

关于国际法医遗传学学会 (ISFG) 对 STR 序列命名建议的解读

刘志勇^{1,2}, 赵凯³, 孙宏钰^{1,2}

(1. 中山大学中山医学院法医学系, 广东 广州 510080;
2. 广东省法医学转化医学工程技术研究中心, 广东 广州 510080; 3. 广东省公安厅刑事技术中心, 广东 广州 510050)

【摘要】基于大规模平行测序 (MPS) 技术的 STR 分型在法医 DNA 检验中的应用日益广泛, 但是 STR 等位基因序列命名规则不统一及自动化程度低的问题阻碍了实验室之间的数据共享与比对, 也不利于基于序列多态性的 STR 数据库构建。国际法医遗传学学会 (ISFG) 近期发布了对 STR 序列命名的建议, 预期将对这个问题的解决有所改善。本文回顾了 ISFG 关于 STR 命名的历程, 对该建议的主要内容进行解读, 并与目前国内司法和公安行业现行的相关规范进行比较, 以期帮助同行理解和使用。

【关键词】STR 命名; 国际法医遗传学学会; 毛细管电泳; 大规模平行测序

【中图分类号】D919.2

【文献标识码】B

【文章编号】1001-5728(2024)03-0361-07

doi:10.13618/j.issn.1001-5728.2024.03.020

Interpretation of the international society of forensic genetics (ISFG)'s recommendations for STR sequence nomenclature. Liu Zhiyong^{1,2}, Zhao Kai³, Sun Hongyu^{1,2} (1. Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China; 2. Guangdong Province Translational Forensic Medicine Engineering Technology Research Center, Guangzhou 510080, China; 3. The Center of Criminal Technology of Guangdong Province, Guangzhou 510050, China)

【Abstract】STR sequence typing based on massively parallel sequencing (MPS) technology is increasingly used in forensic DNA typing, but the problem of inconsistent rules for naming STR sequences and low automation hinder data sharing and comparison between laboratories, and is not conducive to build STR databases based on sequence polymorphisms. The International Society of Forensic Genetics (ISFG) has recently published recommendations for STR sequence nomenclature, which is expected to solve this problem. This paper reviews the ISFG's journey on STR naming, interprets the main content of the recommendations, and compares it with the current relevant standards existing in the domestic justice and public security industries, with the aim of helping peers to understand and use it.

【Key words】STR nomenclature; International society of forensic genetics (ISFG); Capillary electrophoresis (CE); Massively parallel sequencing (MPS)

短串联重复 (short tandem repeat, STR) 是存在于人类基因组的一类遗传标记, 具有高度多态性和遗传稳定性, 被广泛应用于法医学个体识别与

亲子鉴定中。近三十年来, 荧光标记复合扩增结合毛细管电泳 (capillary electrophoresis, CE) 技术的 STR 分型 (简称 CE-STR) 是国内外法医 DNA 分析最主流的检测方法。1997 年, 国际法医遗传学学会 (International Society for Forensic Genetics, ISFG) 的前身国际法医血液遗传学学会 (International Society for Forensic Haemogenetics, ISFH)^[1] 对 STR 序列命名提出建议 (以下简称“ISFG 标准 1997”), 即采用数字化方式命名 STR 等位基因, 例如等位基因 9 表示 STR 重复单位完整重复 9 次, 等位基因 9.3 表示有 9 次完整重复单位和 1 个包含 3 个碱基的不完整重复单位。这种命名方式是将待

