



Surnames and the Y Chromosomes

WANG Chuanchao¹, YAN Shi^{1,2}, LI Hui¹

1. MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 2. CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai 200031, China.

ABSTRACT: Most people get their surnames from their fathers. The surnames therefore well exhibit patrilineal heredity. Strict patrilineal heredity is also obeyed by parts of our genome, the Y chromosome non-recombining portion. Therefore, men sharing a surname are expected to have similar Y chromosomes. However, association between surnames and Y chromosomes is always weak or absent due to the complicated social behaviors of surnames, e.g., various founders for a surname, surname changing, and non-paternity events. To set up better association, the family genealogies in combination with Y chromosomes should be used in historical anthropology studies. Two kinds of markers on Y chromosomes, single nucleotide polymorphism (SNP) and short tandem repeat (STR), are widely used in setting up the Y chromosome genealogy and estimate the time. Y chromosome studies show strong power in resolving many doubtful cases existing for hundreds of years. Application of Y chromosome analyses in the family genealogy studies will be a most important part of historical anthropology researches.

Key words: Surname; Y chromosome; Single nucleotide polymorphism (SNP); Short tandem repeat (STR); Family genealogy.

姓氏与 Y 染色体

王传超¹, 严实^{1,2}, 李辉¹

1. 复旦大学现代人类学教育部重点实验室, 上海 200433; 2. 中国科学院上海计算生物学所, 上海 200031.

摘要: 人们的姓氏大多继承自父亲, 而 Y 染色体是严格的父子相传的基因组片段。所以姓氏与 Y 染色体的遗传应该是平行的, 有共同姓氏的男性可能有相同或相近的 Y 染色体类型。然而, 多起源、改姓、非亲生、从母姓等社会因素弱化了某些姓氏与 Y 染色体的关联, 此时家谱研究可为厘清父系血缘提供线索。Y 染色体上稳定的 SNP 突变可以永远在父系后代中流传, 可以构建可靠的父系基因谱系; 而其上突变较快的 STR 位点又可以用以估算时间。因此, Y 染色体可用于研究很多姓氏宗族的历史, 甚至千百年前的历史疑案。重建姓氏、家谱与 Y 染色体的关系必将成为历史人类学研究的重要内容。

关键词: 姓氏; Y 染色体; 单核苷酸多态(SNP); 短串联重复(STR); 家谱

现代社会中, 几乎每人都有自己的姓氏。一个人的姓氏不仅仅是简单的符号, 还有着丰富的文化、历史、宗族背景。以血缘为脉络的姓氏记录着各家族甚至于各民族的源流, 常被用于寻根溯源、族群识别、婚姻关系等相关研究。数千年来大部分姓氏都从父传递, 而人类基因组中的 Y 染色体更严格地遵循父系遗传, 因此姓氏与 Y 染色体有很好的平行对应关系。随着 Y 染色体上众多遗传标记的发现, 用 Y 染色体结合姓氏分析人类学问题的方法, 在分子人类学领域发挥了重要作用。

姓氏在遗传学上的应用

姓氏最早在中国产生, 其历史可追溯到 5000 年前, 主要来源于远古时代各种图腾

和地名, “氏”为“姓”的分支, “姓”以别婚姻, “氏”以分贵贱。秦汉以后, 姓氏合一, 数量大增[1]。据最新统计, 拥有 13 亿人口的中国目前有 4100 个姓氏[2]。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。至 1997 年, 日本仅 1 亿左右的人口却拥有约 30 万个姓氏[3]。而在西班牙, 子女姓氏沿承父母双方的姓氏, 因而会逐代变化。尽管姓氏的起源与发展演变在不同国家和地区有明显差异, 但大多数情况下姓氏从父遗传却是相同的。

姓氏的基本作用是明血缘、别婚姻, 《左传·僖公二十三年》就有: “男女同姓, 其生不蕃”。姓氏制度也在不同角度上反映了宗法制度的某些内容, 起到了维系和延续宗法制度的社会作用。姓氏最早用于遗传研究

