

对未知关系个体进行系谱推断的自动分析系统

张文杰^{1,2}, 李彩霞², 于迪², 管珊珊¹, 童梦洁^{2,3}, 范虹^{1*}, 刘京^{2,4*}

(1. 陕西师范大学 计算机科学学院, 中国陕西 西安 710119; 2. 公安部物证鉴定中心 法医遗传学公安部重点实验室 现场物证溯源技术国家工程实验室, 中国北京 100038; 3. 合肥市公安局刑警支队, 中国安徽 合肥 230000; 4. 中国政法大学 证据科学教育部重点实验室, 中国北京 100088)

摘要: 基于短串联重复序列(short tandem repeat, STR)和单核苷酸多态性(single nucleotide polymorphism, SNP)的亲缘关系推断在群体遗传学、法医学等领域应用广泛,但目前没有成熟的系统可以实现针对 STR 数据和 SNP 数据的自动批量化处理。为弥补这一空白,项目组开发了一款基于 STR 数据和 SNP 数据的 DNA 系谱推断系统(DNA Genealogy Analyzer Version 1.0, DGA v1.0),该系统可运用 STR 和 SNP 亲缘关系预测算法计算未知亲缘关系群体中的所有亲缘关系。系统核心算法部分基于 Python 语言实现,前端部分采用 HTML 和 CSS 语言编写,整体采用 Python 的 Flask 框架开发。该系统与已有分析方法相比,结果一致性为 100%,运算效率提高了 800%以上,在群体遗传学等领域有重要应用价值。

关键词: 亲缘关系推断; 短串联重复序列(STR); 单核苷酸多态性(SNP); 自动批量化; 法医系谱推断

中图分类号: Q811.4

文献标识码: A

文章编号: 1007-7847(2022)06-0515-07

An Automatic Analysis Software System for Forensic DNA Pedigree Inference

ZHANG Wen-jie^{1,2}, LI Cai-xia², YU Di², GUAN Shan-shan¹, TONG Meng-jie^{2,3},
FAN Hong^{1*}, LIU Jing^{2,4*}

(1. School of Computer Science, Shaanxi Normal University, Xi'an 710119, Shaanxi, China; 2. Institute of Forensic Science, Ministry of Public Security & MPS' Key Laboratory of Forensic Genetics & National Engineering Laboratory for Forensic Science, Beijing 100038, China; 3. Criminal Police Detachment of Hefei Public Security Bureau, Hefei 230000, Anhui, China; 4. Ministry of Education Key Laboratory of Evidence Science, China University of Political Science and Law, Beijing 100088, China)

Abstract: Kinship inference based on short tandem repeat (STR) and single nucleotide polymorphism (SNP) is widely used in population genetics, forensics and other fields, but there is no mature system that can realize automatic batch processing of STR and SNP data. To fill this gap, a DNA pedigree inference system based on STR and SNP data (named DNA Genealogy Analyzer Version 1.0, DGA v1.0) was developed, which can use STR and SNP relationship prediction algorithms to calculate all kinships in the group of unknown relatives. The core algorithm part of the system is implemented based on Python language, the front-end part is written in HTML and CSS language, and the whole is developed using Python Flask framework. Compared with the existing analysis methods, the system has 100% result consistency and its computing efficiency is increased by more than 800%. The system would have important application value in population genetics and other fields.

Key words: kinship inference; short tandem repeat (STR); single nucleotide polymorphism (SNP); automatic

收稿日期: 2022-05-17; 修回日期: 2022-08-31; 网络首发日期: 2022-10-13

基金项目: 国家自然科学基金面上项目(82171870); 陕西省自然科学基金重点项目(2022ZJ-39); 公安部技术研究计划重点项目(2021JSZ15); 中央级公益性科研院所基本科研业务费专项资金资助项目(2021JB004); 法医遗传学公安部重点实验室开放课题(2021FGKFKT07)

作者简介: 张文杰(1997—), 男, 山东沂水人, 硕士研究生; *通信作者: 范虹(1969—), 女, 宁夏平罗人, 博士, 陕西师范大学教授, 主要从事模式识别与智能信息处理研究, Tel: 029-85310161, E-mail: fanhong@snnu.edu.cn; 刘京(1989—), 男, 山东沂水人, 博士研究生, 主检法医师, 主要从事法医遗传学研究, E-mail: biojing@yeah.net.

batching; forensic genealogy inference

(*Life Science Research*, 2022, 26(6): 515~521)