【基金项目】国家自然科学基金面上项目 (81971798)。

【作者简介】刘志勇, 男, 博士研究生, 研究方向: 法医学物证学; E-mail: liuzhy255@mail2.sysu.edu.cn。

赵凯, 男, 硕士, 警务技术正高级任职资格, 主要从事法医学物证学检案与研究; E-mail: 573170268@qq.com。

刘志勇、赵凯同为第一作者。

【通信作者】孙宏钰, 女, 博士, 教授, 主要从事法医学物证学研究; E-mail: sunhy@mail.sysu.edu.cn。

测等位基因的扩增片段长度与等位基因分型标准物 (ladder) 上已确定长度的等位基因片段相比较而得, 本质上是长度多态性检测。STR 等位基因数字化的表示方法简单清晰, 便于数据交流和比对, 成为建立法庭科学 DNA 数据库的重要前提条件, 为打击犯罪做出了巨大贡献。随着大规模平行测序 (massively parallel sequencing, MPS) 技术成熟与成本降低, 越来越多法医 DNA 实验室开始引进该方法进行 STR 研究, 基于复合靶向扩增的商品化 MPS-STR 分型试剂盒陆续通过法医学验证, 例如 ForenSeq DNA Signature Prep Kit (Verogen, 美国)^[2]、MGIEasy Signature Identification Library Prep Kit (MGI Tech, 中国)^[3]、Precision ID GlobalFiler™ NGS STR Panel (Thermo Fisher Scientific, 美国)^[4] 等。与 CE 平台相比, MPS-STR 分型不受荧光种类限制, 可单次检测大量基因座且扩增片段相对较短, 在法医学有特殊的应用价值。另一方面, MPS-STR 分型既能基于等位基因片段长度获得与 CE-STR 分型相兼容的数字化等位基因命名, 又能通过序列多态性分析提高 STR 多态性。随着研究不断深入, STR 序列多态性的价值不断被认可与挖掘, 但基于序列多态性信息如何命名 STR 等位基因则众说纷纭^[5]。2016 年 ISFG DNA 委员会针对该问题提出 8 条原则性建议 (以下简称为“ISFG 标准 2016”)^[6]; 时隔 8 年后, ISFG 近期发布了关于 STR 等位基因序列命名的详细性建议 (以下简称为“ISFG 标准 2024”)^[7]。本文梳理了 ISFG 对于 STR 序列命名建议的制定历程, 对“ISFG 标准 2024”予以解读, 并与国内司法和公安现行行业规范进行比较, 以期帮助同行理解和使用。

1 ISFG 对于 STR 序列命名的早期规定——“ISFG 标准 2016”

2016 年 MPS 技术早期应用于法医学, “ISFG 标准 2016”对 STR 序列命名提出 8 个原则性最低要求^[6]:

- (1) 基于 MPS 的 STR 序列分析软件应允许文本格式的序列字符串导出 / 存储;
- (2) STR 序列应与参考基因组的正向链比对;
- (3) 使用 GRCh38 作为参考基因组;
- (4) 曾经报告为反向链上的基因座需转换为正向链, 并定义起止点;
- (5) 采用全面的 STR 序列命名法, 以确保与未来命名系统相兼容, 并保持与 CE-STR 分型命名的兼容性;
- (6) STR 序列字符串存储应包括侧翼序列, 并

保留基因组起始 / 终止核苷酸坐标元数据;

(7) 全球范围内基于序列的 STR 等位基因频率数据库应使用统一的命名系统;

(8) 新的 STR 复合测序体系应保留以往的 STR 标记, 新标记选择应结合序列多态性群体数据。

在“ISFG 标准 2016”的指引下, 商品化或自建 MPS-STR 序列分型体系一般都输出了 STR 等位基因的全长序列以及与 CE-STR 兼容的数字化名称, 但不同软件产生的 STR 序列命名格式不尽相同, 等位基因长度报告范围具有很大的随意性。另一方面, 随着全球不同人群数据遗传结构的基础性研究日益增多, 对于 STR 序列特征与变异的认识也逐渐深入, 但命名常需人工介入辅助, 自动化程度不理想。这些问题不仅妨碍了不同 STR 复合测序体系、实验室之间的数据交流, 也难以计算统一的基于序列的等位基因频率参数, 阻碍了 MPS-STR 分型的法医学应用。

2 ISFG 对于 STR 序列命名的系列研究

“ISFG 标准 2016”发表之后, 全球多个法医 DNA 实验室持续进行 STR 序列命名相关研究, 产出了以下主要成果:

(1) 法医序列结构指南 (Forensic Sequence STRucture Guide, FSSG)^[8]: 采用“ISFG 标准 2016”推荐的括号格式注释核心重复区域, 提供侧翼区域多态性的位置信息, 校准汇编了大量法医学相关的常染色体 STR、Y-STR 和 X-STR 基因座的基本信息, 可登录 <https://strider.online/nomenclature> 获取该网站;

(2) STR 测序项目 (STR Sequencing Project, STRSeq) 序列目录^[9]: 对每个 STR 特异性序列进行分类和特征描述, 创建了包含 STR 序列区域、重复区域的括号形式注释、侧翼区域多态性、有关测序分析和数据质量, 以及基于长度的等位基因命名等信息的 GenBank 记录, 可登录 <https://www.ncbi.nlm.nih.gov/bioproject/380127> 获取该网站;

(3) 欧洲法庭科学研究院网络组 (European Network of Forensic Science Institutes, ENFSI) 参考数据库 (STRidER)^[10,11]: 用于 STR 质量控制、数据解释和稀有等位基因频率估计等, 搭建了一个法医常染色体 STR 频率数据库, 可登录 <https://strider.online/> 获取该网站。