是在1875年, George Darwin通过分析堂(表)婚得出了英国同姓通婚率和不同阶层的堂(表)近亲通婚率[4]。George Darwin的工作不断被后人丰富和发展, Crow等通过同姓率(isonymy)来计算近亲结婚率[5]。由于居民出生、结婚和死亡等大量相关数据的易得性, 姓氏分布与同姓率被广泛用于研究群体遗传结构、迁徙率等[6-32]。姓氏在流行病学方面也得到应用, 如Abbotts以姓氏为线索从生理、体质特征和社会经济因素等方面分析了移居到苏格兰的爱尔兰人后裔的高死亡率, Polednak发现了美国康涅狄格州的西班牙姓氏的女性有较低乳腺癌发病率等[33-35]。

姓氏与 Y 染色体的父系遗传

虽然姓氏在宏观上被用于分析群体遗传结构, 但是姓氏并不完全遵从父系遗传。就中国的社会情况而言, 收养、继养、入赘、甚至直接改姓, 都会影响姓氏与父系血统的关联程度。另一方面, 中国大多数姓氏起源于春秋时期的各个封国, 当封国内的百姓都以国为姓的时候, 这些同国百姓的血统可能本来就不一致。这就造成了很多比较大的姓氏内部遗传结构不一致。即便这样, 当我们不拘泥于群体中同一姓氏的研究, 而是针对有着明确的历史记载的宗族进行研究, 姓氏无疑还是一个很好的遗传标记。

与姓氏不同, 人类的Y染色体直接代表着父系遗传, 永远是父子相传的, 不会受到任何社会文化和自然因素的影响。人体内有23对染色体, 其中22对常染色体中, 每一对染色体都有一条来自父系, 一条来自母系, 两条染色体在传代过程中对应的部分会发生交换, 从而造成混血的效应, 就是遗传学上说的重组。另一对性染色体包括X染色体和Y染色体。在女性体内, X染色体也是成对的, 分别来自父母双方, 所以也不能避免混血的影响。而在男性体内, 却只有一条来自母亲的X染色体和一条来自父亲的Y染色体, 也就是说男性的Y染色体只能来源于父亲, 所以人体性染色体的遗传方式决定了Y染色体遵从严格的父系遗传(图1)。

Y染色体与X染色体之间是否会发生重

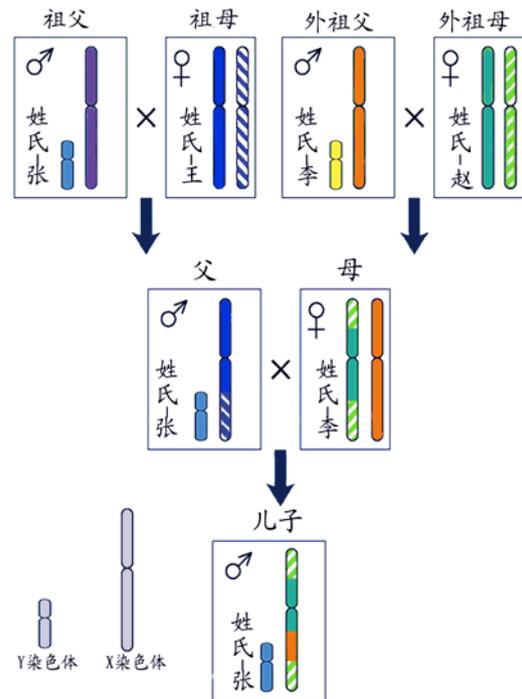


图1 姓氏和Y染色体的共同传递

Fig.1. Parallel Inheritance of Surnames and Y-chromosomes

组呢? 要回答这个问题, 必须先了解Y染色体的结构。人类Y染色体DNA大约包含六千万个碱基对, 其中染色体两端的5%为拟常染色体区域(pseudoautosomal region), 在传代过程中与X染色体相应区段会发生重组, 而主干部分的95%为非重组区域(NRY, non-recombining portion of Y chromosome), 不与任何染色体发生重组(图2)[36,37]。所以, Y染色体主干部分的此特性, 保证了子代能完整地继承父代的Y染色体主干而不受混血影响, 保证了Y染色体主干的严格父系遗传。

