 GM 模式电泳。电压 15 000 V、进样时间 10 s,电泳 30 min。电泳数据应用 GeneMapper® ID-X 1.4 软件分析,读取等位基因的片段大小。

1.2.5 北方汉族群体及家系突变情况调查 采用本研究建立的 21 个 Y-STR 基因座的五色荧光复合扩增体系,对 500 对中国北方汉族父子样本进行检测。直接计数法统计突变个数,用突变个数除以父子对数获得突变率,并利用二项概率分布计算 95% 的置信区间^[11]。在 500 名无关男性样本群体中,用算式(Gene Diversity, GD) $=n(1-\sum Pi^2) / (n-1)$;Pi 表

示等位基因频率, n 代表样本数。计算各基因座的基因多态性,观察每个样本是否存在相同的 Y-STR 单倍型。

2 结果

2.1 复合扩增体系的建立

本文成功设计了 21 对引物,建立了 1 个五色荧光复合扩增体系。在 10 μL 直接扩增 PCR 反应体系中,各基因座引物浓度优化结果为:2.5 $\mu\text{mol/L}$ 各基因座引物在 0.18~0.50 μL 之间。根据优化结果,21 个 Y-STR 基因座均得到有效扩增(图 1),扩增片段长度在 90~370 bp 之间,电泳图谱显示扩增效果好,各基因座之间及多拷贝等位基因之间的峰高基本平衡。除 DYS612 为三核苷酸重复、stutter 峰略高外,其他基因座 stutter 峰低于主峰的 10%。根据

