

姓氏与 Y 染色体基因家谱学研究进展

孙亚男¹ 程宝文² 综述 崔文¹ 审校

(¹ 济宁医学院法医学院, 山东 济宁 272067; ² 云南省公安厅, 云南 昆明 650228)

关键词 姓氏; Y 染色体; 基因家谱学

中图分类号 R394.5, K820.9

文献标识码 A

文章编号 1000-9760(2014)04-119-03

在中国两千多年的封建统治中, 形成了一种“男性才能担当家族传宗接代重任”的封建思想。但是, 现代文明社会已经逐渐摒弃这种“重男轻女”的封建思想, 男女平等才是当今世界的主流。然而, 从遗传学角度看 Y 染色体遵循严格的父系遗传方式, 使得父系亲缘关系传递存在一定的“传宗接代”意义。

姓氏大部分都从父传递, Y 染色体则更严格地遵循父系遗传, 因此姓氏与 Y 染色体之间有很好的平行对应关系。研究姓氏与 Y 染色体之间的关联性对基因家谱学研究具有重要意义, 并且在法医学领域中也有一定的应用价值, 因为它意味着当给定一个姓氏与 Y 染色体特征的联合规律数据库时, 仅根据 DNA 就可进行姓氏的推测, 为案件侦查提供帮助。

1 姓氏的起源与发展

姓氏的起源发展状况在世界各地不尽相同。在日本, 姓氏于公元 5 世纪晚期产生, 但直到 1875 年才被普通平民使用^[1]; 在冰岛, 子女是以父亲的名字为姓氏, 所以其姓氏是世代变化的; 在中国, 姓氏的起源历史可追溯到 5000 年前, 主要来源于远古时代的各种图腾和地名。秦汉以后, 数量大增^[2]。据最新统计, 拥有 13 亿人口的中国目前有 4100 个姓氏^[3]。尽管姓氏的起源时间(中国大约 5000 年前而土耳其仅约为 80 年前)和发展演变在世界各地不尽相同, 但大多数情况下, 姓氏从父遗传的现象却是相同的。

1.1 中国姓氏的分布特点

中国的姓氏分布主要存在两个特点^[4-5]: 一是在历史上的传递是连续和稳定的。漫长的历史进程中, 中国姓氏表现出血缘文化的痕迹与生命遗传物质的特点, 尤其与 Y 染色体的进化具有基本相

同和平行的表现。中国历来有同姓聚居和联宗修谱的习俗, 且婚姻半径较小, 婚娶地域相对固定, 这样易形成同姓人群的分布, 因此, 中国姓氏的分布实际上主要反映了同姓人群的分布规律。另一个特点是存在大姓和小姓两种姓氏状态, 或称为常见姓氏和罕见姓氏。仅占总姓氏量不足 5% 的 100 个常见姓氏集中了 85% 以上的人口, 而占总姓氏量 95% 以上的罕见姓氏仅代表不足 15% 的人口。反映各地人群遗传组成的主要因素是常见姓氏的分布, 决定了中国历史上人群迁移和地域亲缘关系的程度; 而罕见姓氏人群则表现了其地域特色和相对隔离的现象。因此, 中国人群姓氏的分布规律及其与 Y 染色体之间的关联性研究将为探索其在法医学中的应用有重要价值。

2 Y 染色体的遗传学特点

人类 Y 染色体的 95% 为非重组区域(non-recombining portion of Y chromosome, NRY), 不与任何染色体发生重组。因此, Y 染色体的该特性保证了子代能完整地继承父代的 NRY, 保证了其严格的父系遗传。在代代父子相承的传递过程中, Y 染色体也在不断地积累着变化。遗传突变的积累, 使得人类父系遗传体系中 Y 染色体存在一定的差异。NRY 上的单核苷酸多态(SNP) 和短串联重复(STR) 是突变形成的主要的两大类差异。

