



Similarities in Patrilineal Genetics between the Han Chinese of Central China and Chaoshanese in Southern China

ZHANG Fenghuan¹, LI Hui², HUANG Liqun³, LU Yan², HU Shengping^{1,4}

1. Molecular Biology Laboratory, Shantou University Medical College, Shantou GD515031, China; 2. MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 3. Clinical Laboratory, First Affiliated Hospital, Shantou University Medical College, Shantou GD515041, China; 4. Forensic Genetics Laboratory, Shantou University Forensic Medical Service Center, Shantou GD515031, China

ABSTRACT: Y-chromosome short tandem repeat (Y-STR) loci are widely used to trace the origin of the human populations. In a previous study, we used STR analysis to determine that the Chaoshanese of eastern Guangdong (often regarded as a "typical" southern Chinese population) are a mix between the Han Chinese of central China and indigenous ethnic groups of southern China. To further investigate the Y chromosome of the Chaoshanese and the group's genetic affinity with other Chinese populations; we analyzed six specific Y-STR loci in samples taken from 6,292 individuals from 51 populations, including the Chaoshanese, the Minnamese, and various other East Asian populations. The admixture analysis showed that over 85% of the Chaoshanese Y chromosomes is derived from the Han in central China (M_{RH} : 0.86; M_{BE} : 1.19 ± 0.21) while a very small portion is derived from southern indigenous populations. These findings not only confirm that the Chaoshanese originated from central China, but also reveal that the patrilineal genetics of the Chaoshanese can be mainly attributed to Han males from central China. The principal component and phylogenetic analyses show that the Chaoshanese generally clustered with Han populations from southern China, specifically the Minnamese, Taiwanese, Malaysian Chinese, and Singaporean Chinese populations. The phylogenetic tree and the analysis of molecular variance (AMOVA) showed that Chinese populations tended to cluster based on their linguistic families rather than their geographic distribution, suggesting a strong association between the Y-chromosome and linguistic affiliations. These analyses also revealed that the patrilineal boundary between southern and northern populations was not pronounced.

Key words: Han in Central China; Chaoshanese (Teochew and Swatow); Paternal genetic affinity; Southern indigenous population; Y-chromosomal short tandem repeat (Y-STR)

中原汉族是潮汕汉族父系遗传成分的主要贡献者

张凤环¹, 李辉², 黄立群³, 陆艳², 胡盛平^{1,4}

1 汕头大学医学院分子生物学实验室, 广东 汕头 515031; 2 复旦大学生命科学学院现代人类学教育部重点实验室, 上海 200433; 3 汕头大学医学院第一附属医院临床实验室, 广东 汕头 515041; 4 汕头大学司法鉴定中心DNA物证研究室, 广东 汕头 515031

摘要: Y染色体短串联重复序列(Y-STR)位点现已普遍用于追溯人类群体遗传起源的研究。居于广东省东部的潮汕汉族是典型中国南方人群, 本课题组前期通过分析常染色体STR数据证明了潮汕汉族是中原汉族和南方土著居民融合而成的混合人群。为了进一步研究潮汕汉族的父系遗传结构及与其他中国人群的遗传关系, 我们共搜集了含有6个Y-STR位点数据的51个中国人群6292例个体, 其中包括课题组前期研究的潮汕和闽南汉族人群, 及来自其他研究的南北方汉族、藏缅语族、阿尔泰语系、苗瑶语系、南亚语系和侗台语系人群。融合分析结果显示潮汕汉族中85%以上的父系遗传成分源于中原汉族(M_{RH} : 0.86; M_{BE} : 1.19 ± 0.21), 而来自南方土著的微乎其微, 再次证明了潮汕汉族的中原起源, 并进一步提示, 中原汉族男性是潮汕汉族父系遗传成分的主要贡献者; 主成分分析和聚类分析结果显示潮汕汉族大体上与南方汉族聚类, 但在南方汉更接近平南、台湾、马来西亚和新加坡汉族人群; 遗传树和分子方差分析(AMOVA)结果显示中国人群倾向于按照其语系分组聚类, 而非按照其地理分布, 显示父系遗传与语言的紧密关联。这些分析显示父系遗传方面, 南方和北方人群的界线并不明显。

关键词: 中原汉族; 潮汕汉族; 父系遗传关系; 南方土著; Y染色体短串联重复序列

收稿日期: 2010年1月28日 **修回日期:** 2010年2月5日 **联系人:** 胡盛平 sphu@stu.edu.cn

第一作者简介: 张凤环(1984-), 女, 肿瘤分子生物学硕士研究生, shui689@163.com.