所以, 当姓氏已经无法作为追寻祖先的可靠标记的时候, 以现代的分子生物学技术为基础, 研究Y染色体主干的类型分布, 是直接追溯群体或者家族的父系起源的最佳方法, 是验证祖先与后代的父系关联的唯一手段。例如, 在曹操的后人中分析Y染色体特征, 我们就可以了解曹操本人的Y染色体特征。实际上, 在一段有较可信历史记录的时期内, 整个家族的姓氏与父系遗传的关联是可以保证的, 所以家族的姓氏往往与固定的Y染色体类型共同传递, 紧密关联。

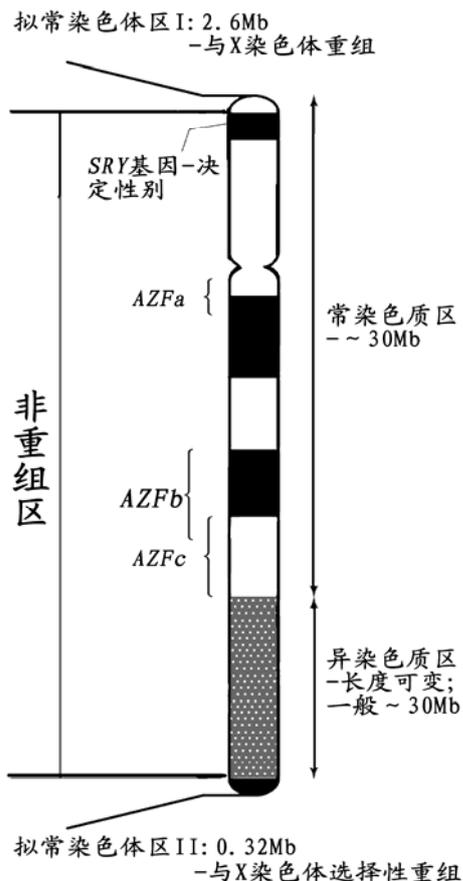


图2 Y染色体模式图

Fig. 2. Idiogram of human Y chromosome

稳定中变化着的 Y 染色体

在一代一代的父子相承的传递过程中，Y染色体也在慢慢地积累着变化。正是因为遗传突变的积累，使得人类父系遗传体系中，距离越远的个体的Y染色体差异也越大。Y染色体上的突变形成的个体差异主要有两大类，单核苷酸多态(SNP)和短串联重复(STR)(图3)。DNA分子由四种碱基(A、T、C、G)按照一定的顺序连接而成，SNP是仅仅一个位置上的碱基类型变化。Y染色体上的同一个SNP在人群中一般只有两种类型。STR则是在染色体的特定区段，由几个碱基组成一个单位重复出现，不同的Y染色体上的同一个STR位置往往有不同的重复拷贝数。SNP和STR由于突变性质和突变速度不同，在分析中有着不同的用途。

要确立父系遗传体系，最重要的前提是祖先的突变可以稳定的保留在后代的Y染

染色体上。SNP突变因为突变速率极低，可以做到在后代中永久地保留，后代只能在祖先的突变基础上积累新的突变，而不会丢失祖先的突变特征。通过比较人类与黑猩猩的Y染色体差异[38]，以及大家系中的Y染色体的差异程度[39]，Y染色体上的SNP突变的速率被计算了出来。每出生一个男子，一个染色体位置上发生SNP突变的概率为大约三千万分之一。实际上由于Y常染色质区的保守性，以及人类历史上大量男子都没有男性后代保留至今的事实，实际的群体中突变率应该低几个数量级。而我们通常研究的是Y染色体非重组区大约三千万个碱基对的常染色质区[40,41]，按照每个碱基对三千万分之一的突变率，这个区段内每个男子平均都会有一个新的突变。这个新的突变随机地出现在Y常染区的任意一个点上，如果这个突变的点上再发生一次突变，那么这个突变就在后代中丢失了，我们就无法通过后代确定祖先的Y染色体突变谱。但是同一个点上先后发生两次突变的概率，按照概率计算方法就是三千万分之一的平方，也就是九百万亿分之一，相对于人类自古以来的人口，这个概率就近似于零。所以我们可以说，绝大多数情况下，祖先的Y染色体上出现的SNP突变特征在后代中能够找到，而后代只能在祖先Y染色体突变谱的基础上增加新的突变(图4)。

