

# Origin of Hakka and Hakkanese :A Genetics Analysis

LI Hui<sup>1</sup>, PAN Wu-Yun<sup>2</sup>, WEN Bo<sup>1</sup>, YANG Ning-Ning<sup>1</sup>, JIN Jian-Zhong<sup>1</sup>, JIN Li<sup>1</sup>①, LU Da-Ru<sup>1</sup>②

(1. Center for Anthropological Studies at Fudan University, Shanghai 200433, China;

2. Institute of Linguistics, Shanghai Teachers University, Shanghai 200234, China)

**Abstract:** Hakka is a distinctive Han Chinese population in Southern China speaking Hakkanese. The origin of Hakka has been controversial. In this report, we analyzed Y chromosomal markers in 148 Hakka males. Principle component analysis of Y-SNP haplotype distribution shows Hakka is clustered strongly with the Han in Northern China, and is also close to She, a Hmong-Mien-speaking population, while the general Southern Han is fairly close to Daic populations. Admixture analysis revealed that the relative genetic contribution 80.2% (Han), 13% (She) and 6.8% (Kam) in Hakka. The network of Y-STR haplotype of M7 individuals in all concerned populations suggested two possible origins of Hmong-Mien contribution in Hakka. One is from Hubei and the other is from Canton. The Kam contribution in Hakka is likely from Kan-Yue, the ancient aborigine of Kiangsi (Jiangxi). The frequency of 9bp-deletion in Region V of mitochondrial DNA of Hakka is 19.7%, which is quite close to She but far from Han. We therefore concluded that genetically the majority of Hakka gene pool shall come from North Han with She contributing the most among all non-Han groups. Regarding the Hmong-Mien character of Hakkanese, the genetic structure of Hakka shows their core may be Kimman, the ancient Hmong-Mien. We hypothesized that a great number of Han people from North China join this population in succession. Southern Chinese dialects, such as Hakkanese may also be those languages of Southern aborigines at first, and turn to extant appearance under the continuance effect of Northern Chinese.

**Key words:** Hakka; genetic structure; Y-chromosome; SNP; STR; She nationality; Hakkanese

## 客家人起源的遗传学分析

李辉<sup>1</sup>, 潘悟云<sup>2</sup>, 文波<sup>1</sup>, 杨宁宁<sup>1</sup>, 金建中<sup>1</sup>, 金力<sup>1</sup>①, 卢大儒<sup>1</sup>②

(1. 复旦大学生命科学学院现代人类学研究中心, 上海 200433; 2. 上海师范大学语言学研究所, 上海 200234)

**摘要:** 客家人是南方讲客家话的特殊汉族民系, 其来源学术界有一定争议。该研究对福建长汀的 148 个客家男子做了遗传分析。从父系遗传的 Y 染色体 SNP 的主成分分析看, 客家人与中原汉族最近, 又偏向于苗瑶语族群中的畲族, 不同于其他南方汉族偏向于侗台语族群。混合分析发现客家人数据结构中汉族结构占 80.2%, 类畲族结构 13%, 类侗族结构 6.8%。各族 M7 个体 Y-STR 单倍型的网络结构分析发现客家人中类苗瑶结构有两个来源, 其一来自湖北, 其一来自广东。客家人之类侗族结构应来自江西土著干越。客家人母系遗传的线粒体 Region V 区段 9 bp 缺失频率为 19.7%, 与畲族很近, 不同于中原汉族。客家人的主要成分应是中原汉人, 畲族是对客家人影响最大的外来因素。与客家话中的苗瑶语特征相印证, 客家人可能是古代荆蛮族的核心成分不断加上中原汉人移民形成的。客家话等南方汉语方言最初也可能是南方原住民语言在中原汉语不断影响下逐渐形成的。

**关键词:** 客家人; 遗传结构; Y 染色体; 单核苷酸多态 (SNP); 短串联重复 (STR); 畲族; 客家话

中图分类号: Q987 文献标识码: A 文章编号: 0379-4172(2003)09-0873-08

收稿日期 2002-08-22; 修回日期 2003-05-25

基金项目: 国家自然科学基金九五计划重大项目(39993420); 复旦大学学生科技创新活动基金(2001B1-603) Priority Proj. No. 39993420, supported by the National Natural Science Foundation of China, and No. 2001B1-603, supported by Fudan Students' Foundation of Science and Technology Creative Projects.]

