

基于有限突变模型和大规模数据的 19 个常染色体 STR 的实际突变率研究

刘志勇², 任贺³, 陈冲⁴, 张京晶⁵, 张晓梦¹, 石妍⁴, 石林玉¹, 陈滢⁴,
程凤¹, 贾莉⁴, 陈曼⁶, 范庆炜⁷, 张家榕¹, 李万婷¹, 王萌春¹,
任子林⁸, 刘雅诚⁴, 倪铭⁸, 孙宏钰², 严江伟¹

1. 山西医科大学法学院, 太原 030001
2. 中山大学中山医学院法医学系, 广州 510080
3. 北京警察学院, 北京 102202
4. 北京通达首诚司法鉴定所, 北京 100192
5. 北京华彦科技有限公司司法鉴定所, 北京 100192
6. 南方医科大学法学院, 广州 510515
7. 川北医学院法医学系, 南充 637000
8. 北京辐射医学研究所, 北京 100850

摘要: 短串联重复序列(short tandem repeat, STR)已广泛用于法医学亲子鉴定和个体识别中, 但 STR 的突变可能会影响其结果的解释。在大多数类似研究中, 由于忽略“隐性”突变现象, STR 的突变率被低估。鉴于此, 为获得更加准确的 STR 实际突变率, 本研究使用 Slooten 与 Ricciardi 提出的有限突变模型和大规模数据, 对 28,313 例(78,739 个体)中国北京汉族已确认亲生关系的亲子鉴定案的 20 个常染色体 STR 基因座(*D3S1358*、*D1S1656*、*D13S317*、*Penta E*、*D16S539*、*D18S51*、*D2S1338*、*CSF1PO*、*Penta D*、*TH01*、*vWA*、*D21S11*、*D6S1043*、*D7S820*、*D5S818*、*TPOX*、*D8S1179*、*D12S391*、*D19S433* 和 *FGA*); 由于有限突变模型中未包含 *D6S1043* 的校正参数, 因此本文实际计算其余 19 个 STR 基因座的突变率)进行了调查。结果发现, 所有基因座均存在突变现象, 总计发生 1665 个突变事件, 包括 1614 个一步突变, 34 个两步突变, 8 个三步突变和 9 个非整步突变。基因座特异性的平均实际突变率在三联体中为 0.00007700 (*TPOX*)~0.00459050 (*FGA*), 在二联体中为 0.00000000 (*TPOX*)~0.00344850(*FGA*)。此外, 本研究还分析了表面和实际突变率、三联体和二联体突变率、父源和母源的突变率之间的关系。研究表明, 实际突变率多大于表面突变率, 而且 μ_1^*/μ_2^* (表面突变率)的比值通常也大于 μ_1/μ_2 (实际突变率) (μ_1^* , μ_1 ; μ_2^* , μ_2 分别是一步和两步的突变率), 即更多的“隐性”突变被释放出来。而且父源和母源的三联体和二联体的突变率也有存在差异。随后, 将这些突变率数据与已发表的中国其他汉族人口的相关研究进行比较, 展现出了 STR 突变率的时间与区域差异。由于样本量大, 本研究中还报告了一些少见的突变事件, 例如同卵双胞胎突变和“假四步突变”等。综上所述, 本研究通过大量数据获得了接近真实的 STR 突变率的估计值, 不仅可为中国法医 DNA 数据库和群体遗传学数据库提供重要的基础数据, 也对开展法医学

收稿日期: 2021-06-02; 修回日期: 2021-08-18

基金项目: 国家自然科学基金重点项目(编号: 82030058)资助[Supported by the National Natural Science Foundation of China (No. 82030058)]

作者简介: 刘志勇, 博士研究生, 研究方向: 法医遗传学。E-mail: liuzhy255@mail2.sysu.edu.cn

任贺, 硕士, 副教授, 研究方向: 法医遗传学。E-mail: snoopy_fr@sohu.com

刘志勇和任贺并列第一作者。

通讯作者: 严江伟, 博士, 教授, 研究方向: 法医基因组学。E-mail: yanjw@sxmu.edu.cn

DOI: 10.16288/j.ycz.21-197

网络出版时间: 2021/8/31 17:30:44

URI: <https://kns.cnki.net/kcms/detail/11.1913.R.20210831.1400.001.html>

个体识别、亲权鉴定和遗传学研究具有重要的意义。

关键词: 常染色体 STR; 有限突变模型; 亲子鉴定; 中国汉族人群; 突变分析

Actual mutational research of 19 autosomal STRs based on restricted mutation model and big data

Zhiyong Liu², He Ren³, Chong Chen⁴, Jingjing Zhang⁵, Xiaomeng Zhang¹, Yan Shi⁴, Linyu Shi¹, Ying Chen⁴, Feng Cheng¹, Li Jia⁴, Man Chen⁶, Qingwei Fan⁷, Jiarong Zhang¹, Wanting Li¹, Mengchun Wang¹, Zilin Ren⁸, Yacheng Liu⁴, Ming Ni⁸, Hongyu Sun², Jiangwei Yan¹

1. School of Forensic Medicine, Shanxi Medical University, Taiyuan 030001, China

2. Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China

3. Beijing Police College, Beijing 102202, China

4. Beijing Tongda Shoucheng Institute of Forensic Science, Beijing 100192, China

5. Beijing Huayan Judicial Authentication Institute, Beijing 100192, China

6. School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China

7. Faculty of Forensic Medicine, North Sichuan Medical College, Nanchong 637000, China

8. Beijing Institute of Radiation Medicine, Beijing 100850, China

Abstract: Short tandem repeat (STR) markers have been widely used in forensic paternity testing and individual identification, but the STR mutation might impact on the forensic result interpretation. Importantly, the STR mutation rate was underestimated due to ignoring the “hidden” mutation phenomenon in most similar studies. Considering this, we use Slocoten and Ricciardi’s restricted mutation model based on big data to obtain more accurate mutation rates for each marker. In this paper, the mutations of 20 autosomal STRs loci (*D3S1358*, *DIS1656*, *D13S317*, *Penta E*, *D16S539*, *D18S51*, *D2S1338*, *CSF1PO*, *Penta D*, *TH01*, *vWA*, *D21S11*, *D6S1043*, *D7S820*, *D5S818*, *TPOX*, *D8S1179*, *D12S391*, *D19S433*, and *FGA*; The restricted model does not include the correction factor of *D6S1043*, this paper calculates remaining 19 STR loci mutation rates) were investigated in 28,313 (Total: 78,739 individuals) confirmed parentage-testing cases in Chinese Han population. As a result, total 1665 mutations were found in all loci, including 1614 one-steps, 34 two-steps, 8 three-steps, and 9 nonintegral mutations. The loci-specific average mutation rates ranged from 0.00007700 (*TPOX*) to 0.00459050 (*FGA*) in trio’s and 0.00000000 (*TPOX*) to 0.00344850 (*FGA*) in duo’s. We analyzed the relationship between mutation rates of the apparent and actual, the trio’s and duo’s, the paternal and maternal, respectively. The results demonstrated that the actual mutation rates are more than the apparent mostly, and the values of μ_1^*/μ_2^* (apparent) are also greater than μ_1/μ_2 (actual) commonly (μ_1^* , μ_1 ; μ_2^* , μ_2 are the mutation rates of one-step and two-step). Therefore, the “hidden” mutations are identified. In addition, the mutations rates of trio’s and duo’s, the paternal and maternal, exhibit significant difference. Next, those mutation data are used to do a comparison with the studies of other Han populations in China, which present the temporal and regional disparities. Due to the large sample size, some rare mutation events, such as monozygotic (MZ) mutation and “fake four-step mutation”, are also reported in this study. In conclusion, the estimation values of actual mutations are obtained based on big data, they can not only provide basic data for the Chinese forensic DNA and population genetics databases, but also have important significance for the development of forensic individual identification, paternity testing and genetics research.

Keywords: autosomal STR; restricted mutation model; parentage testing; Chinese Han population; mutational analysis

STR (short tandem repeat)是人类基因组中非常丰富的一类遗传标记^[1], 由于其高度的多态性^[2,3], 被广泛应用于个体识别、亲子鉴定和群体遗传学等领域。然而, 在大多数 STR 基因座中都可以观察到突变事件, 一般认为复制滑脱是导致 STR 突变的主要机制^[4]。这些突变现象往往会对法医学实践中证据权重评价产生影响, 比如在计算累积父权指数(cumulative paternity index, CPI)时需要特别考虑突变现象, 过高过低的突变率数据都不利于得出客观结论, 因此获得 STR 基因座的准确突变率是非常重要的。

STR 基因座突变率的实际评估计算中, 往往会受抽样群体大小的影响^[5]。尽管已有研究者发表大量关于 STR 基因座突变率的论文, 但基于较少的样本量, 其 STR 突变率的计算结果可能存在偏差。除了以上因素外, 突变率的不同计算方法也会对 STR 实际突变率计算产生影响。在已报道的多数研究中 STR 突变率一般是通过直接计数法来计算的^[5~8], 并且将三联体和二联体的数据进行了合并。然而, 该方法没有考虑“隐性突变(hidden mutation)”现象, 因为人们普遍认为当发生突变时, 突变步数越少越真实, 但实际并非总是如此, 即某些突变会被隐藏。鉴于通过三联体和二联体的计算得到的 STR 基因座突变率有一定差异, 因此合并三联体和二联体的突变数据可能是不合理的。即, 直接计数法对于 STR 的实际突变率的估计有一定偏差。

针对以上抽样群体尺度与计算方法的问题, 本研究采用了大量亲子鉴定数据和 Slooten 与 Ricciardi^[9]的更为严谨的有限突变模型计算方法。基于 28,313 例已确认亲生关系的中国汉族人群亲子鉴定案例, 分别计算了三联体和二联体两种情形的 19 个常染色体 STR 基因座实际突变率。本文详细比较了表面突变率(直接计数法)和实际突变率(有限突变模型), 并与其他多个群体的 STR 突变率进行了对比分析。最后, 本文还讨论了影响 STR 突变率的几个因素, 比如突变年龄与来源、等位基因长度、基因座杂合度等。

1 材料与方法

1.1 数据来源

本研究涉及的基因座突变数据来自 2009~2019

年间中国北京汉族人群的常规亲子鉴定案例的统计结果, 该数据已获得北京通达首诚司法鉴定所许可使用。所有数据均采用匿名化处理, 不涉及个体相关除突变信息以外的其他 STR 基因座具体分型, 也不涉及家系情况。本次统计分析的亲子鉴定案例的 CPI 值均大于 10,000, 达到亲子鉴定的确认标准(未能确认的案例被除外), 且突变的 STR 基因座已经过多个试剂盒验证。总计 28,313 例(78,739 个体)亲子鉴定数据被纳入, 包括 22,113 例三联体(trio's)和 6200 例二联体(duo's) (父亲-孩子: 2342 例; 母亲-孩子: 3858 例)。该数据涉及的 STR 基因座包括 CODIS 系统(*D3S1358*、*D13S317*、*D16S539*、*D18S51*、*CSF1PO*、*TH01*、*vWA*、*D21S11*、*D7S820*、*D5S818*、*TPOX*、*D8S1179*、*FGA*)和 *D1S1656*、*D6S1043*、*Penta E*、*Penta D*、*D2S1338*、*D12S391*、*D19S433*。由于有限突变模型^[9]中未包含 *D6S1043* 基因座的计算校正参数, 因此本研究中共包括除 *D6S1043* 之外其余 19 个 STR 基因座的突变率计算。

1.2 突变率计算

在已确认亲生关系的亲子鉴定案例中, 当发现 STR 等位基因违反孟德尔遗传定律时, 认为该 STR 基因座发生突变事件^[10]。在计算过程中, 突变步数与来源的判定根据 Brinkmann 等^[10]和 Weber 等^[11]的描述, 本研究中所有可疑的 STR 突变事件均再次通过同一试剂盒和其他试剂盒进行检测, 确保分型正确。所有 STR 基因座的三联体和二联体父源母源突变事件的统计结果见附表 1。在突变率计算过程中, 一方面使用直接计数法公式 $\mu^* = n/N$ 计算表面突变率, 其中 n 是观察到的突变事件的数目, N 为减数分裂的发生次数; 另一方面在得到表面突变率后, 通过概率校正因子矩阵 A 校正得到实际突变率(见方程(I))。其中 $A_{\{k, l\}}$ 代表观测到 k 步突变, 实际是 l 步突变的概率。对于任一 $A_{\{k, l\}}$, Slooten 和 Ricciardi^[9]基于 NFI 的参考数据集(2085 个体)的等位基因型频率, 随机组合一对夫妇, 模拟生成子代的基因型, 其中从母亲遗传一个没有突变的等位基因, 但从父亲遗传得到一个实际有 1 步突变的等位基因, 即已知实际突变步数, 然后通过观察得到表面突变步数。上述过程通过计算机模拟重复 100,000 次, 计算不同突变步数的频率, 进而估计最终突变转移概率

