

# Y 染色体 DNA 遗传标记在汉藏语系民族研究中的应用\*

张秀秀<sup>1,2\*\*</sup>, 何 燕<sup>1,2\*\*\*</sup>

(1. 贵州医科大学 地方病与少数民族疾病教育部重点实验室, 贵州 贵阳 550004; 2. 贵州医科大学 贵州省医学分子生物学重点实验室, 贵州 贵阳 550004)

**[摘 要]** Y 染色体 DNA 存在严格的父系遗传和非重组等特性, 其上单核苷酸多态性(Y-SNP)和短串联重复序列(Y-STR)已广泛用于人群或族群的父系遗传背景研究; 本文针对 Y 染色体常用遗传标记 Y-SNP 和 Y-STR 的特性, 综述其在汉藏语系民族的起源、迁徙和融合等相关研究中扮演的角色和重要作用。

**[关键词]** Y-SNP; Y-STR; 民族起源; 民族迁徙; 汉藏语系民族

**[中图分类号]** R34 **[文献标识码]** A **[文章编号]** 1000-2707(2019)04-0373-06

**DOI:**10.19367/j.cnki.1000-2707.2019.04.001

## Application of Y Chromosome DNA Genetic Markers in Ethnic Studies of Sino-Tibetan Languages

ZHANG Xiuxiu<sup>1,2</sup>, HE Yan<sup>1,2</sup>

(1. Key Laboratory of Endemic and Ethnic Diseases, Ministry of Education, Guizhou Medical University, Guiyang 550004, Guizhou, China; 2. Key Laboratory of Molecular Biology, Guizhou Medical University, Guiyang 550004, Guizhou, China)

现代人类走出非洲后, 由于地理的隔离和气候上的差异以及历史长期的积淀<sup>[1]</sup>和语言交流不通<sup>[2]</sup>等因素, 形成许多不同的族群。远古时期, 人类没有文字, 在了解这些族群的演变及融合过程时只能通过口头的传承得到这些族群的起源、变迁过程<sup>[3]</sup>。由于没有具体文字的记载, 从而使得很多族群在生物学起源和迁徙的问题上变得越来越模糊; 在群体的起源以及与其他群体关系的研究中, 相对于语言、历史、考古来说, 分子生物学的方法会相对精准和科学得多<sup>[4]</sup>。因为随着时间推移, 任何语言都会不断的变化, 并可能受到其他语言的影响, 历史记载无法排除主观性的存在<sup>[5]</sup>。考古学数据的可靠性有时会让人怀疑, 而脱氧核糖核酸(deoxyribonucleic acid, DNA)作为人类的遗传物质<sup>[6]</sup>, 虽受到环境的影响会在一定程度发生变异, 但这些变异是有规律且遵循一定速度的, 继而可用

作遗传时间和距离的标尺。可以说, 在人类起源和迁徙过程中, 外部留下的痕迹不多、却可在体内留下了清晰的“足印”: 人群分开的时间短, 遗传信息相似性就大; 分开的时间长, 相似性就小。因此, 分子生物学可以作为一个很好的测量工具, 通过研究人群间的这些有规律可循的变异或多态(即遗传标记), 进而度量人群间的相互关系, 可为错综复杂的人类起源、迁徙和融合等研究带来契机。因此, 利用遗传标记研究不同民族群体的遗传多样性对于了解汉藏语系的起源、迁徙以及相互关系有着重要意义。

## 1 Y 染色体 DNA 遗传标记

人类 Y 染色体 DNA (Y chromosome DNA, Y-DNA) 长约  $5.9 \times 10^4$  kb 的 DNA, 由两端的拟常染

\*[基金项目] 国家自然科学基金资助(31560306); 贵阳市科技计划项目基金资助[(2017)5-9]

\*\* 贵州医科大学 2016 级硕士研究生

\*\*\* 通信作者 E-mail: annieheyan@gmc.edu.cn

网络出版时间: 2019-04-26 网络出版地址: <http://kns.cnki.net/kcms/detail/52.1164.R.20190425.2119.001.html>

色体区(pseudoautosomal regions, PAR)和中间的男性特异区(male specific region of Y chromosome, MSY)组成。拟常染色区(大约占5%)位于Y染色体的两端,在男性减数分裂过程中,拟常染色区可与X染色体重组交换,其余95%为Y染色体男性特异区<sup>[7]</sup>。男性特异区不发生重组呈单倍型传递,在父系遗传中能够忠实地记录传代过程中所产生的突变<sup>[8]</sup>,形成特异的遗传标记,在人类遗传与进化方面意义重大。