2010年2月6日 <http://COMonCA.org.cn/Abs/2010/002.htm> 5

©上海人类学学会 Shanghai Society of Anthropology

中国56个民族的三百多种语言主要分为六大语系：阿尔泰语系(Altaic)、南亚语系(Austro-Asiatic)、南岛语系(Austronesian)、苗瑶语系(Hmong-Mien)、侗台语系(Daic, 也称“百越”)、汉藏语系(Sino-Tibetan, 包括汉和藏缅两个语族)(Ethnologue, <http://www.ethnologue.com>) [1]。目前, 汉族是世界上最大的民族, 其人口总数超过11.8亿。2005年人口普查(www.stats.gov.cn/tjgb/rkpcgb/), 汉族占中国人口的90.56%。据中国历史记载, 汉族是古时华夏部落的后裔, 华夏部落四五千年前形成于黄河流域上游, 壮大于中原地区[1,2]。在过去的两千多年间, 汉族人群及汉族文化逐步扩散到中国南方。百越族是人数最多分布最广的南方土著居民, 广泛居住在中国南方沿海地区(包括潮汕地区)。自秦汉时期由于北方的战乱、饥荒等诸多因素, 汉族向南方迁移, 在汉族文化的巨大影响下, 部分的百越人被同化为汉族(即南方汉族), 部分遗留下来的百越人形成了现在的侗台语系民族[3]。因此, 百越族人群对于研究中国南方人群的遗传结构有着重要意义[4,5]。

潮汕僻处我国的东南隅, 广东省的最东端, 与福建省毗邻, 其地域包括今天的汕头、潮州、揭阳 3 个地级市。潮汕人指居住于该区域并有类似于闽南的独特方言和生活习惯的汉族民系。据潮汕家系及中国历史记载[1,2,6], 现代潮汕人群源于中原汉族, 是由古中原汉族人群与南方土著居民融合而成。公元前 214 年秦始皇派兵南下统一中国, 中原汉族开始迁入潮汕地区。此后又发生多次的中原汉族向南方的迁徙事件, 中原汉族与当地土著居民(百越人)长期融合形成了现在的潮汕人群。本课题组前期研究了潮汕人群白细胞抗原 A、B(HLA-A,B)等位基因的频率分布, 为潮汕的中原起源提供了遗传依据[7]; 同时我们还对潮汕人群的常染色体 STR 位点数据进行了分析, 结果证实了潮汕的中原汉族起源[8]。

人群融合是人类进化研究的焦点之一, 混合人群的研究有助于了解相关人群的遗传结构和历史演化。国内外已有众多学者采用

多种遗传标记对不同地区人群进行分析[9-17]。普遍应用的遗传标记有: 线粒体DNA(mtDNA), 核基因组单核苷酸多态性位点(SNP), 常染色体和 Y 染色体短串联重复序列(STR)等。来自亲本的男女性别比例不平衡是混合人群的普遍现象[9-17], 融合比例与亲本群体的移民时间、最初种族构成有关。我们前期研究已证实了潮汕人群遗传结构的混合特征, 想进一步了解混合中男女性别比例的差异。Y 染色体非重组区的 Y-STR 是显示男性对混合人群遗传成分影响的敏感性标记, 而且已经应用于人类群体关系、群体起源、移民及进化等方面的研究。本研究通过对潮汕人群、闽南人群及搜集筛选的其他中国人群的 Y-STR 数据的分析, 进一步探究潮汕人群的父亲遗传结构及与其他人群之间的关系。