2018 年 ISFG 成立 STRAND (Short Tandem Repeat: Align, Name, Define) 工作组, 并于 2019 年 4 月在伦敦举办 STR 序列命名会议, 讨论了 STR 序

列格式化、STR 基因座起止坐标、参考基因组、人群 STR 序列多态性数据和序列分析软件/工具资源等主题^[12]。之后 STRAND 工作组组织多个实验室继续协调并完善 FSSG、STRSeq 和 STRidER 资源。2021 年, ISFG 组建了一个 DNA 委员会对以往的研究结果进行总结,“ISFG 标准 2024”即为该委员会的最新成果^[7]。

3 ISFG 对于 STR 序列命名的现行规定——“ISFG 标准 2024”

3.1 “ISFG 标准 2024”内容概述

“ISFG 标准 2024”遵循“ISFG 标准 2016”的基本原则,并基于大量群体遗传研究成果制定了新的 STR 序列命名建议。该建议强调 MPS-STR 序列命名要与目前法庭科学 DNA 数据库中 CE-STR 数据具有兼容性,提出了建立序列最小报告范围、等位基因命名自动化解决方案、新 STR 基因座鉴定方法、未来 STR 数据库等建议。具体包括以下 5 条:

(1) 基于序列的 STR 等位基因应保持为序列字符串的形式,比对到当前人类参考基因组的正向链,序列字符串应包括本文定义的最小报告范围(minimum reporting range);

(2) 序列字符串是表示 STR 序列数据的主要格式,但其他简化形式对报告、数据库、数据呈现/描述以及生物信息软件分析等方面很有益。“ISFG 标准 2016”为早期 STR 测序用户提供了一个全面命名法(comprehensive nomenclature),现建议使用更简化的形式。为提供一个最少人工干预的统一命名系统,ISFG 推荐采用 STRNaming 软件进行 STR 序列的括号化重复(bracketed repeat)格式自动输出。此外,基于特定的最小报告范围,还可以使用短小的“序列码(sequence code)”来表示 STR 等位基因序列,方便实验室间比较;

(3) 生物信息学软件应使用标准化资源以促进 STR 序列命名的一致性,鼓励将 ISFG 的 STR 序列命名法和基于序列的等位基因频率计算纳入生物信息学软件;

(4) 探索用于法医分析的新 STR 基因座研究人员应评估既往是否存在该基因座、是否对个人隐私或健康有预测能力,并应遵循建议(1)和(2)确定基因座格式,建立与质量控制数据一致的基因型;

(5) 未来的 STR 数据库可能需要包含存储序列字符和/或能够明确代表序列字符的编码,建议在(3)中所述的资源、法医学界用于分析 STR 序列数据的

生物信息软件、用于管理数据库的软件使用通用的序列编码和命名法。

3.2 “ISFG 标准 2024”要点解读

3.2.1 STR 序列最小报告范围的确定

鉴于每个 STR 基因座核心区域具有明确定义的起止坐标(即用于计算重复单位数目的范围),因此将 MPS 平台获得的等位基因序列统一转化为与 CE 相兼容的数字化名称并不困难,但基于序列的 STR 基因座起止坐标因引物位置而异,不同分析软件的分析区域也长短不一,导致 STR 序列命名不同,阻碍了数据交流,也无法准确计算基于序列的等位基因频率参数。“ISFG 标准 2024”在充分考虑与 CE 数据的兼容性、MPS 获得的信息量和引物设计等问题的基础上,建议统一 STR 序列报告范围,即“最小报告范围”,其确定步骤如图 1 所示。