随着高通量测序技术的兴起, 基于基因组数据进行的亲缘关系推断在群体遗传学、法医系谱学等领域应用越来越广泛。在群体遗传全基因组关联分析(genome-wide association study, GWAS)中, 使用亲缘关系推断技术删除存在亲缘关系的样本, 可降低关联分析结果的假阳性率^[1]; 在法医学中, 亲缘关系推断技术既可用于亲子鉴定^[2], 也可作为案件侦查提供线索^[3]。现阶段用于亲缘关系推断的主要遗传标记是短串联重复序列(short tandem repeat, STR)基因座和单核苷酸多态性(single nucleotide polymorphism, SNP)^[4-6]。STR是人类基因组中广泛分布的一类遗传标记, 因其高鉴别能力, 并且相关试剂盒具有灵敏度高、操作简单和成本较低等特点, 现已成为我国亲缘关系鉴定最常用的遗传标记^[7]。目前, 国内已有一系列基于STR数据的亲缘关系推断行业标准和算法, 例如: 《生物学全同胞关系鉴定技术规范》《法庭科学DNA亲子鉴定规范》等^[8-9], 可预测2级以内亲缘关系。近年来, SNP遗传标记逐渐成为群体遗传领域研究的热点^[10-11], 全基因组SNP数据位点数达到几十万甚至上百万, 利用全基因组SNP数据进行系谱推断的方法为预测相对较远的亲缘关系提供了可能^[12], 可以准确预测5级以内的亲缘关系。“金州杀手”案是第一宗使用系谱推断技术破获的悬案^[13], 此案破获后, 警方利用该技术为两百余例案件提供了侦查线索^[14]。但全基因组SNP数据的分析难度较大, 对使用者的数据处理能力要求较高。

本项目组开发了一套操作简单、交互性强的分析系统——DNA系谱推断系统(DNA Genealogy Analyzer Version 1.0, DGA v1.0), 该系统将复杂的操作流程简单化、自动化和界面化, 使用者无需熟练掌握Python语言和Plink等生物信息软件也可快速使用, 降低了亲缘关系预测难度, 节省了亲缘关系预测时间, 提高了亲缘关系预测效率。

1 材料与方法

1.1 STR数据算法

1.1.1 ITO算法

ITO算法是一种计算两个体间存在全同胞(full sibling, FS)、半同胞(half sibling, HS)、一级堂表亲(first cousin, 1C)和二级堂表亲(second cousin,

2C)等亲缘关系似然比(likelihood ratio, LR)的算法^[15]。对于常染色体STR基因座, 在给定个体基因型时, 通常使用 R 表示个体间存在亲缘关系的机会大小^[15], $R=\psi_0O+\psi_1T+\psi_2I$, 其中 ψ_0 、 ψ_1 和 ψ_2 分别表示两个体出现0个、1个或2个同源基因的概率, O 、 T 、 I 分别表示两个体中一个个体出现0个、1个或2个同源基因时, 另一个体出现相同基因型的概率。ITO算法推断亲缘关系的主要判断标准是亲权指数(paternity index, PI), PI值(I_p)是一个典型的LR。计算LR首先需建立检验假设, 假设两个体存在某种血缘关系, 记作 H_0 , 假设两个体间不存在亲缘关系, 记作 H_1 , 假设的机会(R_0 和 R_1)可根据前述 R 的公式计算^[15], PI值可通过公式 $I_p=R_0/R_1$ 算得。在计算 I_p 时, 需要用到等位基因频率, ITO算法中STR基因座的等位基因频率参考吴薇薇等^[16]的研究。若使用 n 个相互独立的STR基因座对两个体是否存在某种亲缘关系进行预测, 则需将单个STR基因座的 I_p 进行累乘求得累计亲权指数(cumulative paternity index, CPI; 记作 I_{cp}), 公式如下:

$$I_{cp}=I_{p1}\times I_{p2}\times I_{p3}\times I_{p4}\times\cdots\times I_{pn}=\prod_{i=1}^n I_{pi}$$

1.1.2 IBS算法

基于STR数据的状态一致性(identity by state, IBS)算法是在没有双亲数据的支持下鉴定个体间是否存在FS关系的算法, 两个体在同一STR基因座上可能出现相同的等位基因, 这些等位基因被称为状态一致性等位基因^[8]。在1个STR基因座上, 两个体间的状态一致性等位基因个数记作 N_{ibs} , 将 n 个相互独立的STR基因座上的相同等位基因数量进行累加, 记作 N_{ibs} ^[8]。使用常用的19个常染色体STR基因座, 计算得到 N_{ibs} , 根据 N_{ibs} 对两个体间是否存在FS关系给出倾向性意见, 如果 N_{ibs} 大于等于22, 通常给出两个体为FS关系的倾向性意见; 如果 N_{ibs} 小于等于13, 通常给出两个体为无关个体的倾向性意见; 当 N_{ibs} 大于13小于22时, 通常无法给出倾向性意见。

