

“法医系谱学瓜熟蒂落”解读

朱波峰^{1,2}, 郭瑜鑫², 李淑瑾³, 刘超⁴, 丛斌^{3*}

1. 南方医科大学法医学院, 广州 510515;

2. 西安交通大学口腔医院, 陕西省颅颌面精准医学研究重点实验室, 西安 710004;

3. 河北医科大学法医学院, 石家庄 050017;

4. 广州市刑事科学技术研究所, 广州 510030

* 联系人, E-mail: hbydcngbin@126.com

随着在美国被称为“金州杀手”(“golden state killer”)的连环奸杀案嫌疑人落网, 法医系谱学研究作为一种新型的侦查技术手段受到广泛关注. 该技术被美国 *Science* 杂志评选为 2018 年十大科学突破之一.

2018 年 4 月, 72 岁的美国加州前警察约瑟夫·迪安杰洛(Joseph James DeAngelo)被指控在 1974~1986 年期间犯下约 50 起强奸案和 12 起谋杀案, 时隔 30 多年, 警方终于将这个连环杀手缉拿归案(<https://www.nytimes.com/2018/04/25/us/golden-state-killer-serial.html>). 警方将封存已久的样本 DNA 数据上传至商业化的基因数据库 GEDmatch, 匹配至约瑟夫的远亲, 通过还原其所在家系, 最终认定他是凶手^[1]. 在迪安杰洛被捕后不足一月, 警方又应用相同的方法, 抓获了一名在 1987 年杀害了一对年轻加拿大夫妇的凶手(<https://www.sciencenews.org/article/genetic-sleuthing-murder-suspect-cold-case-talbott>). 连续两起陈年旧案的成功破获, 仅仅是法医系谱学技术应用的开端. 在 2018 年 4~8 月, 短短 4 个月, 至少有 13 起应用该技术成功破获的案例. 应用法医系谱学短期内成功侦破大量案件成为犯罪侦查领域的重大突破, 这意味着该技术趋于成熟.

由于常染色体上的基因在遗传过程中存在分离、重组的现象, 因此不同层级亲缘关系的个体之间共享的同源(identity-by-descent, IBD)片段的长度不同. 例如, 当两个体间共享 IBD 片段长度大于 700 cM 时, 意味着两者存在一代表亲或更近的亲缘关系; 当两者间共享 IBD 片段长度大于 100 cM 时, 两者间的关系则更接近于三代表亲或更近的亲缘关系^[1]. 三角测量(triangulation)技术被广泛地应用于多个领域, 从心理学到政治学均有涉及. 2004 年, 遗传系谱学家比尔·赫斯特(Bill Hurst)引入“遗传三角”(“genetic triangulation”)的概念, 用于描述利用两个或多个已知数据点测定 Y 染色体短串联重复序列(short tandem repeat, STR)或线粒体 DNA 祖先单倍型的方法^[2]. 如今, 结合常染色体 DNA 的共享 IBD 片段信息, 该技术也被用于构建系谱树. 每个“遗传三角”由 3 点构成, 包括 A: 样本 DNA; B: 匹配列表上的任意样本; C: 与 A 和 B 两样本有共享 IBD 片段的第三人. 利用该方法可以在 DNA 数据库



丛斌 中国工程院院士, 获何梁何利科学与技术进步奖, 全国十大“新时代最美司法鉴定人”称号. 在人类基因组 DNA 多态性研究及法医学应用方面取得重要成果. 在广泛软组织损伤, 应激性损伤等复杂死因的鉴定及毒品依赖等方面进行了系列研究, 为国家出台“非法证据排除”制度性规定

提供了科学依据; 发现 CCK-8 具有抗炎, 降低肺动脉高压, 改善微循环, 对树突状细胞、巨噬细胞抗原呈递及 Th 细胞平衡具有调节作用. 获国家科技进步一等奖 1 项、二等奖 2 项.

中找到多个样本 B 和样本 C, 并构建多个“遗传三角”, 结合多个三角的结果, 就有可能描绘出样本 A 的系谱树(<https://www.familytreemagazine.com/premium/triple-play-dna-matches-triangulation/>).