步骤二中选择核心重复序列上下游 25 个核苷酸的范围是考虑该距离通常包含了 STRSeq 项目中观察到的 STR 侧翼序列变异分布范围^[9],也有助于

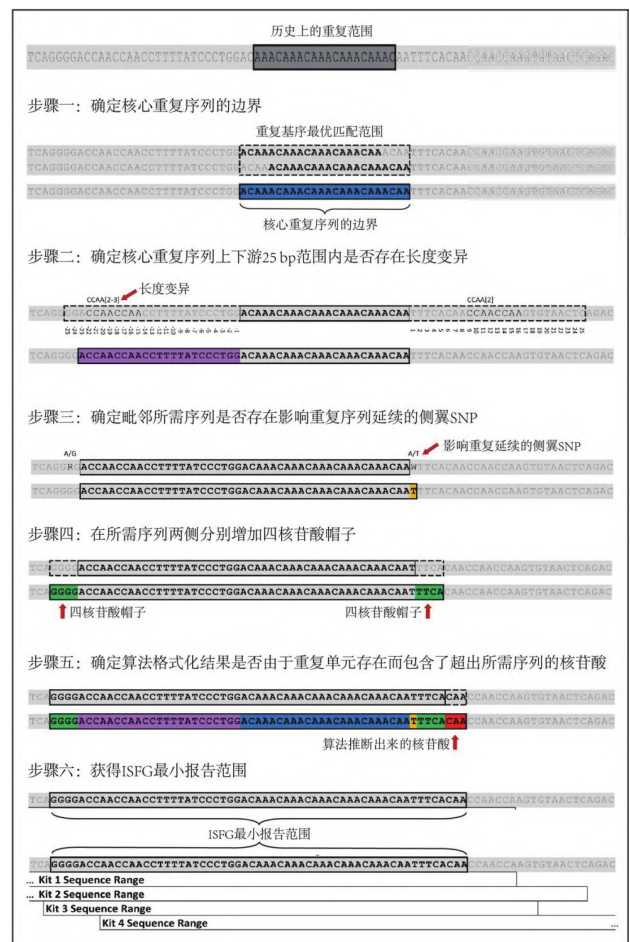


图 1 STR 序列最小报告范围的确定步骤(改编自^[7])
Fig. 1 Steps to determine the minimum reporting range of STR sequences (adapted from^[7])

提示避开在长度变异区域设计 PCR 引物, 进而减少可能出现的等位基因丢失和 / 或与 CE-STR 分型不一致现象。统一 STR 序列最小报告范围不仅有助于提供足够的侧翼区域来区分核心重复序列的末端, 也有助于统一不同引物扩增体系、不同分析软件的分型结果。目前法医学中常用 STR 序列的最小报告范围在 FSSG (https://strider.online/downloads/FSSG_v6.xlsx) 中持续更新, 可作为参考使用。ISFG 建议实验室仍然保存序列字符串的原始记录, 但可使用最小报告范围进行数据交流, 如果现有 STR 测序体系未覆盖最小报告范围, 则应重新设计引物以满足此要求。

3.2.2 STR 序列命名的自动化

用括号格式注释重复序列是法医 STR 序列命名常用的格式, 将序列字符串的重复区域压缩成“可读”的描述性格式, 即由重复基序和次数表示, 但命名过程常需人为干预, 这不仅降低分型效率, 也增加了命名不一致的可能。2019 年 Hoogenboom 等提出 STRNaming 软件, 极大地提高了 STR 序列命名的自动化程度^[12,13]。该软件基于最长不间断延伸 (longest uninterrupted stretch, LUS) 策略, 从给定的参考序列中提取 STR 基因座的基本特征 (基序、侧翼、最长不间断重复区域、用于计算 CE 长度的校正因子等), 并将提取到的特征用于目标 STR 序列, 之后通过预先设定的一组参数评价全部搜索结果, 最终将 STR 等位基因序列自动化、标准化处理形成统一的括号化格式, 该软件可开源获取 (<https://fdstools.nl/strnaming/> 或 <https://github.com/Jerrythafast/strnaming>)。STRNaming 软件进行 STR 序列命名的基本格式为: CE 等位基因长度命名_按 5' 到 3' 的顺序用括号格式列出重复单元和次数以及 < 8 nt 的中断序列_侧翼变异, 例如“CE12_AC[5]ATCT[12]_+1CT > -”, 既缩短了字符串的书写或储存长度, 又保持了序列的“可读性”与“可溯源性”。STRNaming 软件实现了将 STR 等位基因序列自动处理为统一的括号化格式的功能, 为“ISFG 标准 2024”所推荐使用。

除了用括号注释重复序列字符串格式外, “ISFG 标准 2024”还推荐使用短小的“序列码”来促进 STR 序列命名自动化。2020 年, Young 等发布了一种使用哈希函数 SHA-256 辅以其他算法生成 STR 序列码的方法, 将每条 DNA 序列转换为唯一的序列标识符 (sequence identifier, SID)^[14]。无论序列长度是短到 1 个碱基还是长到整条染色体, 该方法均可将其转化成为由 26 个英语字母组合构成的 55 个