作者简介: 李辉(1978-), 男, 博士研究生, 研究方向: 分子人类学。E-mail: LH@cableplus.com.cn; Tel: (021) 65642419

① 通讯作者: 金力(1963-), 男, 研究方向: 人类分子遗传学。E-mail: jin-li@fudan.edu; Tel: (021) 65642419

② 通讯作者: 卢大儒(1965-), 男, 研究方向: 人类分子遗传学。E-mail: ludr@fudan.edu; Tel: (021) 65642799

客家人是指分布在我国南方各省的讲客家话的一个汉族民系,大约有 65 624 000 人口。在各国的华侨中也有很大比例的客家人。闽、粤、赣交界地区是客家人最集中的地区,其中长汀可能是客家人形成过程中的一个重要居住地<sup>[1]</sup>,是客家人中的主流。关于客家人的来源,很早就有人探究。三四十年代,针对当时存在的某些“客家非汉族论”、“客家为汉族与苗、瑶、僮、畲等族的混血种说”,罗香林把客家人讲成是“中原最纯正的正统汉人的后裔<sup>[2]</sup>”,而实际上,各地汉族在演变过程中,与外族混血是很普遍的现象,客家人也很难例外,所以其方言中许多常用词与苗瑶语、壮侗语有渊源关系<sup>[3]</sup>。关于客家人形成年代的观点也很多,袁家骅<sup>[4]</sup>认为“东晋永嘉以后,客家先民受战乱所迫,先后经历了几次的大迁徙运动”。

客家话是客家人的一个象征,是一种特殊的汉语方言。刘镇发<sup>[5]</sup>与刘纶鑫<sup>[6]</sup>都著文指出,严格意义上的客家话应该界定为明末清初嘉应州一带的方言以及从这些地方迁出去的客家人所操的方言。刘纶鑫还认为,这种方言不仅在明末清初以前,甚至在罗氏所谓前 3 次北民南迁以前已经形成,所以它还应该包括与嘉应方言有关的闽西 7 县和赣南 18 个县市的方言。郑张尚芳分析了与客家话相当接近的浙南畲话,指出浙南畲话不是畲族学的客家话,有可能是两者都从湖南迁向广东,畲族在唐代又迁到浙南,此过程中两者都学了古中国的南部官音,然后又因居住邻近而相互交融<sup>[7]</sup>。所以客家话研究是汉语研究中的一个重要范例,对于弄清大部分南方汉语方言的原始核心是汉语还是南方民族语有关键作用。

人群的起源涉及到其文化的来源、语言的产生等方面的人群特征的形成,而最根本的是人自身血缘的来历,这就要依靠遗传学的研究方法。由于技术手段的限制,早期对人群的遗传调查是通过人群的外在遗传特征,比如血型、盯聆、球蛋白等进行的<sup>[8]</sup>,这种不是无选择压力的指标,携带的人群分化的信息量极少,又受混血等因素影响,很难有群体特异性。随着 DNA 检测技术的不断完善,直接通过分析 DNA 分子特征来研究群体性质的工作发展了起来。线粒体 DNA 的研究持续了较长时间,它单纯的母系遗传特征避免了混血影响,较多的突变类型又可籍以区分个体,所以研究现代人群线粒体可以追溯其女性祖先的遗传特征。全世界人群的线粒体

研究得出了著名的“非洲夏娃学说<sup>[9]</sup>”。但是线粒体 DNA 的分子量相当小,使得其信息量有限,突变速度过快又难以把握其变化历史,群体间母系交流的过频又使群体的特征不明显,这些原因使目前的线粒体研究得出了一些不确的结论<sup>[10]</sup>,难以深入。但是线粒体始终是研究母系遗传的重要依据。纯父系的 Y 染色体非重组区弥补了线粒体的各种缺点,成为现阶段人群遗传研究的最重要指标。其中的单核苷酸突变(SNP)稳定慢速,各突变位点组合的单倍型构成了一个单根的系统树,把人类的父系血统也追溯到了非洲。SNP 单倍型在各民族系统间也有特异性的分布,有一定的民族标志作用<sup>[11]</sup>。Y 染色体上的另一种标记,短串联重复(STR)又具有较快的突变速度,能分析人群系统内部较短的演变历史<sup>[12]</sup>。本研究通过 Y-SNP 检测客家人的族属成分,又用部分 Y-STR 分析其发生渊源。

## 1 材料和方法

### 1.1 研究对象

福建长汀各地的客家男子 148 人的血样。各个体经查没有可知的外来血统,都是较纯正的客家人。由于个体 Y 染色体与整个家族理论上应一致,所以样本对总体的代表性较强。148 人互相没有较近的亲缘关系,其中涉及的姓氏有刘 13、钟 11、王 10、陈 9、张 9、李 9、丘 7、黄 6、谢 6、罗 5、马 5、付 4、胡 4、廖 3、林 3、涂 3、修 3、杨 3、童 3、曹 2、邓 2、郭 2、兰 2、卢 2、曾 2、蔡、戴、丁、范、韩、华、江、康、赖、吕、毛、彭、上官、沈、汤、巫、吴、肖、许、严、俞、郑,涵盖了绝大部分的长汀客家较大的姓氏<sup>[1]</sup>,基本可以代表长汀客家人。

### 1.2 实验方法

用常规的酚-氯仿法抽出血样中的 DNA。用核酸聚合酶链式反应(PCR)扩增线粒体 Region V 区段、Y 染色体各 SNP 多态位点(M130, YAP, M15, M89, M9, M122, M7, M134, M119, M110, M95, M88, M45, M120)和 STR 位点(DYS19, DYS389-1, DYS389-2, DYS390, DYS392)。Y 染色体 SNP 位点除了 YAP 和 M15 外,采用二步扩增法,在巢式扩增引物中引入针对 SNP 的酶切位点,用限制酶切割 PCR 产物。用琼脂糖凝胶电泳检测线粒体 Region V,