虽然法医系谱学在近期才被重视, 但是利用家系研究推断犯罪嫌疑人方法早已不新鲜. 早在 2006 年, *Science* 杂志就已经报道了通过匹配亲属 DNA 来追查犯罪嫌疑人的可能性^[3]. 2016 年 8 月, 震惊全国的“甘肃白银连环杀人案”告破, 正是警方通过对 Y 染色体 DNA 的家系排查, 与高氏家系比中, 最终将犯罪嫌疑人抓获. 先前的案件中, 利用家系推测犯罪嫌疑人一般仅限于一至两代的近亲^[1], 或者是拥有相同 Y 染色体基因分型的亲属. 而在“金州杀手”案件中, 警方利用的是犯罪嫌疑人的父系和母系的三代以内堂表亲(关系如图 1 所示), 构建犯罪嫌疑人的家系, 追溯到犯罪嫌疑人. 这意味着, 如果在 DNA 数据库中有犯罪嫌疑人的表亲或者亲缘关系更近的亲属, 则有可能快速锁定犯罪嫌疑人.

1 法医系谱学的应用领域

法医系谱学是在 2005 年由杜克大学科林·菲茨帕特

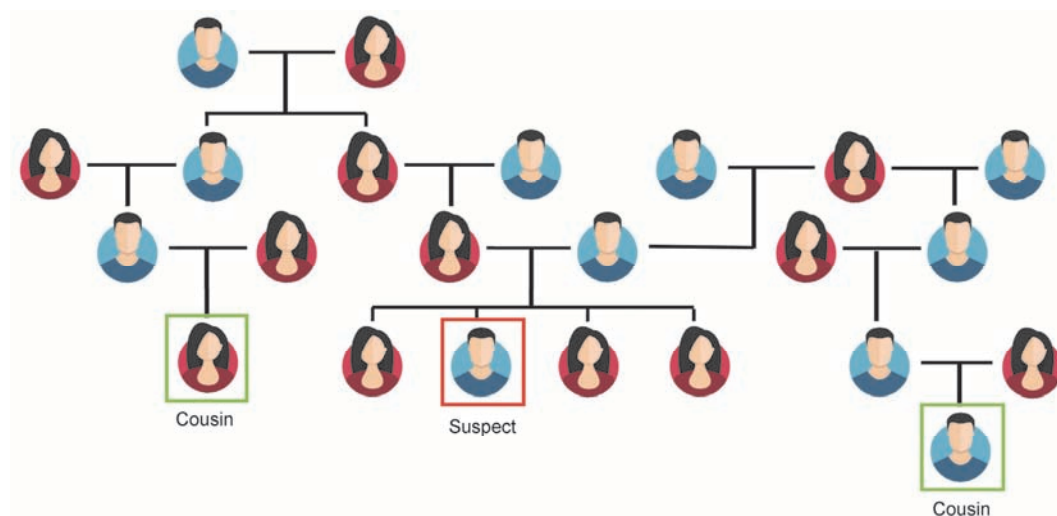


图1 (网络版彩色)“金州杀手”案系谱图

Figure 1 (Color online) Family tree for “golden state killer”

里克教授(Colleen Fitzpatrick)提出的,起初是用于描述解决复杂亲缘关系的科学技术手段,后来该名词又被赋予了更多法律层面的含义(<https://www.forensicmag.com/article/2016/12/forensic-genealogy-real-story>).与传统系谱学相同,两者都是基于DNA数据库中的基因信息和相关技术来研究亲缘关系,包括多种DNA检测技术的建立和DNA数据库信息挖掘等.两者的不同之处在于,法医系谱学分析结果必须满足亲缘关系鉴定标准,才能供司法鉴定使用.因此较传统系谱学,法医系谱学可以更多地被应用于证据鉴识层面.