一、材料和方法

1. 人群数据搜集

潮汕和闽南两人群的数据来自本课题组前期研究结果[18,19], 其他人群数据通过公用数据库或国内外同行评议期刊等途径获取。人群数据的筛选标准是: 1, 截止2008年7月1日之前发表在国内外同行评议期刊上的人群数据; 2, 人群样本数不少于30例; 3, 人群的STR数据既包括以行政区划(省、直辖市、自治区、特别行政区、省辖市)进行界定的数据, 也包括以民系进行界定(如潮汕人、闽南人)的数据; 4, 数据中必须包括6个Y-STR位点(DYS19, DYS389I, DYS390, DYS391, DYS392和DYS393)的完整单倍型数据。按照以上标准筛选出的人群中, 同一种族或行政区域出现不止一套的Y-STR单倍型人群数据。针对这种相同人群有两套或两套以上单倍型频率数据的情况, 首先利用Arlequin软件根据这几套人群数据的Y-STR单倍型频率计算遗传距离 F_{ST} 值和相关 P 值[20], 若相同人群相互之间的 F_{ST} 值无统计学差异($P > 0.05$), 则对这些人群的单倍型频率数据用加权平均法进行合并作为该地区人群的最终数据, 反之,

表 1 本研究所用人群

Tab.1. Populations information

地理分组	人群	地区	语系分类	参考文献	
北方	汉族	东北	汉	[21,22]	
	汉族	辽宁	汉	[23]	
	汉族	宁夏	汉	[24]	
	汉族	天津	汉	[25]	
	汉族	河南	汉	[26]	
	汉族	山东	汉	[27]	
	汉族	北京	汉	[28]	
	回族-1	宁夏	汉	[29]	
	回族-2	宁夏	汉	[30]	
	朝鲜族	吉林	阿尔泰	[31]	
	撒拉族	青海	阿尔泰	[32]	
	蒙古族	内蒙古	阿尔泰	[33]	
	维吾尔族	新疆	阿尔泰	[34]	
	东乡族	甘肃	阿尔泰	[35]	
	藏族-1	青海	藏缅	[36]	
	南方	汉族	湖南	汉	[37]
		汉族	四川	汉	[38,39]
		汉族	云南	汉	[40]
		汉族	浙江	汉	[41]
		汉族	香港	汉	[42]
汉族		台湾	汉	[43]	
汉族		潮汕	汉	[18]	
汉族		闽南	汉	[19]	
汉族		新加坡	汉	[44,45]	
汉族		马来西亚	汉	[46]	
藏族-2		西藏	藏缅	[47]	
藏族-3		西藏	藏缅	[48]	
藏族-4		西藏	藏缅	[49]	
纳西族		云南	藏缅	[50]	
彝族		云南	藏缅	[51]	
土家族		重庆	藏缅	[52]	
瑶族		广西	苗瑶	[53]	
俫人		广西	南亚	[5]	
青仡佬		广西	侗台	[5]	
拉基人		云南	侗台	[5]	
木佬人		贵州	侗台	[5]	
红仡佬		贵州	侗台	[5]	
黎族-杞		海南	侗台	[5]	
布央人		云南	侗台	[5]	
仡隆人		海南	侗台	[5]	
高栏山仔		广西	侗台	[5]	
临高人		海南	侗台	[5]	
五色人		广西	侗台	[5]	
水族		广西	侗台	[5]	
莫家人		贵州	侗台	[5]	
仡佬族		广西	侗台	[5]	
毛南族		广西	侗台	[5]	
标人		广东	侗台	[5]	
伴僮人		贵州	侗台	[5]	
查家人		海南	侗台	[5]	
僂傣人		上海	侗台	[5]	

则这些人群将作为独立的人群进行分析。

最终, 本研究筛选出51个人群[5,18,19,21-53], 包括6292例个体, 其共有单倍型1697种。这些人群若按照北纬30°(大致以长江为界)的地理界线划分[54], 可分为16个北方人群和35个南方人群; 若按照语系划分, 可分为19个汉族人群(包括7个北方汉族群体, 10个南方汉族群体, 2个回族群体), 7个藏缅语系人群, 5个阿尔泰语系人群, 1个藏缅语系人群, 1个南亚语系人群和18个侗台语系人群。各人群的地理分布和语系分组详见表1。在地理位置上中原地区属于北方区域, 但是为了与本课题组前期研究[7,8]保持一致, 我们在本研究中仍然特指中原地区。