$A_{\{k, 1\}}$ 。该模型假设:(1)当发生不确定来源的突变时,认为父源与母源发生的概率相同,均为 0.5;(2)当观察到的突变事件可以同时解释为相同的突变增加或者突变减少时,该模型方法不予区分。本文采用模型中下列矩阵方程(I)计算实际突变率 μ 。

$$\begin{bmatrix} \mu 0^* \\ \mu 1^* \\ \mu 2^* \\ \vdots \\ \mu k^* \end{bmatrix} = \begin{bmatrix} 1 & A_{1,0} & A_{2,0} & \cdots & A_{k,0} \\ 0 & A_{1,1} & A_{2,1} & \cdots & A_{k,1} \\ 0 & 0 & A_{2,2} & \cdots & A_{k,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & A_{k,k} \end{bmatrix} \cdot \begin{bmatrix} \mu 0 \\ \mu 1 \\ \mu 2 \\ \vdots \\ \mu k \end{bmatrix} \quad \text{方程(I)}$$

在“方程(I)”中, μk^* 为表面为 k 步的突变率, $\mu 0^*$ 为表面不发生突变的概率; μk 代表实际为 k 步的突变率, $\mu 0$ 为实际不发生突变的概率; $A_{k,1}$ 是表面突变步数为 k , 而实际步数为 1 的概率; μ^* 为基因座特异性表面突变率; μ 为基因座特异性实际突变率。为了便于理解该模型, 方程(II)列出了方程(I)的展开式。以 TH01 为例, 统计计算得到了表面突变率 μk^* 后, 再根据 Slooten 和 Ricciardi^[9]提供的矫正矩阵 $A_{\{k, 1\}}$, 可以结合方程(II)反解出实际突变率 μk , 其具体的计算过程见附表 2。本文采用在线计算器 (<https://www.numberempire.com/matrixbinarycalculator.php>) 进行矩阵求逆。

$$\begin{cases} \mu 0 + \mu 1 A_{1,0} + \mu 2 A_{2,0} + \mu 3 A_{3,0} = \mu 0^* \\ \mu 1 A_{1,1} + \mu 2 A_{2,1} + \mu 3 A_{3,1} = \mu 1^* \\ \mu 2 A_{2,2} + \mu 3 A_{3,2} = \mu 2^* \\ \mu 3 A_{3,3} = \mu 3^* \end{cases} \quad \text{方程(II)}$$

2 结果与分析

2.1 突变分析

本项研究中, 在 28,313 例确认的亲亲子鉴定案例(22,113 例三联体和 6200 例二联体)的 20 个常染色体 STR 基因座中共发现了 1665 个突变事件, 所有基因座均观察到突变现象。基于有限突变模型计算得到的三联体和二联体的基因座特异性实际突变率和总突变率数据显示(表 1, 不包含 *D6S1043* 基因座): 三联体基因座特异性平均实际突变率范围为 0.00007700 (*TPOX*)~0.00459050 (*FGA*), 二联体为 0.00000000 (*TPOX*)~0.00344850 (*FGA*); *FGA* 基因座具有较高的

父源突变率(0.00767100)和总突变率(0.00459050), *Penta E* 具有较高的母源突变率(0.00237900), 与 Jin 等^[12]基于直接计数法的研究结果一致。

同卵双胞胎(monozygotic, MZ)是由单个精子和单个卵子结合形成受精卵后, 分裂成两个胚胎形成的, 其发生率约为 1/250^[13]。从理论上讲, 同卵双胞胎拥有相同的基因组, 然而研究者在 STR^[14]、拷贝数变异(CNV)^[15]和单核苷酸多态性(SNP)^[16]中观察到了一些差异。本次研究中, 在 10 对 MZ 中发现了相同突变, 包括 7 例男性和 3 例女性(附表 3), 突变基因座为 *D6S1043* (2 例)、*Penta E* (2 例)、*D18S51* (2 例)、*FGA*、*D21S11*、*D12S391* 和 *CSFIPO*, 且均为一步突变。在已公开发表的论文中, 尚未报道同时在 MZ 中发现不一致的 STR 突变, 但 Wang 等^[17]报道了一对 MZ 中其中一个个体的 *vWA* 基因座出现三等位基因模式的案例。在亲子鉴定实践中, 一般较少会在同一个体观察到多个突变。如果观察到, 其可能会干扰鉴定人得出亲子关系的正确结论。在本次研究的 28,313 例亲子鉴定案件中, 共发现 27 例孩子的基因型中可观察到 2 个突变的事件, 发生率为 0.00095363 (27/28,313, 20 个常染色体 STR 基因座) (附表 4A); 同时也发现 3 例孩子的基因型中可观察到 3 个突变的事件(附表 4B)。在这 30 例的多基因座突变案例中, 本研究使用了总计超过 39 个基因座的多个 STR 试剂盒对亲子关系进行了确认。

2.2 表面与实际突变率的比较

基于直接计数法和有限突变模型, 计算得到每个 STR 基因座的表面突变率(μ^*)与实际突变率(μ), 三联体的相关突变率数据见附表 5, 二联体相关数据见附表 6。在使用有限突变模型的矩阵计算(方程(I))中, 本研究发现有三个负值的实际突变率, 即三联体组的 *D13S317* 和 *D19S433* 基因座的 $M\mu 2$; 二联体组的 *D3S1358* 基因座的 $F\mu 1(M, F$ 分别代表母源和父源突变)。为了查找原因, 本研究分析了 *D13S317* 和 *D19S433* 基因座在三联体中的突变特征, 发现母源突变中可观察到一步和三步突变, 但未观察到两步突变; 同理, 二联体组的 *D3S1358* 基因座仅在父源中观察到两步突变, 但未观察到一步突变。即表明, 实际突变率负值可能是在某 STR 基因座可观察到高步数突变, 但未观察到某个低步数突变的情况

表 1 三联体与二联体 STR 基因座特异性实际突变率

Table 1 The total STR loci-specific actual mutation rates in trio's and duo's

基因座	三联体			二联体		
	F μ	M μ	平均 μ	F μ	M μ	平均 μ
CSF1PO	0.00449100	0.00089800	0.00269450	0.00379500	0.00115200	0.00247350
D12S391	0.00674500	0.00188500	0.00431500	0.00532800	0.00000000	0.00266400
D13S317	0.00197200	0.00089900	0.00143550	0.00310500	0.00094300	0.00202400
D16S539	0.00246100	0.00073900	0.00160000	0.00164200	0.00049800	0.00107000
D18S51	0.00524900	0.00140400	0.00332650	0.00309400	0.00075100	0.00192250
D19S433	0.00222300	0.00123300	0.00172800	0.00158100	0.00096000	0.00127050
D1S1656	0.00264300	0.00120200	0.00192250	0.00195900	0.00000000	0.00097950
D21S11	0.00255700	0.00201239	0.00228470	0.00203300	0.00123400	0.00163350
D2S1338	0.00273800	0.00047700	0.00160750	0.00120300	0.00000000	0.00060150
D3S1358	0.00160539	0.00073400	0.00116970	0.00149600	0.00097800	0.00123700
D5S818	0.00274700	0.00092600	0.00183650	0.00610000	0.00061700	0.00335850
D7S820	0.00325000	0.00050900	0.00187950	0.00149800	0.00045500	0.00097650
D8S1179	0.00443800	0.00081000	0.00262400	0.00368100	0.00000000	0.00184050
FGA	0.00767100	0.00151000	0.00459050	0.00573600	0.00116100	0.00344850
Penta D	0.00182500	0.00049400	0.00115950	0.00243100	0.00048900	0.00146000
Penta E	0.00513100	0.00237900	0.00375500	0.00525600	0.00158700	0.00342150
TH01	0.00022611	0.00010000	0.00016306	0.00077600	0.00047100	0.00062350
TPOX	0.00007700	0.00007700	0.00007700	0.00000000	0.00000000	0.00000000
vWA	0.00600600	0.00108800	0.00354700	0.00074900	0.00090900	0.00082900
Total	0.06405550	0.01937639	0.04171595	0.05146300	0.01220500	0.03183400

“ μ ”表示实际突变率;“F”表示父源;“M”表示母源。

下发生的(即非连续的突变步数现象)。换言之,即这几个负值可能在提示潜在低步数突变应该会被观察到,但实际中并没有观察到,这已经超过了有限突变模型的可解释范畴。因此在数据处理过程中,把三个负值实际突变率设为零。如图 1 所示,在父源和母源突变之间有一个明显的界限,特别是在三联体的一侧(图 1A),这可能提示三联体与二联体的计算得到的 STR 突变率是有差异的,尽管这种差异可能来自于三联体与二联体本身,或者来自于样本量的差异,但可间接说明以往直接合并三联体与二联体数据进行计算存在不合理之处。此外,从该图也可发现在三联体和二联体中,TH01 和 TPOX 基因座的表面和实际突变率都较低。与之相反的是,D12S391 和 FGA 基因座在三联体中的表面和实际突变率均较高,而 D21S11 和 Penta E 基因座在二联体中的表面和实际突变率较高。为了进一步分析有限

突变模型的校正效果,绘制了实际突变率与表面突变率比值图(μ/μ^*) (图 2)。该图表明,除父源的 D3S1358 和 TH01($\mu/\mu^*\approx 1$)和母源的 D21S11 基因座($\mu/\mu^*\approx 1$)外,其余基因座的突变率均得到了较大程度的矫正($\mu/\mu^*>1$)。就矫正程度而言,二联体组的基因座的效应表现得更为明显(与 Slooten 和 Ricciardi^[9]的研究结果相似),尤其是 CSF1PO 和 D5S818 基因座。也就是说有限突变模型通过参数矫正的方式,将更多的隐性的突变挖掘出来。然而,在亲子鉴定过程中,需要计算有突变率参与的父权指数时,目前一般采用直接计数法计算得到平均 STR 突变率(比如男性取 0.002;女性取 0.001~0.0005)来进行。若为了更严谨的计算,可采用三联体和二联体的父源、母源或者 STR 基因座特异性突变率,当然也可以使用本次研究基于有限突变模型计算获得的 STR 基因座特异性实际突变率。从理论上讲,计算单个基因座