*M130*, *YAP* 等 PCR 产物和其他 Y-SNP 酶切产物。用 373 测序仪分析 STR 位点扩增产物长度。

### 1.3 数据分析

统计线粒体 Region V 区段 9 bp 缺失频率,直接与各地数据比较。

采用宿兵的命名系统<sup>[13]</sup>同几个 Y-SNP 单倍型的频率,与已报道的国内主要民族和地理上较近的人群<sup>[12,14~16]</sup>作比较。用 SPSS10.0 软件作主成分分析<sup>[17]</sup>观察各群体间的亲缘远近,分析客家人的族属。根据 Bertorella 的原理计算客家人 Y-SNP 数据中包含的汉族、畲族、侗族数据结构的比例<sup>[18]</sup>。把客家人与各地汉族的 Y-SNP 作主成分分析,寻找与客家人较近的汉族群体。

把客家人中的 *M7* 个体的 Y-STR 单倍型与各地其他群体中 *M7* 个体的 Y-STR 单倍型用 Network3.0 软件<sup>[19]</sup>画网络结构图,分析客家人中 *M7* 成分的来源。

## 2 结 果

### 2.1 客家人的 Y-SNP 单倍型及其与各民族群体的主成分分析

在客家人中检出了 11 种单倍型,其中比例最高的是 H6、H8 和 H5,其余类型都很微量。为了分析客家人的族属,表 1 中把客家人的数据与文献报道的其他群体数据<sup>[12,14~16]</sup>作了比较。

表 1 客家人与国内其他群体的 Y-SNP 单倍型频率(%)  
Table 1 Y-SNP haplotypes' frequencies(%) of Hakka and other populations in China

群 体 Population	样本量 Size	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14
客家 Hakka	148	2.70	0.68		2.03	14.19	54.05	4.05	16.22	2.03	0.68	2.70			0.68
土家 Tujia	10	10.00				20.00	30.00	10.00		20.00			10.00		
彝 Lolo	14			14.28		42.87	21.43		7.14			14.28			
北方汉 Har(N.)	82	8.50			2.40	22.00	29.00		23.00	9.80					4.90
南方汉 Har(S.)	283	7.77	1.41		1.41	12.02	24.38	1.77	27.56	17.31		3.53	0.71	1.41	
白 Bai	13	15.40		7.70		15.40	30.80		30.80						
纳西 Naxi	18					77.80						16.70			5.50
藏 Tibetan	8		12.50	25.00	12.50				50.00						
瑶 Mien	14	7.14			7.10	21.40	14.29		35.71				14.30		
畲 She	11	18.18				9.09	18.18	27.27	18.18			9.09			
侗 Kam	20	15.00					15.00		15.00	25.00	10.00	20.00			
壮 Zhuang	29	3.60		3.60	7.10	3.60	3.60		25.00	17.90		25.00	10.70		
马桥 Dangdaic	74	2.08		6.25			4.17		27.08	39.58	12.50	8.33			
阿美 Amis	6									100.00					
泰雅 Atayal	24						29.20	4.20	4.20	54.20	8.30				
布农 Bunun	9									11.11	66.67		22.22		
排湾 Paiwan	11								18.20	54.50	27.30				
花苗 Hmong	40	7.69		7.69	7.69	38.46	15.38	7.69				15.38			
六甲 Lujia	32	11.11				7.41	40.74		37.04			3.70			

土家族和客家人显然是差别较大的。客家人最主要的 H5、H6 和 H8 与南北汉族最为相似。客家人中有 H7,这在与之地理上最近的少数民族畲族中比例最高。客家人的 H9、H10 和 H11 结构与侗族最为相似。中国台湾原住民和其他民族的结构与客家人

较远。客家人中 H6 最多,比表 1 中其他群体比例高的多。而广西特殊的汉族群体六甲人也有特别多的 H6。六甲人虽不讲客家话,但也被称为“客”<sup>[16]</sup>。

根据 Y-SNP 单倍型频率,对各群体作主成分分析,得到的各主成分所占数据方差比例见表 2。

表 2 各族人群 Y-SNP 单倍型频率主成分分析得到各主成分的数据方差比例(%)  
Table 2 Variances percentage of components from principle component analysis of Y-SNP haplotypes' frequencies of ethnic populations

主成分 Component	1	2	3	4	5	6	7	8	9	10	11
方差比例 Percentage of variance	38.377	23.407	13.107	9.433	5.334	3.989	2.897	1.7	0.997	0.489	0.181

前 3 主成分占到方差的 74.9% ,对原始数据的代表性较大 ,已包含了大部分的信息。根据表 3 中的各人群对应主成分值 ,用前 3 主成分作坐标轴构建直角三维坐标系 ,各群体在坐标系中的分布见图 1。