由多个SNP突变构成的一种突变谱被称为一种单倍型。例如图4中就有5个SNP突变，陆续构成5种单倍型。其中1型是其他单倍型的祖先型，其他单倍型都是后代型。祖先型与所有后代型合称为一个单倍群。一个家族的所有Y染色体理论上都属于一个单倍群，因为其中所有的男性都应该来自同一个祖先。

当然，单倍群的概念可大可小。大而言之，全世界的Y染色体都属于一种单倍群，都来自二十多万年前一个东非晚期智人男子[42,43]。进而，全世界又可以分为20种主干单倍群，编号从A到T。最古老的A和B单倍群都没有走出非洲，C和D单倍群最早来到了澳洲和亚洲，E单倍群来到了亚洲又回到非洲，F单倍群衍生出GHIJ等单倍

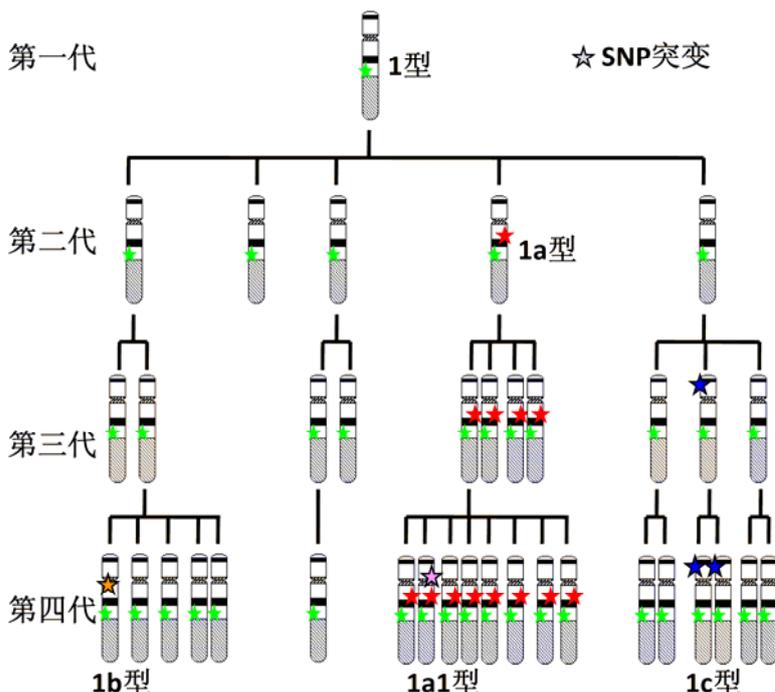


图 4 Y 染色体突变谱构成单倍型

Fig.4. Y chromosome haplotypes formed by accumulation of SNP mutations

表 1 不同文献估算的 STR 突变率 Tab.1. Y chromosome STR mutation rates estimated in different studies

突变率类型	作者	发表时间	突变率/(位点·代)
家系突变率	Heyer et al. [44]	1997	2.1×10^{-3}
家系突变率	Caglià et al. [45]	1997	$0.27-1.1 \times 10^{-3}$
家系突变率	Jobling et al. [46]	1999	2.1×10^{-3}
家系突变率	Kayser et al. [47]	1997	3.2×10^{-3}
家系突变率	Kayser et al. [48]	2000	2.8×10^{-3}
进化突变率	Forster et al. [49]	2000	2.6×10^{-4}
家系突变率	Holtkemper et al. [50]	2001	4.0×10^{-3}
进化突变率	Zhivotovsky et al. [51]	2004	6.9×10^{-4}

可由公式 $t = -Ne \ln(1 - V/Ne\mu)$ 推算出某特定 Y-SNPs 发生的大致时间。公式中，Ne 是有效群体大小， μ 是突变率，ln 是自然对数，V 是观察到的群体中的某一 STR 数值的方差，计算得到的 t 是经历的世代数，再乘以每一世代的年数即可得到时间 [54]。