字符的 SID 码, 不同序列生成不同的 SID。SID 可通过在线生成 (<https://nichevision.github.io/sid.js/>); 按照 STR 序列最小报告范围生成的 STR 序列 SID 码包含在 STRSeq 的 GenBank 记录中, 可整合到生物信息学软件中进行搜寻和查找。“ISFG 标准 2024”建议使用 STR 序列 SID 码的前 5 个字符来唯一识别各个 STR 序列等位基因; 如果同时使用基因座名称和 CE 等位基因数字名称, 则使用 SID 码的前 4 个字符即可。例如, TH01 基因座基于最小报告范围的等位基因“7”的 DNA 序列为“ATGGTGAATGAATGAATGAATGAATGAATGAGGGA”, 生成的 SID 码为“UCYLMZAZPQEEOSMFIEIRFFSYHYOLNKVYAVUYKYWEUVDINCBXFDDMXB”; “基因座名称+SID”格式可记录为: TH01_UCYLM, “基因座名称+CE 等位基因名称+SID”格式可记录为: TH01_7_UCYL。将 STR 等位基因转化为 SID 的命名方法对于提高序列命名效率、减少数据库存储空间、混合物分析均有益处^[14]。

3.2.3 STR 序列命名的标准化资源

“ISFG 标准 2024”强调将持续协调与更新 FSSG、STRSeq 和 STRidER 等 STR 序列多态性命名相关资源, 以促进该建议实施。

FSSG 已完成基于 GRCh38 的注释, 更新了商品化 STR 试剂盒中包含的常染色体、Y 染色体和 X 染色体 STR 基因座信息 (https://strider.online/downloads/FSSG_v6.xlsx), 包括: 传统的括号格式 (2016 年~2022 年)、STRNaming 括号格式 (2023 年以后)、STR 序列最小报告范围、常见商品化试剂盒的 STR 序列范围等。

STRSeq 是 STR 序列多态性命名的重要参考资源, 目前正在努力更新超过 2 500 份 STRSeq 的 GenBank 记录, 内容主要为: 更新 STR 序列范围以适应试剂生产商推荐范围和 ISFG 推荐的最小报告范围、更新括号格式的重复结构以符合 ISFG 的建议、添加 STR 序列最小报告范围的 SID 码等。目前, STRBase (<https://strbase.nist.gov>) 也在整合链接 STRSeq 和 GenBank 记录。

STRidER 正在开发对用户上传的 STR 序列数据 (至少包括 STR 序列最小报告范围) 进行质量控制的功能。即将提交的序列与 STRSeq 保存的记录进行比较, 若二者不匹配时, 将对未知等位基因的测序深度、序列范围、侧翼区域多态性和相关序列 (phylogenetic context) 进行评估, 通过此评估的等位基因序列将被创建为新的 GenBank 记录。此外, STRidER 平台也试图开发基于 STR 序列最

小报告范围进行序列多态性等位基因频率估计的功能。

3.2.4 新基因座的确认与命名

在探索新的 STR 基因座用于法医学实践时，应避免选择可能表征个人隐私或健康信息的基因座，也应仔细评估该基因座过去是否已被命名。可以利用 HGM Workshop 11 report、NCBI Probe 数据库 (<https://ftp.ncbi.nih.gov/pub/ProbeDB>)、WebSTR 数据库 (<http://webstr.ucsd.edu>)、国际遗传系谱学会 (International Society of Genetic Genealogy, ISOGG) Wiki 和相关资源中的信息 (https://isogg.org/wiki/Wiki_Welcome_Page) 以及相关文献 [15-18] 来确定 STR 基因座是否已经被命名。对于新基因座的命名仍采用与早期相兼容的方式，即 D#S# 命名 (例如 D21S11) 格式，其中“D”代表 DNA，随后的数字或字母对应染色体编号，后面一个字母表示基因座的复杂性 (“S”代表唯一的 DNA 片段，“Z”代表单条染色体有重复的 DNA 片段，“F”代表多条染色体上有同源序列家族的 DNA 片段)，随后的数字为唯一性编号 [19]。研究人员如有需求，可联系 ISFG 的 STRAND 工作组和 / 或 strseq@nist.gov 协助进行新基因座确认和命名。

3.2.5 STR 序列数据库的建设

目前全球 DNA 数据库已包含数以亿计基于长度多态性的 STR 分型数据，重新检测这些样本将其转化为基于序列多态性分型不合理也没必要。“ISFG 标准 2024”建议通过实施与旧标准相兼容的序列多态性 STR 命名规则来实现序列多态性分型与长度多态性分型的关联。新的数据库应存储 STRNaming 软件命名的等位基因最小报告范围的字符串或者 SID 序列码，而不是实际序列，这有利于减少所需的字符空间和提高数据隐私性。若存储空间允许，建议存储最小报告范围的 STR 序列用于检索和数据交换。在未来一段时间内，确保基于 STR 序列多态性的数据库与现有的基于长度多态性的数据库相兼容是至关重要的，是在现有 STR 长度多态性数据库中纳入序列多态性数据还是开发与现有 STR 长度多态性数据库对接的并行序列多态性数据库？数据库开发者需要权衡成本效益进行抉择。