2.1 Y-SNP 的遗传特点

Y-SNP 的多态主要是碱基转换、缺失或者插入(也包括长片段的插入, 如 Alu), 其突变率相当低(约 2×10^{-8} 次/代), 所以每次突变都具备很好的代表性; 另外, 其发生 DNA 回复突变(recurrent mutation) 的概率更低, 几乎忽略不计, 且更不会有向多种状态发生突变的情况。因此, 在 NRY 上产生的一系列突变能够准确地反映其漫长历史中的

演化过程。

单倍群(haplogroup)是一组相关联的 Y-SNPs 位点的组成。定义单倍群的 Y-SNPs 在进化过程中的出现有时间顺序。因此,可以用单倍群来构建 Y 染色体系统发育树。通过高效液相色谱技术,Underhill 等^[6-7]发现了大量的 Y-SNPs,大大地推动了其在人类学研究中的应用。随着更多的 Y-SNPs 被发现和研究,目前已经构建了较为详细的世界人群 Y 染色体系统发育树^[8],并在不断更新中(<http://www.isogg.org/tree/index.html>)。各单倍群及其分支在世界人群中的分布显示了很强的地理特异性。研究发现,主要有 D-M174、C-M130、N-M231、O-M175 单倍群分布于东亚^[9],其中 O-M175 单倍群主要有 O1-M119、O2-M95 和 O3-M122^[8] 3 个亚群,O3-M122 单倍群为东亚和东南亚的主要单倍群^[10]。

2.2 Y-STR 的遗传特点

Y-STR 的多态主要是每个 STR 位点上重复单位的拷贝数变化,其突变率较高(约 6.9×10^{-4} /代),有利于估算不同 Y 染色体之间的分化时间,因此 STR 位点成了 Y 染色体上的“时钟”。同 Y-SNP 相似,不同位置上的数个 Y-STR 可以构成单倍型^[11]。通过分析 Y 染色体单倍型的多样性程度可以计算群体的共祖时间。假设某个一步(single-step)突变模型,且群体的有效群体大小稳定,就可由公式 $t = -Ne \ln(1 - V/Ne\mu)$ 推算出某个 Y-SNP 位点的发生大致时间。公式中,Ne 为有效群体大小, μ 是突变率,ln 是自然对数,V 为群体中某 STR 被观察到的数值的方差,计算得到的 t 是经历的世代数,乘以每一世代的年数即可得到发生时间^[12]。

3 姓氏与 Y 染色体之间的关联性研究

姓氏传递的连续和稳定才能保证姓氏与 Y 染色体的基本相同和平行的传递关系。通过分析和比较宋朝、明朝和当代中国姓氏的分布曲线、同姓率和地域人群间的亲缘关系,袁义达等^[4-5,13-14]发现 1000 多年前的姓氏分布和当代基本一致。比较法国 19 世纪末和 1975 年的若干姓氏的分布情况,Legay 等^[15]也发现了姓氏传递的稳定性。这些研究证实各国的姓氏传承是相对稳定的,为姓氏和 Y 染色体的关联分析提供了理论依据。

同姓氏会提高出现相同 Y 染色体单倍型的机率,并且 Y 染色体单倍型相同的概率随着姓氏在

人群中分布频率的减少而增加^[16]。Balanovska 等^[17]提出了一种新的同姓人群 Y 染色体单倍群频率的统计计算方法,并认为研究同一姓氏人群的关联性需联合分析 Y-SNP 与 Y-STR 两种遗传标记。Melissa Gymrek 等也报道了可以通过 Y-STRs 检测或基因家谱学数据库查询推断姓氏^[18]。通过分析南欧姓氏与 Y 染色体之间的关联性,采用 17 个 Y-STR 基因座进行统计分析表明哥伦布并非生于意大利的热那亚,而是来自西班牙的加泰罗尼亚^[19]。Begona Martinez-Cruz 等^[20]联合 Y-SNP 和 Y-STR 分析了罗马尼亚瓦拉吉亚王朝时期的贵族姓氏巴萨拉布,表明在历史长流中存在大量的外姓人群改姓巴萨拉布。Y 染色体分析还可提供某一姓氏家谱中未记录的基因混合事件信息,且在深入的系谱遗传学研究中,姓氏的选择对于真正土著群体的选择发挥着重要的作用^[21]。意大利学者 Alessio Boattini 等,通过一系列的研究,包括采用一种亲缘关系矩阵和特征映射的遗传统计方法来表明以姓氏为基础的抽样程序可以扩展到非隔离群体^[22];以创建者姓氏作为判断样本祖先来源的标准,利用 Y-SNPs 与 Y-STRs 遗传标记研究,支持基于姓氏选择的 Arbereshe 人样本可以作为阿尔巴尼亚创建者群体的 Y 染色体遗传多态性的代表^[23];并通过对意大利的姓氏数据研究确定了一种通过姓氏构建祖先人群结构的方法^[24]。