以 Y 染色体上 STR 的总突变率二分之一来估算，几乎每个人可以构成独特的单倍型。但是，由于突变是一步一步发生的，父系亲缘关系越近的个体之间的 STR 单倍型越相似，一个纯粹由父系传递的姓氏应有相近的 STR 单倍型。理论上，有了足够数量的 Y 染色体 SNP 和 STR 后，通过调查一个姓氏

宗族内的男性的单倍型，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。

姓氏与 Y 染色体关联实践分析

在实际应用中，姓氏与 Y 染色体是否具有基本相同的和平行的表现还要看姓氏传递是否连续和稳定。多项研究证实各国的姓氏的传承是相对稳定的。袁义达等 [55-58] 分析和比较了宋朝、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，发现 1000 多年前的姓氏分布和当代基本一致。Legay 等 [59] 比较了法国 19 世纪末的

和1975年的若干姓氏的分布,也发现了其传递的稳定性。这些研究为姓氏和Y染色体关联分析提供了理论依据。Giraldo等[60]发现哥伦比亚的4个有“卫星式”Y染色体的家族有3个是同姓氏的。Sykes等[62]检测DYS19、DYS390、DYS391、DYS393四个Y-STR位点,对48名随机的Sykes姓氏英国男性个体进行Y-STR单倍型分析,发现21名个体(43.8%)的4个Y-STR单倍型相同,而该单倍型在对照组的139名随机的英国男性个体中却未检出;这反映了Sykes姓氏比较单一的起源。其余的Sykes姓氏个体的Y-STR单倍型与对照组差异不明显,若不考虑Y-STR基因变异的累积, Sykes家族的姓氏变更率为每代1.3%。与Sykes等只选一个姓氏作研究不同, King等[63]用150对随机选出的男性(每对两人有一个共同的英国姓氏)来分析,也发现同一姓氏,出现相同的Y染色体单倍型的机率提高;并且随着姓氏在人群中分布频率的减少, Y染色体单倍型相同的机率增加。Hill等[64]和Wilson等[65]的研究也表明爱尔兰、奥克尼群岛男性的Y染色体单倍型分类与姓氏显著相关。

利用Y染色体来检测历史上的家族关系疑案,有多项成功的案例,较有意思的是Foster[66]和Skorecki[67]的研究。1802年,美国第三任总统Thomas Jefferson因被怀疑与女仆Sally Hemings有过孩子而遭起诉。此后,人们一直对此事争论不休,而Foster用Y染色体回答了这个问题。Foster比较了Jefferson的叔叔、Sally的大儿子和最小儿子的男性后代Y染色体YAP、一些SNP、STR及小卫星MSY1等多态位点,得出结论Jefferson是Sally的最小儿子的生父。Y染色体不但能够解决数百年的疑案,还能追溯到数千年前的历史。Skorecki等就证实了圣经中的传说。圣经中记载犹太人中的祭司是由犹太教的第一祭司Aaron开始按血缘代代相传,而身为德系犹太人祭司的Skorecki发现他与一个西班牙系犹太人祭司的体质特征差别很大,这让他寝食难安。Skorecki就和研究Y染色体的专家Hammer教授合作以YAP和DYS19来分析犹太教祭司的单倍型,结果显示德系和西班牙系犹太祭司们与非

祭司的犹太人相比有较近的亲缘关系,也就是说祭司们可跨越3300年追溯到一个共同的父系祖先。Y染色体的分析与圣经故事的完美契合着实让人吃惊。

对于中国的姓氏与Y染色体的相关性,也有许多研究见诸报道。吴东颖等[68]应用Y染色体的DYS19、DYS390、DYS391和YCAII四个STR,分析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和分布均无显著差异。邓志辉等[69]通过对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。这说明,汉族的大姓内部基本没有同源性,相关Y染色体研究只能在明确的姓氏宗族中开展。