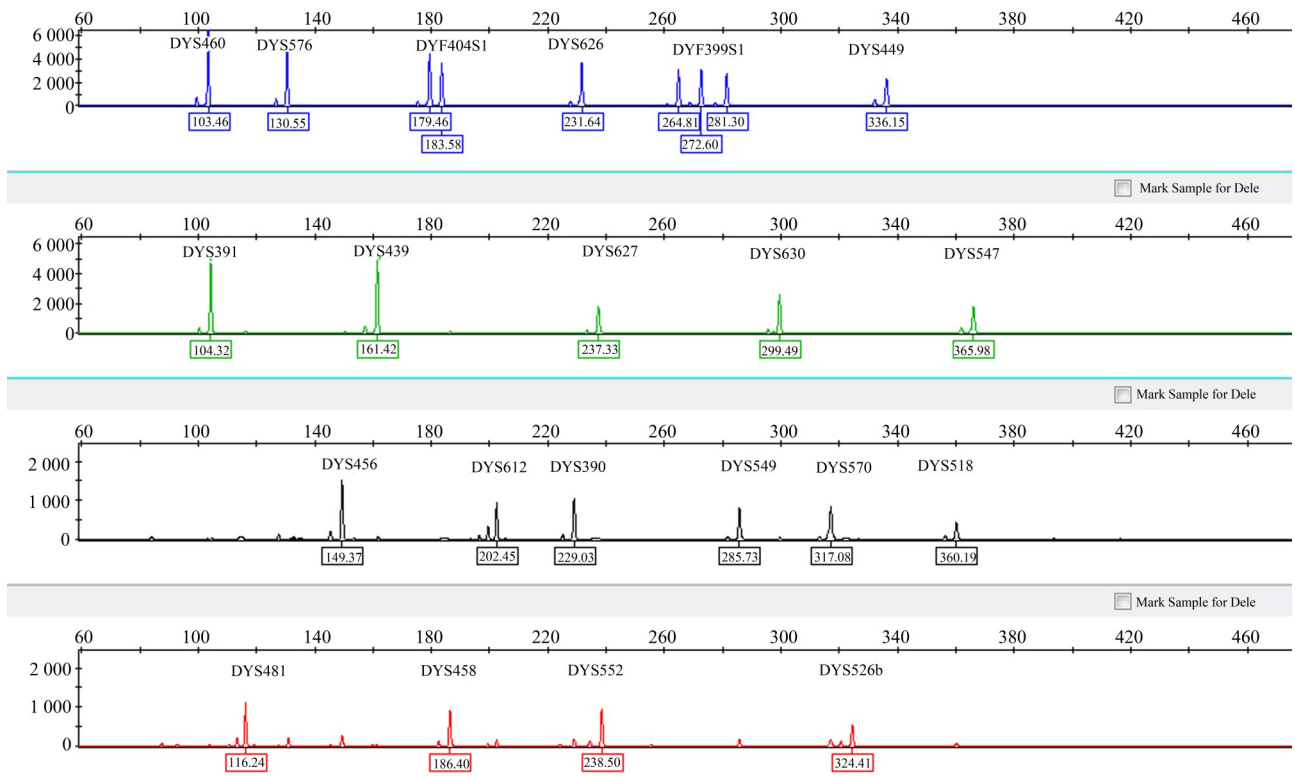


图1 21个Y-STR基因座复合扩增优化图谱

Fig. 1 A genotyping of a sample with 21 Y-STR Optimized amplification system

DYS526a/b 序列特点,设计一对引物仅扩增 DYS526b 基因座(图2)。

2.2 基因多态性结果

在500个北方汉族无关个体中,本次建立的21个Y-STR检测体系,未发现相同的单倍型。基因多态性结果请见表2。在一些多拷贝Y-STR基因座上发现异常分型,见图3-5。

2.3 父子突变率调查结果

500个父子对中观察到16对出现2个基因座的突变,1对出现3个基因座的突变,未发现超过3个基因座同时出现突变的情况。其他具体突变的情况见表2。

3 讨论

Ballantyne 等^[12]通过系列研究表明13个RM Y-STR组合(DYS570、DYS576、DYS518、DYS526a/b、DYS626、DYS627、DYS449、DYS547、DYS612、DYF387S1、DYF399S1、DYF403S1 a/b 和 DYF404S1)无论在多态性还是个体识别能力方面均明显强于当前常用的17个Y-STR的Yfiler系统(美国AB公司)。我们前期的研究中发现:①

DYF387S1、DYF403S1b、DYS526a 在中国汉族群体突变率没有到达快速变异基因座的标准;② DYF403S1a/b 基因座单一引物扩增效率高、没有杂峰,但只要与其它基因座复合扩增,产生很多非特异性产物,难以纳入20多个基因座的复合检测体系中;③ Ballantyne 等^[12]公布的DYS526a/b引物,其中一条引物能与DNA模板的两段序列结合(DYS526a引物与模板有两个碱基的错配),产生两个最少相差180bp的等位基因,由于错配和大片段扩增缺乏优势的原因,DYS526a和DYS526b扩增效率都不高。鉴于DYS526b基因座序列包含了DYS526a基因座,在本次体系构建中我们设计引物仅能检测DYS526b基因座(图2),可提高其在复合扩增体系中扩增效率。本次纳入复合扩增体系的基因座既要优选在中国群体突变率高的,又要考虑复合扩增效率和减少非特异性峰的影响,最终建立的21个Y-STR五色荧光复合扩增体系,每个基因座均获得了准确分型,复合体系扩增DNA分型结果稳定、均衡,具有种属特异性。

法医学应用的Y-STR基因座一般为单拷贝,但有少部分Y-STR基因座为多拷贝。其原因是因为在Y染色体重复回文区域有重复拷贝存在^[13]或

一对引物能与不同序列结合^[14]产生多个等位基因。本体系中 DYF404S1 为二拷贝基因座, DYF399S1 为三拷贝基因座, 相对于单拷贝 Y-STR 基因座, 多拷贝 Y-STR 基因座具有较高的遗传多态性, 存在多个等位基因, 有助于法医学个人识别和亲缘关系鉴定。然而, DYF404S1 和 DYF399S1 会出现异常分型, 在 500 个无关男性样本的 DNA 分型图谱中, DYF404S1 基因座上出现三等位基因 14 例(图 3), DYF399S1 基因座出现四等位基因 10 例(图 4)和五等位基因 2 例(图 5), 且有 5 个样本在这两个基因座上同时出现稀有拷贝数等位基因分型。出现异常多等位基因 DNA 分型提示在混合斑鉴定的案件中, 对于嫌疑人数目的确定会存在困难, 需结合其他检验结果综合分析。