图 1 中 ,侗台语群体和中国台湾原住民聚在上方 ,苗瑶语群体聚在下方 ,而汉藏语群体在中部。客家人最接近北方汉族 ,也较近畲族。而南方汉族则较偏向侗台语群体。

### 2.2 客家人群体父系组成的混合比例分析

由于客家人的 Y-SNP 单倍型数据中有与汉族、畲族、侗族较相似的结构 ,我们假设与这 3 个民族相似的古代人群混合构成了现代的客家人。根据 Bertorelle 等设计的方法<sup>[18]</sup> ,我们计算出了客家人父系遗传结构中 3 个人群结构的混合比例 ,汉族约 80.2% ,畲族约 13% ,侗族约 6.8%。可见 ,虽然客家人中汉族成分占绝大多数 ,但其他来源都大于 5% ,也不可忽视。

表 3 各族群体对应主成分值

Table 3 Component data of ethnic populations

主成分	Component	1	2	3	4	5
客家	Hakka	0.855	-0.225	0.095	-0.394	0.090
土家	Tujia	0.697	0.008	0.607	-0.213	-0.055
彝	Lolo	0.569	-0.521	0.376	0.423	0.096
北方汉	Han(N.)	0.957	-0.132	0.026	0.024	0.112
南方汉	Han(S.)	0.956	0.210	-0.144	0.036	-0.012
白	Bai	0.890	-0.231	-0.297	-0.074	0.105
纳西	Naxi	0.254	-0.488	0.483	0.653	0.097
藏	Tibetan	0.264	0.047	-0.827	0.287	0.079
瑶	Mien	0.703	-0.228	-0.410	0.305	0.233
畲	She	0.548	-0.261	-0.162	-0.205	-0.352
侗	Kam	0.538	0.644	0.033	0.042	-0.171
壮	Zhuang	0.392	0.434	-0.338	0.495	-0.260
马桥	Dangdaic	0.400	0.852	-0.109	0.281	0.035
阿美	Amis	0.166	0.836	0.383	0.154	-0.192
泰雅	Atayal	0.476	0.731	0.410	-0.181	-0.051
布农	Bunun	-0.243	0.358	0.134	-0.073	0.794
排湾	Paiwan	0.193	0.907	0.149	0.182	0.214
花苗	Hmong	0.424	-0.595	0.499	0.402	-0.070
六甲	Lujia	0.893	-0.111	-0.350	-0.222	0.076

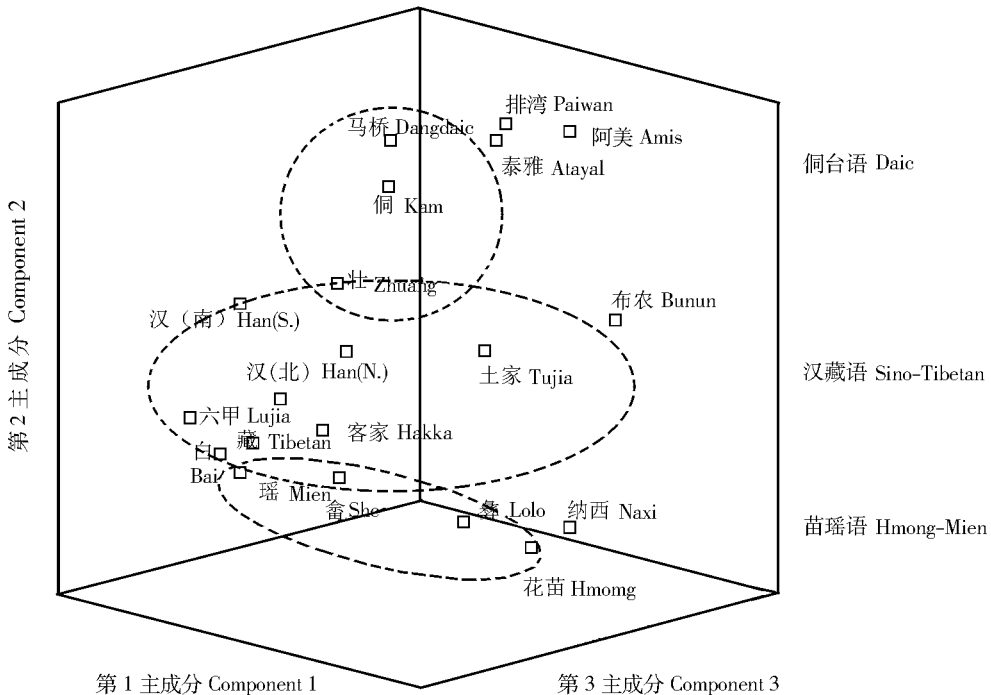


图 1 Y-SNP 单倍型频率主成分分析前 3 主成分坐标系各族群体散点图

Fig.1 Principle component map of 19 ethnic populations in China

表 4 汉族人群 Y-SNP 单倍型主成分分析各主成分的数据方差比例 (%)

Table 4 Variances percentage of components from principle component analysis of Y-SNP haplotypes' frequencies of Han Chinese populations