Y染色体上常用的遗传标记有单核苷酸多态性(Y chromosome single nuclear polymorphism, SNP, Y-SNP)和短串联重复序列(Y chromosome short tandem repeat, Y-STR)两种<sup>[9]</sup>,其中Y-SNP突变速率较低,可反映较为久远的人群父系事件;Y-STR突变速率较高,多用于评估较近的历史事件。由于单一遗传标记所传达的信息量较少,分子遗传学研究中,往往是以单倍型(haplotype)甚至单倍群(haplogroup)来对人群的进化和迁徙展开研究。单倍型是单倍体基因型的简称,在遗传学上是指在同一染色体上进行共同遗传的多个基因座上等位基因的组合<sup>[10]</sup>,如由多个突变位点构成的一种突变谱。一般情况下,单倍型可由SNP或STR等遗传标记来进行确认,在分子进化的研究中,单倍群(或称单倍型类群)是指一组类似的单倍型,它们有一个共同的单核苷酸多态性祖先<sup>[11]</sup>,即祖先单倍型与所有后代单倍型合称一个单倍群。多年来,科学家们通过对世界多地人群的Y-SNP和Y-STR等遗传标记的研究发现,Y染色体单倍群分布具有很强的地理特异性<sup>[12]</sup>。研究不同地域及不同人群的Y染色体单倍群的频率分布,可用于推测人群进化、迁移及历史<sup>[13]</sup>。2003年Jobling MA等<sup>[14]</sup>修订了人类Y染色体单倍群系统发育树,即YCC(2003),图中根据Y染色体单倍型类群的不同可把全部现代智人分为18个单倍群,用从A到R的18个字母代表;18个单倍群下,又可分为多个更小的“子”单倍群,用其上“父”单倍群的字母和数字及其他字母来表示(图1)。谱系树中的父节点代表的对应基因突变是所有子节点共有的,但反之不然。例如单倍群D对应的基因突变是M174,单倍群D1对应的基因突变除了“父”单倍群D的M174外还有“子”单倍群D1特有的M15。YCC(2003)谱系图同时还推导出了全球Y单倍群图<sup>[15]</sup>,见图2。图中展现了世界各地人群的Y染色体单倍群的主要分属类型,并可根据各单倍群对应基因突变

出现的时间<sup>[16]</sup>,推测出相应人群的迁徙路径<sup>[17]</sup>。

## 2 汉藏语系民族的形成

在人类起源和迁徙等研究中,语言学是一种较为科学的方式。根据语言的发展和演变、直接而明显的关联,对语言进行归类的方法称为语言系属分类。语言系属分类主要依据语言语音、词汇、语法规则之间某些对应关系,把具有相似的语言归于同一类语群,这种语群称为同族语言即“语族”。按“语族”之间的某些对应关系,又归在一起,这类同类语族称为同系语言即“语系”<sup>[18]</sup>。其中东亚主要为汉藏语系,其下包含汉语族、藏缅语族、苗瑶语族、壮侗语族等<sup>[19]</sup>,见图3。

关于人类起源,目前较公认的“非洲起源说”认为,十多万年前,现代人共同起源于非洲<sup>[20]</sup>,部分Y染色体上带有M168位点突变的人群在9万多年前走出非洲,后到了中东与尼安德特人的祖先有一些基因上的交流<sup>[21]</sup>。6万年前携带M130突变的人群沿南线顺时针迁徙进入印度<sup>[22]</sup>沿着印度洋进入东南亚<sup>[23]</sup>,成为了现代亚洲人的祖先,5万年前其中一支前往东南亚岛屿和太平洋诸岛产生支系C2-M38和C4-M210,而后,C3-M48和C3-M407向北进入东亚、中亚等地区<sup>[8]</sup>。一部分人在3万年前,抵达我国珠江流域形成百越民族,还有一部分人,大约于1.5万年前由云南进入中国,到达了河套地区和黄河中上游的盆地形成了汉藏语系民族<sup>[23]</sup>。单倍群O3-M122(O3a1c-002611, O3a2c1 \*-M134, O3a2c1a-M117等)在汉藏语系民族几乎所有人群中都非常普遍<sup>[24]</sup>,是汉藏人群中主要的父系单倍群<sup>[25]</sup>,因此,在揭示这些群体的源流方面比其他单倍群更具信息性。Su等<sup>[26]</sup>也发现几乎所有的汉藏人群在单倍群O3-M122存在高频分布,并推测古羌人群是汉藏语系民族的祖先。在西藏以东的羌族群体中发现O3a2c1a-M117有很高的频率<sup>[27]</sup>,提示这个地区很有可能是汉藏群体的起源地。下面对汉藏语系民族各语族(汉语族、侗台语族、苗瑶语族和藏缅语族)进行一一探讨。