汉族大姓氏内部的不一致,有很多可能的原因。在理想的情形下,每种姓氏都有一个唯一来源,即该姓氏的奠基者只是一人或是有相同Y染色体单倍型的多人,在姓氏传承过程中没有发生过干扰(改姓、非亲生等),此时一种姓氏可以被一种SNP和STR的单倍型来鉴定[70]。但是中国的大多数姓氏起源并不单一。周朝的姓氏大多是以封国为氏,后改为姓。比如曹国的王室后代姓曹,但是其仆役后人也可以姓曹,甚至整个封国内所有百姓后代都可以姓曹。而曹国内的百姓来源本来就是多样的,有着各种各样的Y染色体。所以中国的姓氏总体上内部父系血缘不一致。另外,犹如STR单倍型随时间而演化出越来越多的类型一样,姓氏在传承过程中经历的时间越长,其受到的社会干扰越多,显示出的差异也越大。在中国,姓氏有近5000年的历史,来源复杂,且存在避祸改姓、避讳改姓、过继改姓、皇帝赐姓与贬姓、少数民族用汉姓等等问题。举个简单的例子,中国的100个大姓中有53个据称起源于姬姓[58]。如此,研究中国的姓氏难度极大,但是中国编修家谱的传统对厘清这份繁复的血缘关系有很大帮助。

家谱是一种以表谱形式记载某一同宗共祖以血缘关系为主体的家族世系繁衍兼

及其他方面情况的特殊图书体裁。也就是说,入谱者必须是同宗共祖,即使同姓,若不同祖,也不能修入一部家谱之中[71]。在中国的广大农村,人们一直有着同姓聚居的习俗,加上婚姻半径较小,由家谱确定的某一地域内同姓人群,可以认为是有相同或相近Y染色体的父系隔离群体,这也就为分子人类学分析Y染色体DNA多样性提供了极好的研究模型。然而,某些家谱里有假托、借抄的内容[72],因此对于家谱资料的应用必须审慎。但是在Y染色体检验这种无可辩驳的科学证据面前,任何家谱都可以得到检验和修正。姓氏、家谱和Y染色体的关联研究必然成为研究中国人起源与演变的重要方式,开创历史人类学研究的新篇章。