主成分	Component	1	2	3	4	5	6	7	8	9
方差比例	Percentage of Variance	65.892	15.148	9.564	6.551	1.298	0.607	0.402	0.357	0.118

## 2.3 客家人的 Y-SNP 单倍型与各地汉族的主成分分析

许多省市的汉族群体的 Y-SNP 已作过调查,我们选取了样本量大于 10 的 19 个省市的汉族<sup>[13]</sup>与长汀客家人作主成分分析,得到的各主成分占数据方差的比例见表 4。

前 3 个主成分的累计方差占 90.6%,对原始数据的代表度比与其他民族分析高得多。因为汉族内部类型相对集中,数据的可比性强得多。根据表 5 数值用前 3 主成分构建图 2 坐标系。

图 2 中各地汉族人群可以分辨出中原组、华东组、湖广组和西南组。福建其他汉族基本归入湖广组,而客家人归入了中原组,与河北、山西汉族最接近。

表 5 汉族群体对应主成分值

Table 5 Component data of Han Chinese populations

主成分	Component	1	2	3	4
客家	Hakka	0.811	0.298	-0.491	-0.077
黑龙江	Heilongjiang	0.849	-0.507	-0.118	-0.059
辽宁	Liaoning	0.187	0.546	0.464	0.640
河北	Hebei	0.924	0.038	-0.376	0.032
山东	Shandong	0.981	0.058	-0.004	0.118
山西	Shanxi	0.648	0.568	-0.490	0.110
甘肃	Gansu	0.621	-0.754	0.180	0.088
新疆	Xinjiang	0.621	-0.754	0.180	0.088
河南	Henan	0.857	0.448	-0.101	0.195
安徽	Anhui	0.905	0.355	0.167	-0.029
浙江	Zhejiang	0.875	0.061	0.309	-0.349
江苏	Jiangsu	0.962	0.195	0.158	0.056
上海	Shanghai	0.840	0.373	0.325	-0.127
湖北	Hubei	0.977	-0.079	0.086	-0.110
四川	Sichuan	0.721	-0.272	0.254	0.526
江西	Jiangxi	0.917	0.056	0.184	0.171
湖南	Hunan	0.939	0.058	-0.119	-0.070
福建	Fujian	0.916	-0.191	-0.280	-0.066
云南	Yunnan	0.813	-0.550	0.080	-0.010
广西	Guangxi	0.119	0.365	0.776	-0.492
广东	Guangdong	0.869	0.058	-0.146	-0.406