根据罗卡交换定律^[4],“凡两个物体接触,必会产生转移”.任何犯罪现场都必然会遗留与犯罪嫌疑人有关的生物痕迹,即使是再狡猾的犯罪嫌疑人,在实施犯罪行为时也会留下蛛丝马迹.目前,由于受数据库信息量及比对方式的限制,警方虽然已经获得犯罪嫌疑人的DNA信息,但因DNA数据库缺乏足够人员数据信息,直接比中的概率小,所以无法获取犯罪嫌疑人足够的身份信息.虽然我国的法医DNA数据库经历十余年的发展,已形成较大规模,但是与我国人口总基数及群体数相比,仍然无法达到人群的全面覆盖.然而,当犯罪嫌疑人的DNA数据在数据库中沒有直接匹配上时,可以使用法医系谱学技术,只需要在数据库中匹配到犯罪嫌疑人的表亲或者亲缘关系更近的亲属,就有可能据此来认定犯罪嫌疑人.该技术的应用,无疑将会充分有效地发挥现有DNA数据库鉴识效能.

除了认定犯罪嫌疑人以外,法医系谱学还可应用于打击拐卖儿童、寻找失踪人口等领域.结合目前我国建立的“全国公安机关查找被拐卖/失踪儿童DNA数据库”,可为救助被拐卖/失踪儿童提供强大的技术支持.法医系谱学家也将该技术应用于其他与法律相关的案件,例如协助寻找

失踪的遗产继承人、进行遗嘱认证、追踪房产所有权人等.

2 应用法医系谱学的优势

2.1 提高DNA数据库利用效能

目前我国的DNA数据库已形成一定规模,全国已建成590多个法医DNA实验室,全国DNA数据库快速比对平台等多系统联动应用,形成了有中国特色的完整先进的DNA数据应用体系(<http://www.mps.gov.cn/n2254098/n4904352/c6312936/content.html>).按照数据来源和用途可以分为前科DNA库、现场检材DNA数据库、未知名尸体DNA数据库、失踪人员亲属DNA数据库和基础DNA数据库^[5].即使我国已建立起较大规模的DNA数据库体系,面对庞大的人口基数,DNA数据库的覆盖度仍然有限.目前已有的DNA数据库所包含的信息多来自有犯罪前科的个体,因此,没有犯罪前科的嫌疑人的DNA比对仍然是难题.法医系谱学研究为解决这一困境打开了局面,相对于直接匹配犯罪嫌疑人或其一至两代近亲,间接找到其远亲的概率更高,由此增加了寻找犯罪嫌疑人的可能性.

2.2 提高办案效率,保证案件相关当事人信息安全

从开始侦查到最终锁定犯罪嫌疑人往往有一定的办案流程.案发后,技术人员对犯罪现场进行反复复查,对可能提取DNA的相关物证进行检测,并将案件现场获取检材的DNA分型结果与数据库中的DNA分型进行比对.若未匹配成功,警方通常会对现场周围人群及与案件相关人员进行大规模排查,以期发现可能匹配的犯罪嫌疑人.

利用法医系谱学策略,将增加DNA入库比对并追溯犯罪嫌疑人的概率,很大程度上可以避免后期的大规模摸

排,节约办案成本.2018年 *Science* 杂志最新发表的数据统计显示,以欧洲人遗传背景的DNA数据库为例,当数据库中存有大约300万个个体的DNA数据(个体总数约占该人群成年人总数的2%),则该人群超过99%的人口都可以在数据库中匹配到至少一个三代表亲,超过65%的人口可以在数据库中匹配到至少一个二代表亲^[1].虽然该统计是以欧洲人遗传背景为例,但是仍具有可参考性.随着DNA数据库数据的不断增加,加之法医系谱学技术的应用,我国有望实现DNA数据的人口全面覆盖,从而达到建立全民DNA数据库的目的.

大规模排查需要耗费大量的警力资源,同时需要花费较长时间完成,并且在排查过程中,案件的相关信息一旦泄露,犯罪嫌疑人则很可能望风而逃,躲避排查,从而会增加破案难度.此外,大规模排查易引起社会恐慌,影响周围群众的生活.结合法医系谱学技术的办案手段,提高办案效率的同时,也增强了案件信息的保密工作,将是今后案件突破的一大利器.