致谢

本研究得到复旦大学文科科研推进计划资金支持。

参考文献

1. 范又琪 (1999) 姓氏起源. 武汉: 长江出版社. 3-4.
2. 袁义达, 邱家儒 (2007) 中国姓氏·三百大姓. 上海: 华东师范大学出版社.
3. 丹羽基二 (1997) 日本苗字大辞典. 东京: 芳文馆.
4. Darwin GH (1875) Marriages between first cousins in England and their effects. *J Statist Soc* 38:153-184.
5. Crow JF, Mange JF (1965) Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugenics Q* 12:199-203.
6. Roberts DF, Rawling CP (1974) Secular trends in genetic structure: an isonymic analysis of Northumberland parish records. *Ann Hum Biol* 1:393-410.
7. Morton NE, Smith C, Hill R, Frackiewicz A, Law P, Yee S (1976) Population structure of Barra (Outer Hebrides). *Ann Hum Genet* 39:339-352.
8. Shin EH, Yu EY (1984) Use of Surnames in Ethnic Research: The Case of Kims in the Korean-American Population. *Demography* 21: 347-360.
9. Lasker GW, Kaplan BA (1985) Surnames and genetic structure: Repetition of the same pairs of names in married couples, a measure of subdivision of the population. *Hum Biol* 55:431-440.
10. Mascie-Taylor CGN, Lasker GW (1985) Geographical distribution of common surnames in England and Wales. *Ann Hum Biol* 12:397-401.
11. Pinto-Cisternas J, Pineda L, Barrai I (1985) Estimation of inbreeding by isonymy in Iberoamerican populations: an extension of the method of Crow and Mange. *Am J Hum Genet* 37: 373-385.
12. Barrai I, Barbujani G, Beretta M, Maestri I, Russo A, Formica G, Pinto-Cisternas J (1987) Surnames in ferrara: distribution, isonymy and levels of inbreeding. *Ann Hum Biol* 14:415-423.
13. Piazza A, Rendine S, Zei G, Moroni A, Cavalli-Sforza LL (1987) Migration rates of human populations from surname distributions. *Nature* 329:714-716.
14. Barrai I, Formica G, Barale R, Beretta M (1989) Isonymy and migration distance. *Ann Hum Genet* 53:249-262.
15. Holloway S, Sofaer JA (1989) Coefficients of relationship by isonymy within and between the regions of Scotland. *Hum Biol* 61:87-97.
16. Mascie-Taylor CGN, Lasker GW (1990) The Distribution of Surnames in England and Wales: A Model for Genetic Distribution. *Man New Series* 25 :521-530.
17. Sokal R, Harding R, Mascie-Taylor CGN (1992) A spacial analysis of 100 surnames in England and Wales. *Ann Hum Biol* 19:445-476.
18. Rodriguez-Larralde A, Pavesia A, Scapolia C, Conterio F, Siriaand G, Barraia I (1994) Isonymy and the genetic structure of Sicily. *J Biosoc Sci* 26:9-24.
19. Mourrieras B, Darlu P, Kochez J (1995) Surname distribution in France: A distance analysis by distorted geographical map. *Ann Hum Biol* 22:183-198.
20. Barrai I, Scapolia C, Beretta M, Nesti C, Mamolini E, Rodriguez-Larralde A (1996) Isonymy and the genetic structure of Switzerland. I. The distribution of surnames. *Ann Hum Biol* 23:431-455.
21. Barrai I, Scapoli C, Beretta M (1997) Isolation by distance in Germany. *Hum Genet* 100:684.
22. Rodriguez-Larralde A, Barrai I, Nesti C (1998) Isonymy and isolation by distance in Germany. *Hum Biol* 70:1041-1056.
23. Barrai I, Rodriguez-Larralde A, Mamolini E (1999) Isonymy and isolation by distance in Italy. *Hum Biol* 71:947-961.
24. Barrai I, Rodriguez-Larralde A, Mamolini E (2000) Elements of the surname structure of Austria. *Ann Hum Biol* 27:607-622.
25. Rodriguez-Larralde A, Morales J, Barrai I (2000) Surname frequency and isonymy structure of Venezuela. *Am J Hum Biol* 12:352-362.
26. Barrai I, Rodriguez-Larralde A, Mamolini E (2001) Isonymy structure of USA population. *Am J Phys Anthropol* 114:109-123.
27. Barrai I, Rodriguez-Larralde A, Manni F (2002) Isonymy and isolation by distance in the Netherlands. *Hum Biol* 74:263-283.
28. 徐立群, 李辉, 奚慧峰, 金力 (2002) 上海郊区姓氏和通婚分析. *遗传学报* 29:666-673.
29. Rodriguez-Larralde A, Gonzales-Martin A, Scapoli C, Barrai I (2003) The Names of Spain: A Study of the Isonymy Structure of Spain. *Am J Phys Anthropol* 121:280-292.
30. Dipierri JE, Alfaro EL, Scapoli C, Mamolini E, Rodriguez-Larralde A, Barrai I (2005) Surnames in Argentina: A Population Study Through Isonymy. *Am J Phys Anthropol* 128:199-209.
31. Jorde LB, Morgan K (2005) Genetic structure of Utah Mormons: Isonymy analysis. *Am J Phys Anthropol* 72:403-412.
32. Tay JSH, Yip WCL (2007) The estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient. *Ann Hum Genet* 48:185-194.
33. Cook D, Hewitt D, Milner J (1972) Uses of the surname in epidemiologic research. *Am J Epidemiol* 95:38-45.
34. Abbotts J, Williams R, Smith GD (1999) Association of medical, physiological, behavioural and socio-economic factors with elevated mortality in men of Irish heritage in West Scotland. *J Publ Health Med* 21:46-54.
35. Polednak AP (1996) Estimating breast cancer incidence in Hispanic women in Connecticut, 1989-1991. *Ethn Health* 1:229-235.
36. Quintana-Murci L, Krausz C, McElreavey K (2001) The human Y chromosome: function, evolution and disease. *Forensic Sci Intl* 118:169-181.
37. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110: 118-124.
38. Kuroki Y, Toyoda A, Noguchi H, Taylor TD, Itoh T, Kim DS, Kim DW, Choi SH, Kim IC, Choi HH (2006) Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nat Genet* 38:158-167.