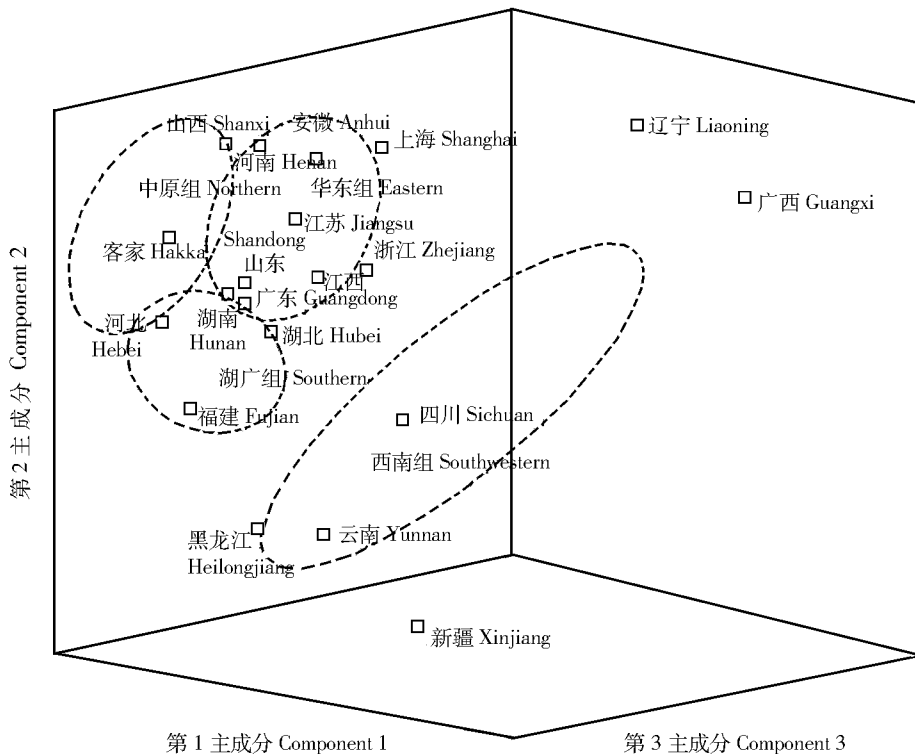


图 2 客家人与各地其他汉族群体主成分分析前 3 主成分三维散点图

Fig.2 Principle component map of Hakka and other Han Chinese populations

## 2.4 客家人 M7 个体 Y-SNP 单倍型网络结构分析

在客家人中发现了 6 个 M7 个体,对其 Y 染色

体上 5 个 STR 位点检测发现构成了 5 种单倍型。与我们调查的国内其他群体中的 M7 个体的单倍型一起作网络结构分析,结果见图 3。

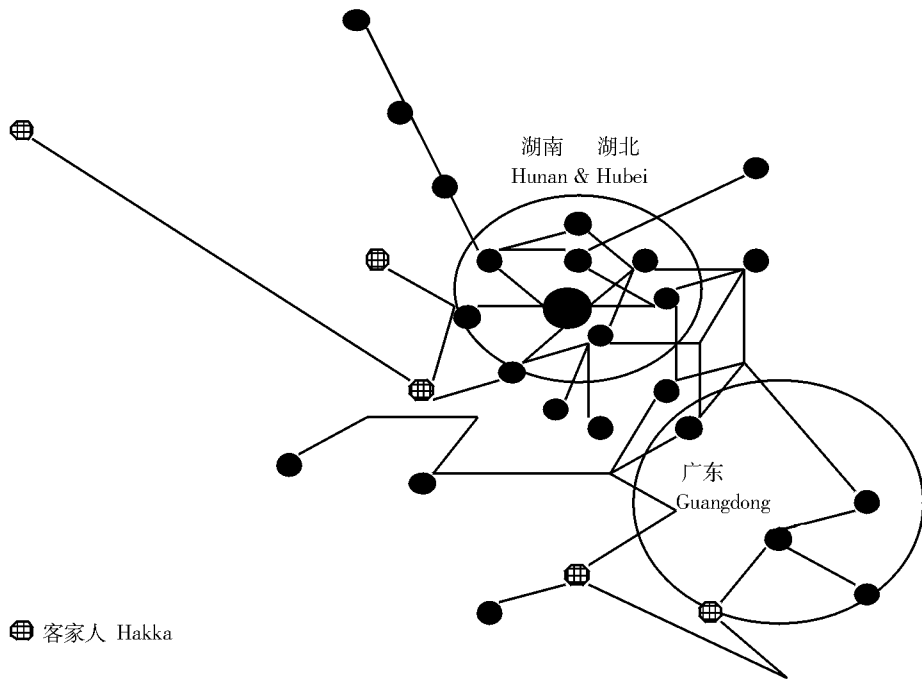


图3 客家人和其他民族群体  $M7$  个体 STR 网络结构图

Fig.3 Network structure of  $M7$  individuals of Hakka and other ethnic populations

图3中可见,客家人  $M7$  个体有两种发生位置。其一是直接从湖南湖北的各少数民族中衍生而来,需要补充说明的是,与这几个客家人相连的主要是土家族个体。其二是从广东八排瑶衍生而来,与畲族相似。

### 2.5 Region V 9 bp 缺失提示的客家人母系遗传结构

对客家人的线粒体 Region V 区段的检测发现有 19.7% 左右的 9 bp 缺失(单拷贝)个体。中国 9 bp 缺失的比例<sup>[20]</sup>在苗瑶语群体中最高,特别是湖南广西之间高达 30%~50%,而中原汉族则维持在 10% 上下。畲族的比例为 15%,客家人比畲族稍高,但没有显著差异,而明显高于中原汉族。

## 3 讨论

### 3.1 客家人可能的来源和迁徙路线

从 Y-SNP 数据看来,客家人与中原汉族最接近。分析得到的 80.2% 汉族结构说明客家人中大多数男性可能来源于中原汉族。所以罗香林的观点

是有一定根据的,但并不全面。另外,我们也注意到客家人相当高的 H6 比例是与中原汉族不同的,这不知是客家人在迁徙过程中形成的,还是在中原时就与其他部落不同。但是南方确实有许多汉族特殊群体 H6 比例偏高,这是个有待探讨的现象。

如果客家人真的来自中原,他们应当是通过湖北、江西进入闽西的。在通过湖北时,融入了部分当地荆蛮土著的  $M7$  突变个体。当然也无法排除这部分  $M7$  个体是后来个别地迁入客家人地区的可能性。江西地区的土著据记载为干越,后来的山都木客据考也属于干越<sup>[21]</sup>。侗族自称“干”(Kam),也有一种来自江西的传说<sup>[22]</sup>,很可能是干越的后裔。所以客家人中 6.8% 的与侗族相似的结构,也很可能来自其迁徙路线上江西的古代干越族,与其他福建民系不同。其他民系含有 H12 而没有发现 H9、H10、H11,最可能是闽越后代的成分<sup>[23]</sup>。而干越和闽越虽同属百越,但百越的遗传结构在相对一致性之外内部也有明显的差异<sup>[24]</sup>,所以这两支也可能有很大不同。客家人没有进入福建腹地,所以没有接触到闽越族,而江西和闽西的干越则部分融入了客家人。当然大部分客家人从中原经湖北、江西进入闽西的推测还要全面考察和验证。