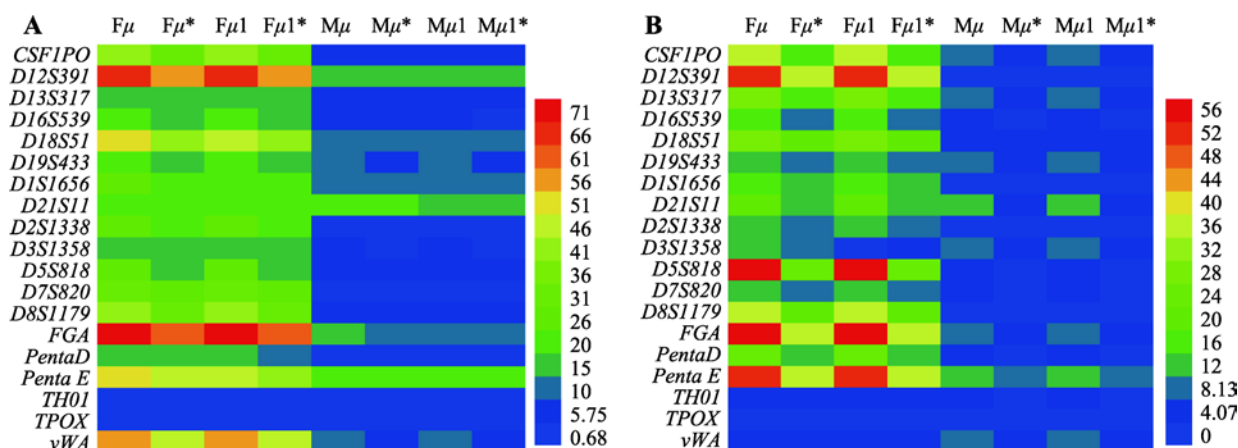


图 1 三联体和二联体表面和实际突变率热图

Fig. 1 The heatmap of apparent and actual mutation rates in trio's and duo's

A: 三联体情形; B: 二联体情形; “ μ^* ”表示表面突变率; “ μ ”表示实际突变率。

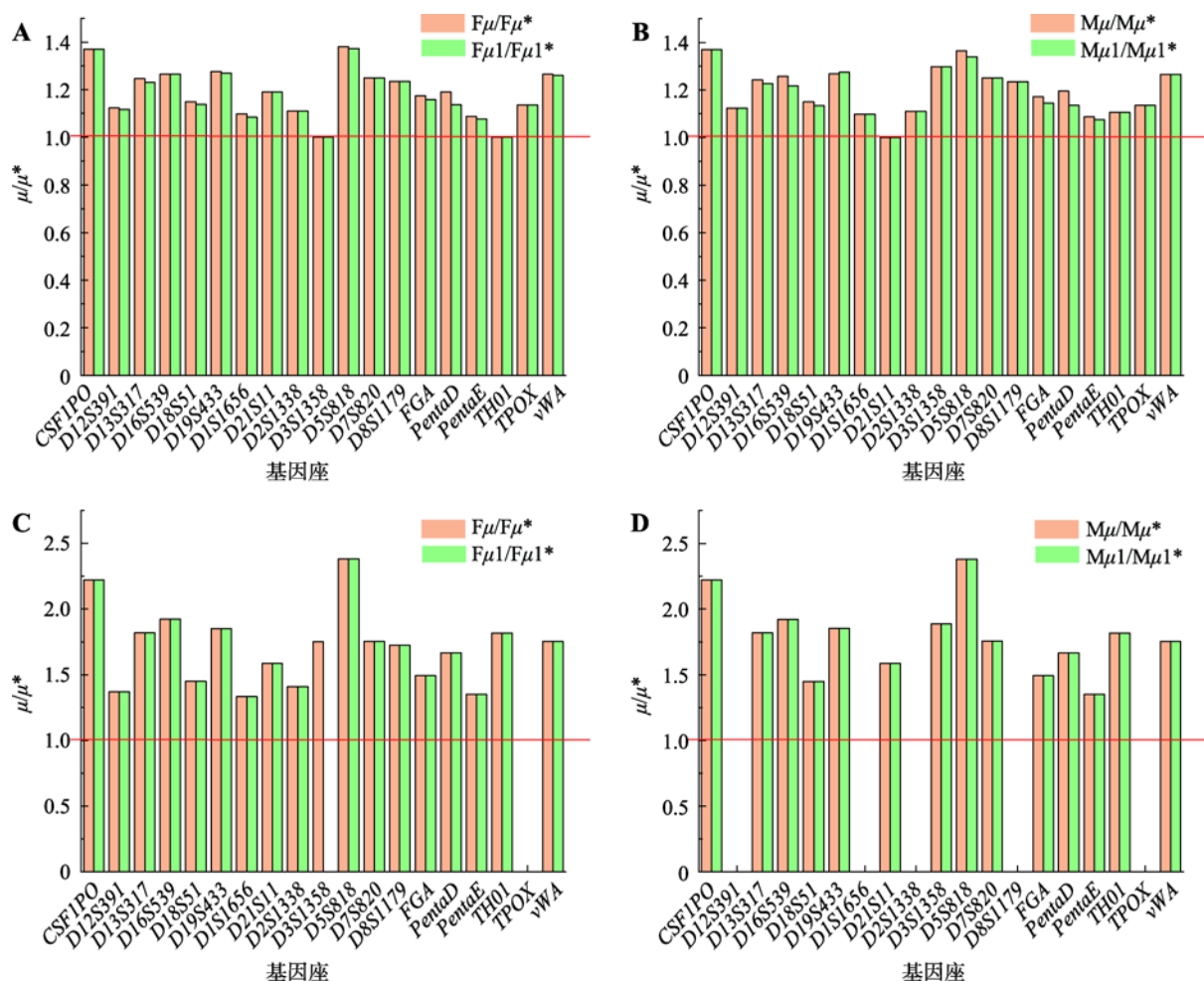


图 2 表面突变率与实际突变率比较

Fig. 2 The comparisons of apparent and actual mutation rates

“A”和“B”代表三联体组父源(F)和母源(M)的 μ/μ^* 比值;“C”和“D”分别代表二联体组父源(F)和母源(M)的 μ/μ^* 比值;空白处表示比式的分母为零的情况。

父权指数时, 采用上述两种算法得到的突变率, 结果是会存在差异, 但对总体的鉴定结论的影响有限。

2.3 三联体和二联体组突变率的比较

在以往基于直接计数法研究中, 一般将三联体和二联体组的数据进行合并计算突变率。然而, 三联体和二联体得到的突变率水平是有一定差异, 其在 2.1 和 2.2 中已进行部分阐述。在大多数 STR 基因座中, 三联体父源的实际突变率值大于二联体的父源相关值(图 3, 绿色>黑色); 三联体母源的实际突变率值大于二联体的母源相关值(图 3, 粉色>蓝色); 它们的比值表现出相似的规律(图 3), 特别是在 *vWA* 基因座。与此同时, 本研究也发现存在相反趋势的基因座, 如三联体组的 *D13S317*、*D5S818* 和二联体组的 *CSF1PO*、*D3S1358*、*TH01* 等, 因此将三联体和二联体的数据合并是不合理的, 它们之间的差异是存在且不宜忽略的。

2.4 父源和母源突变率的比较

在 1665 个突变事件中, 共观察到 1178 例父源突变和 293 例母源突变事件, 同时有 194 例突变来源不明(父源或母源), 其中父源突变占多数, 为 70.75% (图 4A)。在三联体组中, 总 $F\mu$ 为 0.06405550, $M\mu$

为 0.01937639 (表 1); 二联体中, 总 $F\mu$ 为 0.05146300, $M\mu$ 为 0.01220500 (表 1)。从以上数据可以得知, 父源突变率一般较母源突变率高, 在三联体和二联体组中分别约为 3.3 倍和 4.2 倍, 其与以往的研究结果一致^[11,12]。Jónsson 等^[18]对冰岛 1548 个三联体进行的新生突变(*de novo* mutations, DNMs)研究也表明

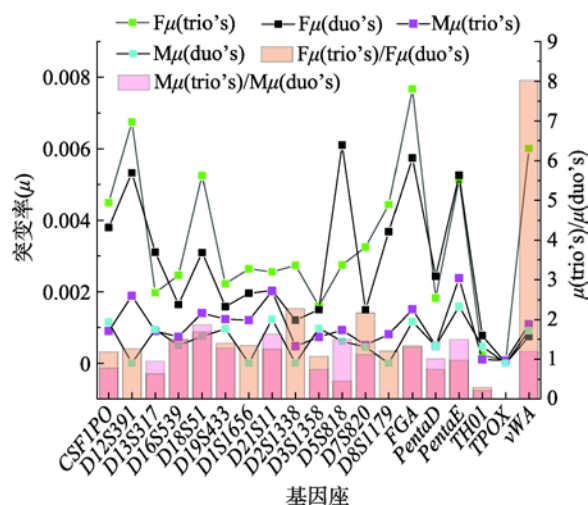


图 3 三联体和二联体中 $F\mu$ 、 $M\mu$ 及其比值的比较

Fig. 3 The comparisons of $F\mu$, $M\mu$ and their ratios in trio's and duo's

“trio's”表示三联体组; “duo's”表示二联体组; 空白处表示比式的分母为零的情况。

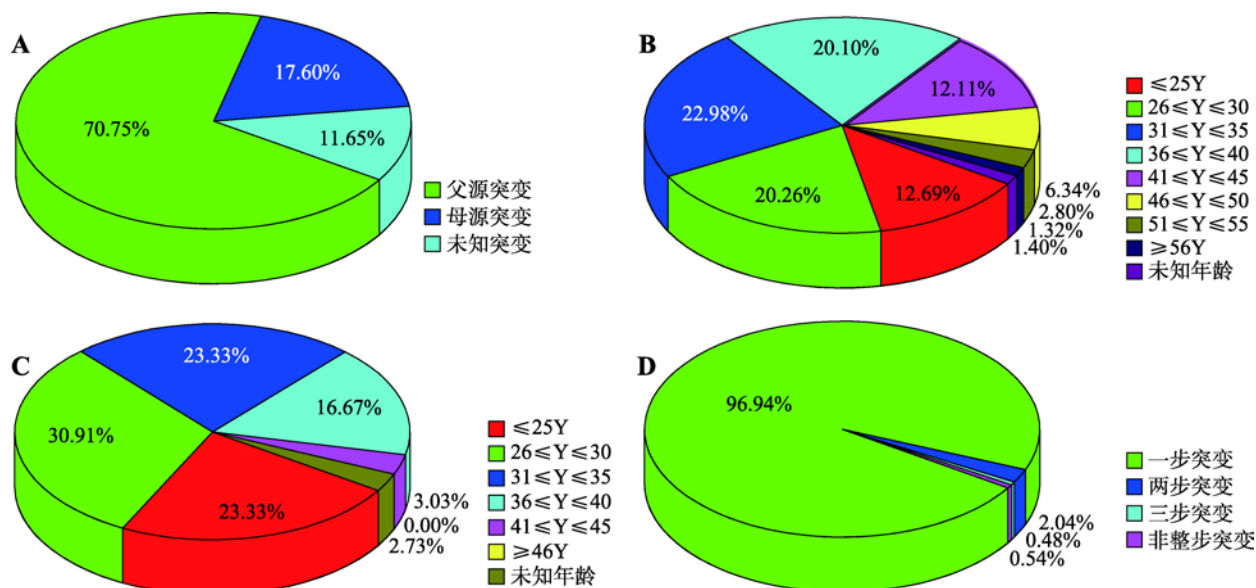


图 4 本次研究中突变事件来源、年龄和步数的比较

Fig. 4 The comparisons of mutation origin, age, and step in this research

A: 突变来源分布; B: 父源突变年龄分布; C: 母源突变年龄分布; D: 突变步数分布。

父源的突变数量约是母源的 4 倍,这也侧面反映了本次研究的结果。本研究还比较了不同突变来源的基因座特异性突变率(图 5)。如图 5A 显示,三联体中绝大多数基因座不同突变来源计算得到的突变率具有一定差异性,除 *D21S11* 和 *TPOX* 基因座外,父源突变率是母源突变率 2 倍以上。而二联体组也显示了相似的趋势(图 5B)。这可能是男性与女性形成配子时经历的细胞分裂次数不同,而男性需要产生更多的配子的缘故^[10]。可以预知,使用突变来源特异性基因座突变率将有助于亲子关系等法医学鉴定得出的更加客观而准确结论。另外,为了说明不同突变来源的年龄范围,本研究制作了父源(图 4B)和母源(图 4C)突变年龄的扇形图,从中可以看到父源突变的年龄范围比母源突变的年龄范围更广,这可能是由于男性在较长时间内可以保持一定的生殖能力,而女性不可以的原因。

2.5 突变步长与突变方向分析

在所有的突变事件中,共观察到 1614 个一步突变、34 个两步突变、8 个三步突变和 9 个非整步突变。一步突变占总突变事件的 96.94%,是主要的突变类型,逐步突变模型(stepwise mutation model, SMM)可以解释一步突变为主的现象^[19,20]。在其他非一步突变中,两步突变占 2.04%,三步突变占 0.48%,而非整步突变占 0.54% (图 4D)。发生非整步突变的基因座分别为: *D19S433* (3 例)、*D21S11* (2 例)、

DIS1656、*D6S1043*、*TH01* 和 *D12S391*; 突变来源为父亲的 6 例,母亲的 1 例,未能确定突变来源的 2 例;非整步的突变方向,其中 7 例突变减少,1 例突变增加,另有 1 例不能确定突变方向;关于非整步突变的步数,0.1、0.2、0.3 步均有观察到,其中有 5 例含有 0.2 步突变,2 例有 0.3 步突变,1 例有 0.1 步突变,另有 1 例为 0.3 或者 0.1 步突变。就本研究观察到的 9 例非整步突变而言,突变减少和包含 0.2 步突变是占多数的,详细信息展现在附表 7 中。为了进一步分析有限突变模型对于 STR 基因座等位基因不同突变步数的突变率影响,本研究计算了三联体中发生一步突变与两步突变的比值(图 6)。在父源(图 6A)中, μ_1^*/μ_2^* 的值普遍大于 μ_1/μ_2 ,特别是在 *vWA* 基因座(本次研究: $\mu_1^*/\mu_2^*\approx 104$; $\mu_1/\mu_2\approx 73$),这表明由于有限突变计算模型的使用, μ_1/μ_2 之间的差异正在减小。这些规律与 Slooten 与 Ricciardi^[9]的研究相似($\mu_1^*/\mu_2^*\approx 75$; $\mu_1/\mu_2\approx 58$),也就是说更多的高步数隐性突变被挖掘出来。在母源(图 6B)基本上也显示了这一规律,但发现了具有相反趋势的基因座,如 *D21S11*、*FGA* 和 *Penta E*,根据该模型作者的解释^[9],这可能是由于模型本身不完美的原因造成的。