吴东颖等^[25]应用 Y 染色体的 DYS19、DYS390、DYS391 和 YCAII4 个 STR 基因座,分析了 50 例中国汉族王姓男性 DNA 样本,结果与随机样本的单倍型的频率分布无显著差异。邓志辉等^[26]对深圳地区李姓、王姓和张姓的无关男性个体 Y-STR 单倍型遗传多态性分析表明,该地区此三姓氏无关男性个体 Y-STR 单倍型的遗传多态性丰富,与以往的汉族无关男性群体遗传资料比较差异不显著。侯伟光^[27]对山东地区张、王、李、赵、刘、陈、孔、孟、偃、么 10 家姓氏 17 个 Y-STRs 进行了多态性检测,分析 10 家姓氏谱系渊源关系表明,10 家姓氏 17 个 Y-STR 基因座的等位基因频率分布与山东人群分布总体上相似,且遗传关系主要分 3 支,偃姓、赵姓为第 1 支,李姓、孟姓为第 2 支,其余为第 3 支。

4 展望

在同一姓氏内研究 Y 染色体单倍型差异,表明姓氏的传递受到多种因素(包括姓氏多起源,非

父传递,基因漂流)的影响^[28]。遗传学中姓氏和 Y 染色体的联合分析,使其在法医学方面也具有潜在的应用价值,由此形成的“基因家谱学”也迅速引起了公众的关注。由于 Y 染色体的突变是一步步发生的,父系亲缘关系越近的个体之间的 Y 染色体单倍型越相似,一个纯粹由父系传递的姓氏应有相近的 Y-STR 单倍型。理论上讲,足够数量的 Y-SNP 和 Y-STR 信息,通过调查一个姓氏宗族内的 Y 染色团体信息,就能够很清楚的构建其家族 Y 染色体的谱系树,乃至编写一部清晰的基因家谱。

参考文献:

- [1] Yasuda N. Studies of isonymy and inbreeding in Japan[J]. Hum Biol, 1983, 55(2): 263-276.
- [2] 范义琪. 百家姓姓氏起源[M]. 武汉: 长江文艺出版社, 1999: 3-4.
- [3] 袁义达, 邱家儒, 张瑞, 等. 中国姓氏 三百大姓: 群体遗传和人口分布(上)[M]. 上海: 华东师范大学出版社, 2007: 5-6.
- [4] 袁义达, 张诚, 马秋云, 等. 中国人姓氏群体遗传-I. 姓氏频率分布与人群遗传分化[J]. 遗传学报, 2000, 27(6): 471-476.
- [5] 袁义达, 张诚, 杨焕明. 中国人姓氏群体遗传-II. 姓氏传递的稳定性与地域人群的亲缘关系[J]. 遗传学报, 2000, 27(7): 565-572.
- [6] Underhill PA, Shen P, Lin AA, et al. Y chromosome sequence variation and the history of human populations[J]. Nat Genet, 2000, 26(3): 358-361.
- [7] Underhill PA, Passarino G, Linet AA, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations[J]. Ann Hum Genet, 2001, 65(Pt 1): 43-62.
- [8] Karafet TM, Mendez FL, Meilerman MB, et al. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree[J]. Genome Res, 2008, 18(5): 830-838.
- [9] 钟华. 东亚现代人 Y 染色体群体结构和史前迁移[D]. 北京: 中国科学院研究生院, 2010: 52-56.
- [10] Shi H, Dong YL, Wen B. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122[J]. Am J Hum Genet, 2005, 77(3): 408-419.
- [11] Kayser M, Caglia A, Corach D, et al. Evaluation of Y-chromosomal STRs: a multicenter study[J]. Int J Legal Med, 1997, 110(3): 1029-1033.
- [12] Su B, Xiao J, Underhill P, et al. Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age[J]. Am J Hum Genet, 1999, 65(6): 1718-1724.
- [13] 袁义达, 金锋, 张诚, 等. 宋朝中国人的姓氏分布与群体结构分化[J]. 遗传学报, 1999, 26(3): 187-197.