1.1.3 PTT算法

三联体亲子鉴定(parentage testing of trios, PTT)算法是指被检测男子、孩子生母与孩子或者被检

测女子、孩子生父与孩子的亲子鉴定算法^[9],通过对单个 STR 基因座上的等位基因进行分析,计算单个 STR 基因座的 I_p 和非父或非母排除概率(power of exclusion, PE; 记作 E_p)。其将 n 个相互独立的 STR 基因座通过累乘求得 I_{cp} 和累计非父或非母排除概率(cumulative power of exclusion, CPE; 记作 E_{cp}),并根据 I_{cp} 和 E_{cp} 大小对是否存在亲子关系给出倾向性意见。PTT 算法的主要判断标准是 I_{cp} , I_{cp} 越大,存在亲子关系的可能性越大。

1.2 SNP 数据算法

1.2.1 IBS 算法

基于 SNP 数据的 IBS 算法是等位基因频率估计的状态一致性共享统计量算法^[7]。该算法假设各 SNP 标记间相互独立,通过计算每个样本中 SNP 标记的等位基因频率和 IBS 共享等位基因数量,计算亲缘关系系数 ϕ ,并通过 ϕ 和零 IBD 共享统计量 π_0 推断亲缘关系等级, ϕ 越大表示亲缘关系越近,而 π_0 越大,表示亲缘关系越远。该算法能准确预测 1~4 级内的亲缘关系,并且其受位点检出率的影响较小。

1.2.2 IBD 算法

共祖片段(identity-by-descent, IBD)长度算法通过检测各个体从一个共同祖先继承的相同 DNA 片段长度和数量^[7],判断亲缘关系远近。其首先对 SNP 数据中的冗余信息进行过滤,筛选可用位点,并将筛选后的数据进行格式转化,然后进行同源染色体分离,查找各染色体上的 IBD 匹配片段并将其合并,最后通过计算 IBD 长度和数

量预测个体间亲缘关系。

1.3 测试样本

1.3.1 样本来源

样本来源于实验室志愿者,共计 5 个家系 253 份样本,参考刘京等^[8]的法医系谱亲缘关系等级对其进行亲缘关系等级划分。其中,SNP 数据可用样本为 253 份,总共 4 925 对亲缘关系,包括双胞胎(monozygotic twin, MZ)、亲子(parent-off-spring, PO)、全同胞(FS)、2 级(2nd)、3 级(3rd)、4 级(4th)、5 级(5th)、6 级(6th)、7 级(7th)、8 级(8th)、9 级(9th)等亲缘关系,253 份 SNP 数据样本的亲缘关系等级分布如图 1A 所示;STR 数据可用样本来自两个家系,共 41 份样本,有 133 对亲缘关系,包括亲子(PO)、全同胞(FS)、2 级(2nd)、3 级(3rd)、4 级(4th)、5 级(5th)、6 级(6th)等亲缘关系,41 份 STR 数据样本的亲缘关系等级分布如图 1B 所示。所有样本在采集前均签署知情同意书,本研究通过了公安部物证鉴定中心伦理论证(2021-006)。

1.3.2 DNA 提取与检测

所有测试样本均先使用 MagAttract M48 DNA Manual 试剂盒(Qiagen 公司,德国)提取 DNA,再使用 NanoDrop 2000c 分光光度计(Thermo Scientific 公司,美国)进行 DNA 定量。

使用 PowerPlex 21 System 法医检测试剂盒(Promega 公司,美国)对提取产物进行常染色体 STR 分型检测;使用 WeGene V2 基因芯片(安澜智能有限公司)进行全基因组 SNP 检测,经数据格式转换和处理后获得约 60 万常染色体 SNP 位点分型。