尽管有限突变计算模型中不考虑突变方向(突变减少/增加),但本研究仍然统计了等位基因减少和增加的分布情况。在 1665 次突变事件中(附表 8),总计有 706 次突变减少(mutation loss),703 次突变

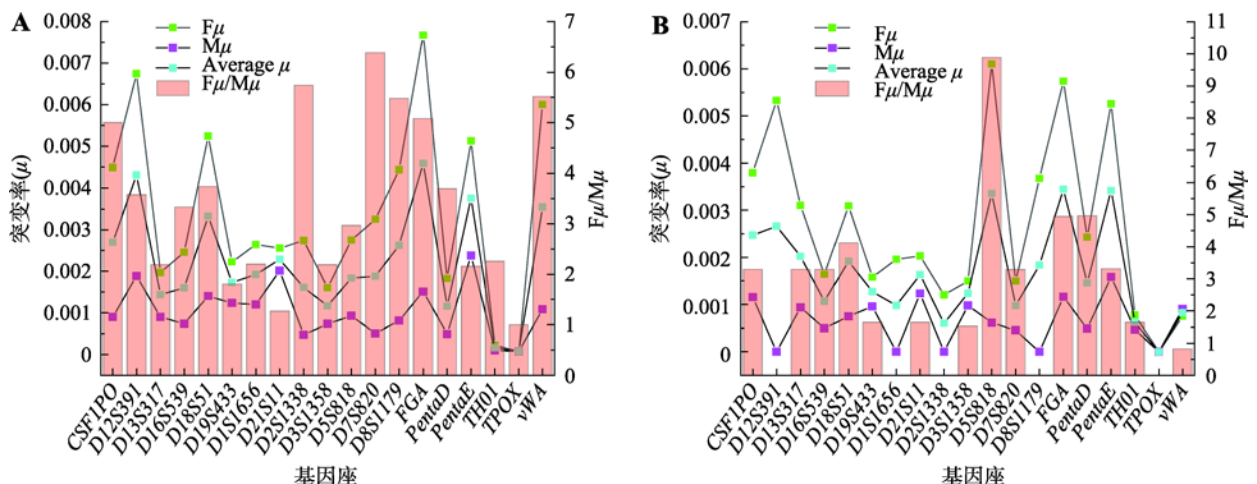


图 5 不同突变来源的基因座特异性突变率比较

Fig. 5 The comparisons of loci-specific mutation rates in different mutation origins

A: 三联体情形; B: 二联体情形; 空白处表示比值的分母为零的情况。

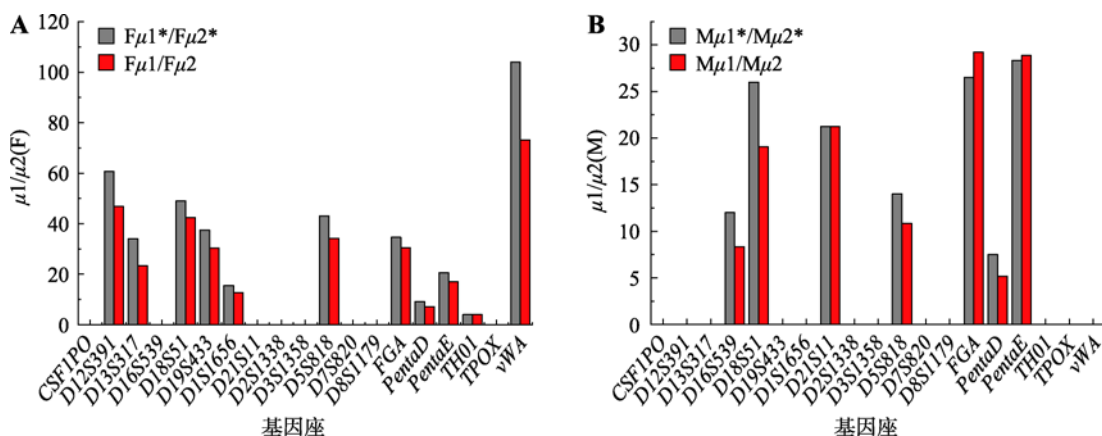


图 6 三联体父源与母源的突变步长分析

Fig. 6 The mutation step analysis of the paternal and maternal origins in trio's

A: 父源情形; B: 母源情形; 空白处表示比式的分母为零的情况。

增加(mutation gain), 256 次为不确定突变方向或者分组; 突变减少与增加(loss/gain)之比约为 1 : 1, 与 Lan 等^[21]的研究相似。Brinkmann 等^[10]报道认为突变的发生与等位基因的长度存在一定关联, 长的等位基因比短的等位基因更易于发生突变; Shao 等^[5]认为较长的等位基因倾向于发生突变减少事件, 而短的等位基因则更倾向于发生突变增加事件。为了探究等位基因长度与突变方向之间的关系, 参照 Ge 等^[22]和 Xu 等^[23]的方法, 将每个 STR 基因座的等位基因根据长度分为三组, 即较短(short group)、中等(medium group)、较长(long group)等位基因组, 详细的统计数据见附表 8。从表中可计算得知在以上三个分组中, 总的突变次数分别为 105、1042、397, 其中中等等位基因长度组发生的突变事件最多(约为 67.49%), 较长等位基因长度组次之(约为 25.71%), 而较短等位基因长度组最少(约为 6.80%), 也即中等及较长等位基因更易于发生突变, 较短的等位基因发生突变较少, 这与 Xu 等^[23]的研究结论一致; 较短等位基因组突变减少与增加(loss/gain)的比值约为 0.49 (<1), 中等组为 0.86, 较长组为 1.72 (>1), 本研究结果也表现出等位基因“短变长, 长变短”的规律, 这也得到 Shao 等^[5]、Ge 等^[22]、Xu 等^[23]、Dupuy 等^[24]的支持, 其可能与 STR 的序列结构有关。

2.6 杂合度与实际突变率的关系

一般认为, 杂合度高的基因座其突变率也较高, 也即突变率高可能是基因座杂合度高的一个原因。

据此根据已有文献报道的杂合度数据^[25], 分析了每个 STR 基因座的杂合度和实际突变率的关系, 使用 Origin 软件绘制线性拟合图(图 7), Pearson's $r=0.68454$ 。结果也相似地表明, 杂合度较低的 STR 基因座突变率普遍较低, 如 TH01、TPOX 基因座, 这与前人的研究相一致^[12]。

3 讨论

不同的突变率计算方法结果可能存在潜在的偏差, 比如传统的直接计数法在处理突变数据时未考虑到“隐性突变”现象^[26,27]。此处以两个确认的三

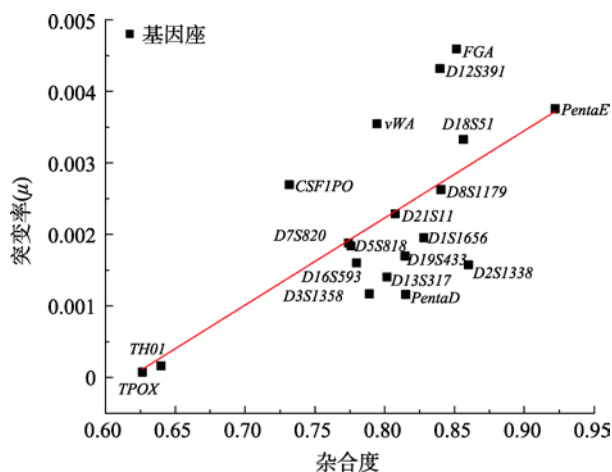


图 7 19 个常染色体 STR 基因座杂合度与实际突变率的关系

Fig. 7 The relationship between heterozygosity and actual mutation rates for the 19 autosomal STR loci

联体亲子鉴定为例进行说明(下文“a”为正整数)。

案例一: 某 STR 基因座母亲的分型是 $M=(a, a+1)$, 孩子是 $C=(a, a+1)$, 父亲是 $F=(a-1, a+1)$, 假设真实的遗传情况为孩子 C 的“a+1”来自母亲 M 的“a+1”, “a”来自父亲 F 的“a-1”或者“a+1”一步突变后的“a”, 即实际发生一步突变。从表面上看, 可能会认为孩子 C 的“a+1”遗传自父亲 F 的“a+1”, 而“a”遗传自母亲 M 的“a”, 此时表面上会认为没有突变发生, 即“突变发生有无认识偏差”型隐性突变(Type I 型); 案例二: 某 STR 基因座母亲的分型是 $M=(a+1, a+1)$, 孩子是 $C=(a, a-1)$, 父亲是 $F=(a, a-1)$, 假设真实遗传情况为孩子 C 的“a-1”来自母亲 M 的“a+1”两步突变后的“a-1”, “a”来自父亲 F 的“a”, 即实际发生两步突变。从表面上看, 可能会认为孩子 C 的“a-1”遗传自父亲 F 的“a-1”, “a”遗传自母亲 M 的“a+1”一步突变后的“a”, 此时表面上会认为发生一步突变, 因为当不知道突变真相时, 一般判断为较短的突变步数, 即“真实突变步数认识偏差”型隐性突变(Type II 型)。类似地, 隐性突变也可以出现在二联体中。在上述案例中, 表面观察到的突变被定义为表面突变(apparent mutation, ApM), 而不可见的真实突变被定义为实际突变(actual mutation, AcM)。因此, 如果不考虑隐性突变(比如直接计数法), 突变率可能就会被低估。尽管研究者无法直接从表面的 STR 分型突变现象中洞察真实的突变, 但 Slooten 和 Ricciardi^[9]通过突变模拟的方式研究了该隐性突变现象, 并提供了可接近真实突变率的计算模型和矫正参数。该模型根据突变的等位基因是否超过 STR 等位基因分型标准品(ladder)的范围 L, 可分为有限突变模型(restricted mutation model)和无限突变模型(unrestricted mutation model), 而真实的生物学过程可能介于这两个模型之间。当不知道突变的等位基因是否超过 Ladder 时, 无限突变模型可能是合适的。然而, 基于大量数据研究调查发现^[9], 突变的等位基因不包括在 Ladder 中是一种非常罕见的现象, 本次研究的数据也证明这一点。综合上述, 鉴于使用简单的有限突变模型是具有一定合理性的, 本研究拟采用该模型计算实际 STR 突变率。

基于以上有限突变模型计算获得的 STR 基因座实际突变率, 本研究一方面为了观察其是否有时间

差异, 收集了 Yan 等^[28]于 2006 年报道的与本研究具有相同来源的北京汉族突变率数据, 这组数据调查了 6532 例亲子鉴定案例中的 19,754 个样本的 17 个常用 STR 基因座特异性突变率, 这些基因座均包含在本次研究中; 另一方面为了观察本研究的北京汉族与中国其他汉族人群突变率的地域差异, 收集了中国其他地区或省(华北、华东、华南地区和河南、上海、贵州、河北、云南、福建、广东等省)汉族群体的突变率数据(附表 9)。如各基因座突变率比较的散点图所示(图 8), 在时间变化方面: 共享的 17 个 STR 基因座中, 除 *D21S11*、*TH01*、*TPOX* 和 *Penta D* 外, 其余 13 个基因座的突变率在 14 年间中增幅较大(红色方框与黑色圆点), 尤其是 *D18S51* 和 *Penta E* 的突变率增加将近一倍; 在地域变化方面, 不同汉族群体间的 STR 基因座特异性突变率具有一定差异, 特别是在 *vWA*、*FGA*、*D21S11*、*D8S1179*、*Penta E* 和 *D12S391* 基因座, 这些差异可能来自于多个地域来源样本的遗传结构不同^[29,30], 或是使用群体样本量不同, 或使用的计算模型方法不同。