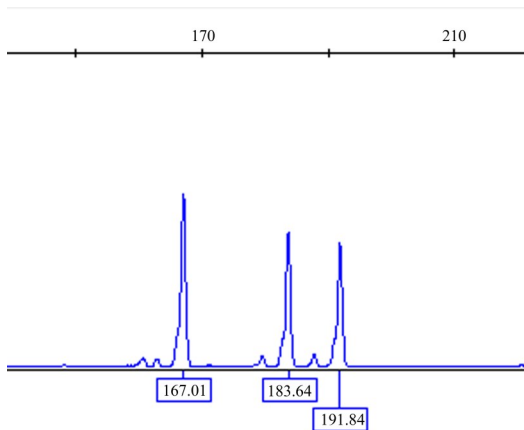


图 3 DYF404S1 上出现三等位基因分型

Fig. 3 Distinctive tri-allelic pattern in DYF404S1

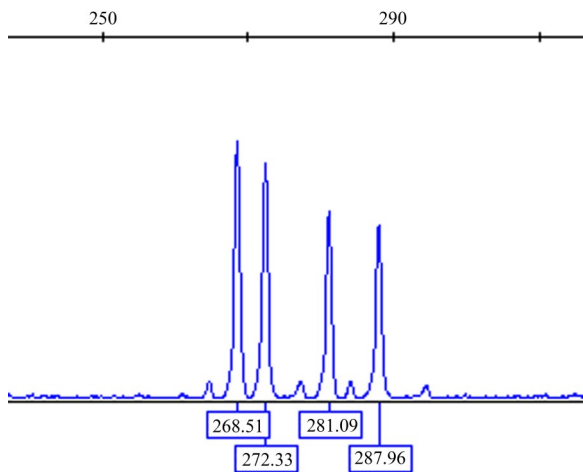


图 4 DYSF399S1 上出现四等位基因分型

Fig. 4 Distinctive tetra-allelic pattern in DYSF399S1

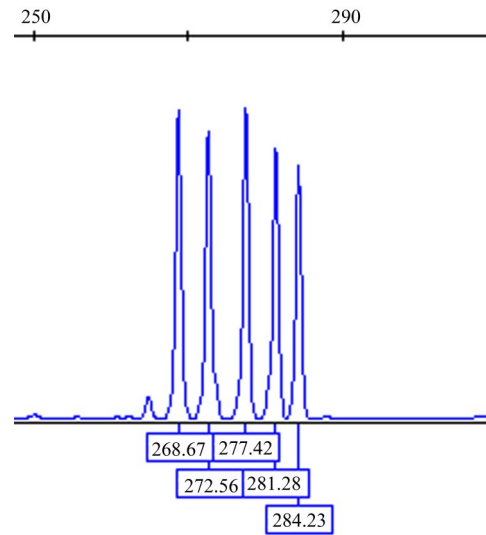


图 5 DYSF399S1 上出现五等位基因分型

Fig. 5 Distinctive five-allelic pattern in DYSF399S1

快速变异的 Y-STR 基因座核心序列结构复杂、重复次数多, 需要更多的空间来容纳这些基因座, 故而本文建立了五色荧光复合扩增体系, 使用了 6-FAM、HEX、TAMRA、ROX 和 ILS550 荧光染料组合。根据在北方汉族人群中的调查结果, 显示 21 个 Y-STR 基因座的基因多态性除 DYS391 为 0.402 3 外, 其余 20 个 Y-STR 在 0.611 8 以上, 在 500 个无关男性个体中, 未观察到有相同单倍型的情况出现。在 500 个父子对中大部分基因座的突变率远高于常染色体 STR 及目前法医学常用的 Y-STR 基因座, 突变率最高达到 7.4×10^{-2} , 发生在 DYF399S1 基因座上。观察 500 对北方汉族父子的 21 个 Y-STR 基因座 DNA 分型图谱, 共出现 134 次等位基因突变, 其中一步突变出现 127 次, 两步突变出现 7 次, 符合逐步突变机制。增加突变与减少突变数目相同, 均为 67 次。本次研究突变率 $\geq 1.00 \times 10^{-2}$ 的 12 个 Y-STR 基因座中有 10 个基因座属于复合或复杂核心序列, 核心序列重复次数较多, 体现了基因座的突变率受重复序列结构和等位基因重复次数的影响, 重复序列结构复杂度越高、平均重复次数越多的基因座, 突变率相对更高^[15-16]。Ballantyne 等^[4]提出重复区毗邻的不可变序列长度对基因座的突变率有一定影响, 不可变序列越长则会增加 DNA 合成时复制滑脱的机会, 本次研究中重复区毗邻的不可变序列长度大于 30 bp 的基因座有 9 个 (DYF404S1、DYS439、DYS449、DYS518、DYS526b、DYS547、DYS552、DYS626 和 DYS627),

其中突变率 $\geq 6.00 \times 10^{-3}$ 有8个,验证了上述理论。DYS630和DYS390基因座虽不在Ballantyne等列出的RM Y-STR之中,但在北方汉族群体中突变率达到快速变异标准。本研究在DYS391、DYS456基因座上未发生突变,而在南方汉族群体^[7]中的突变率为 3.3×10^{-3} 和 4.4×10^{-3} 。另外,DYS549、DYS481、DYS552、DYS460突变率接近Y-STR的平均突变率,为 2.0×10^{-3} ,分析可能是由于本次调查父子对样本量的限制,同时也体现了Y-STR突变率受不同地域、人群的差异的影响^[17]。

