

线粒体 DNA 高变区测序比对技术应用

桑志明, 周一涵, 石卫胜

河南周口三川司法鉴定中心, 河南 周口 466000

DOI:10.61369/MRP.2026080028

摘 要 : 线粒体 DNA 高变区由于有着母系遗传、拷贝数高、进化速率快等特点, 在法医学个体识别、物种亲缘鉴定、群体遗传研究中有着重要的应用价值。本文从技术解决方案的思路出发, 对线粒体 DNA 高变区测序比对技术的适用性层面进行了系统的分析, 说明了样本提取纯化、PCR 扩增、测序平台、序列比对算法等核心技术参数, 提出了包含实验流程设计、质量控制、数据分析、结果验证的技术实施方案。研究说明, 通过改善反应体系以及比对方案, 可以增加高变区序列数据的准确性以及重现性, 为相关领域给出可靠的技术支持。

关 键 词 : 线粒体 DNA; 高变区; 测序比对

Application of Mitochondrial DNA Hypervariable Region Sequencing and Alignment Technology

Sang Zhiming, Zhou Yihan, Shi Weisheng

Henan Zhoukou Sanchuan Forensic Identification Center, Zhoukou, Henan 466000

Abstract : Mitochondrial DNA hypervariable regions have important application value in forensic individual identification, species phylogenetic identification, and population genetics research due to their characteristics of maternal inheritance, high copy number, and fast evolution rate. This article systematically analyzes the applicability of mitochondrial DNA hypervariable sequencing alignment technology from the perspective of technical solutions. It explains the core technical parameters such as sample extraction and purification, PCR amplification, sequencing platform, and sequence alignment algorithm, and proposes a technical implementation plan that includes experimental process design, quality control, data analysis, and result verification. Research has shown that by improving the reaction system and comparison scheme, the accuracy and reproducibility of high variability region sequence data can be increased, providing reliable technical support for related fields.

Keywords : mitochondrial DNA; high variability zone; sequencing comparison

引言

线粒体 DNA 高变区是分子标记的重要靶标, 序列多态性给个体识别、进化研究给予了丰富的信息来源。伴随测序技术的一直发展, 高变区测序比对法已经变成了法医学和生物分类学中常用的手段。本文主要论述线粒体 DNA 高变区测序比对技术的适用性、关键技术、实施方案, 目的是建立一套有工程应用价值的技术解决方案。

一、技术适用性分析

(一) 法医学个体识别适用性

线粒体 DNA 高变区在法医学个体识别中有着独有的优势。由于线粒体 DNA 是母系遗传的, 所以其高变区序列在毛发、骨骼等核 DNA 含量较低的检材中依旧可以有较高的检出率。对于高度降解的生物样本, 高变区短片段扩增方略可以有效地避免长片段模板断裂造成的检测失败。核 DNA STR 分型不能得到有效结果的情况下, 线粒体 DNA 高变区序列比对可以给个体识别给出补充性的遗传学证据, 其鉴别能力取决于群体中单倍型频率分布。

(二) 物种亲缘鉴定适用性

线粒体 DNA 高变区在物种亲缘鉴定上具有明确的适用范围。高变区序列在种内个体间保持相对保守的遗传背景, 种间存在明显的序列差异, 所以它变成区分物种分类地位的分子标记。在哺乳动物亲缘鉴定中, 用不同物种高变区序列的碱基替换率及插入缺失特征来建立系统发育关系树。线粒体 DNA 存在异质性以及核线粒体片段插入现象, 单一高变区序列在近缘物种鉴定中容易产生误判, 需要融合多个高变区片段或者核基因标记来融合判断。

(三) 群体遗传研究适用性

线粒体 DNA 高变区在群体遗传学研究中有着稳定的适用基

础。高变区序列突变积累速率比编码区快，可以在较短的进化时间尺度上记录群体的遗传分化事件。对不同地理群体高变区序列的单倍型多样性、核苷酸多样性指数实行分析，可以追踪母系遗传谱系的迁徙路线、分化历史。线粒体 DNA 高变区对样本量要求较低，小样本群体也能得到有着统计意义的单倍型频率数据，所以在珍稀物种或者古 DNA 群体遗传研究中有较强的技术适应性。