39. Xue YL, Wang QJ, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao YL, MacArthur DG, Quail MA, Carter NP, Yang HM, Tyler-Smith C (2009) Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree. *Curr Biol* 19:1453-1457.
40. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
41. Venter JC (2001) The sequence of the human genome. *Science* 291:1304-1351.
42. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598-612.
43. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonn -Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd K, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner P (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358-361.
44. Heyer E, Puymirat J, Dietjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799-803.
45. Cagli  A, Novelletto A, Dobosz M, Malaspina P, Ciminelli BM, Pascali VL (1997) Y-chromosome STR loci in Sardinia and continental Italy reveal island-specific haplotypes. *Eur J Hum Genet* 5:288-92.
46. Jobling MA, Bouzekri N, Taylor PG (1998) Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). *Hum Mol Genet* 7:643-653.
47. Kayser M, Cagli A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Knijff P, Roewer L (1997) Evaluation of Y chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125-133.
48. Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger K, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am J Hum Genet* 66:1580-1588.
49. Forster P, R hl A, L nnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A Short Tandem RepeatBased Phylogeny for the Human Y Chromosome. *Am J Hum Genet* 67:182-196.
50. Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B (2001) Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Hum Mol Genet* 10: 629-633.
51. Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *Am J Hum Genet* 74:50-61.
52. Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* 28(2):e8.
53. Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449-456.
54. Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza LL, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y chromosome Evidence for a northward migration of modern humans in East Asia during the last Ice Age. *Am J Hum Genet* 65: 1718-1724.
55. 袁义达,张诚,马秋云,杨焕明(2000)中国人姓氏群体遗传 I. 姓氏频率分布与人群遗传分化. *遗传学报* 27:471-476.
56. 袁义达,张诚,杨焕明(2000)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. *遗传学报* 27:565-572.
57. 袁义达,金锋,张诚(1999)宋朝中国人的姓氏分布与群体结构分化. *遗传学报* 26: 187-197.
58. 袁义达,张诚(2002)中国姓氏:群体遗传与人口分布. 上海:华东师范大学出版社.
59. Legay JM, Vernay M (2000) The distribution and geographical origin of some French surnames. *Ann Hum Biol* 27:587-605.
60. Giraldo A, Mart nez I, Guzm n M (1981) A family with a satellited Yq chromosome. *Hum Genet* 57:99-100.
61. Schmid M, Haaf T, Eva Solleder, Schempp W, Leipoldt M, Heilbronner H, Schmid M (1984) Satellited Y chromosomes: structure, origin and clinical significance. *Hum Genet* 67:72-85.
62. Sykes B, Irven C (2000) Surname and the Y-chromosome. *Am J Hum Genet* 66: 1417-1419.
63. King TE, Ballereau SJ, Schurer KE, Jobling MA (2006) Genetic signatures of coancestry within surnames. *Curr Biol* 16: 384-388.
64. Hill EW, Jobling MA, Bradley DG (2000) Y chromosome variation and Irish origins. *Nature* 404:351-352.
65. Wilson JF, Weiss DA, Richards M (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 98:5078-5083.
66. Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Rene Mieremet, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. *Nature* 396: 27-28
67. Skorecki K, Selig S, Blazer S, Bradman R, Bradman N, Warburton PJ, Ismajlowicz M, Hammer MF (1997) Y chromosomes of Jewish priests. *Nature* 385:32
68. 吴东颖,马素参,刘明,黄尚志,刘春芸(2000)应用Y染色体多态标记对汉族王姓亲缘关系的研究. *人类学学报* 19:132-137.
69. 邓志辉,李茜,王大明,高素青,曾健强(2007)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. *遗传* 29: 1336-1344.
70. Jobling MA (2001) In the name of the father: surnames and genetics. *Trends Genet* 17:353-357.
71. 王鹤鸣(2009)中国家谱总目. 上海:上海古籍出版社.
72. 赵世瑜(2006)祖先记忆、家园象征与族群历史—山西洪洞大槐树传说解析. *历史研究* 1:49-64.