### 3.2 畲族可能是客家人最大的外来影响

为了辨认汉藏语各群体受侗台、苗瑶不同的影响,前文各族主成分分析结果特意把侗台和苗瑶拉向两端观察。大部分南方汉族显然偏向侗台语群体,可能受百越影响比苗瑶影响大得多。而客家人则不同,更多地偏向了苗瑶群体。所以,客家人中虽然也有百越成分,但其受畲族影响大得多。畲族是唐宋时期从广东迁来的苗瑶类群体,对于福建越族土著和先到汉民而言,同客家人一样是“客”。这两个群体之间接触甚密也很自然,所以北上经客家地区的畲族大多改用近客家话的汉语方言<sup>[25]</sup>,而客家人也涵入了很大部分与畲族一样的广东苗瑶群体的遗传结构。

母系血统比父系血统的交流频繁得多,所以父系结构上客家人 13% 的畲族成分对客家总体遗传结构而言可能是很保守的估计。从母系的线粒体 Region V 的 9 bp 缺失情况看来,客家人与北方汉族差异颇大,而基本与畲族一致。可能这两个群体之间的通婚曾普遍发生。

另外,长汀是客家人的中心地区,所以汉族成分必然是相对高的。在边远地区的客家人中的汉族成分比例就不能保证那么高了。很可能存在有些地区的客家人完全或主要来自畲族。这要更多的田野调查来验证。特别是其他客家人聚居区,比如广东梅州、江西赣州等地的客家人,其遗传结构也需要进一步研究,不能认为必然和福建长汀的一致。

### 3.3 客家人遗传结构对客家话发生研究的启示

客家话中含有相当多的苗瑶语词汇,而很少有侗台语词汇。这些词汇都是不常用的基本词,不是因为汉人借用苗瑶语,很可能是客家人的最早的成分——古代的荆蛮族为了与汉人打交道,借用了大多数汉语词汇,而保存那些出现频率不高的词语。南方方言中有大量无法找到汉语来源的语词。合理的解释是原始客家语并不是中原地区的汉族移民带到南方来的,而是当地的原住民荆蛮人在学习汉语的过程中形成的混合语。在开始的时候,可能带有更多的荆蛮话土语,后来受汉语的不断影响,汉语的成分越来越多,就成了真正的汉语方言,只有少部分荆蛮土语还残存在现代的客家话中。因此推测客家人最早的成分可能是荆蛮族,故而在客家人的遗传结构中会有部分来自湖南湖北的苗瑶成分。

但是,如果认为原始客家语并不是中原移民带来的,又怎么解释客家人的遗传分析汉族为绝大多数,又与中原汉人最接近呢?客家话的形成过程中,当原始客家话在荆蛮人中形成以后,尽管每个时代都有大量的中原移民加入,因而改变了这个语言社团的遗传组成,但是他们都是陆陆续续移入的,一代一代地融进客家的语言社团。每移进一个家庭,他们的子女这一辈就学会了客家话,完全融入并壮大了客家语言社团。所以语言的特征保留了,但是遗传结构发生了很大变化。客家话也不断地变化,但是主要是受汉语权威方言的影响。在历史上的某个时期,客家人与畲族一起南迁到赣闽粤交界,长期的接触使畲族人接受了客家话。

所以客家人的遗传结构与其语言特征相印证,提示这一民系可能是在荆蛮族的核心上不断加入汉族移民形成的。这也支持与传统的语言史观不同的观点,认为大部分的南方汉语方言,最初可能是南方原住民在中原汉语不断影响下逐渐形成的<sup>[26]</sup>。

### 参考文献 (References):

- [ 1 ] CHEN Zhi-Ping. Six branch of Han in Fujian. Fuzhou: Fujian Press, 2000.  
陈支平. 福建六大民系. 福州: 福建出版社, 2000.
- [ 2 ] CHEN Yun-Dong. Hakkanese. Taipei: Liensa Press, 1981.  
陈运栋. 客家人. 中国台北: 连亚出版社, 1981.
- [ 3 ] DENG Xiao-Hua. Relationships between Hmong-Mien, Daic and Hakkanese. *Chinese Journal of Ethnic Linguistics*, 1999(3): 42~49.  
邓晓华. 客家话跟苗瑶壮侗语的关系问题. 民族语文, 1999(3): 42~49.
- [ 4 ] YUAN Jia-Hua. A summary of Han dialects. Beijing: Character Reformation Press, 1989.  
袁家骅. 汉语方言概要. 北京: 文字改革出版社, 1989.
- [ 5 ] LIU Zhen-Fa. Hakka: A misunderstood history and a historical misunderstanding. Beijing: Academic Research Press, 2001.  
刘镇发. 客家——误会的历史、历史的误会. 北京: 学术研究杂志社, 2001.
- [ 6 ] LIU Guan-Xin. Out of error. Proceedings for the 50<sup>th</sup> celebration of Journal of Chinese Linguistics 2002.  
刘纶鑫. 走出误区. 《中国语文》建刊 50 周年国际学术研讨会论文集, 2002.
- [ 7 ] ZHENG-ZHANG Shang-Fang. The characteristic and origin of She language in south Zhejiang. Proceedings for international symposium of dialects in Southeastern China, 2001.  
郑张尚芳. 浙南畲族话的特点及其来源分析. 中国东南方言国际研讨会论文集, 2001.
- [ 8 ] XIAO Chun-Jie, DU Ruo-Fu, Cavalli-Sforza L L, Minch E. Principle