### 2.1 Y染色体多态位点在汉语族中的研究

汉族和回族归属于汉语族,从父系遗传学角度来看,居住在黄河中上游地区的氏羌人群中出现单倍群O3a2c1 \*-M134和O3a2c1a-M117形成了炎黄部落,以及东向分支携带O3a1c-002611发展成

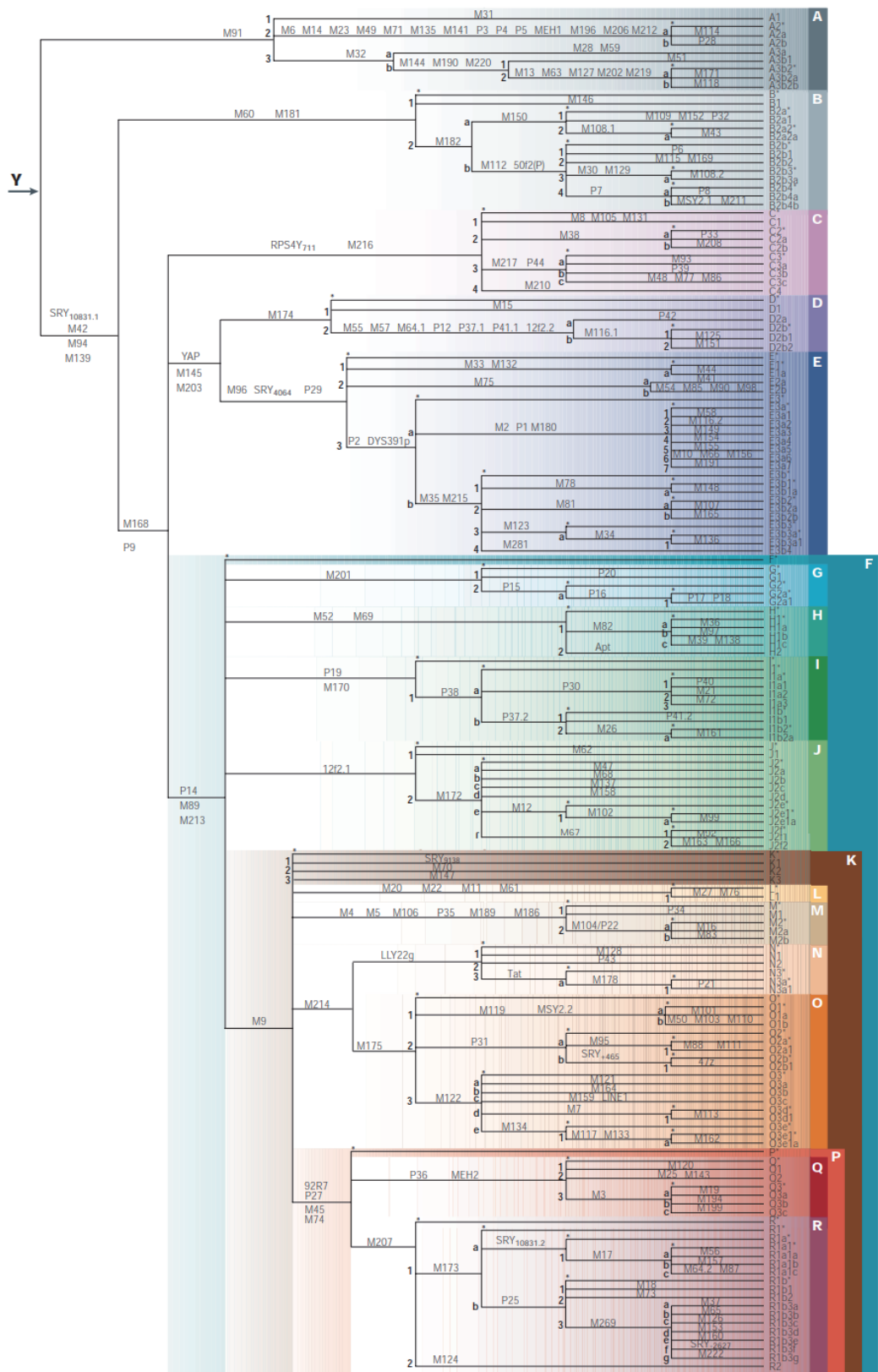


图 1 单倍群划分树谱

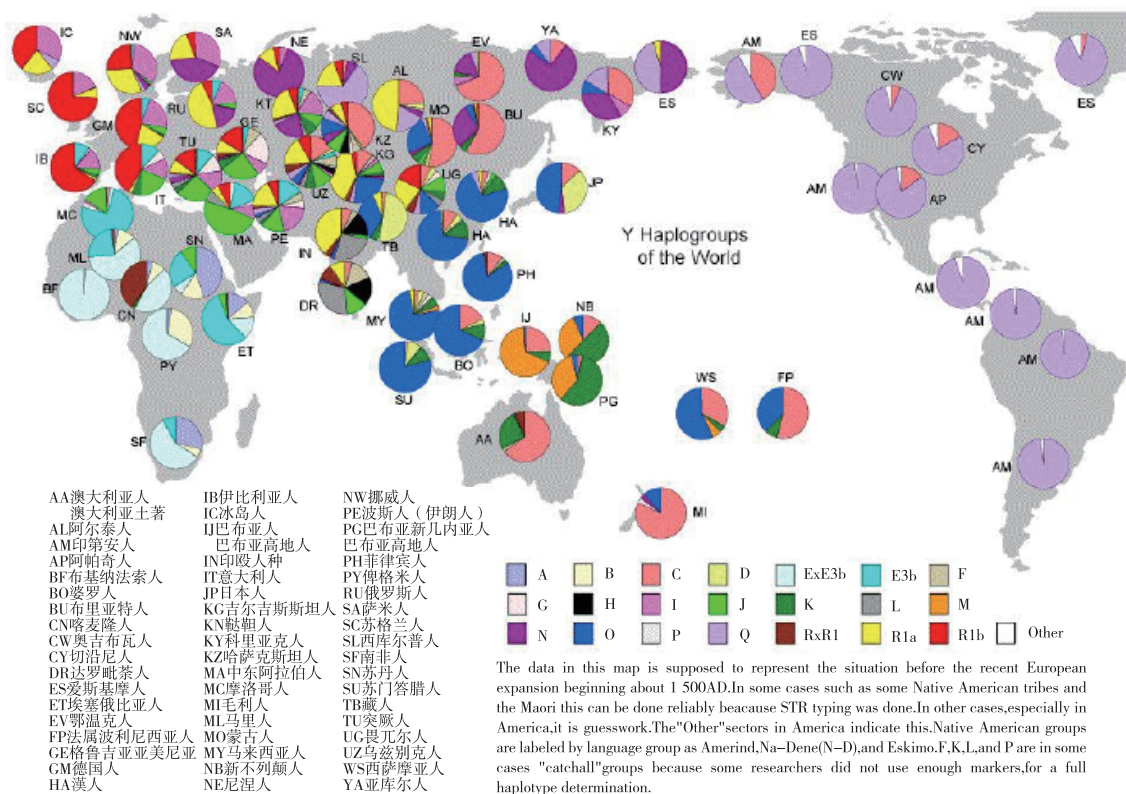


图 2 Y 染色体单倍群世界分布图

传结构上非常相似。

## 2.2 Y 染色体多态位点在苗瑶语族中的研究

苗族、瑶族、畲族归属于苗瑶语族。苗瑶语族主要分布在中国南部及东南亚地区,孟高棉语族和苗瑶语族在 Y 染色体遗传上有非常高的相似度,XIAO Y C 等<sup>[33]</sup>通过对 47 个群体进行分析发现苗瑶语族与孟高棉族群在遗传上密切相关,在两个群体中发现 O3a4-M7 的频率都很高。此外,单倍群 O3a4 在其他群体中几乎不存在<sup>[34]</sup>。梁祚仁<sup>[35]</sup>对 14 个民族 17 个 Y-STR 进行遗传数据分析并根据遗传距离进行系统进化树的绘制,在进化树上明显的看到,苗瑶语族(苗族、瑶族)与壮侗语族(仫佬族、侗族、佬族、水族、壮族、毛南族)以及汉语族(汉族)之间的遗传关系较近。