2.3 提高犯罪嫌疑人匹配的成功率

利用Y染色体DNA追踪犯罪嫌疑人家系是目前警方侦破案件的常用手段,如“甘肃白银连环杀人案”的成功告破,正是利用了Y染色体追踪犯罪嫌疑人家系.但是利用Y染色体DNA受诸多条件限制.第一,犯罪嫌疑人需为男性,该技术对女性犯罪嫌疑人无任何作用;第二,相较于常染色体,Y染色体上可用的遗传标记更有限,比对效能相对较低;第三,部分Y染色体上的遗传标记突变率高,从而易发生假比中、假排除,易误导案件侦查方向;第四,基于成本考虑,目前检验相同数目Y染色体STR基因座的成本远高于常染色体STR.在“金州杀手”案件破获的过程中,警方没有利用Y染色体溯源,而是利用常染色体匹配到了犯罪嫌疑人母系的一个女性远亲及其父系的一个男性远亲,从而追溯到犯罪嫌疑人,因此数据库中的常染色体DNA数据在此案件的破获中起到至关重要的作用.

亚尼夫·埃利克(Yaniv Erlich)等人基于直接面向消费者(direct to customer, DTC)常染色体基因检测数据库中的128万个个体常染色体DNA数据进行了家系匹配的统计分析.由于同源片段的长度越长,说明两者之间的亲缘关系越近.埃利克在进行DNA匹配的过程中,设置了不同的同源片段匹配长度.结果显示在DTC数据库中,当同源片段长度达到100 cM及以上时,大约60%的样本可以找到相关匹配^[1].与“金州杀手”案件匹配到三代以内堂表亲类似,匹配100 cM的同源片段长度的水平就约相当于第三代表亲或更近的亲缘关系.该匹配的成功率远远高于利用Y染色体DNA的匹配效率.

综合法医系谱学应用的多种优势,基于法医系谱学技术进行常染色体的家系推断可以和作为Y染色体DNA推

断家系联合应用,以便更高效地追溯犯罪嫌疑人.

3 应用法医系谱学的潜在问题

3.1 伦理问题

近年来,在美国先后出现了多个商业化的DNA数据库,如23andMe,用于帮助用户分析DNA,提供潜在疾病的风险评估;Ancestry及GEDMatch均为族谱推断网站,可用于帮助寻找家族系谱来源(<https://www.sciencenews.org/article/genetic-genealogy-forensics-top-science-stories-2018-yir>).超过1200万的美国人加入了基因检测的消费潮流,只需要把唾液样本送到23andMe或Ancestry这样的公司,就可以了解疾病风险,探索家族起源(<https://www.sciencenews.org/article/family-dna-ancestry-tests-review-comparison>).已经接受过DNA检测的人,可以将自己的检测结果上传到公共数据库,匹配来自不同公司检测的DNA数据,以便寻找可能存在亲缘关系的人.“金州杀手”案件就是利用GEDMatch中的DNA数据进行匹配,从而锁定犯罪嫌疑人.

随着该案件的侦破,法医系谱学技术及DNA数据库的使用所引发的伦理问题随之而来(<https://www.geek.com/tech/golden-state-killer-dnasearch-raises-major-data-privacy-concerns-1739868/>).英国遗传系谱学专家德比·肯尼特指出,“这种技术潜在地影响了所有人,并不只是网站数据库里的那些人”.利用法医系谱学可以推断已上传信息用户的多位亲属的基因信息,尽管这些亲属并未自愿在网上共享自己的DNA及身份信息,也就是说,一位用户将自己的DNA数据上传至数据库中,则意味着与该用户存在遗传相关性的亲属的部分DNA信息也被公开.执法部门是否可以在未经允许的情况下,对犯罪嫌疑人亲属的身份相关信息进行调查,这些亲属的隐私是否得到有效保护,这些都是目前争论的焦点.并且这些DNA数据包含了个人的遗传信息,有可能泄露个体的遗传特征谱的信息,因此该技术的应用,涉及伦理、法律、道德以及人权等多个层面的问题.