（四）降解样本检测适用性

线粒体 DNA 高变区在降解样本检测中有着很强的技术适用性。生物样本在环境因素的影响下发生 DNA 链断裂时，线粒体 DNA 的拷贝数高达核 DNA 的数百倍到数千倍，短片段高变区序列还是可以在严重降解的模板中成功扩增。对古代骨骼、甲醛溶液固定组织、火灾现场炭化样本设计扩增片段长度为 150bp 到 300bp 的引物对，可以大幅度提高 PCR 反应的成功率。高变区序列不需要繁复的等位基因分离步骤，在模板量很低的情况下也能直接测序得到可读序列。

二、关键技术参数

（一）样本提取纯化技术

线粒体 DNA 高变区测序的第一步就是样本提取纯化，该技术会直接影响到后面 PCR 扩增的效率以及测序数据的质量。血液样本用酚氯仿抽提法或者磁珠法都可以得到较高的总 DNA 纯度，磁珠法由于操作自动化程度高、避免有机溶剂残留，更适用于批量检材处理。骨骼样本需要经过脱脂以及脱钙预处理，用 EDTA 溶液在 4℃ 条件下振荡脱钙 48—72h，转速 150rpm，可以释放骨组织中的 DNA 分子。提取完成后，借助紫外分光光度计检测 DNA 浓度及纯度，要求 A260/A280 比值在 1.8—2.0 区间，A260/A230 比值大于 2.0，浓度不低于 10 ng/μL。对于微量样本，用低吸附离心管和带滤芯吸头，防止 DNA 在转移过程中非特异性损失。

（二）PCR 扩增技术参数

PCR 扩增是线粒体 DNA 高变区测序的主要步骤，反应体系参数要经过严格改良。反应总体积一般设定为 25 μL 或 50 μL，其中模板 DNA 加入量为 1—5 ng，10×PCR 缓冲液 2.5 μL 或 5.0 μL，dNTP 混合物浓度为 200 μM，上下游引物终浓度各为 0.2—0.4 μM。高变区 I 及高变区 II 的扩增，一般用 400—600bp 的引物设计，退火温度根据引物 T_m 值来确定，一般为 55—60℃，延伸时间为 1min/kb，一般为 1min。由于线粒体 DNA 高变区存在碱基替代热点，反应体系中加入适量的 BSA 可以消除 PCR 抑制剂的干扰，MgCl₂ 浓度控制在 1.5—2.5 mM。热循环参数中，预变性温度 95℃ 持续 5 min，变性温度 94—95℃ 持续 30 s，循环数设定为 30—35 个，末次延伸 72℃ 持续 7 min。

（三）测序平台技术参数

测序平台的选择直接决定线粒体 DNA 高变区序列的读取长度以及数据通量。目前常用的测序平台分为 Sanger 测序仪和高通量测序仪两类，技术参数差别较大。如表 1 所示，Sanger 测序平台的读取长度为 500—800bp，适合于已知扩增片段的定向测序，运

行时间约为 3—5h，单次反应成本低，数据准确性高于 99.9%。高通量测序平台 Illumina MiSeq，读取长度为 150—300bp，单次运行可产生数百万条序列读段，适合混合样本或者异质性分析，运行时间约为 24—48h，单次运行成本高但是均摊到单样本后有着规模效应。法医学常规检案中，Sanger 测序由于操作流程标准化程度高、数据解读直观，依然是线粒体 DNA 高变区测序的首选平台。对低频异质性位点做深度测序的情况下，高通量测序平台能给予更高的序列包含深度。

表 1 常用测序平台技术参数对比

技术参数	Sanger 测序平台	高通量测序平台
读取长度 /bp	500—800	150—300
数据通量 / 条	1—2	1 × 10 ⁶ —1 × 10 ⁸
运行时间 /h	3—5	24—48
单次成本 / 元	50—100	5000—10000
碱基准确性 /%	99.9	99.0—99.9
适用场景	单一样本定向测序	多样本混合测序