### 2.3 Y 染色体多态位点在藏缅语族中的研究

中国西南地区喜马拉雅南麓西起巴基斯坦,东到尼泊尔、越南等国家,分布着大约 300 多种藏缅语族语言,有的语言还跨境分布<sup>[36]</sup>,各族群之间存在唇齿相依的族源关系,藏缅群体源于中国西北的古老族群,在过去 2 000 ~ 3 000 年里,由于受到秦朝和其他王朝的压迫,他们被迫从中国西北地区南迁<sup>[37]</sup>。几乎所有的藏缅群体中单倍群 O3 \* 和 O3e 都存在高频分布<sup>[18]</sup>,然而,单倍型类群 O3a1c-

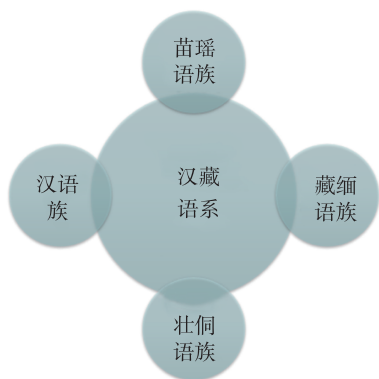


图3 汉藏语系民族主要的四大类语族

为东夷部落<sup>[28]</sup>。炎黄部落与东夷部落一起逐渐发展成为一个被称为汉族的大人群<sup>[28]</sup>。O3a2c1 \*-M134, O3a2c1a-M117 和 O3a1c-002611 作为 O3-M122 的亚群分别占汉族的 12% ~ 17%<sup>[29]</sup>, O3a2c1 \*-M134 和 O3a2c1a-M133 经常在一起出现,在汉族中尤其常见<sup>[30]</sup>。

汉族名称正式使用是在汉朝时期<sup>[31]</sup>。任贺<sup>[32]</sup>通过南方汉族与北方汉族样本的结果进行比较,发现 Y 染色体上的基因分布随着地域不同,遗传结构也存在着差异。而李辉等<sup>[19]</sup>分析了 28 个地区汉族群体显示北方汉族和南方汉族在父系遗



002611 被发现在藏缅语族群体中频率非常低,这表明该谱系可能没有参与藏缅语族种群的形成<sup>[38]</sup>。韩建利<sup>[39]</sup>对云南大理白族男性人群 43 个 Y-SNP 位点进行遗传学分析,发现单倍群 O3-M122、O3a1c-002611 在白族的分布频率分别为 0.428 1、0.097 1,与其他学者的研究结果相符。

2.4 Y 染色体多态位点在壮侗语族中的研究

壮族、侗族、水族、黎族、傣族、布依族、仫佬族和毛南族归属于壮侗语族<sup>[40]</sup>(图 4)。目前学术界在壮侗语族源流问题上认为,壮侗语族与古代的百越族同源<sup>[41]</sup>,O2a1-M95 与 O1-M119 被认为是壮侗语族的高频单倍群<sup>[42]</sup>。

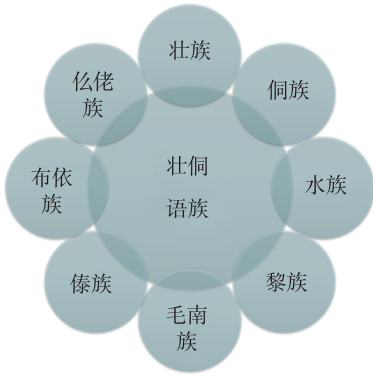


图 4 壮侗语族的 8 个民族

在新石器时代,中国东南部人口扩张时壮侗语族群体衍化出单倍群 O- M175<sup>[33]</sup>,O2a1-M95 谱系在大约 2~4 万年起源于东亚南部的壮侗语族<sup>[43]</sup>。YANG Z 等<sup>[44]</sup>对云南少数民族 Y 染色体单倍群进行遗传学分析,发现水族在单倍群 O3a2c1 -M134 具有高频分布(55.0%),O2a1-M95 为壮族和布依族的主要单倍群,频率分布分别为 38.3% 和 41.7%。