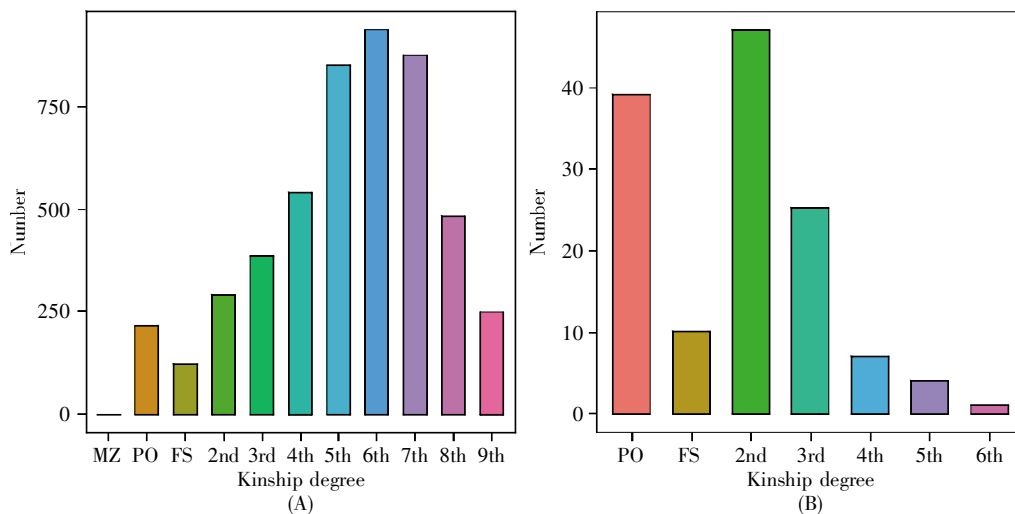


图 1 样本间各亲缘关系数量分布

(A) 253 份 SNP 数据的亲缘关系等级分布; (B) 41 份 STR 数据的亲缘关系等级分布。

Fig.1 The quantity distribution of each kinship degree among samples

(A) The kinship rank distribution of 253 SNP data; (B) The kinship rank distribution of 41 STR data.

2 系统简介

2.1 系统设计思路

DGA v1.0 包括注册登录模块、系统简介模块、亲缘关系预测模块以及用户管理模块,其中亲缘关系预测模块为系统的核心模块。将基于 STR 和 SNP 数据的各种算法分别进行封装、简化运行、调试等操作,实现在系统页面选择上传数据,一键运行相应算法,最后预测结果自动下载的功能,具体功能模块如图 2 所示。

2.2 系统功能实现

DGA v1.0 包括前端交互页面和后台算法两部分,核心算法部分使用 Python 语言实现,前端部分采用 HTML 和 CSS 语言编写,整体采用 Python 的 Flask 框架进行开发,注册登录模块、系统简介模块、亲缘关系预测模块以及用户管理模块等四大模块功能均已实现。DGA v1.0 将亲缘关系算法模块化,降低了算法使用的复杂度,同时前端交互界面可以让用户有更好的使用体验,使系统使用起来方便简洁。

2.3 系统输入与输出

DGA v1.0 中包含基于 STR 数据和 SNP 数据的多种算法,不同算法输入有所不同。针对 STR 数据的 IBS、ITO、PTT 算法的输入数据为表格文件,文件中需包含样本信息、STR 基因座和等位基因等信息。针对 SNP 数据的 IBS、IBD 算法的输入

数据是全基因组 SNP 分型文件,包括 fam 文件、bed 文件和 bim 文件, fam 存储的是样本家系等信息, bed 存储基因型信息, bim 文件存储每个遗传变异(通常是 SNP)的相关信息。

用户在进行亲缘关系预测时需上传数据,并选择相对应的算法进行计算,运算成功后,可在系统页面查看结果文件,该结果文件将待预测亲缘关系的样本信息、所存在的亲缘关系全部列出。

3 结果与分析

3.1 STR 数据准确性评估

3.1.1 ITO 算法准确性评估

选取 41 份样本中的 FS、HS、1C 和 2C 共 94 对亲缘关系进行准确性验证,比较预测结果与真实亲缘关系(表 1)。从比较结果可以看出, DGA v1.0 的 ITO 算法在 1 级和 2 级亲缘关系预测中准确率较高,且置信区间准确率可达 100%。由于 3 级和 5 级亲缘关系样本数量太少,所以无法正确评估 ITO 算法在 3 级和 5 级亲缘关系中的预测准确率,需要增加样本继续验证。

3.1.2 IBS 算法准确性评估

使用 DGA v1.0 中的 STR-IBS 算法对 41 个样本中的 10 对 FS 关系进行准确性评估(图 3),结果显示 10 对 FS 关系的 N_{IBS} 均大于等于 22,准确率为 100%,说明 STR-IBS 算法可以很好地用于 FS 关系预测。