4 “ISFG 标准 2024”与国内相关标准的比较

4.1 中国司法鉴定行业对于 STR 序列命名的标准——“SF 标准 2018”

我国司法部于 2018 年发布了《SF/Z JD0105011-2018 法医学 STR 基因座命名规范》(以

下简称为“SF 标准 2018”) [20]，对于 STR 基于长度多态性和序列多态性的等位基因命名进行了简要规定，分别与“ISFG 标准 1997”和“ISFG 标准 2016”的相关建议一致。该标准的 STR 等位基因序列命名格式同“ISFG 标准 2016”中的全面命名法，包括了基因座名称、等位基因名称 (基于重复单位的重复次数)、参考基因组版本号、染色体号、核心重复序列起始和终止核苷酸位置、基序结构及侧翼序列变异 [20]。如图 2 示例，D13S317 基因座等位基因的 CE 分型为 12，与等位基因 11 的参考序列相比，侧翼序列在 82148001 和 82148069 两个位置 (GRCh38) 分别存在 SNP，基于“SF 标准 2018”可获得全面命名的序列字符串格式：D13S317 [12]-GRCh38-Chr13:82148025-82148068 [TATC]₁₂ 82148001-A; 82148069-T。

4.2 中国公共安全行业对于 STR 序列命名的标准——“GA 标准 2020”

我国公安部于 2020 年发布了《GA/T 1694-2020 序列多态 STR 等位基因命名规则》(以下简称“GA 标准 2020”) [21]，创新性地提出了基于通式结构和命名细则相结合的 STR 序列命名方法，详细描述了一套 STR 等位基因序列命名规则。图 3 展示了“GA 标准 2020”的命名通式，对于图 2 中示例的 D13S317 的等位基因 12 序列，按照“GA 标准 2020”的命名规则可表示为：12[-24sA/+1sT]。

4.3 国内外 STR 序列命名标准的比较

“ISFG 标准 2024”、“SF 标准 2018”、“GA 标准

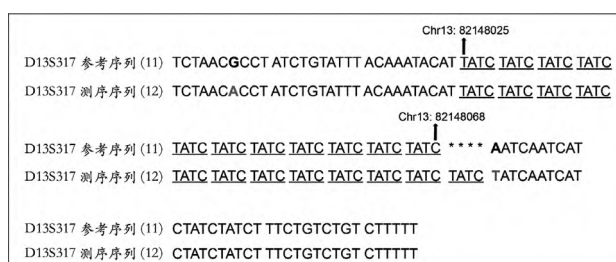


图 2 “SF 标准 2018”的 STR 序列命名 (改编自 [20])
Fig. 2 STR sequence nomenclature for SF standard 2018 (adapted from [20])

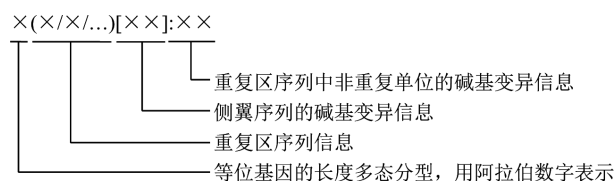


图 3 “GA 标准 2020”的 STR 序列命名 (摘自 [21])
Fig. 3 STR sequence nomenclature for GA standard 2020 (excerpted from [21])

表 1 “ISFG 标准 2024、SF 标准 2018 和 GA 标准 2020”对于 STR 序列命名的比较

Table 1 Comparisons of STR sequence nomenclature for ISFG standard 2024, SF standard 2018 and GA standard 2020