3.2 推断准确性

由于法医系谱学的应用,美国警方在近期接连侦破多个悬案,该技术也被称为新一代的“破案神器”.但是仍然有专家对该技术应用的准确性提出了质疑^[6].即使是目前应用较为成熟的Y染色体DNA,仍然有出现误判而导致冤假错案的可能性.例如,台湾冤狱平反协会公布,在一次案件侦破中,由于Y染色体STR的错误匹配,把陈龙琦当作罪犯在台湾监禁4年之久(<http://twinnocenceproject.org/CaseDetail.php?CaseId=18>).在进行远亲匹配的过程中,有亲缘关系的人数随着亲缘关系的距离变远呈增长趋势,也就是说,一个人有近1000个四代表亲,而五代表亲则多达

5000 个,而在这些人中,只有小部分人的 DNA 存在于 DNA 数据库中^[6]。因此,相对于 Y 染色体 DNA 的应用,利用法医系谱学技术推断犯罪嫌疑人的远亲则更加复杂,从而会影响推断的准确性。

除了该技术本身的复杂性以外,所利用的数据库本身的安全性和真实性也是需要考虑的^[7]。法医系谱学家对于犯罪嫌疑人远亲的匹配,完全基于现有的数据库。对 DNA 数据库的高度依赖性,有可能导致该技术的应用出现偏差。即使专家已经采用较为完善的算法以避免发生错误,但是如果最基础的数据来源出现问题,则会产生完全错误的结果。DNA 数据库中的数据假如被黑客恶意篡改,或者被非法访问,后果不堪设想。并且 DNA 数据库数据本身的准确性也是应考虑的因素。例如,不同来源的商业化的 DNA 数据库所上传的 DNA 数据信息,有可能来自不同试剂检测,参照不同的 DNA 检测标准,造成不同公司所提供的 DNA 检测结果存在差异。因而该技术的应用也对 DNA 数据库的真实性、安全性和准确性提出了较高的要求。

4 应对潜在问题及应用前景展望

尽管法医系谱学结合 DNA 数据库的法医学应用会带来诸多伦理问题和误判风险的争议,但是这些并不能成为放弃使用新技术的理由,面对争议,唯有找到更多的应对策略,进一步明确应用标准,并提升防控风险。

首先,应进一步明确法医系谱学应用范围,并严格把控对 DNA 数据库的访问权限。对于有可能出现的超出案件本

身以外的推断,以及对于非法访问数据库甚至恶意篡改数据库信息等行为的惩罚写入立法,以此保护案件相关人员的隐私,降低对犯罪嫌疑人亲属及其他所涉及人员的影响。

其次,构建 DNA 数据库时选择与表型如发色、肤色、身高等人体表型特征不相关的内含子区域位点,以防泄露个人表型遗传特征信息^[8]。也可借鉴美国 FBI 的 DNA 数据库管理规定,数据库中只有索引编号和 DNA 数据信息,而相关的个人详细身份信息则储存在各州,如果要同时获取两种信息,则需要经过两个部门同时许可,这样大大提高了数据安全性,降低了被恶意篡改的风险。

再次,制定统一的应用法医系谱学推断远亲的各项参数标准规范,避免因法医系谱学家主观误判造成错误的匹配。鉴于我国多民族多遗传背景的现状,进行法医系谱学应用时应考虑不同民族遗传背景的影响,适当对推断算法进行调整优化,以确保匹配的准确性。

最后,鉴于应用法医系谱学推断远亲的复杂性,建议目前仅作为其他已具备国家/行业标准的认定方法的辅助方法,用于缩小案件排查范围,为警方提供 DNA 线索。

虽然法医系谱学的应用仍然存在争议,例如:可能会暴露无关人群的基因信息;是否会为系谱中涉及的人群带来影响;安全性和准确性是否能够得到有效的保证;这些都是今后亟待解决的问题。无论面临多少挑战,随着法医系谱学研究的日渐成熟,该技术将为犯罪案件的侦破,失踪人口的查寻等带来新的解决思路,为法医学多个领域难题的攻破开辟新途径。