（四）序列比对算法技术

序列比对算法是实行线粒体 DNA 高变区序列分析的重要方法。获得原始测序峰图或者读段数据之后，需要凭借序列拼接以及比对来确定样本的高变区单倍型。对于 Sanger 测序数据，使用 CodonCode 或者 SeqMan 软件实行序列拼接，最小匹配碱基数设为 20bp，重叠区域一致性阈值大于 90%。拼接完成后，用 consensus 序列及修订后的剑桥参照序列 rCRS 实行比对，用局部比对算法如 Smith-Waterman 算法，匹配得分为 +1，错配罚分为 -3，空位开放罚分为 -5，空位延伸罚分为 -2。对高通量测序数据用 BWA-MEM 算法将读段映射到参照序列上，种子长度设为 19bp，最大错配数设为 4。比对结果经过 SAMtools 格式转换之后，用 IGV 可视化工具检查高变区核心位点的包含深度和碱基分布频率，保证主要碱基频率大于背景噪声阈值。

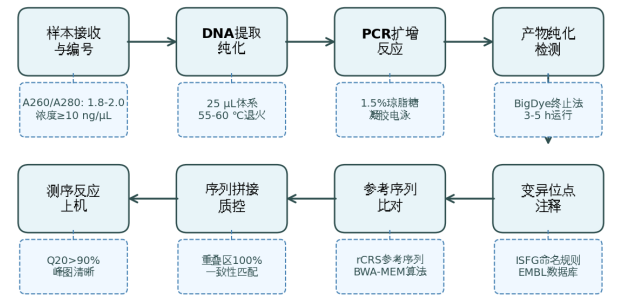


图 1 线粒体 DNA 高变区测序比对技术流程图

图 1 为线粒体 DNA 高变区测序比对技术流程，共 8 个连续环节，各个环节用标准化操作实现技术衔接。样本接收环节建立唯一的编号体系，保证样本可以追溯。DNA 提取纯化环节要求 A260/A280 比值在 1.8—2.0 区间，浓度不低于 10 ng/μL。PCR 扩增环节使用 25 μL 反应体系，退火温度 55—60℃。产物纯化检测环节用 1.5% 琼脂糖凝胶电泳验证后进入测序反应，Sanger 测序运行时间 3—5h。数据分析阶段依次完成序列拼接质控、参照序列比对、变异位点注释，序列拼接要求重叠区 100% 一致，参照比对使用 rCRS 序列以及 BWA-MEM 算法，变异注释按照 ISFG 命

名规则, 比对 EMBL 数据库。该流程图给出了各个环节的技术参数接口, 给实验室标准化操作给予可视化的依据。

三、技术实施方案

(一) 实验流程设计

线粒体 DNA 高变区测序比对的实验流程要遵照标准化及可追溯的原则。整体流程分为样本接收、DNA 提取、PCR 扩增、产物纯化、测序反应、数据分析和报告出具 7 个连续环节。样本接收环节要记录样本类型、重量或者体积、保存条件、送检时间, 建立唯一的样本编号体系。DNA 提取环节根据样本类型选择相应的提取试剂盒, 每批次实验设置阳性对照以及阴性对照各 1 份。PCR 扩增环节使用巢式 PCR 或者半巢式 PCR 方针, 第一轮扩增使用外侧引物包含高变区全长, 第二轮扩增使用内侧引物分段获取短片段序列, 两轮 PCR 产物分别用 1.5% 琼脂糖凝胶电泳验证。测序反应环节按照平台操作规程实行 BigDye 终止法测序或者文库建立测序, 产物纯化后上机检测。数据分析环节由两名技术人员独立实行序列比对和单倍型判定, 结果一致后才能进入报告出具阶段。

(二) 质量控制方案

质量控制方案是保证线粒体 DNA 高变区测序数据可靠性的必要手段。试剂方面所有的 PCR 引物及测序引物都要经过 HPLC 纯化, 防止未合成片段干扰反应的特异性。在实验操作上, PCR 前区以及 PCR 后区物理隔离, 实验台面用 10% 次氯酸钠溶液擦拭, 紫外线照射 30min 以上。人员层面操作人员应佩戴一次性帽子、口罩、无粉手套, 防止外源 DNA 污染。数据层面每批次实验设置试剂空白对照、阳性对照、阴性对照, 阳性对照用已知单倍型的标准 DNA 样本, 空白对照用去离子水代替模板 DNA。当对照样本出现预期结果偏差时, 该批次全部样本都必须重新检测。另外建立室内质控图记录每月实验成功率、污染率指标, 成功率低于 95% 或者污染率高于 2% 时启动原因排查程序。