项目	“ISFG 标准 2024”	“SF 标准 2018”	“GA 标准 2020”
报告范围	重复区序列 + 侧翼序列, 定义了最小报告范围	重复区序列 + 侧翼序列, 未规定报告范围	重复区序列 + 侧翼序列, 未规定报告范围
命名形式	简化的括号格式序列字符串和 (或) 序列码	采用 ISFG 标准 2016 建议获得全面命名法的序列字符串	按照自定义命名通式和命名细则获得简化的序列字符串
命名举例 *	STRNaming: CE12_TATC[13]AATC[1]ATCT[3] SID: GNOQCFOWUQXXTZCXUUSDF TRDMTZXCZZKTVVMOOYMTFYQI BRGWBHYGE 或 D13S317_GNOQC 或 D13S317_12_GNOQ	D13S317 [12]-GRCh38- Chr13:82148025-82148068 [TATC]12 82148001-A; 82148069-T	12[-24sA/+1sT]
可读性	具有人可读性	具有人可读性	具有人可读性
溯源性	可还原序列	可还原序列	可还原序列
自动化	可使用开源软件进行命名	暂未有公开的配套软件发布	暂未有公开的配套软件发布
质量控制平台	STRidER 平台可进行命名质量控制	STRidER 平台可进行命名质量控制	暂未有公开的命名质量控制平台
CE 数据兼容性	可兼容	可兼容	可兼容

*: 均以“SF 标准 2018”中的 D13S317 等位基因 12 序列为例 (图 2)。对于“ISFG 标准 2024”, 使用 ISFG 确定的 D13S317 基因座最小报告范围 (GRCh38 Chr13:82148021-82148104), 序列为“ACATTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCACTATCTATCTTCTCTGCTGCTTTTT”。

2020”对于 STR 序列命名的比较详见表 1。

5 结语

在 DNA 测序技术日新月异发展的时代, STR 测序逐渐成为常规技术手段, 基于序列多态性的 STR 等位基因序列如何科学、有效地命名一直是法医学界悬而未决的问题, 阻碍了数据交流、共享及 DNA 数据库的构建。“ISFG 标准 2024”在调研多个实验室大量 STR 序列多态性数据的基础上, 就此问题提出了较全面的解决方案, 并提供了开源的分析资源及软件, 预期将促进基于序列多态性的 STR 数据交流和数据库建设。该标准可以为国内同行提供参考和借鉴, 推动我国 STR 序列多态性分析的标准化、自动化建设进程。

致谢 感谢来自“法医物证前沿”微信公众号编辑部的南方医科大学法医学院陈曼博士、安徽医科大学法医学系方雅婷博士、贵州医科大学法医学院张晗博士对 ISFG 标准 2024 翻译时所作的贡献。

参考文献

[1] Lincoln P J. DNA Recommendations--Further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems[J]. Forensic Sci Int, 1997, 87(3): 181-184.

- [2] Guo F, Liu Z, Long G, et al. High-resolution genotyping of 58 STRs in 635 northern Han Chinese with MiSeq FGx® forensic genomics system[J]. Forensic Sci Int Genet, 2023, 65: 102879.
- [3] Li R, Shen X, Chen H, et al. Developmental validation of the MGIEasy signature identification library prep kit, an all-in-one multiplex system for forensic applications[J]. Int J Legal Med, 2021, 135(3): 739-753.
- [4] Tao R, Qi W, Chen C, et al. Pilot study for forensic evaluations of the Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ System[J]. Forensic Sci Int Genet, 2019, 43: 102147.
- [5] 康克莱, 叶健, 季安全, 等. 法医学二代测序标准化进展与展望[J]. 中国法医学杂志, 2020, 35(1): 73-77.
- [6] Parson W, Ballard D, Budowle B, et al. Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements[J]. Forensic Sci Int Genet, 2016, 22: 54-63.
- [7] Gettings K B, Bodner M, Borsuk L A, et al. Recommendations of the DNA commission of the International Society for Forensic Genetics (ISFG) on short tandem repeat sequence nomenclature[J]. Forensic Sci Int Genet, 2024, 68: 102946.
- [8] Phillips C, Gettings K B, King J L, et al. “The devil’s