亲子鉴定实践中, 有时可能遇到假突变现象(fake mutation), 其应引起足够的重视。例如, 本次研究中在一例确认的三联体亲子鉴定中 *D13S317* 基因座发现其基因型不符合遗传规律的现象, 称之为“假四步突变(fake four-steps mutation)”: 假设父 AF=8, 母亲 M=12 和孩子 C=8 (图 9)。当使用四个 STR 扩增试剂盒(PowerPlex[®]21 (Promega, 美国)、

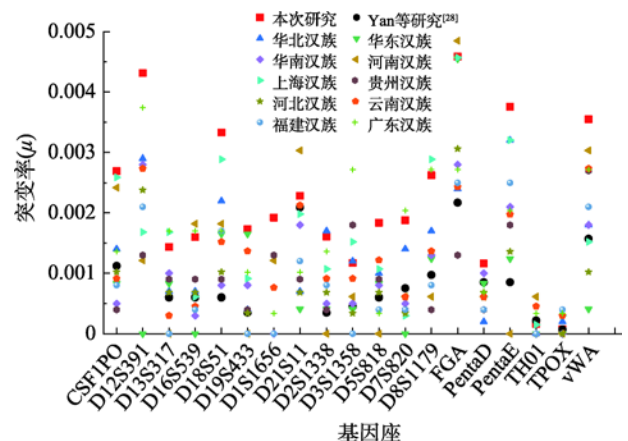


图 8 中国 10 个不同地区或省汉族人群突变率比较

Fig. 8 The Han population mutation rate comparisons of 10 different regions or provinces in China

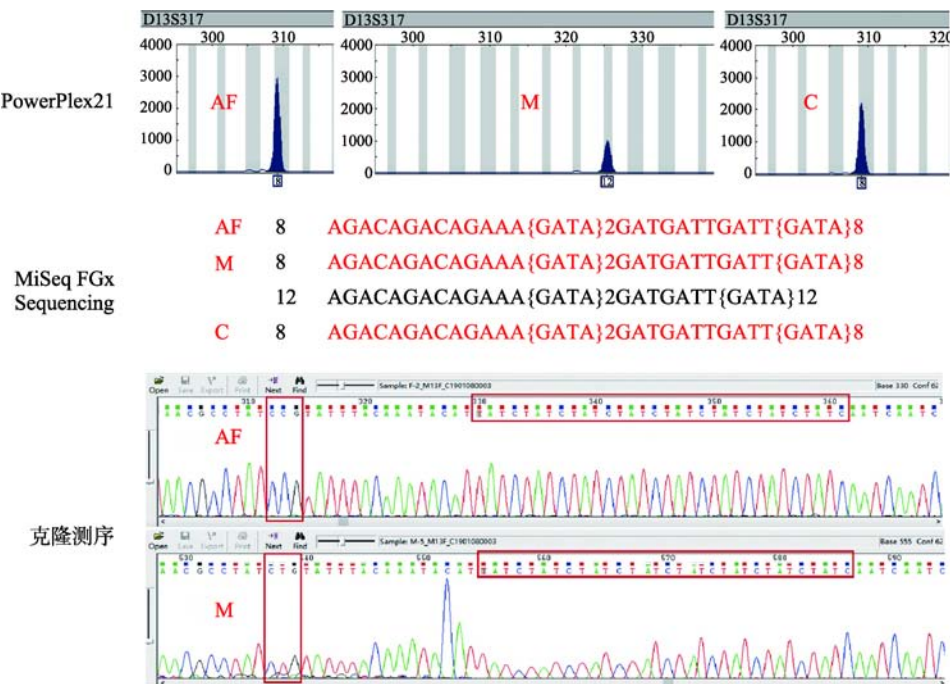


图 9 基于三种方法对 *D13S317* 基因座的测序结果

Fig. 9 The sequencing results of *D13S317* locus using three methods

Microreader™21 (阅微基因, 北京)、Microreader™23sp-B (阅微基因, 北京)、AGCU 17+1 system (中德美联, 无锡)确认基因型后, 得到相同的结果。起初, 怀疑母亲和孩子之间可能发生了四步突变, 即母亲的等位基因“12”突变为“8”。为了进一步证实这个猜想, 首先使用 ForenSeq DNA Signature Prep Kit (Verogen, 美国)在 MiSeq FGx 仪器(Illumina, 美国)中进行测序, 并使用 ForenSeq 通用分析软件(Verogen, 美国)分析原始数据。测序结果显示(图 9), 母亲的等位基因“8”发生了丢失, 其序列与孩子的等位基因“8”是一致的。为了继续寻找母亲等位基因丢失的原因, 使用 primer5 软件设计了 *D13S317* 基因座的引物(F: TGGAAGCTGCATAGAC; R: ACTCCATACAAATGCAAGAGA), 并进行了克隆测序(图 9)。如红色方框所示(图 9), 母亲的等位基因“8”的核心区上游第 17 bp 处发生单碱基突变, 由正常的“C”突变为“T”。以上研究表明, *D13S317* 基因座引物结合区的突变可能是多个试剂盒未能成功扩增出母亲等位基因“8”的原因。实际中, 真正的四步突变是非常少见的, 当遇到可疑的四步突变时, 测序有时是很必要的。

本研究在 28,313 例确认的亲子鉴定案例中, 共

发现 1665 个突变事件。基于以上突变数据和有限突变模型, 分别计算了三联体和二联体案例中的父源与母源的 19 个常染色体 STR 基因座的特异性实际突变率。与直接计数法计算突变率相比, 本次的计算结果揭示了更多的隐性突变现象, 因此获得的突变率数值理论上更加接近真实情况。本研究不仅可为中国法医 DNA 数据库和群体遗传学数据库提供重要的基础数据, 也对开展法医学个体识别、亲权鉴定和遗传学研究具有重要的意义。

附录:

附加材料详见文章电子版 www.chinagene.cn。

参考文献(References):

[1] Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol*, 2003, 4(2): R13. [DOI]

[2] Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations.

- Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics, *Int J Legal Med*, 1997, 110(4): 175–176. [DOI]
- [3] Butler JM. Short tandem repeat typing technologies used in human identity testing. *Biotechniques*, 2007, 43(4): 2–5. [DOI]
- [4] Kornberg A, Bertsch LL, Jackson JF, Khorana HG. Enzymatic synthesis of deoxyribonucleic acid, XVI. oligonucleotides as templates and the mechanism of their replication. *Proc Natl Acad Sci USA*, 1964, 51(2): 315–323. [DOI]
- [5] Shao CC, Lin MX, Zhou ZH, Zhou YQ, Shen YW, Xue AM, Zhou HG, Tang QQ, Xie JH. Mutation analysis of 19 autosomal short tandem repeats in Chinese Han population from Shanghai. *Int J Legal Med*, 2016, 130(6): 1439–1444. [DOI]
- [6] Li HX, Peng D, Wang Y, Wu RG, Zhang YM, Li R, Sun HY. Evaluation of genetic parameters of 23 autosomal STR loci in a southern Chinese Han population. *Ann Hum Biol*, 2018, 45(4): 359–364. [DOI]
- [7] Wang HD, Kand B, Su N, He M, Zhang B, Guo YX, Zhu BF, Liao SX, Zeng ZS. Evaluation of the genetic parameters and mutation analysis of 22 STR loci in the central Chinese Han population, *Int J Legal Med*, 2017, 131(1): 103–105. [DOI]
- [8] Qu N, Zhang XC, Liang H, Ou XL. Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population. *Forensic Sci Int Genet*, 2019, 38: e16–e17. [DOI]
- [9] Slooten K, Ricciardi F. Estimation of mutation probabilities for autosomal STR markers. *Forensic Sci Int Genet*, 2013, 7(3): 337–344. [DOI]
- [10] Brinkmann B, Klitsch M, Neuhuber F, Hühne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet*, 1998, 62(6): 1408–1415. [DOI]
- [11] Weber JL, Wong C. Mutation of human short tandem repeats. *Hum Mol Genet*, 1993, 2(8): 1123–1128. [DOI]
- [12] Jin B, Su Q, Luo HB, Li YB, Wu J, Yan J, Hou YP, Liang WB, Zhang L. Mutational analysis of 33 autosomal short tandem repeat (STR) loci in southwest Chinese Han population based on trio parentage testing. *Forensic Sci Int Genet*, 2016, 23: 86–90. [DOI]
- [13] Taylor MJ, Fisk NM. Prenatal diagnosis in multiple pregnancy. *Baillieres Best Pract Res Clin Obstet Gynaecol*, 2000, 14(4): 663–675. [DOI]
- [14] Wurmb-Schwark NV, Schwark T, Christiansen L, Lorenz D, Oehmichen M. The use of different multiplex PCRs for twin zygosity determination and its application in forensic trace analysis. *Leg Med(Tokyo)*, 2004, 6(2): 125–130. [DOI]
- [15] Xu YJ, Li TT, Pu T, Cao RX, Long F, Chen S, Sun K, Xu R. Copy number variants and exome sequencing analysis in six pairs of Chinese monozygotic twins discordant for congenital heart disease. *Twin Res Hum Genet*, 2017, 20(6): 521–532. [DOI]
- [16] Weber Lehmann J, Schilling E, Gradl G, Richter DC, Wiehler J, Rolf B. Finding the needle in the haystack: differentiating “identical” twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Sci Int Genet*, 2014, 9: 42–46. [DOI]
- [17] Wang LF, Yang Y, Zhang XN, Quan XL, Wu YM. Tri-allelic pattern of short tandem repeats identifies the murderer among identical twins and suggests an embryonic mutational origin. *Forensic Sci Int Genet*, 2015, 16: 239–245. [DOI]
- [18] Jónsson H, Sulem P, Kehr B, Kristmundsdóttir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, Ward LD, Arnadóttir GA, Helgason EA, Helgason H, Gylfason A, Jonasdóttir A, Jonasdóttir A, Rafnar T, Frigge M, Stacey SN, Magnusson OT, Thorsteinsdóttir U, Masson G, Kong A, Halldorsson BV, Helgason A, Gudbjartsson DF, Stefansson K. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, 2017, 549(7673): 519–522. [DOI]
- [19] Valdes AM, Slatkin M, Freimer NB. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics*, 1993, 133(3): 737–749. [DOI]
- [20] Klitsch M, Dauber EM, Ricci U, Cerri N, Immel UD, Kleiber M, Mayr WR. Haplotype studies support slippage as the mechanism of germline mutations in short tandem repeats. *Electrophoresis*, 2010, 25(20): 3344–3348. [DOI]
- [21] Lan Q, Wang HD, Shen CM, Guo YX, Yin CY, Xie T, Fang YT, Zhou YS, Zhu BF. Mutability analysis towards 21 STR loci included in the AGCU-21 kit in Chinese Han population. *Int J Legal Med*, 2018, 132(5): 1287–1291. [DOI]
- [22] Ge JY, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R. Mutation rates at Y chromosome short

- tandem repeats in Texas populations, *Forensic Sci Int Genet*, 2009,3 (3): 179–184. [\[DOI\]](#)
- [23] Xu W, Wang YQ, Zhang DD, Wang DX, Zhou L, Ye XL, Zhu CG, Shi YZ. Mutation analysis of 21 autosomal short tandem repeats in Han population from Hunan, China. *Ann Hum Biol*. 2019, 46(3): 254–260. [\[DOI\]](#)
- [24] Dupuy BM, Stenersen M, Egeland T, Olaisen B. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci, *Hum Mutat*, 2004, 23(2): 117–124. [\[DOI\]](#)
- [25] Xie BB, Chen L, Yang YR, Lv YX, Chen J, Shi Y, Chen C, Zhao HY, Yu ZL, Liu YC, Fang XD, Yan JW. Genetic distribution of 39 STR loci in 1027 unrelated Han individuals from northern China. *Forensic Sci Int Genet*, 2015, 19: 205–206. [\[DOI\]](#)
- [26] Chakraborty R, Stivers DN, Zhong YX. Estimation of mutation rates from parentage exclusion data: applications to STR and VNTR loci. *Mutat Res*, 1996, 354(1): 41–48. [\[DOI\]](#)
- [27] Vicard P, Dawid AP. A statistical treatment of biases affecting the estimation of mutation rates. *Mutat Res*, 2004, 547(1–2): 19–33. [\[DOI\]](#)
- [28] Yan JW, Liu YC, Tang H, Zhang QX, Huo ZY, Hu SN, Yu J. Mutations at 17 STR loci in Chinese population. *Forensic Sci Int*, 2006, 162(1–3): 53–54. [\[DOI\]](#)
- [29] Lu DJ, Liu QL, Wu WW, Zhao H. Mutation analysis of 24 short tandem repeats in Chinese Han population. *Int J Legal Med*, 2012, 126(2): 331–335. [\[DOI\]](#)
- [30] Qian XQ, Yin CY, Ji Q, Li K, Fan HT, Yu YF, Bu FL, Hu LL, Wang JW, Mu HF, Haigh S, Chen F. Mutation rate analysis at 19 autosomal microsatellites. *Electrophoresis*, 2015, 36(14): 1633–1639. [\[DOI\]](#)