- component analysis of gene frequencies in populations of China. *Science in China Series C* 2000, 30(4):434~442.
- 肖春杰,杜若甫, Cavalli-Sforza L L, Minch E. 中国人群基因频率的主成分分析. *中国科学 C 辑*, 2000, 30(4):434~442.
- [ 9 ] Cann R L, Stoneking M, Wilson A C. Mitochondria DNA and human evolution. *Nature*, 1987, 325:31~36.
- [ 10 ] Wang L, Oota H, Saitou N, Jin F, Matsushita T, Ueda S. Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. *Mol Biol Evol*, 2000, 17(9):1396~1400.
- [ 11 ] Jin L, Su B. Natives or immigrants: modern human origin in East Asia. *Nature Reviews Genetics*, 2000, 1(2):126~134.
- [ 12 ] Su B, Xiao C, Dekra R, Seielstad M T, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L L, Chakraborty R, Jin L. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet*, 2000, 107:582~590.
- [ 13 ] Su B, Xiao J, Underhill P, Dekra R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet*, 1999, 65(6):1718~1724.
- [ 14 ] Su B, Jin L, Underhill P, Martinson J, Saha N, McGarvey S T, Shriver M D, Chu J, Oefner P, Chakraborty R, Dekra R. Polynesian origins: insights from the Y chromosome. *Proc Natl Acad Sci USA*, 2000, 97:8225~8228.
- [ 15 ] LI Hui. A Genetic Study on Genes of Human Bones in Graves. In: The CPAM of Shanghai. Maqiao report on Excavation between 1993 and 1997. Shanghai: Shanghai Fine Arts Publisher, 2002, 57~68.
- 李辉. 墓葬人骨的遗传基因研究. 上海市文物管理委员会, 马桥 1993~1997 年发掘报告. 上海: 上海书画出版社, 2002, 57~68.
- [ 16 ] LI Hui, HOU Jing-Rong, YANG Ning-Ning. A Molecular Anthropologic Study on the Origin of Liujia People in Guangxi. *Journal of Guangxi University for Nationalities*, 2002, 24(5):38~43.
- 李辉, 侯井榕, 杨宁宁. 广西六甲人来源的分子人类学分析. *广西民族学院学报(哲学社会科学版)*, 2002, 24(5):38~43.
- [ 17 ] Cavalli-Sforza L L, Menozzi P, Piazza A. The History and Geography of Human Genes. Princeton: Princeton Univ Press, 1994.
- [ 18 ] Bertorelle G, Excoffier L. Inferring admixture proportions from molecular data. *Mol Biol Evol*, 1998, 15:1298~1311.
- [ 19 ] Bandelt H J, Forster P, Rael A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*, 1999, 16:37~48.
- [ 20 ] Yao Y G, Watkins W S, Zhang Y P. Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of east and southeast Asia. *Hum Genet*, 2000, 107(5):504~512.
- [ 21 ] JIANG Bing-Zhao, SHI Yi-Long, HUANG Xiang-Chun. Cliff coffins on Long-Fu Hill and Daic ethnic culture. Changchun: Jilin People's Press, 2001.
- 蒋炳钊, 石奕龙, 黄向春. 龙虎山崖葬与百越民族文化. 长春: 吉林人民出版社, 2001.
- [ 22 ] SHI Lin. Comparison studies on Kam-Daic. Tianjin: Tianjing Ancient Books Press, 1997.
- 石林. 侗台语比较研究. 天津: 天津古籍出版社, 1997.
- [ 23 ] LI Hui. Preliminary Analyses of Genetic Structure of Daic Populations. Proceedings of The International Symposium of Anthropological Studies. Shanghai: Center for Anthropological Studies at Fudan University, 2002, 89~94.
- 李辉. 百越族群遗传结构分析的初步思路. 现代人类学国际研讨会论文集. 上海: 复旦大学现代人类学研究中心, 2002, 89~94.
- [ 24 ] LI Hui. On dividing of Baiyue people's hereditary frame. *Study of Ethnics in Guangxi*, 2002, 7(4):26~31.
- 李辉. 百越遗传结构的一元二分迹象. *广西民族研究*, 2002, 7(4):26~31.
- [ 25 ] HE Jun-Fang. Another mention of language conversation among nationalities in China. *Ethnology Research*, 1999, (3):45~52.
- 何俊芳. 也论我国民族的语言转用问题. *民族研究*, 1999, (3):45~52.
- [ 26 ] PAN Wu-Yun. Characteristic of Hakkanese and origin of Chinese southern dialects. The 11<sup>th</sup> annual for international symposium of Chinese linguistics, 2002.
- 潘悟云. 客家话的性质——兼论南方汉语方言的形成历史. 第 11 届国际汉语语言学年会论文, 2002.

(责任编辑:刘波勇)