推荐阅读文献

- 1 Erlich Y, Shor T, Pe'er I, et al. Identity inference of genomic data using long-range familial searches. *Science*, 2018, 362: 690–694
- 2 Kerchner C F. Triangulation Method for Deducing the Ancestral Haplotype in Y-DNA Surname Projects. *Genetic Genealogy DNA Testing Dictionary*. 2004
- 3 Bieber F R, Brenner C H, Lazer D. Human genetics—Finding criminals through DNA of their relatives. *Science*, 2006, 312: 1315–1316
- 4 Mistek E, Fikiet M A, Khandasammy S R, et al. Toward Locard's exchange principle: Recent developments in Forensic trace evidence analysis. *Anal Chem*, 2019, 91: 637–654
- 5 Liu Q, Ke J, Zhang M M. Thoughts on the construction of DNA database for forensic science in China (in Chinese). *Legal Syst Soc*, 2018, 28: 161–163 [刘泉, 柯技, 张苗苗. 对未来我国法庭科学 DNA 数据库建设的思考. *法制与社会*, 2018, 28: 161–163]
- 6 Court D S. Forensic genealogy: Some serious concerns. *Forensic Sci Int-Gen*, 2018, 36: 203–204
- 7 Hazel J W, Clayton E W, Malin B A, et al. Is it time for a universal genetic forensic database? *Science*, 2018, 362: 898–900
- 8 de Londras F, Dzehtsiarou K. II. Grand chamber of the European Court of Human Rights, A, B & C v Ireland, decision of 17 December 2010. *Int Comp Law Q*, 2013, 62: 250–262

Summary for “‘法医系谱学瓜熟蒂落’解读”

New genetic sleuthing tool: Forensic genealogy

Bofeng Zhu^{1,2}, Yuxin Guo², Shujin Li³, Chao Liu² & Bin Cong^{3*}

¹ School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China;

² Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an 710004, China;

³ College of Forensic Medicine, Hebei Medical University, Shijiazhuang 050017, China;

⁴ Guangzhou Forensic Science Institute, Guangzhou 510030, China

* Corresponding author, E-mail: hbydcongbin@126.com

Recently, a new genetic sleuthing tool, forensic genealogy, was used to track down Joseph James DeAngelo suspected of being the “golden state killer” in the US. Since then, decades-old cold cases and some fresh crimes have been solved after genetic genealogy searches. While the description of the genomic analysis steps taken to find DeAngelo appeared to be the latest groundbreaking technology, investigators found genetic relatives could link even distant relatives accurately by locating identity-by-descent (IBD) segments after querying direct-to-customer (DTC) genomic database. In addition, in order to solve those cases, powerful triangulation techniques combined with some traditional genealogical methods in forensic genealogy were used to identify shared segments for validating the links, and the details were introduced.

Forensic genealogy is the study of identification and kinship in legal contexts including finding family references for DNA identification, identifying missing persons, locating heirs for probate matters and more. The advantages for the forensic genetics community of introducing the new technology were discussed. Such as, improving the effectiveness of the existing DNA database to identify suspects in crime, reducing the impact of “innocent” individuals due to the investigation and protect security of case information, and possessing higher success rates than with surname inference from the Y chromosome.

Also, the ethical issues and risks of false identification should be taken into account. Legal and privacy experts worry that genetic genealogy searchers will put numerous people under police scrutiny and will never guarantee total privacy of their DNA information. Although the detailed linking information available in a DTC genealogy database is private to those directly concerned, which is guaranteed by the companies, they also warn that legal processes may force them to reveal personal information. Meanwhile, using this approach to uncover relatives may not be that simple if the relationship is more distant, the number of familial relationships increases exponentially with a number of generations, and only a small proportion of them will be available on any given DNA database.

There remain implementation issues that would need to be debated by the public and ultimately resolved by government. First of all, to make laws to strictly control access to genomic data for privacy protection of relevant personnel. Secondly, to avoid using phenotypic markers, such as the markers in relation to hair color, stature and different human diseases, in creating the DNA database. Thirdly, to normalize the formulas and parameters for linking distant relatives to prevent mismatches caused by subjective misjudgments of forensic genealogists. Finally, to apply the technology as an auxiliary tool for other identification methods at present, due to its complexity.

Even though the use of forensic genealogy will bring many ethical issues and risks of miscalculation, it is not the reason to abandon the application of this new technology. More standards and policies should be put in place.

forensic genealogy, golden state killer, DNA database, ethnics

doi: 10.1360/N972019-00266