2. 统计分析

利用 Phylip 3.6 软件[55]根据人群的单倍型频率数据计算群体间遗传距离 D 值[56], 并构建比较人群的邻接(neighbor-joining, NJ)遗传树[57], bootstrap 法评估树状图的可靠性; 用 SPSS13.0 软件对人群单倍型频率数据进行主成分分析 (Principal component analysis), 以观察群体间遗传关系; 为进一步探究汉族人群主成分值与其地理分布的相关性, 分别以第一和第二主成分值作为等高值, 用 Surfer8.0 (www.goldensoftware.com)软件绘制汉族人群 Y-STR 主成分的地理分布图; 用 Arlequin 3.1 软件[20]分别按照人群的地理分组和语系分组进行分子方差分析, 评估群体分化程度; 用 ADMIX 2.0[58]和 LEADMIX[59]软件进行融合分析, 分别计算 M_{BE} [60]和 M_{RH} [61]统计值, 评估中原汉族和南方土著居民对于潮汕人群的遗传贡献率。在融合分析中, 河南汉族代表中原汉族作为一个亲本群体, 而另一南方亲本的代表采用文波等人[14]的方法, 选用黎族和仡隆两个基本未受汉族遗传影响的南方土著人群[5]单倍型频率的算术平均值。我们假设人群融合时间始于 2300 年前中原汉族(当时的秦军)第一次迁入潮汕地区[1,2,6]。

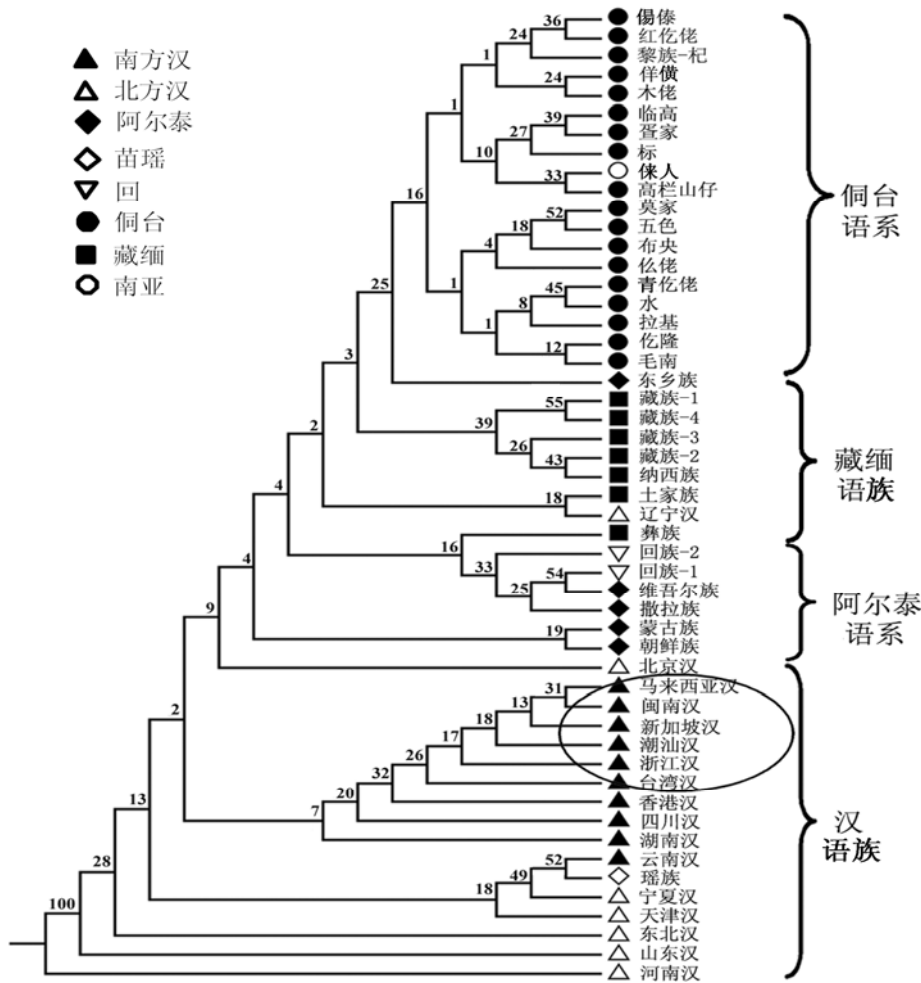


图1 东亚各族群的遗传进化树图 进化树图节点处的数值是经过1000次迭代之后的bootstrap值(百分比)