- [14] 袁义达, 张诚. 中国姓氏: 群体遗传与人口分布[M]. 上海: 华东师范大学出版社, 2002: 32-36.
- [15] Legay JM, Vernay M. The distribution and geographical origin of some French surnames[J]. Ann Hum Biol, 2000, 27(6): 587-605.
- [16] King TE, Ballereau SJ, Schurer KE, et al. Genetic signatures of coancestry within surnames[J]. Curr Biol, 2006, 16(4): 384-388.
- [17] Balanovska EV, Romanov AG, Balanovsky OP. Namesakes or Relatives? Approaches to Investigating the Relationship between Y Chromosome Haplogroups and Surnames[J]. Molecular Biology, 2011, 45(3): 473-485.
- [18] Gymrek M, McGuire AL, Golan D, et al. Identifying personal genomes by surname inference[J]. Science, 2013, 339(6117): 321-324.
- [19] Martinez-Gonzalez LJ, Martinez-Espin E, Alvarez JC, et al. Surname and Y chromosome in Southern Europe: a case study with Colom/Colombo[J]. Eur J Hum Genet, 2012, 20(2): 211-6.
- [20] Martinez-Cruz B, Ioana M, Calafell F, et al. Y-Chromosome Analysis in Individuals Bearing the Basarab Name of the First Dynasty of Wallachian Kings[J]. PLoS ONE, 2012, 7(7): e41803.
- [21] Larmuseau MH, Vanoverbeke J, Gielis G, et al. In the name of the migrant father-Analysis of surname origins identifies genetic admixture events undetectable from genealogical records[J]. Heredity, 2012, 109(2): 90-95.
- [22] Boattini A, Pedrosi ME, Luiselli D, et al. Dissecting a human isolate: Novel sampling criteria for analysis of the genetic structure of the Val di Scalve (Italian Pre-Alps)[J]. Annals of Human Biology, 2010, 37(4): 604-609.
- [23] Boattini A, Luiselli D, Sazzini M, et al. Linking Italy and the Balkans. A Y-chromosome perspective from the Arbereshe of Calabria[J]. Annals of Human Biology, 2011, 38(1): 59-68.
- [24] Boattini A, Lisa A, Fiorani O, et al. General Method to Unravel Ancient Population Structures through Surnames, Final Validation on Italian Data[J]. Human Biology, 2012, 84(3): 235-270.
- [25] 吴东颖, 马素参, 刘明, 等. 应用 Y 染色体多态标记对汉族王姓亲缘关系的研究[J]. 人类学学报, 2000, 19(2): 132-137.
- [26] 邓志辉, 李茜, 王大明, 等. 深圳无偿献血人群中李姓、王姓和张姓无关男性个体 Y-STR 单倍型遗传多态性[J]. 遗传, 2007, 29(11): 1336-1344.
- [27] 侯伟光. 山东汉族 17 个 Y 染色体 STR 遗传多态性调查及姓氏渊源研究[D]. 辽宁: 辽宁医学院, 2011, 53-58.
- [28] King TE, Jobling MA. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames[J]. Mol Biol Evol, 2009, 26(5): 1093-1102.

(收稿日期 2014-02-15)