**法医物证多人身份鉴定问题**

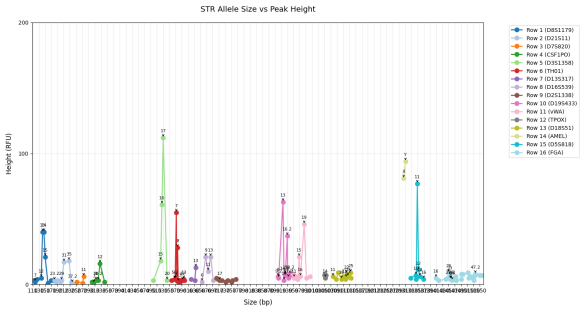
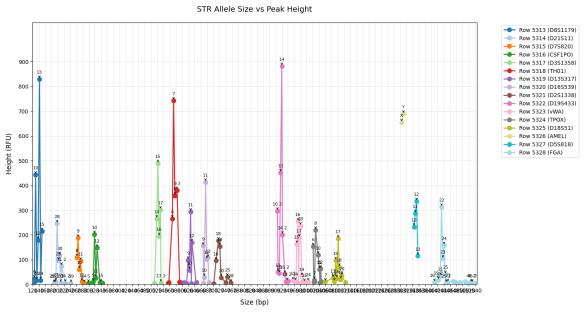
犯罪现场法医物证鉴定是关系到国家安全、公共安全、人民生命财产安全和社会稳定的重大问题。目前法医物证鉴定依赖DNA分析技术不断提升。DNA检验的核心是STR（Short Tandem Repeat，短串联重复序列）分析技术，STR的核心序列重复次数存在个体差异多态性，因此STR也被称为细胞的DNA指纹。

STR基因座是染色体上一个特定的物理位置，等位基因是同一基因座上不同表现形式的DNA序列。在STR图谱中，每一个主峰代表一个等位基因，其size表示该STR等位基因的DNA片段长度，不同size对应不同的等位基因，height是峰高，反映该等位基因的DNA量，可用于判断样本是否为混合样本。每个个体在每个基因座上拥有两个等位基因，基因型指的是个体基因座上等位基因的组合，不同个体的基因型常不同，因此可检查特定基因座上的基因型来进行身份鉴定。

对于多人犯罪案件中的DNA物证，其混合STR图谱的分析是案件侦破的关键。混合STR图谱中包含了多人的DNA信息，需要正确分析混合数据中各组分的贡献者构成和比例，作为案情判断依据之一。

请你们团队研究解决下面问题：

问题1 混合STR图谱分析的首要问题是判断贡献者人数。贡献者人数的正确与否决定着分析结果的准确率。依据附件1中混合STR图谱数据（如图1所示）设计算法或模型，用于识别某一混合样本中的贡献者人数，并评估其准确性。

(a)2人混合图谱 （b）5人混合图谱

图1 不同人数混合图谱示意图

问题2 在分析出贡献者人数后，还需要判断各贡献者的混合比例。当贡献者比例接近时，等位基因可能重叠，导致误判基因型。明确比例有助于更精准地分析混合图谱。依据附件2中混合ST图谱数据（如图2所示）设计算法或模型，用于识别某一混合样本中的贡献者比例，并评估其准确性。

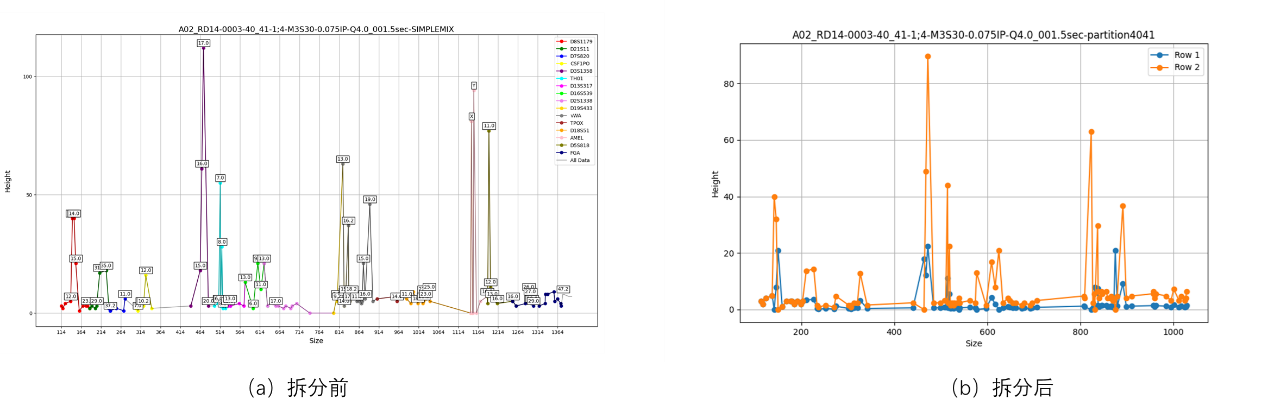


图2 2人混合图谱拆分示意图

问题3 根据附件1与附件2的混合STR图谱数据以及附件3中各个贡献者的基因型，设计算法或模型，用于推断某一混合STR图谱中各个贡献者对应的基因型，并评估其准确性。

问题4 依据附件4中混合STR图谱数据（如图3所示）设计算法或模型，用于减少混合样本中噪声的干扰，以提高混合样本分析的准确性。

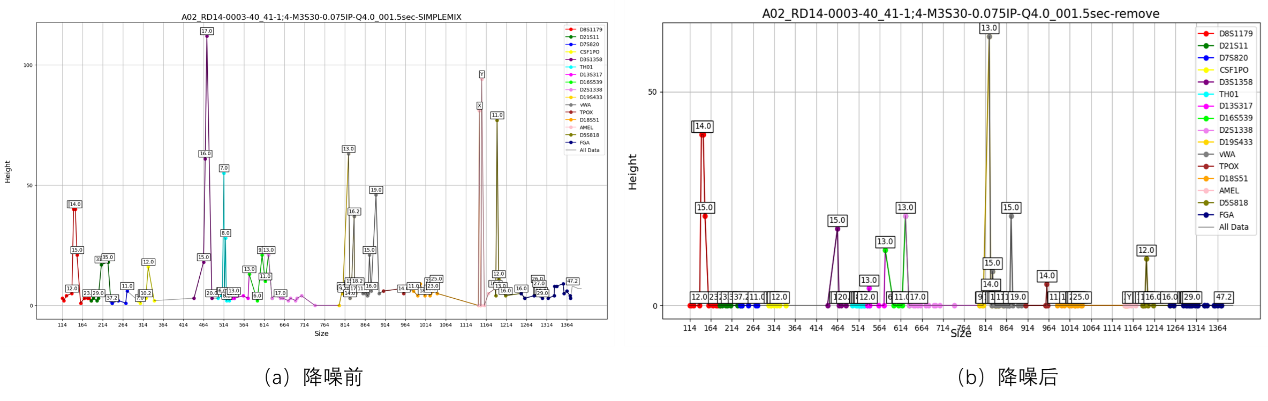


图3 2人混合图谱降噪示意图

数据集及其说明见附件：

链接:https://pan.baidu.com/s/1aNpk0oONWA6w7JR7-PYGFg?pwd=3uu6 提取码: 3uu6