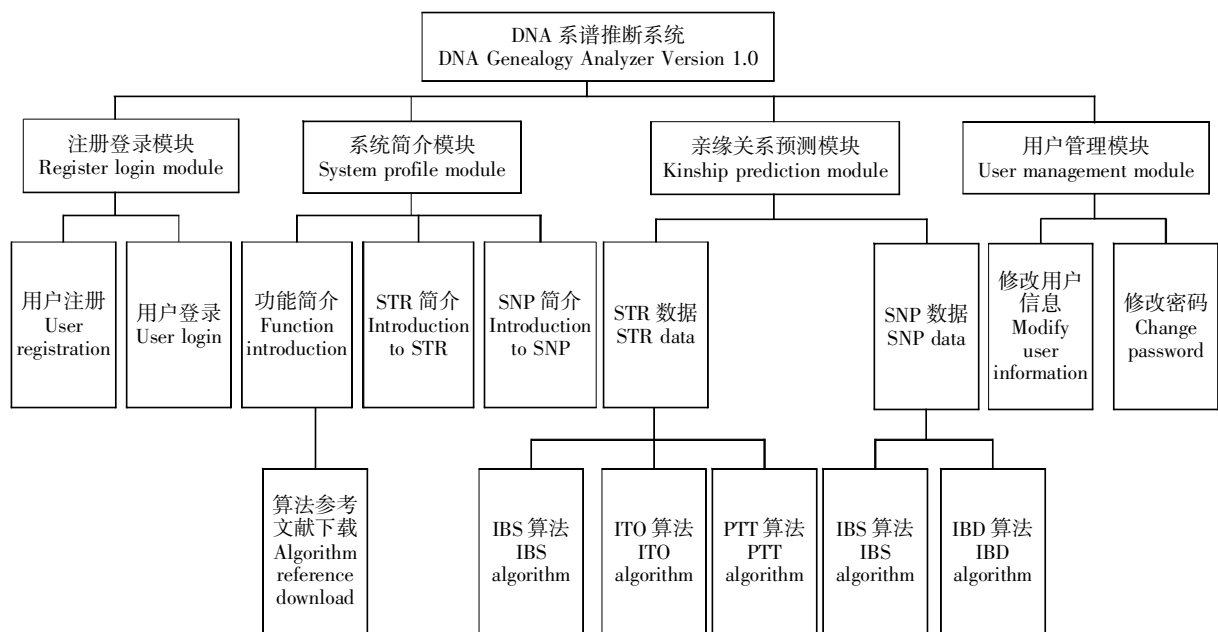


图 2 系统功能图

Fig.2 System function diagram

表 1 ITO 算法预测 41 个样本 STR 数据亲缘关系的准确性评估
Table 1 The prediction accuracy of kinship relationship of STR data from 41 samples by ITO algorithm

Real kinship	Predicted kinship				AC	CIA
	FS (1)	HS (2)	1C (3)	2C (5)		
1	10				100.0%	100.0%
2	7	32	8		68.1%	100.0%
3		6	10	9	40.0%	64.0%
4				7		
5			1	3	75.0%	75.0%
6				1		

注: AC 表示准确率,即正确预测两个相关个体真实亲属关系程度的概率; CIA 表示置信区间准确率,即两个相关个体被预测在其真实亲属关系的 1 级范围内,且不考虑 7 级以上或未知关系的概率。

Notes: AC represents the accuracy rate, that is, the probability of correctly predicting two related individuals as the degree of their real kinship; CIA represents the accuracy rate of confidence interval, that is, the probability that two related individuals is predicted to be within the 1st-degree range of their real kinship, and above 7th-degree or unknown relationship is not taken into consideration.

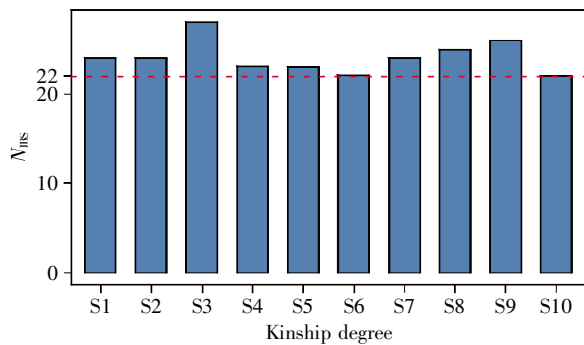


图 3 STR-IBS 算法预测结果
红色虚线为 $N_{IBS}=22$, 是判断两个个体为 FS 关系的阈值。

Fig.3 STR-IBS algorithm prediction results
The red dotted line is $N_{IBS}=22$, which is the threshold for judging the relationship between two individuals as FS.

3.1.3 PTT 算法准确性评估

使用 DGA v1.0 中的 PTT 算法对 41 个样本中的 39 对 PO 关系进行亲缘关系预测。从预测结果可得, 39 对 PO 关系被全部预测出来, 平均 I_{cp} 可达 $1.45E+12$ 。PTT 算法在预测 PO 关系时准确率达到 100%, 说明 PTT 算法能够很好地用于 PO 关系预测。

3.2 SNP 数据准确性评估

3.2.1 SNP 数据 IBS 算法准确性评估

为测试 DGA v1.0 中 SNP-IBS 算法的准确性, 使用 253 份样本的 SNP 数据进行亲缘关系预测, 将 DGA v1.0 预测结果与真实亲缘关系进行比较 (图 4)。结果显示, SNP-IBS 算法在 1 级亲缘关系中的预测准确率可达 100%, 在 2 级和 3 级亲缘关系中的预测准确率在 90% 以上, 但从 4 级开始, 随亲缘关系等级增加, 准确率明显下降。