Fig.1. The neighbor-joining phylogenetic tree of the East Asian populations. The numbers in the nodes are percentage bootstrap values estimated from 1,000 replicates.

二、结果

1. 不同人群间的遗传关系

根据 51 个人群的 Y-STR 单倍型数据计算 D 值并构建邻接遗传树(图 1), 图 1 中显示, 所有人群均是按照语系聚类而不是根据其地理分布, 这种现象在主成分分析图(图 2A)中也很明显。图 2A 中, 汉族和侗台语系人群明显地被第一和第二主成分划分, 并各自聚类。在汉族的聚类组中, 第一主成分将南方汉族和北方汉族人群分开, 但分界并不是很明显。有趣的是, 河南汉族距离南方汉族较近, 这与我们前期分析 HLA[16]和常染色 STR[17]的研究结果是一致的。除汉族和侗台语系以外的其他语系人群在图 2 中分布比较分散, 可能是因为分析中这些语系的人群数目较

少。以上结果均显示了中国人高度统一的父系遗传结构及其与语系分类的相关性。

为了进一步研究潮汕汉族与其他 16 个汉族人群的遗传关系, 我们对汉族人群的 Y-STR 单倍型频率进行了主成分分析(图 2B), 构建了邻接遗传树(图 3), 并绘制了汉族人群的第一和第二主成分地理分布图(图 4)。总体上, 潮汕汉族与南方汉族聚在一起, 但相对于其他南方汉族人群, 潮汕汉族更接近于闽南、台湾、新加坡和马来西亚汉族(图 1 和图 3)。潮汕汉族和这 4 个人群的遗传距离 D 值分别是 0.000006, 0.000008, 0.000008 和 0.00001。这些人群的紧密关系在图 4 中也得到了印证, 其主成分值均在同一等级。在北方汉族中, 河南汉族与南方汉族的距离 ($D=0.000009$), 是潮汕汉族与各北方汉族的

表 2 51 个中国人群的分子方差分析 Tab.2. AMOVA of 51 Chinese populations

不同分组	差异来源		
	组间(F_{CT}, P)	组内人群之间 (F_{SC}, P)	人群内部(F_{ST}, P)
地理分组	0.08 (0.00077, 0.016)	0.84 (0.00839, 0.000)	99.08 (0.00916, 0.000)
语系分组	0.56 (0.00563, 0.000)	0.55 (0.00549, 0.000)	98.89 (0.01110, 0.000)

遗传距离中的最小值，这在一定程度上提示了中原汉和南方汉的亲密关系。而且，图4中显示潮汕-闽南主成分值接近中原汉的主成分值。

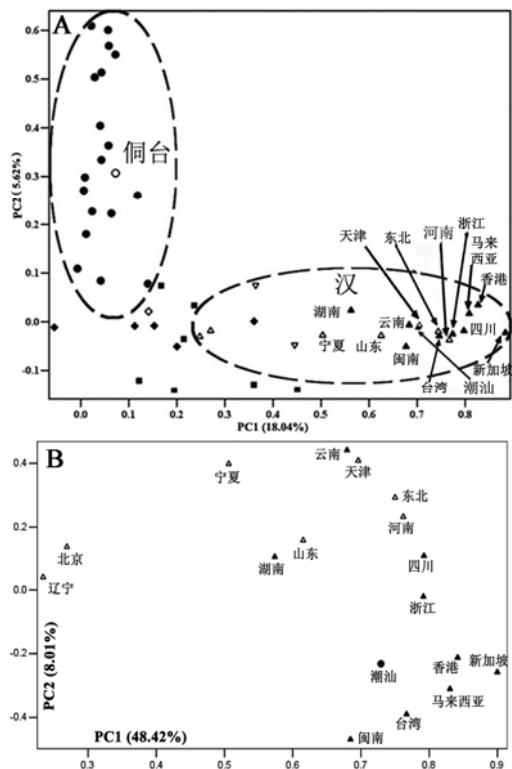


图2 主成分分析图 A. 东亚各民族; B. 汉族
Fig.2. Principal component plot. A. East Asian populations; B. Han Chinese