综上所述,从 Y 染色体父系遗传角度分析,汉藏语系民族的父系起源不是单一的,而是多元的,汉藏语系民族是语族间、民族间以及与其他语系、语族、民族之间发生基因交融的结果;在某一地域的同一民族或语族人群可以认为有相同或相近的 Y 染色体父系遗传结构。Y 染色体与人群的关联研究必然成为研究群体起源和演变、迁徙的重要方式,Y 染色体非重组区多态性在鉴别民族的起源和迁移研究中具有较大的准确度和可操作性,必将继续发挥其巨大作用。

3 展望

随着生命科学各领域尤其是分子生物学的突

破性进展,人类学、民族学、遗传学等的目光已经从语言学、历史学、考古学转移到了分子生物学,Y-STR 和 Y-SNP 可以为各语族提供近期和早期的遗传学证据,为各个语族的源流提供分子生物学依据,其目标将是建成一个综合的、系统的 DNA 数据库(DNA database),为相关研究提供可参考的依据。

4 参考文献

[1] 夏国恩,蓝勋. 融入民族文化的高等财经教育新型人才培养模式探究——以广西壮乡文化为例[J]. 经济研究参考, 2016,(70):75-79.

[2] 吕裕彬. 方言、族群和经济相互关系[D]. 南宁:广西大学, 2017.

[3] 曹端波. 苗族古歌中的苗族祖居地与民族迁徙[J]. 贵州师范学院学报, 2014,30(2):1-9.

[4] 李锋. 东亚现代人来源的考古学思考:证据与解释[J]. 人类学学报, 2018,37(2):176-191.

[5] 胥志刚. 浅论历史研究的客观性问题[J]. 黑龙江史志, 2015,(9):357-362.

[6] 翁屹,张翻. 如何发现 DNA 是生命的遗传物质[J]. 哈尔滨工业大学学报(社会科学版), 2010,12(2):1-7.

[7] 黎宇飞. Y-STR 与姓氏的关联性及其法医学和遗传学意义的研究[D]. 沈阳:中国医科大学,2012.

[8] 廖世玉. 南亚语系人群的父系起源和史前迁徙的遗传学研究[D]. 合肥:安徽大学, 2015.

[9] 韩建利. 云南白族男性人群 Y-SNP 遗传多态性研究[D]. 昆明:昆明医科大学, 2014.

[10] 孙明霞. CYP21A2 基因单核苷酸多态性与汉族痤疮的关联性研究[D]. 上海:复旦大学, 2012.

[11] 张永科. 利用 Y 染色体非重组区多态探究新疆阿勒泰地区图瓦人与邻近人群的遗传学关系[D]. 杭州:中山大学, 2009.

[12] P V, MJ I, L V, et al. Characterization of the iberian Y chromosome haplogroup R-DF27 in northern spain[J]. Forensic Science International Genetics, 2016,27:142.

[13] 刘树虎,热比亚木·巴克,伊力哈木·乃扎木,等. 和田维吾尔族人群 Y 染色体遗传多样性分析[J]. 生物技术, 2017,27(1):38-44.

[14] JOBLING M A, TYLER-SMITH C. The human Y chromosome: an evolutionary marker comes of age[J]. Nature Reviews Genetics, 2003,4(8):598-612.

[15] WEI W, AYUB Q, XUE Y, et al. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping[J]. Forensic Science In-