(责任编辑: 朱波峰)

附表 1 中国汉族 20 个常染色体 STR 基因座的突变数和案例数

Supplementary Table 1 The number of mutations and cases in the 20 STR loci studied in Chinese Han Population

基因座	三联体突变数				三联体 案例数	二联体突变数			二联体案例数	
	父源	母源	父源或母源	总计		父源	母源	总计	父亲孩子二联体	母亲孩子二联体
<i>CSF1PO</i>	65	7	15	87	22,113	4	2	6	2342	3858
<i>D12S391</i>	113	24	21	158	20,562	8	0	8	2057	3406
<i>D13S317</i>	30	11	10	51	22,113	4	2	6	2342	3858
<i>D16S539</i>	38	8	10	56	22,113	2	1	3	2342	3858
<i>D18S51</i>	96	22	10	128	22,113	5	2	7	2342	3858
<i>D19S433</i>	36	19	5	60	22,113	2	2	4	2342	3858
<i>D1S1656</i>	29	11	8	48	13,712	2	0	2	1361	2149
<i>D21S11</i>	44	41	7	92	22,113	3	3	6	2342	3858
<i>D2S1338</i>	50	5	9	64	22,113	2	0	2	2342	3858
<i>D3S1358</i>	29	6	13	48	22,113	2	2	4	2342	3858
<i>D6S1043</i>	54	8	5	67	20,562	0	2	2	2057	3406
<i>D5S818</i>	40	11	8	59	22,113	6	1	7	2342	3858
<i>D7S820</i>	51	3	13	67	22,113	2	1	3	2342	3858
<i>D8S1179</i>	75	10	9	94	22,113	5	0	5	2342	3858
<i>FGA</i>	135	18	19	172	22,113	9	3	12	2342	3858
<i>Penta D</i>	30	7	3	40	20,562	3	1	4	2057	3406
<i>Penta E</i>	92	40	10	142	20,562	8	4	12	2057	3406
<i>TH01</i>	4	1	2	7	22,113	1	1	2	2342	3858
<i>TPOX</i>	1	1	1	3	22,113	0	0	0	2342	3858
<i>vWA</i>	97	11	16	124	22,113	1	2	3	2342	3858
Total	1109	264	194	1567	22,113	69	29	98	2342	3858

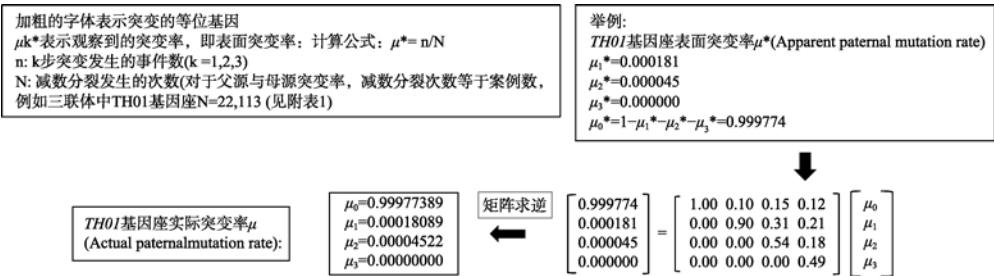
由于有限突变模型中未包含 *D6S1043* 基因座的计算校正参数, 因此我们的实际突变率计算共包括除 *D6S1043* 之外其余 19 个 STR 基因座。

附表 2 基于有限突变模型三联体组 TH01 基因座实际突变率计算实例

Supplementary Table2 The detail mutation rate calculation example of the TH01 locus in trio's

基因座	孩子 基因型	父亲	母亲	突变 来源	突变 步数	F1*	F2*	F3*	M1*	M2*	M3*
<i>TH01</i>	8, 9	9	9	父亲 母亲	1	0.5	0	0	0.5	0	0
<i>TH01</i>	8, 9	9	9	父亲 母亲	1	0.5	0	0	0.5	0	0
<i>TH01</i>	9	8	6, 9	父亲	1	1	0	0	0	0	0
<i>TH01</i>	6, 11	9, 10	6, 9	父亲	1	1	0	0	0	0	0
<i>TH01</i>	9	7	9	父亲	2	0	1	0	0	0	0
<i>TH01</i>	8	6, 9	8, 9	父亲	1	1	0	0	0	0	0
<i>TH01</i>	7, 10	7	9	母亲	1	0	0	0	1	0	0
总计(n)						4	1	0	2	0	0
突变*						0.000181	0.000045	0	0.00009	0	0

F1*, F2* F3*: 代表突变来源为父亲, 突变步数为 1, 2, 3 的情况. M1*, M2* M3*: 代表突变来源为母亲, 突变步数为 1, 2, 3 的情况。



附表 3 10 例同卵双生子 STR 基因座突变描述

Supplementary Table 3 The STR mutated description of monozygotic twins from 10 cases

案例	孩子	试剂盒	父亲年龄	母亲年龄	孩子性别	突变基因座	孩子分型	父亲分型	母亲分型	突变来源	突变步数
案例 1	MZ1	AB-Identifiler, AGCU21+1	30	30	男	<i>FGA</i>	22,24	21,23	22,24	未知	± 1
	MZ2		30	30	男	<i>FGA</i>	22,24	21,23	22,24	未知	± 1
案例 2	MZ1	Goldeneye 20A, AGCU21+1	33	28	男	<i>D6S1043</i>	13, 20	13, 18	18, 21	母亲	-1
	MZ2		33	28	男	<i>D6S1043</i>	13, 20	13, 18	18, 21	母亲	-1
案例 3	MZ1	Goldeneye 20A	21	23	女	<i>Penta E</i>	15,25	17,24	12,15	父亲	+1
	MZ2		21	23	女	<i>Penta E</i>	15,25	17,24	12,15	父亲	+1
案例 4	MZ1	Goldeneye 20A	26	24	男	<i>D18S51</i>	12,16	14,15	12,15	父亲	+1
	MZ2		26	24	男	<i>D18S51</i>	12,16	14,15	12,15	父亲	+1
案例 5	MZ1	Goldeneye 20A, AGCU21+1	40	39	女	<i>D21S11</i>	30,33.2	31,32.2	30,31.2	父亲	+1
	MZ2		40	39	女	<i>D21S11</i>	30,33.2	31,32.2	30,31.2	父亲	+1
案例 6	MZ1	Goldeneye 20A	44	40	男	<i>D6S1043</i>	18,21.3	12,22.3	11,18	父亲	-1
	MZ2		44	40	男	<i>D6S1043</i>	18,21.3	12,22.3	11,18	父亲	-1
案例 7	MZ1	PowerPlex®21	32	31	女	<i>Penta E</i>	16,17	16	16,18	未知	± 1
	MZ2		32	31	女	<i>Penta E</i>	16,17	16	16,18	未知	± 1
案例 8	MZ1	PowerPlex®21	42	24	男	<i>D12S391</i>	17,22	18,22	18,19	母亲	-1
	MZ2		42	24	男	<i>D12S391</i>	17,22	18,22	18,19	母亲	-1
案例 9	MZ1	PowerPlex®21	38	32	男	<i>CSF1PO</i>	12,14	10,13	12	父亲	+1
	MZ2		38	32	男	<i>CSF1PO</i>	12,14	10,13	12	父亲	+1
案例 10	MZ1	PowerPlex®21	31	31	男	<i>D18S51</i>	15,17	14,16	14,15	父亲	+1
	MZ2		31	31	男	<i>D18S51</i>	15,17	14,16	14,15	父亲	+1

“MZ” 表示同卵双胞胎(monozygotic twins); “-” 表示突变减少; “+” 表示突变增加; “ \pm ” 表示突变增加或者减少。

附表 4A 同一个孩子观察到 2 个突变事件的 27 例案件

Supplementary Table 4A The 27 cases of 2 mutation events observed in one child

案例	试剂盒	父亲年龄	母亲年龄	孩子性别	突变基因座	孩子分型	父亲分型	母亲分型	突变来源	突变步数
案例 1	AB-Identifiler	24	22	女	<i>D3S1358</i>	15	15,16	16	母亲	-1
					<i>D19S433</i>	12,14.2	13,14	14,14.2	父亲	-1
案例 2	Goldeneye 20A	25	22	女	<i>D18S51</i>	16	13,15	13,16	父亲	+1
					<i>D12S391</i>	18,21	18,20	18,20	未知	+1
案例 3	Goldeneye 20A	31	29	男	<i>D7S820</i>	9, 12	11	9, 10	父亲	+1
					<i>CSF1PO</i>	12, 13	10, 13	13	父亲	-1
案例 4	Goldeneye 20A	26	27	男	<i>CSF1PO</i>	13,14	12,14	9,14	未知	± 1
					<i>FGA</i>	22,24	23,26	24,24.2	父亲	-1
案例 5	Goldeneye 20A	55	25	女	<i>D3S1358</i>	16,18	16,17	15,16	父亲	+1
					<i>D12S391</i>	19,20	18,21	19	父亲	-1

续附表

案例	试剂盒	父亲年龄	母亲年龄	孩子性别	突变基因座	孩子分型	父亲分型	母亲分型	突变来源	突变步数
案例 6	Goldeneye 20A	26	22	女	<i>D6S1043</i>	12	11,13	12,13	父亲	±1
					<i>D8S1179</i>	14	14,15	13,15	母亲	±1
案例 7	Goldeneye 20A	37	38	女	<i>D12S391</i>	20,23	21	18,23	父亲	-1
					<i>FGA</i>	22,23	22,24	22,25	父亲	±1
案例 8	Goldeneye 20A	24	19	男	<i>D5S818</i>	11,14	10,14	10,12	母亲	±1
					<i>D6S1043</i>	14,17	13,19	17,18	父亲	+1
案例 9	PowerPlex®21	30	27	女	<i>D21S11</i>	30,31	31,31.2	31,33	未知	-1
					<i>D2S1338</i>	23,24	18,24	19,24	未知	-1
案例 10	PowerPlex®21	35	23	女	<i>D21S11</i>	30,32.2	29	31.2,32.2	父亲	+1
					<i>D12S391</i>	18,22	20,21	18,19	父亲	+1
案例 11	PowerPlex®21	40	34	男	<i>vWA</i>	17,18	19	14,17	父亲	-1
					<i>FGA</i>	22,24	21,23	21,24	父亲	±1
案例 12	PowerPlex®21	33	NA	女	<i>penta E</i>	10,22	16,23	NA	父亲	-1
					<i>D5S818</i>	13	11,12	NA	父亲	+1
案例 13	PowerPlex®21	29	25	女	<i>D8S1179</i>	12	11,13	12,13	父亲	±1
					<i>FGA</i>	23,24	22.2,25	20,23	父亲	-1
案例 14	PowerPlex®21	37	28	女	<i>D18S51</i>	13,21	13,15	19,22	母亲	-1
					<i>vWA</i>	18	16,19	18	父亲	-1
案例 15	PowerPlex®21	39	NA	女	<i>Penta E</i>	16,17	10,18	NA	父亲	-1
					<i>D18S51</i>	14,19	18,20	NA	父亲	±1
案例 16	PowerPlex®21	24	25	女	<i>D13S317</i>	9,11	8,10	9,10	父亲	+1
					<i>Penta E</i>	12,15	13,16	16	母亲	+1
案例 17	PowerPlex®21	30	32	男	<i>D18S51</i>	14	14	13,17	母亲	+1
					<i>FGA</i>	23,25	21,23	18,22	父亲	-1
案例 18	PowerPlex®21	24	21	女	<i>TPOX</i>	8,11	8,10	8	父亲	+1
					<i>D8S1179</i>	13,15	15,17	15,18	父亲	-2
案例 19	PowerPlex®21	28	28	女	<i>D6S1043</i>	14,17	17	13,18	母亲	+1
					<i>D16S539</i>	11,14	10,11	9,12	父亲	+1
案例 20	PowerPlex®21	37	33	男	<i>Penta E</i>	11,27	13,26	11,20	父亲	+1
					<i>Penta D</i>	10	9	10,13	父亲	+1
案例 21	PowerPlex®21	28	25	女	<i>D13S317</i>	10,12	10,11	8,10	父亲	+1
					<i>D21S11</i>	30,31	30,31	29,32	母亲	+1
案例 22	PowerPlex®21	26	23	女	<i>D1S1656</i>	10,15	15	11	母亲	-1
					<i>D13S317</i>	18,26	19,27	18,19	父亲	-1
案例 23	PowerPlex®21	39	28	女	<i>D16S539</i>	10,12	9,13	12	父亲	+1
					<i>D21S11</i>	30.2,31.2	28,30.2	29,32.2	母亲	-1
案例 24	PowerPlex®21	45	41	女	<i>D16S539</i>	11	9,12	11,12	父亲	-1
					<i>D8S1179</i>	13,18	13,17	12,13	父亲	+1
案例 25	PowerPlex®21	28	29	女	<i>Penta D</i>	11,14	8,9	12,14	父亲	+2
					<i>D21S11</i>	30,31.2	29,30	30	未知	+1.2
案例 26	PowerPlex®21	30	28	女	<i>D13S317</i>	9	9	8	母亲	+1
					<i>D5S818</i>	12,14	10,13	9,12	父亲	+1
案例 27	PowerPlex®21	40	33	男	<i>D1S1656</i>	14,16	15,17.3	12,14	父亲	+1
					<i>D13S317</i>	8,13	9,12	8,12	父亲	+1