2. 分子方差分析

所有研究人群根据地理位置分为南、北方两组(地理分组)，根据语系分为6组：阿尔泰语系、南亚语系、南岛语系、苗瑶语系、侗台语系和汉藏语系(语系分组)。分子方差分析结果(表 2)显示不同变异程度及不同分组水平上的三种 F 统计值，概括了 Y-STR 多态性在地理和语系两种不同分组情况下的差异。图 1 和图 2 中，相同语系的人群聚类，如果这种聚类有一定的遗传基础，那么在分子方差分析中，语系分组的组间差异应比地理分组的组间差异大。如预期结果，尽管两种分组的 F_{CT} 统计值均有显著性差异

($P < 0.05$)，但 Y 染色体的分化程度在语系组间(0.56%; $F_{CT} = 0.00563, P < 0.001$)比在地理组间(0.08%; $F_{CT} = 0.00077, P = 0.016$)明显的多，提示人群按语系分组时存在较大程度上的组间分化。相反，同组内人群间的差异程度在地理分组上占 0.84% ($F_{SC} = 0.00839, P < 0.001$)，而语系分组时仅占 0.55% ($F_{SC} = 0.00549, P < 0.001$)。这些结果显示中国人群的父系遗传结构与其语系分类的相关性比与其地理位置的相关性更强。

3. 潮汕汉族的融合分析

历史记录[1,2]和我们前期常染色体 STR 数据分析结果[8]均显示潮汕汉族的父系遗传结构有两个潜在的亲本起源：中原汉族和百越人群。因河南覆盖了古中原地区[16,17]，我们假定河南汉族代表中原汉族作为融合分析的一个亲本；通常认为遗传结构属于侗台人群的黎族和仡隆人[62]两个人群作为一个南方亲本代表，因这两个人群世代居于海南岛上，琼州海峡将海南岛与大陆隔开，使得该岛上居民长期隔离，很少与外来种族人群发生融合[5,62]，所以黎族和仡隆人群的遗传结构最能代表南方土著居民的遗传组成。表 3 中显示，两种软件计算的融合分析结果高度一致。潮汕汉有较大成分(大于 85%)的中原汉族起源 ($M_{RH}: 0.8614; M_{BE}: 1.1868 \pm 0.2054$)，而仅有小部分源于南方土著居民 ($M_{RH}: 0.1386; M_{BE}: -0.1868 \pm 0.2054$)。这个结果证明潮汕汉族的父系遗传结构几乎全部来源于中原汉族，而南方土著的成分则是微乎其微。

表 3 潮汕人群的融合分析
Tab.3. Admixture analysis of Chaoshanese

统计方法	亲本群体	
	中原汉族	南方土著
混合比例 (Leadmix, M_{RH})	0.8614	0.1386
混合比例 (Admix, M_{BE})	1.1868 ± 0.2054	-0.1868 ± 0.2054