(三) 数据分析流程

数据分析流程包含原始数据质控、序列拼接、参照比对、变异注释、单倍型判定 5 个步骤。原始数据质控环节, Sanger 测序峰图 Q20 值大于 90% 的碱基比例大于 80%, 高通量测序数据去除接头序列和低质量读段, 质量阈值设为 Q30。序列拼接环节, 正

向及反向测序读段需要至少 100bp 的有效重叠区域, 重叠区一致性为 100% 才可以合并。参照比对环节, 将拼接后的 consensus 序列和 rCRS 做全局比对, 允许有 1-2bp 的空位差异。变异注释环节, 记录高变区 I 的 16024-16365bp 区间及高变区 II 的 73-340bp 区间内所有的碱基替换以及插入缺失位点, 变异命名按照 ISFG 命名规则。单倍型判定环节, 将检出的变异位点以及 EMBL 单倍型数据库实行比对, 匹配率大于 95% 就可以赋予相应的单倍型编号。对于存在异质性位点的样本, 记录主要碱基以及次要碱基的峰高比例或者读段比例。

(四) 结果验证标准

结果验证标准是保证线粒体 DNA 高变区测序比对结论准确性的最后技术关口。验证工作分为实验重复验证及数据库比对验证两个方面。实验重复验证要求, 首次检出的罕见变异或者新单倍型, 用另一对独立设计的引物对同一样本实行重新扩增和测序, 两次独立实验得到的序列一致性要达到 100%。数据库比对验证要求, 将检出的高变区序列上传到 GenBank 或者 MitoFish 等公共数据库实行 BLAST 检索, 检索结果中最高相似度序列的覆盖率要大于 98%, 一致性要大于 99%。对于法医学应用, 还要计算随机匹配概率, 即该单倍型在目标群体中的出现频率, 频率数据来源于已发表的群体调查文献或者在线频率数据库。当随机匹配概率小于 0.5% 的情况下, 就可以认为该线粒体 DNA 高变区序列在个体识别中有着足够的排除能力。验证借助的序列数据以及实验记录要存档保存, 保存期限不少于 10 年。

四、结语

线粒体 DNA 高变区测序比对技术属于分子生物学范畴内的核心分析手段, 技术成熟度已经可以满足常规工程应用的要求。本文从适用性分析、核心技术参数、实施方案三个方面建立了完整的解决方案, 突出标准化流程和质量控制对数据可靠性的保障作用。实际应用中根据样本类型、检测目的合理选择测序平台及比对方略, 防止技术方法的盲目使用。伴随测序成本一直减少, 生物信息学算法不停改良, 线粒体 DNA 高变区测序比对技术在更多领域将发挥出它的应用价值, 相关技术规范以及质量标准的更进一步精进, 也会使该技术规范化应用。

参考文献

- [1] 张浩, 李佳倩, 陈凯欣, 等. 中华鳖“鄱湖 1 号”线粒体基因组全序列组装及分析 [J/OL]. 动物学杂志, 1-15[2026-06-01].<https://doi.org/10.13859/j.cjz.202625085>.
- [2] 杨雅云, 蒋欣悦, 费亮, 等. 松油烯-4-醇对鸡胚原代心肌细胞氧化应激和线粒体生物发生的影响 [J]. 中国兽医学报, 2026, 46(5): 1083-1090. DOI: 10.16303/j.cnki.1005-4545.2026.05.16.
- [3] 朱婷婷, 尚碧月, 吕李飞, 等. 温脾益气化饮方对心肌梗死后心衰大鼠线粒体氧化应激及心室重塑的影响 [J/OL]. 时珍国医国药, 1-8[2026-06-01].<https://link.cnki.net/urlid/42.1436.r.20260510.2132.044>.
- [4] 彭飞, 徐源, 刘诗琪, 等. 三仁汤调控线粒体自噬抑制 cGAS-STING/NF- κ B 通路改善非酒精性脂肪性肝病的机制研究 [J/OL]. 中国医院药学杂志, 1-10[2026-06-01].<https://link.cnki.net/urlid/42.1204.r.20260424.1858.002>.
- [5] 钱雪松. 联合应用线粒体 SNP 和线粒体 DNA 高变区测序排除同一母系鉴定 1 例 [J]. 法医学杂志, 2021, 37(4): 602-604.