附表 4B 同一个孩子观察到 3 个突变事件的 3 例案件

Supplementary Table 4B The 3 cases of 3 mutation events observed in one child

案例	试剂盒	父亲年龄	母亲年龄	孩子性别	突变基因座	孩子分型	父亲分型	母亲分型	突变来源	突变步数
案例 1	Goldeneye 20A, MicroreaderTM23sp ID Sysrem	47	39	男	<i>D8S1179</i>	12,14	13	14	父亲	-1
					<i>D17S1290</i>	15,19	15,20	15,18	未知	±1
					<i>FGA</i>	22,23	24	22,24	父亲	-1
案例 2	PowerPlex®21, MicroreaderTM23sp ID Sysrem	39	26	男	<i>D6S1043</i>	12,17	17,20	13,14	母亲	-1
					<i>D8S1132</i>	18,21	18,19	22		-1
					<i>FGA</i>	19,24	20,25	19,25	父亲	-1
案例 3	PowerPlex®21, MicroreaderTM23sp ID Sysrem	34	33	男	<i>D2S1338</i>	19,21	19	19,20	母亲	+1
					<i>D10S1435</i>	13	13	12,14		±1
					<i>FGA</i>	21,25	22,24	21,23	父亲	+1

“-”表示突变减少; “+”表示突变增加; “±”表示突变增加或者减少; “NA”表示无相关数据。

附表 5A 三联体组的表面突变率

Supplementary Table 5A The apparent mutation rates in trio's

基因座	$F\mu^*$	$F\mu1^*$	$F\mu2^*$	$F\mu3^*$	$M\mu^*$	$M\mu1^*$	$M\mu2^*$	$M\mu3^*$
<i>CSF1PO</i>	0.00327861	0.00327861	0.00000000	0.00000000	0.00065572	0.00065572	0.00000000	0.00000000
<i>D12S391</i>	0.00600623	0.00590896	0.00009727	0.00000000	0.00167785	0.00167785	0.00000000	0.00000000
<i>D13S317</i>	0.00158278	0.00153756	0.00004522	0.00000000	0.00072356	0.00067833	0.00000000	0.00004522
<i>D16S539</i>	0.00194456	0.00194456	0.00000000	0.00000000	0.00058789	0.00054267	0.00004522	0.00000000
<i>D18S51</i>	0.00456745	0.00443178	0.00009044	0.00004522	0.00122100	0.00117578	0.00004522	0.00000000
<i>D19S433</i>	0.00174106	0.00169584	0.00004522	0.00000000	0.00097228	0.00092706	0.00000000	0.00004522
<i>D1S1656</i>	0.00240665	0.00226079	0.00014586	0.00000000	0.00109393	0.00109393	0.00000000	0.00000000
<i>D21S11</i>	0.00214806	0.00214806	0.00000000	0.00000000	0.00201239	0.00192195	0.00009044	0.00000000
<i>D2S1338</i>	0.00246461	0.00246461	0.00000000	0.00000000	0.00042961	0.00042961	0.00000000	0.00000000
<i>D3S1358</i>	0.00160539	0.00160539	0.00000000	0.00000000	0.00056528	0.00056528	0.00000000	0.00000000
<i>D5S818</i>	0.00198978	0.00194456	0.00004522	0.00000000	0.00067833	0.00063311	0.00004522	0.00000000
<i>D7S820</i>	0.00260028	0.00260028	0.00000000	0.00000000	0.00040700	0.00040700	0.00000000	0.00000000
<i>D8S1179</i>	0.00359517	0.00359517	0.00000000	0.00000000	0.00065572	0.00065572	0.00000000	0.00000000
<i>FGA</i>	0.00653462	0.00626328	0.00018089	0.00009044	0.00128883	0.00119839	0.00004522	0.00004522
<i>Penta D</i>	0.00153195	0.00133742	0.00014590	0.00004863	0.00041338	0.00036475	0.00004863	0.00000000
<i>Penta E</i>	0.00471744	0.00449859	0.00021885	0.00000000	0.00218850	0.00206692	0.00007295	0.00004863
<i>TH01</i>	0.00022611	0.00018089	0.00004522	0.00000000	0.00009044	0.00009044	0.00000000	0.00000000
<i>TPOX</i>	0.00006783	0.00006783	0.00000000	0.00000000	0.00006783	0.00006783	0.00000000	0.00000000
<i>vWA</i>	0.00474834	0.00470312	0.00004522	0.00000000	0.00085922	0.00085922	0.00000000	0.00000000

附表 5B 三联体组的实际突变率

Supplementary Table 5B The actual mutation rates in trio's

基因座	$F\mu$	$F\mu1$	$F\mu2$	$F\mu3$	$M\mu$	$M\mu1$	$M\mu2$	$M\mu3$
<i>CSF1PO</i>	0.00449100	0.00449100	0.00000000	0.00000000	0.00089800	0.00089800	0.00000000	0.00000000
<i>D12S391</i>	0.00674500	0.00660400	0.00014100	0.00000000	0.00188500	0.00188500	0.00000000	0.00000000
<i>D13S317</i>	0.00197200	0.00189100	0.00008100	0.00000000	0.00089900	0.00083200	-0.00004300	0.00011000

续附表

基因座	F μ	F μ 1	F μ 2	F μ 3	M μ	M μ 1	M μ 2	M μ 3
D16S539	0.00246100	0.00246100	0.00000000	0.00000000	0.00073900	0.00066000	0.00007900	0.00000000
D18S51	0.00524900	0.00504600	0.00011900	0.00008400	0.00140400	0.00133400	0.00007000	0.00000000
D19S433	0.00222300	0.00215200	0.00007100	0.00000000	0.00123300	0.00118200	-0.00002400	0.00007500
DIS1656	0.00264300	0.00245100	0.00019200	0.00000000	0.00120200	0.00120200	0.00000000	0.00000000
D21S11	0.00255700	0.00255700	0.00000000	0.00000000	0.00201239	0.00192195	0.00009044	0.00000000
D2S1338	0.00273800	0.00273800	0.00000000	0.00000000	0.00047700	0.00047700	0.00000000	0.00000000
D3S1358	0.00160539	0.00160539	0.00000000	0.00000000	0.00073400	0.00073400	0.00000000	0.00000000
D5S818	0.00274700	0.00266900	0.00007800	0.00000000	0.00092600	0.00084800	0.00007800	0.00000000
D7S820	0.00325000	0.00325000	0.00000000	0.00000000	0.00050900	0.00050900	0.00000000	0.00000000
D8S1179	0.00443800	0.00443800	0.00000000	0.00000000	0.00081000	0.00081000	0.00000000	0.00000000
FGA	0.00767100	0.00725200	0.00023800	0.00018100	0.00151000	0.00137300	0.00004700	0.00009000
Penta D	0.00182500	0.00152100	0.00021400	0.00009000	0.00049400	0.00041400	0.00008000	0.00000000
Penta E	0.00513100	0.00484700	0.00028400	0.00000000	0.00237900	0.00222200	0.00007700	0.00008000
TH01	0.00022611	0.00018089	0.00004522	0.00000000	0.00010000	0.00010000	0.00000000	0.00000000
TPOX	0.00007700	0.00007700	0.00000000	0.00000000	0.00007700	0.00007700	0.00000000	0.00000000
vWA	0.00600600	0.00592500	0.00008100	0.00000000	0.00108800	0.00108800	0.00000000	0.00000000

“F μ ”表示父源突变率；“M μ ”表示母源突变率；“F μ ”与“M μ ”后的数字表示突变步数；标红的数字表示计算出的实际突变率为负值，详见文中。

附表 6A 二联体组表面突变率

Supplementary Table 6A The apparent mutation rates in duo's

基因座	F μ^*	F μ 1*	F μ 2*	F μ 3*	M μ^*	M μ 1*	M μ 2*	M μ 3*
CSF1PO	0.00170794	0.00170794	0.00000000	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000
D12S391	0.00388916	0.00388916	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D13S317	0.00170794	0.00170794	0.00000000	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000
D16S539	0.00085397	0.00085397	0.00000000	0.00000000	0.00025920	0.00025920	0.00000000	0.00000000
D18S51	0.00213493	0.00213493	0.00000000	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000
D19S433	0.00085397	0.00085397	0.00000000	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000
DIS1656	0.00146951	0.00146951	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D21S11	0.00128096	0.00128096	0.00000000	0.00000000	0.00077760	0.00077760	0.00000000	0.00000000
D2S1338	0.00085397	0.00085397	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D3S1358	0.00085397	0.00042699	0.00042699	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000
D5S818	0.00256191	0.00256191	0.00000000	0.00000000	0.00025920	0.00025920	0.00000000	0.00000000
D7S820	0.00085397	0.00085397	0.00000000	0.00000000	0.00025920	0.00025920	0.00000000	0.00000000
D8S1179	0.00213493	0.00213493	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
FGA	0.00384287	0.00384287	0.00000000	0.00000000	0.00077760	0.00077760	0.00000000	0.00000000
Penta D	0.00145843	0.00145843	0.00000000	0.00000000	0.00029360	0.00029360	0.00000000	0.00000000
Penta E	0.00388916	0.00388916	0.00000000	0.00000000	0.00117440	0.00117440	0.00000000	0.00000000
TH01	0.00042699	0.00042699	0.00000000	0.00000000	0.00025920	0.00025920	0.00000000	0.00000000
TPOX	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
vWA	0.00042699	0.00042699	0.00000000	0.00000000	0.00051840	0.00051840	0.00000000	0.00000000

附表 6B 二联体组的实际突变率

Supplementary Table 6B The actual mutation rates in duo's

基因座	Fμ	Fμ1	Fμ2	Fμ3	Mμ	Mμ1	Mμ2	Mμ3
CSF1PO	0.00379500	0.00379500	0.00000000	0.00000000	0.00115200	0.00115200	0.00000000	0.00000000
D12S391	0.00532800	0.00532800	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D13S317	0.00310500	0.00310500	0.00000000	0.00000000	0.00094300	0.00094300	0.00000000	0.00000000
D16S539	0.00164200	0.00164200	0.00000000	0.00000000	0.00049800	0.00049800	0.00000000	0.00000000
D18S51	0.00309400	0.00309400	0.00000000	0.00000000	0.00075100	0.00075100	0.00000000	0.00000000
D19S433	0.00158100	0.00158100	0.00000000	0.00000000	0.00096000	0.00096000	0.00000000	0.00000000
DIS1656	0.00195900	0.00195900	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D21S11	0.00203300	0.00203300	0.00000000	0.00000000	0.00123400	0.00123400	0.00000000	0.00000000
D2S1338	0.00120300	0.00120300	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
D3S1358	0.00149600	-0.00155400	0.00305000	0.00000000	0.00097800	0.00097800	0.00000000	0.00000000
D5S818	0.00610000	0.00610000	0.00000000	0.00000000	0.00061700	0.00061700	0.00000000	0.00000000
D7S820	0.00149800	0.00149800	0.00000000	0.00000000	0.00045500	0.00045500	0.00000000	0.00000000
D8S1179	0.00368100	0.00368100	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
FGA	0.00573600	0.00573600	0.00000000	0.00000000	0.00116100	0.00116100	0.00000000	0.00000000
Penta D	0.00243100	0.00243100	0.00000000	0.00000000	0.00048900	0.00048900	0.00000000	0.00000000
Penta E	0.00525600	0.00525600	0.00000000	0.00000000	0.00158700	0.00158700	0.00000000	0.00000000
TH01	0.00077600	0.00077600	0.00000000	0.00000000	0.00047100	0.00047100	0.00000000	0.00000000
TPOX	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
vWA	0.00074900	0.00074900	0.00000000	0.00000000	0.00090900	0.00090900	0.00000000	0.00000000