3.2.2 SNP 数据 IBD 算法准确性评估

为验证 DGA v1.0 中 IBD 算法的准确性, 使

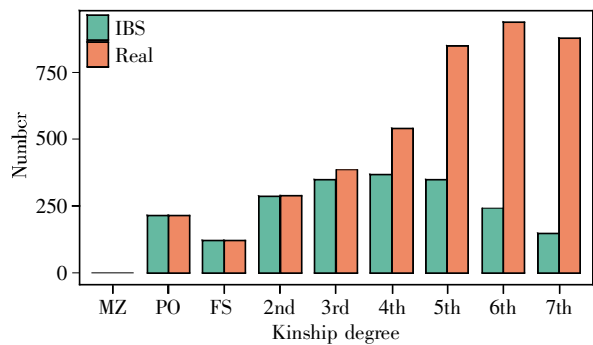


图 4 SNP-IBS 算法对 253 份样本 SNP 数据的预测结果与真实亲缘关系的比较

Fig.4 Comparison between the prediction results of the SNP-IBS algorithm on the SNP data of 253 samples and the real relationship

用 DGA v1.0 对 253 份样本的 SNP 数据进行亲缘关系预测, 获得亲缘关系预测结果。从图 5 可知, 与已知真实亲缘关系比较, IBD 算法在 1 级亲缘关系中的预测准确率可以达到 100%, 在 3 级和 4 级亲缘关系中的预测准确率在 75% 以上, 在 2 级亲缘关系中的预测准确率较低, 只有 68%。同时, 我们计算了预测结果的置信区间准确率 (图 5), 结果显示, IBD 算法在 1 级到 5 级的置信区间准确率高于 95%, 随亲缘关系等级增加, 置信区间准确率明显下降。

3.3 系统效率评估

从测试样本中取不同数量的 STR 和 SNP 样本数据, 对 DGA v1.0 的亲缘关系预测时间和人工预测亲缘关系的时间进行比较, 统计结果如表 2 和表 3 所示。从表中可知, 随着样本数量的增加, DGA v1.0 系统的速度优势越来越明显, 而人工预测亲缘关系, 在预测时间上明显落后的同时, 预测结果也存在偏差。综合表 2 和表 3 中的预测

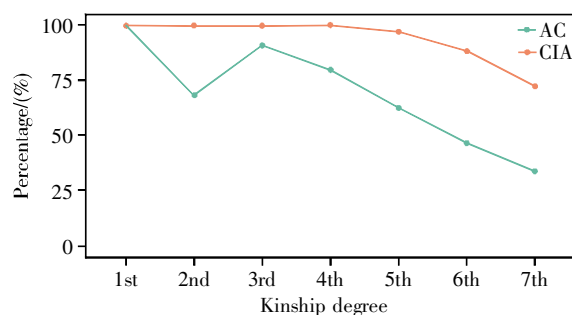


图5 IBD算法对253份样本SNP数据的预测准确性

AC表示准确率,即正确预测两个相关个体真实亲属关系程度的概率;CIA表示置信区间准确率,即两个相关个体被预测在其真实亲属关系的1级范围内,且不考虑7级以上或未知关系的概率。

Fig.5 Prediction accuracy of IBD algorithm on the SNP data of 253 samples

AC represents the accuracy rate, that is, the probability of correctly predicting two related individuals as the degree of their real kinship; CIA represents the accuracy rate of confidence interval, that is, the probability that two related individuals is predicted to be within the 1st-degree range of their real kinship, and above 7th-degree or unknown relationship is not taken into consideration.

时间可得, DGA v1.0 总体运算效率较人工预测提高了800%以上, 尤其在使用STR数据时远超800%。

4 讨论

目前, 在人类遗传学和法医系谱学领域基于STR数据和SNP数据预测亲缘关系的研究应用广泛。我们尝试将两种不同遗传标记的应用结合, 开发出一套基于SNP和STR数据的集成算法系统——DGA v1.0, 该系统可部署在服务器上运行和使用, 实现了对STR数据和SNP数据的批量