- in the detail”: release of an expanded, enhanced and dynamically revised forensic STR sequence guide[J]. *Forensic Sci Int Genet*, 2018, 34: 162-169.
- [9] Gettings K B, Borsuk L A, Ballard D, et al. STRSeq: a catalog of sequence diversity at human identification short tandem repeat loci[J]. *Forensic Sci Int Genet*, 2017, 31: 111-117.
- [10] Bodner M, Bastisch I, Butler J M, et al. Recommendations of the DNA commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal short tandem repeat allele frequency databasing (STRidER)[J]. *Forensic Sci Int Genet*, 2016, 24: 97-102.
- [11] Bodner M, Parson W. The STRidER report on two years of quality control of autosomal STR population datasets[J]. *Genes (Basel)*, 2020, 11(8): 901.
- [12] Gettings K B, Ballard D, Bodner M, et al. Report from the STRAND working group on the 2019 STR sequence nomenclature meeting[J]. *Forensic Sci Int Genet*, 2019, 43: 102165.
- [13] Hoogenboom J, Sijen T, van der Gaag K J. STRNaming: Generating simple, informative names for sequenced STR alleles in a standardised and automated manner[J]. *Forensic Sci Int Genet*, 2021, 52: 102473.
- [14] Young B, Faris T, Armogida L. A nomenclature for sequence-based forensic DNA analysis[J]. *Forensic Sci Int Genet*, 2019, 42: 14-20.
- [15] Willems T, Gymrek M, Highnam G, et al. The landscape of human STR variation[J]. *Genome Res*, 2014, 24(11): 1894-1904.
- [16] Phillips C. A genomic audit of newly-adopted autosomal STRs for forensic identification[J]. *Forensic Sci Int Genet*, 2017, 29: 193-204.
- [17] Pemberton T J, Sandefur C I, Jakobsson M, et al. Sequence determinants of human microsatellite variability[J]. *BMC Genomics*, 2009, 10(1): 612.
- [18] Hanson E K, Ballantyne J. Comprehensive annotated str physical map of the human Y chromosome: forensic implications[J]. *Leg Med*, 2006, 8(2): 110-120.
- [19] Willard H F, Skolnick M H, Pearson P L, et al. Report of the committee on human gene mapping by recombinant DNA techniques (Part 1 of 5)[J]. *Cytogenet Genome Res*, 1985, 40(1/4): 360-385.
- [20] 中华人民共和国司法部. 法医学 STR 基因座命名规范 (SF/Z JD0105011-2018)[S]. 2018.
- [21] 中华人民共和国公安部. 序列多态 STR 等位基因命名规则 (GA/T 1694-2020)[S]. 2020.

(收稿日期: 2024-01-31)

(上接第 338 页)

- cannabis sativa L. (hemp)[J]. *Molecules*, 2018, 23(10): 26-39.
- [7] Montero L, Meckelmann S W, Kim H, et al. Differentiation of industrial hemp strains by their cannabinoid and phenolic compounds using LC × LC-HRMS[J]. *Anal Bioanal Chem*, 2022, 414(18): 5445-5459.
- [8] 陈丹丹, 彭兴盛, 郑荣. 基于大麻二酚及 Δ^9 -四氢大麻酚含量的化妆品质量研究 [J]. *药物分析杂志*, 2022, 42(2): 346-351.
- [9] 朱敏航, 周靖, 徐云辉, 等. HPLC 法测定国内工业大麻花叶原料中四氢大麻酚、大麻二酚和大麻二酚酸的含量 [J]. *中国医药工业杂志*, 2022, 53(1): 95-100.
- [10] 赵志东, 余丽宇, 孟娇, 等. 固相萃取 - 液相色谱 - 串联质谱法测定尿液中四氢大麻酚和 Δ^9 -四氢大麻酚的含量 [J]. *理化检验 (化学分册)*, 2023, 59(10): 1156-1161.
- [11] 陈顺琴, 唐美, 黄江, 等. 超声提取 - 超高效液相色谱 - 串联质谱法测定毛发中常见毒品及代谢物的含量 [J]. *理化检验 (化学分册)*, 2022, 58(1): 58-64.

(收稿日期: 2023-11-13)

(上接第 360 页)

- 69-72, 76.
- [10] 张保生. 刑事错案及其纠错制度的证据分析 [J]. *中国法学*, 2013(1): 90-102.
- [11] 刘鑫, 蔡衡衡. 微量 DNA 证据解释对事实认定的影响和挑战——评《误导性 DNA 证据 - 司法误判的原因》[J]. *中国法医学杂志*, 2020, 35(5): 559-562.
- [12] Giannelli P C. The admissibility of laboratory reports in criminal trial: the reliability of scientific proof[J]. *Ohio State Law J*, 1988, 49(3): 671-700.
- [13] Imwinkelried E J, 著. 王进喜, 等译. 科学证据的秘密与审查 [M]. 北京: 中国人民大学出版社, 2020: 47.
- [14] *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 509 U.S. 579 (1993).
- [15] 常林. 谁是司法鉴定的“守门人”? ——关于司法鉴定管理问题的决定实施五周年成效评析 [J]. *证据科学*, 2010(18): 618-632.
- [16] 陈如超. 司法鉴定管理体制改革的的方向与逻辑 [J]. *法学研究*, 2016(1): 187-208.

(收稿日期: 2023-10-14)