“Fμ”表示父源突变率; “Mμ”表示母源突变率; “Fμ”与“Mμ”后的数字表示突变步数; 标红的数字表示计算出的实际突变率为负值, 详见文中。

附表 7 非整步突变事件汇总

Supplementary Table 7 The total nonintegral mutations events

基因座	基因型			突变来源	可能的突变步数
	孩子	父亲	母亲		
D19S433	13	14.2	13, 15.2	父亲	-1.2
D19S433	13, 16	15, 16.2	13, 14		-0.2
D19S433	13	15.2	-		-2.2
DIS1656	16	16.3 , 17	15, 16		-0.3
D6S1043	17, 23.3	18, 24	11, 17		-0.1
D21S11	30, 32	29, 30.2	32, 32.2	母亲	-0.2
TH01	9	6, 9	9.3		-0.3
D21S11	30, 31.2	29, 30	30	未知	+1.2
D12S391	19.3 , 20	19 , 20	20 , 21		+0.3/-0.1

加粗的字体表示突变的等位基因。

附表 8 等位基因突变增加与减少事件汇总

Supplementary Table 8 The total mutation events of allele mutation gain and loss

基因座	较短组 Short group					中等组 Medium group					较长组 Long group					未确定分组	
	AR	NG	NL	NU	UM	AR	NG	NL	NU	UM	AR	NG	NL	NU	UM	Undeterminable group	Total
<i>CSF1PO</i>	7-9.1	1	0	0	1	9.2-11.1	6	4	0	10	12-15	31	28	3	62	20	93
<i>D12S391</i>	14-18.3	8	11	1	20	19-23	36	64	8	108	24-28	8	7	0	15	23	166
<i>D13S317</i>	5-7	0	0	0	0	8-12	22	11	2	35	13-15	5	12	0	17	5	57
<i>D16S539</i>	6-8	0	0	0	0	9-12	21	10	4	35	13-16	6	11	1	18	6	59
<i>D18S51</i>	10-13	9	0	0	9	13.2-20	54	32	5	91	21-25	15	18	0	33	2	135
<i>D19S433</i>	9-11.2	2	0	0	2	12-16	24	25	2	51	16.2-18.2	3	8	0	11	0	64
<i>D1S1656</i>	10-12	0	1	0	1	13-17.3	20	20	2	42	18-19.3	1	4	0	5	2	50
<i>D21S11</i>	27-29	8	7	0	15	29.2-33.1	26	31	3	60	33.2-35.2	2	12	0	14	9	98
<i>D2S1338</i>	16-19.2	5	1	0	6	19.3-23.3	21	7	0	28	24-26	9	19	1	29	3	66
<i>D3S1358</i>	12-13	0	0	0	0	14-18	21	22	9	52	19-20	0	0	0	0	0	52
<i>D6S1043</i>	9-12	2	0	0	2	12.2-18	13	15	4	32	18.2-22.3	13	19	2	34	1	69
<i>D5S818</i>	7-9	0	0	0	0	9.2-13	27	16	13	56	13.2-15	2	8	0	10	0	66
<i>D7S820</i>	7-9	2	1	1	4	9.1-11	5	8	2	15	11.1-14	12	19	3	34	17	70
<i>D8S1179</i>	8-10	4	0	0	4	11-15	34	19	11	64	16-18	7	22	1	30	1	99
<i>FGA</i>	13-18.2	1	1	0	2	19-24.2	51	53	18	122	25-28	15	29	3	47	13	184
<i>Penta E</i>	5-15	16	7	5	28	16-21.3	54	37	6	97	22-33.1	7	14	0	21	8	154
<i>PentaD</i>	5-9	5	1	0	6	10-18	18	12	5	35	19-22	1	0	0	1	2	44
<i>TH01</i>	5-7	1	0	0	1	8-12	2	6	0	8	13-15	0	0	0	0	0	9
<i>TPOX</i>	7-8	0	1	0	1	8.1-11.2	1	0	0	1	12-13	0	1	0	1	0	3
<i>vWA</i>	14-15	1	1	1	3	16-19	41	40	19	100	19.3-21	4	11	0	15	9	127
<i>Total</i>	—	65	32	8	105	—	497	432	113	1042	—	141	242	14	397	121	1665

AR: allele range (等位基因长度范围); NG: number of mutation gain (突变增加的数量); NL: number of mutation loss (突变减少的数量); NU: number of uncertain mutation (不能确定突变方向的数量); NM: number of total mutation (突变总数)。

附表 9 中国其他省或区汉族人群的突变率汇总

Supplementary Table 9 The Han populations mutation rates from other provinces or regions in China

基因座	本次研究	Yan 等研究 ^[1]	华北汉族 ^[2]	华东汉族 ^[2]	华南汉族 ^[3]	河南汉族 ^[4]	上海汉族 ^[5]	贵州汉族 ^[6]	河北汉族 ^[7]	云南汉族 ^[8]	福建汉族 ^[9]	广东汉族 ^[9]
	<i>n</i> =78739	<i>n</i> =19754	<i>n</i> =16612	<i>n</i> =4494	<i>n</i> =6303	<i>n</i> =2475	<i>n</i> =12413	<i>n</i> =3879	<i>n</i> =5006	<i>n</i> =10941	<i>n</i> =3279	<i>n</i> =4094
<i>CSF1PO</i>	0.0027	0.0011	0.0014	0.0008	0.0005	0.0024	0.0026	0.0004	0.0010	0.0009	0.0008	0.0014
<i>D12S391</i>	0.0043	—	0.0029	0.0000	0.0028	0.0012	0.0017	0.0013	0.0024	0.0027	0.0021	0.0037
<i>D13S317</i>	0.0014	0.0006	0.0007	0.0008	0.0010	0.0000	0.0017	0.0009	0.0007	0.0003	0.0000	0.0017
<i>D16S539</i>	0.0016	0.0006	0.0007	0.0000	0.0003	0.0018	0.0006	0.0009	0.0007	0.0005	0.0004	0.0017
<i>D18S51</i>	0.0033	0.0006	0.0022	0.0017	0.0008	0.0018	0.0029	0.0009	0.0010	0.0015	0.0017	0.0017
<i>D19S433</i>	0.0017	0.0004	0.0004	0.0017	0.0008	0.0000	0.0009	0.0004	0.0003	0.0014	0.0000	0.0010
<i>D1S1656</i>	0.0019	—	—	—	0.0000	0.0012	—	0.0013	—	0.0008	0.0000	0.0003
<i>D21S11</i>	0.0023	0.0021	0.0007	0.0004	0.0018	0.0030	0.0020	0.0009	0.0007	0.0021	0.0012	0.0010
<i>D2S1338</i>	0.0016	0.0004	0.0017	0.0004	0.0005	0.0000	0.0011	0.0004	0.0007	0.0009	0.0008	0.0014
<i>D3S1358</i>	0.0012	0.0005	0.0012	0.0004	0.0005	0.0006	0.0015	0.0018	0.0003	0.0009	0.0000	0.0027

续附表

基因座	本次研究 <i>n</i> =78739	Yan 等研究 ^[1] <i>n</i> =19754	华北汉族 ^[2] <i>n</i> =16612	华东汉族 ^[2] <i>n</i> =4494	华南汉族 ^[3] <i>n</i> =6303	河南汉族 ^[4] <i>n</i> =2475	上海汉族 ^[5] <i>n</i> =12413	贵州汉族 ^[6] <i>n</i> =3879	河北汉族 ^[7] <i>n</i> =5006	云南汉族 ^[8] <i>n</i> =10941	福建汉族 ^[9] <i>n</i> =3279	广东汉族 ^[9] <i>n</i> =4094
<i>D5S818</i>	0.0018	0.0006	0.0010	0.0008	0.0008	0.0000	0.0011	0.0009	0.0007	0.0012	0.0004	0.0003
<i>D7S820</i>	0.0019	0.0008	0.0014	0.0000	0.0005	0.0006	0.0003	0.0004	0.0003	0.0006	0.0004	0.0020
<i>D8S1179</i>	0.0026	0.0010	0.0017	0.0012	0.0013	0.0006	0.0029	0.0004	0.0014	0.0014	0.0008	0.0027
<i>FGA</i>	0.0046	0.0022	0.0024	0.0046	0.0028	0.0049	0.0046	0.0013	0.0031	0.0024	0.0025	0.0027
<i>Penta D</i>	0.0012	0.0009	0.0002	0.0008	0.0010	0.0006	0.0006	0.0004	0.0007	0.0006	0.0004	0.0007
<i>Penta E</i>	0.0038	0.0009	0.0032	0.0012	0.0021	0.0000	0.0032	0.0018	0.0014	0.0020	0.0025	0.0020
<i>TH01</i>	0.0002	0.0002	0.0000	0.0000	0.0000	0.0006	0.0002	—	0.0000	0.0005	0.0000	0.0003
<i>TPOX</i>	0.0001	0.0001	0.0002	0.0000	0.0000	0.0000	0.0003	—	0.0000	0.0003	0.0004	0.0003
<i>vWA</i>	0.0035	0.0016	0.0018	0.0004	0.0018	0.0030	0.0015	0.0027	0.0010	0.0027	0.0021	0.0027
<i>Total</i>	0.0013	—	0.0013	0.0008	—	—	0.0016	—	0.0163	0.0012	0.0008	0.0014

本次研究中显示为三联体平均突变率; “*n*” 表示样本量; “—” 表示无相关数据。

附录涉及的参考文献(References):

- [1] Yan JW, Liu YC, Tang H, Zhang QX, Huo ZY, Hu SN, Yu J. Mutations at 17 STR loci in Chinese population. *Forensic Sci Int*, 2006, 162(1–3): 53–54.
- [2] Qian XQ, Yin CY, Ji Q, Li K, Fan HT, Yu YF, Bu FL, Hu LL, Wang JW, Mu HF. Mutation rate analysis at 19 autosomal microsatellites. *Electrophoresis*, 2015, 36(14): 1633–1639.
- [3] Li HX, Peng D, Wang Y, Wu RG, Zhang YM, Li R, Sun HY. Evaluation of genetic parameters of 23 autosomal STR loci in a Southern Chinese Han population, *Ann Hum Biol*, 2018, 45(4): 359–364.
- [4] Wang. HD, Kang B, Su N, He M, Zhang B, Guo YX, Zhu BF, Liao. SX, Zeng ZS. Evaluation of the genetic parameters and mutation analysis of 22 STR loci in the central Chinese Han population. *Int J Legal Med*, 2016, 131(1): 103–105.
- [5] Shao CC, Lin MX, Zhou ZH, Zhou YQ, Shen YW, Xue AM, Zhou HG, Tang QQ, Xie JH. Mutation analysis of 19 autosomal short tandem repeats in Chinese Han population from Shanghai. *Int J Legal Med*, 2016, 130(6): 1439–1444.
- [6] Yang MQ, Ren Z, Ji JY, Zhou H, Zhang HL, Dai JL, Wang J, Huang J. Population genetic data and mutations of 22 autosomal STR loci in Guizhou Han population. *Forensic Sci Int Genet*, 2017, 29: e29–e30.
- [7] Sun LJ, Shi K, Tan L, Zhang Q, Fu LH, Zhang XJ, Fu GP, Li SJ, Cong B. Analysis of genetic polymorphisms and mutations at 19 STR loci in Hebei Han population. *Forensic Sci Int Genet*, 2017, 31: e50–e51.
- [8] Zhang XF, Liu LL, Xie RF, Wang GY, Shi Y, Gu T, Hu LP, Nie SJ. Population data and mutation rates of 20 autosomal STR loci in a Chinese Han population from Yunnan Province, Southwest China. *Int J Legal Med*, 2018, 132(4): 1083–1085.
- [9] Qu N, Zhang XC, Liang H, XL Ou. Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population. *Forensic Sci Int Genet*, 2019, 38: e16–e17.