化和自动化处理, 提高了亲缘关系推断的效率, 达到了系统开发的目的。

本研究基于测试数据验证了DGA v1.0预测的准确性, 结果显示, DGA v1.0在保证准确性的同时, 对时效性要求较高的现场案件取证分析, 可以起到很好的辅助作用。对于STR数据, STR-IBS算法和PTT算法可对1级亲缘关系中的FS和PO关系进行预测, 且预测准确率可达100%。传统ITO算法只能计算单一关系 I_{cp} , 根据 I_{cp} 判断是否存在亲缘关系, 预测效率低。而经本项目组改进, 系统中的ITO算法可同时计算4种亲缘关系的 I_{cp} 并比较大小, 确定个体间更倾向于哪一种亲缘关系, 提高了算法的预测效率。通过使用已知亲缘关系的样本数据测试改进后的ITO算法可得, 改进后ITO算法的亲缘关系预测速度大幅度提高, 所得预测结果准确率较传统ITO算法有明显提升, 进一步挖掘了STR数据亲缘关系推断的潜力。

对于全基因组SNP数据, 百万数量级的位点分型弥补了STR基因座数量有限的缺点。基于SNP数据的两种算法在预测1级亲缘关系时准确率都可达100%, 并且SNP-IBS算法还可以在保证准确率的前提下, 对1级亲缘关系中的FS、MZ和PO关系进行区分预测, 可以与STR-IBS算法和PTT算法的预测结果进行交互验证。与基于STR数据的算法相比较, SNP-IBS算法能够预测3级以内亲缘关系, 且准确率可达90%以上, 此类结果与之前项目组研究结果^[17]一致。IBD算法在预测2级亲缘关系时准确率只有68%, 相比较1级与3级准确率明显下降。通过比较预测结果与

表2 基于STR数据的系统与人工预测时间的比较

Table 2 Comparison of time consumption between the system and manual prediction on STR data

List	Sample size	Routine time (STR-IBS)	DGA v1.0 time (STR-IBS)	Routine time (ITO)	DGA v1.0 time (ITO)	Routine time (PTT)	DGA v1.0 time (PTT)
1	1	2 min	0.1 s	5 min	0.2 s	7 min	0.2 s
2	15	30 min	0.5 s	75 min	0.8 s	105 min	0.8 s
3	30	58 min	0.7 s	150 min	1.2 s	210 min	1.1 s
4	41	75 min	1.2 s	205 min	1.5 s	287 min	1.8 s

表3 基于SNP数据的系统与人工预测时间的比较

Table 3 Comparison of time consumption between the system and manual prediction on SNP data

List	Sample size	Routine time (SNP-IBS)	DGA v1.0 time (SNP-IBS)	Routine time (IBD)	DGA v1.0 time (IBD)
1	1	8 min	50 s	5 h	100 min
2	100	30 min	64 s	10 h	120 min
3	200	45 min	80 s	12 h	150 min
4	253	63 min	90 s	18 h	180 min

实际亲缘关系,我们发现,有30%的2级关系被预测为3级,这表明IBD算法在使用IBD片段长度判断亲缘关系等级时的区间阈值仍需优化。我们尝试使用Parabon公司的亲缘关系等级划分区间来进行预测^[9],结果显示,上述2级亲缘关系被预测为3级亲缘关系的情况得到明显改善,2级亲缘关系的预测准确率达到97%。虽然IBD算法在2级到5级的亲缘关系预测准确率只有60%到80%,但其置信区间准确率可达95%以上,能够区分个体间是否存在亲缘关系,且可用于远亲预测。

DGA v1.0的优势在于能够批量处理STR和SNP两种不同类型的基因型数据,极大程度上缩短了亲缘关系预测时间。项目组同时开发了DGA v1.0单机版,除需要大量计算资源的IBD算法外,其余算法单机版均已全部实现,亲缘关系预测速度和准确率与DGA v1.0一致。DGA v1.0单机版无需注册,并且将算法所依赖环境完全封装,可以直接安装并离线使用,有较大推广和实用价值。DGA v1.0的功能模块具有良好的拓展性,项目组下一步将增加线粒体SNP母系亲缘关系推断等算法,丰富DGA v1.0的功能,使其更好地应用于遗传分析和母系亲缘关系鉴定。

总之,DGA v1.0界面友好、操作简便,用户可快速熟悉和使用,同时该系统集成了多种算法的计算流程并实现了数据处理操作的自动化和批量化,提高了计算速度,为现场取证分析和案件侦查提供了技术支持,预计其在群体遗传学、生物信息学以及法医系谱学等领域将发挥重要作用。

参考文献(References):