应用这些高变异Y-STR基因座进行法医学亲子鉴定及个人识别,更易于区分具有父系亲缘关系的男性个体,有助于解决目前法医学应用Y-STR基因座只能排除不能认定的瓶颈。但需要注意的是,由于高变异Y-STR突变率较高,在亲子鉴定中对有父系亲缘关系的男性个体容易错误排除,故不能完全取代现有Y-STR试剂盒,但可以作为CODIS常染色体和常规低突变Y-STR系统基因座的补充,具有良好的法医应用价值。

参考文献

- [1] Parson W, Niederstätter H, Lindinger A, et al. Y-STR analysis on DNA mixture samples—results of a collaborative project of the ENFSI DNA working group[J]. *Forensic Sci Int Genet*, 2008, 2(3): 238–242.
- [2] Budowle B, Ge J, Low J, et al. The effects of Asian population substructure on Y-STR forensic analyses[J]. *Leg Med (Tokyo)*, 2009, 11(2): 64–69.
- [3] Heyer E, Puymirat J, Dieltjes P, et al. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees[J]. *Hum Mol Genet*, 1997, 6(5): 799–803.
- [4] Ballantyne KN, Goedbloed M, Fang R, et al. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications[J]. *Am J Hum Genet*, 2010, 87(3): 341–353.
- [5] Boattini A, Sarno S, Bini C, et al. Mutation rates and discriminating power for 13 rapidly-mutating Y-STRs between related and unrelated individuals[J]. *PLoS One*, 2016, 11(11): e0165678.
- [6] Yuan L, Chen W, Zhao D, et al. Mutation analysis of 13 RM Y-STR loci in Han population from Beijing of China[J]. *Int J Legal Med*, 2019, 133(1): 59–63.
- [7] 王瑛, 张楚楚, 李燃, 等. 中国南方汉族群体27个Y-STR基因座的遗传多态性及突变研究[J]. *中山大学学报(医学科学版)*, 2015, 36(5): 650–656.
Wang Y, Zhang CC, Li R, et al. Analysis of genetic polymorphisms and mutations at 27 Y-STR loci in southern Chinese Han population[J]. *J Sun Yat-sen Univ (Med Sci)*, 2015, 36(5): 650–656.
- [8] 吴微微, 王怀锋, 郝宏蕾, 等. 中国汉族人群46个Y-STR基因座多态性与突变调查[J]. *中国法医学杂志*, 2015, 30(3): 256–259; 264.
Wu WW, Wang HF, Hao HL, et al. An analysis of genetic polymorphisms and mutation at 46 Y-STR loci in Chinese Han population[J]. *Chin J Forensic Med*, 2015, 30(3): 256–259; 264.
- [9] 翁玮霞. 17个Y-STR基因座单倍型调查、突变研究与法医学应用[D]. 南方医科大学, 2013: 30.
Weng WX. The haplotype frequency, mutation research and forensic applications of 17 Y-STR loci[D]. Southern Medical University, 2013: 30.
- [10] Lang M, Ye Y, Li JN, et al. Comprehensive mutation analysis of 53 Y-STR markers in father-son pairs[J]. *Forensic Science International: Genetics Supplement Series*, 2017, 6: e57–e58.
- [11] Exact Binomial and Poisson Confidence Intervals (Revised 05/25/2009) [2021-3-1]. <http://statpages.org/confint.html>.
- [12] Ballantyne KN, Keerl V, Wollstein A, et al. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages[J]. *Forensic Sci Int Genet*, 2012, 6(2): 208–218.
- [13] Diederich M, Martín P, Amorim A, et al. A case of double alleles at three Y-STR loci: forensic implications[J]. *Int J Legal Med*, 2005, 119(4): 223–225.
- [14] 尚蕾, 莫晓婷, 杨帆, 等. 多拷贝Y-STR基因座在法庭科学领域的研究[J]. *刑事技术*, 2018, 43(2): 97–103.
Shang L, Mo XT, Yang F, et al. Progress towards forensic research and application of multi-copied Y-STR loci[J]. *Forensic Science and Technology*, 2018, 43(2): 97–103.
- [15] Brinkmann B, Klitsch M, Neuhuber F, et al. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat[J]. *Am J Hum Genet*, 1998, 62(6): 1408–1415.
- [16] Jochens A, Caliebe A, Rosler U, et al. Empirical evaluation reveals best fit of a logistic mutation model for human Y-chromosomal microsatellites[J]. *Genetics*, 2011, 189(4): 1403–1411.
- [17] Kayser M, Kittler R, Erler A, et al. A comprehensive survey of human Y-chromosomal microsatellites[J]. *Am J Hum Genet*, 2004, 74(6): 1183–1197.

(编辑 祁方昉)