- ternational Genetics, 2013, 7(6):568–572.
- [16] GAUDENCIO S, COSTA H A, AMORIM A, et al. Y-SNPs genetic variation in the Caucasian population living in southern Portugal. *Forensic Science International: Genetics Supplement Series*, 2015, (5):e344–e345.
- [17] 刘树虎. 新疆三个隔离人群 Y 染色体和线粒体遗传多样性分析[D]. 乌鲁木齐:新疆大学, 2018.
- [18] 熊霄, 杨希. 从地理语言学和语系的角度审视汉语与英语的差异[J]. *湖北函授大学学报*, 2015(14).
- [19] 李辉, 金力. Y 染色体与东亚族群演化[M]. 上海科学技术出版社, 2015.
- [20] VIGILANT L, STONEKING M, HARPENDING H, et al. African populations and the evolution of human mitochondrial DNA. [J]. *Science*, 1991, 253(5027): 1503–1507.
- [21] GREGORY M D, KIPPENHAN J S, EISENBERG D P, et al. Neanderthal-derived genetic variation shapes modern human cranium and brain[J]. *Sci Rep*, 2017, 7(1): 6308.
- [22] UNDERHILL P A, PASSARINO G, LIN A A, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. [J]. *Annals of Human Genetics*, 2012, 65(1): 43–62.
- [23] 石无鱼. 中国人的起源[J]. *大科技(科学之谜)*, 2010, (4): 24–25.
- [24] WANG C C, YAN S, QIN Z D, et al. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611[J]. *Journal of Systematics & Evolution*, 2013, 51(3): 280–286.
- [25] EXCOFFIER L, LAVAL G, SCHNEIDER S. Arlequin (version 3.0): An integrated software package for population genetics data analysis[J]. *Evol Bioinform Online*, 2005, 1(4A): 47–50.
- [26] SU B, XIAO C, DEKA R, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas [J]. *Human Genetics*, 2000, 107(6): 582–590.
- [27] 孙宏开. 汉藏语系语言的共同创新[J]. *民族语文*, 2014(2): 3–13.
- [28] WANG C C, WANG L X, SHRESTHA R, et al. Genetic structure of Qiangic populations residing in the western Sichuan corridor[J]. *PLoS ONE*, 2014, 9(8): e103772.
- [29] YAO X, TANG S, BIAN B, et al. Improved phylogenetic resolution for Y-chromosome haplogroup O2a1c-002611 [J]. *Scientific Reports*, 2017, 7(1): 1146.
- [30] TREJAUT J A, POLONI E S, YEN J C, et al. Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia [J]. *Bmc Genetics*, 2014, 15(1): 77.
- [31] 孙大虹. 毛主席是我们家里人云南 26 个民族人文情怀影像纪实[J]. *青年与社会*, 2012(3): 88–97.
- [32] 任贺. 南方汉族群体 Y-STR 和 Y-SNP 系统的研究和法医学应用[D]. 太原:山西医科大学, 2007.
- [33] XIAO Y C, ZHEN D Q, BO W, et al. Human migration through bottlenecks from southeast asia into east asia during last glacial maximum revealed by Y chromosomes [J]. *Plos One*, 2011, 6(8): e24282.
- [34] ZHENDONG Q, BO W, et al. Human migration through bottlenecks from southeast asia into east asia during last glacial maximum revealed by Y chromosomes[J]. *PLoS ONE*, 2011, 6(8): e24282.
- [35] 梁梓仁. 广西 10 个少数民族 Y 染色体 17 个 STR 基因座的遗传多态性[D]. 广州:广州医学院, 2009.
- [36] 孙宏开. 跨喜马拉雅的藏缅语族语言研究[J]. *民族学刊*, 2015, 6(2): 69–76.
- [37] 谢选华. 从 Y 染色体遗传结构看藏缅语族人群的起源[D]. 上海:复旦大学, 2004.
- [38] YAN S, WANG C C, ZHENG H X, et al. Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers[J]. *Plos One*, 2014, 9(8): e105691.
- [39] 韩建利. 云南白族男性人群 YSNP 遗传多态性研究[D]. 昆明:昆明医科大学, 2016.
- [40] 黄桂秋, 吕妍. 骆越族群人类起源神话母题探析[J]. *广西师范学院学报(哲学社会科学版)*, 2013, 34(4): 10–14.
- [41] 洪琳, 王珏. 黄山地名里的千古之谜——歙县八个村名里的“那”[J]. *黄山学院学报*, 2017, 19(1): 70–75.
- [42] SHI H, DONG Y L, WEN B, et al. Y-Chromosome evidence of southern origin of the east asian-specific haplogroup O3-M122[J]. *American Journal of Human Genetics*, 2005, 77(3): 408–419.
- [43] ZHANG X, LIAO S, QI X, et al. Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent[J]. *Scientific Reports*, 2015, 5(4): 1–8.
- [44] YANG Z, DONG Y, GAO L, et al. The distribution of Y chromosome haplogroups in the nationalities from Yunnan Province of China[J]. *Annals of Human Biology*, 2005, 32(1): 8.

(2019-01-18 收稿, 2019-03-26 修回)  
中文编辑: 刘平; 英文编辑: 周凌