- [1] XU Z L, TAYLOR J A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies[J]. *Nucleic Acids Research*, 2009, 37 (Suppl. 2): W600-W605.
- [2] 王皓, 尹秋丹, 徐菁鹤. DNA遗传标记在亲子鉴定中的应用[J]. 世界最新医学信息文摘(WANG Hao, YIN Qiu-dan, XU Jing-he. Application of DNA genetic markers in paternity testing[J]. *World Latest Medicine Information*), 2018, 18(A3): 65.
- [3] 刘京, 马咪, 魏以梁, 等. 法医SNP系谱推断技术助破14年久冷案[J]. 刑事技术(LIU Jing, MA Mi, WEI Yi-liang, et al. Forensic SNP genealogical inference helping ferret out a 14-year-long unsolved cold case[J]. *Forensic Science and Technology*), 2021, 46(6): 652-656.
- [4] LEE H, CHEN L. Inference of kinship using spatial distributions of SNPs for genome-wide association studies[J]. *BMC Genomics*, 2016, 17: 372.
- [5] ZHANG Q Z, ZHOU Z, WANG L, et al. Pairwise kinship testing with a combination of STR and SNP loci[J]. *Forensic Science International: Genetics*, 2020, 46: 102265.
- [6] 刘聪, 肖超, 黄钰洁, 等. 新型连锁遗传标记的法医学研究进展[J]. 中国法医学杂志(LIU Cong, XIAO Chao, HUANG Yujie, et al. Advances in studies on the new type of linkage genetic marker[J]. *Chinese Journal of Forensic Medicine*), 2018, 33(3): 269-272.
- [7] 刘京, 季安全, 李彩霞, 等. 法医系谱分析研究进展[J]. 刑事技术(LIU Jing, JI An-quan, LI Cai-xia, et al. Research progress of forensic genealogy analysis[J]. *Forensic Science and Technology*), 2019, 44(3): 189-194.
- [8] 中华人民共和国司法部. 生物学全同胞关系鉴定技术规范: SF/T 0117-2021[S]. 北京: 中国标准出版社(The Ministry of Justice of the People's Republic of China. Specification of biological full sibling testing: SF/T 0117-2021[S]. Beijing: China Standard Press), 2021.
- [9] 中华人民共和国公安部. 法庭科学DNA亲子鉴定规范: GA/T 965-2011[S]. 北京: 中国标准出版社(The Ministry of Public Security of the People's Republic of China. Regulation for parentage testing laboratories of forensic: GA/T 965-2011[S]. Beijing: China Standard Press), 2011.
- [10] CHEUNG E Y Y, PHILLIPS C, EDUARDOFF M, et al. Performance of ancestry-informative SNP and microhaplotype markers[J]. *Forensic Science International: Genetics*, 2019, 43: 102141.
- [11] KLING D, PHILLIPS C, KENNETT D, et al. Investigative genetic genealogy: current methods, knowledge and practice[J]. *Forensic Science International: Genetics*, 2021, 52: 102474.
- [12] KLING D. On the use of dense sets of SNP markers and their potential in relationship inference[J]. *Forensic Science International: Genetics*, 2019, 39: 19-31.
- [13] PHILLIPS C. The golden state killer investigation and the nascent field of forensic genealogy[J]. *Forensic Science International: Genetics*, 2018, 36: 186-188.
- [14] KENNETT D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes[J]. *Forensic Science International*, 2019, 301: 107-117.
- [15] 陆惠玲, 杨庆恩. 用ITO法计算两个体间的血缘关系机会[J]. 中国法医学杂志(LU Hui-ling, YANG Qing-en. Using ITO method to calculate the chance of consanguinity between two bodies[J]. *Chinese Journal of Forensic Medicine*), 2002, 17(3): 188-191.
- [16] 吴微微, 刘冰, 郝宏蕾, 等. 中国28个省/区汉族人群41个STR基因座多态性数据分析[J]. 中国法医学杂志(WU Weiwei, LIU Bing, HAO Hong-lei, et al. The study and analysis of polymorphisms of 41 STR loci in Chinese Han population from 28 provinces/regions[J]. *Chinese Journal of Forensic Medicine*), 2016, 31(1): 27-32.
- [17] 管珊珊, 张文杰, 魏以梁, 等. 基于IBS算法预测亲缘关系准确性研究[J]. 生物化学与生物物理进展(GUAN Shan-shan, ZHANG Wen-jie, WEI Yi-liang, et al. Accuracy research on the kinship relationship prediction by IBS algorithm[J]. *Progress in Biochemistry and Biophysics*), 2022, 49(3): 591-599.
- [18] 刘京, 郭志芳, 魏以梁, 等. 法医SNP系谱推断技术破获古汉墓被盜案1例[J]. 法医学杂志(LIU Jing, GUO Zhi-fang, WEI Yi-liang, et al. Forensic SNP genealogy inference technology uncovers a case of the theft of an ancient Han tomb[J]. *Journal of Forensic Medicine*), 2021, 37(4): 600-602.
- [19] GREYTAK E M, MOORE C, ARMENTROUT S L. Genetic genealogy for cold case and active investigations[J]. *Forensic Science International*, 2019, 299: 103-113.