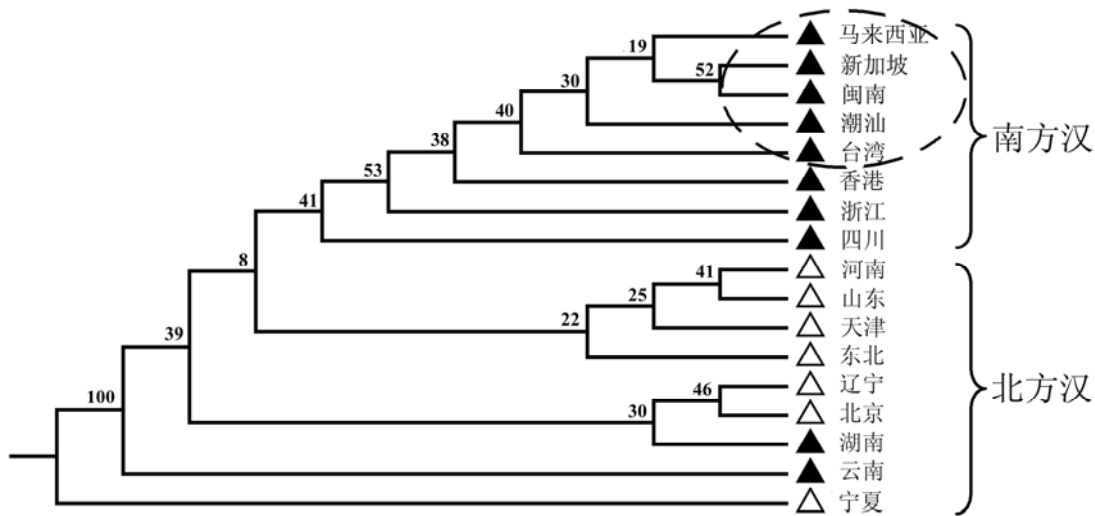


图3 汉族人群的遗传进化树 进化树图节点处的数值是经过1000次迭代之后的bootstrap值(百分比)

Fig.3. The phylogenetic tree of Han populations based on the haplotype frequencies of six Y-STR loci. The numbers in the nodes are percentage bootstrap values estimated from 1,000 replicates.

三、讨论

通过计算两种不同的统计值[60,61]评估两亲本对潮汕汉族的相对遗传贡献率,我们首次发现潮汕汉族的父系遗传结构有相当大成分来源于中原汉族,而南方土著的贡献率较小(表3)。本研究的结果再次证实之前报道[8]的潮汕人的中原起源,而且显示中原汉族是潮汕父系遗传成分的主要贡献者,说明潮汕汉族和中原汉族有着高度相似的父系遗传成分。

在南方汉族人群的研究中已有报道:南方汉族有着与北方汉族高度相似的父系遗传成分(北方汉族的男性是南方汉族父系基因库的主要贡献者),而其来自南北亲本的母亲系遗传成分相当(南北亲本群体均是南方汉族母系基因库的贡献者)[16]。这种机制被解释为由于“定向婚配(directional mating)”所产生的明显的性别偏向性人群融合[63]。潮汕汉族属于南方汉族(图 1-3)[6,7,8],因此我们估计潮汕汉族的形成也存在相似的融合过程。潮汕汉也和其他南方汉一致,有着来自中原汉和南方土著相同比例的母亲系遗传成分(本课题组未发表的数据)。不同世界人群的研究已经提供了证据:“定向婚配”主要由移民性

别比例控制,受文化习俗如配偶选择、语言和社会特征等的影响[64,65]。上述事实均可在潮汕相关的历史和人口统计信息中得到查证。潮汕地区的移民史中,中原汉族向潮汕地区大规模的南迁移民主要是男性士兵(大多是青壮年男子)组成的军队[6,16,17,66]。移民中如此不均衡的高比例男性势必会影响配偶选择,而当地土著女性成为主要的婚配对象[67]。在文化实践方面,汉族南迁时文化的传播伴随着人口的扩张[14]。由于掌握了较为先进的科技和文化力量,汉人毋庸置疑地占据主导地位并广泛影响原来的土著居民[2]。而且,古代中国严格执行从夫居(婚后妻子居住在丈夫家中)的婚姻模式和父系世袭制(由男子继承氏族),这也大大促进了汉族男性融入古潮汕人中,因此现代潮汕汉族保持着和中原汉族亲本相同的父系遗传结构。实际上,当时的政府也鼓励士兵移民在战后定居潮汕地区时,与当地女性居民婚配[66]。这种通婚是军事移民与原住居民直接交往的重要形式,客观上有利于缓和军事移民和原住民之间的矛盾,并加强文化的交流和民族的融合。值得一提的是,现在的潮汕汉族仍相当程度地保留着来自中原汉族的文化习俗[66]。

通过对除了 Y 染色体遗传标记以外其他

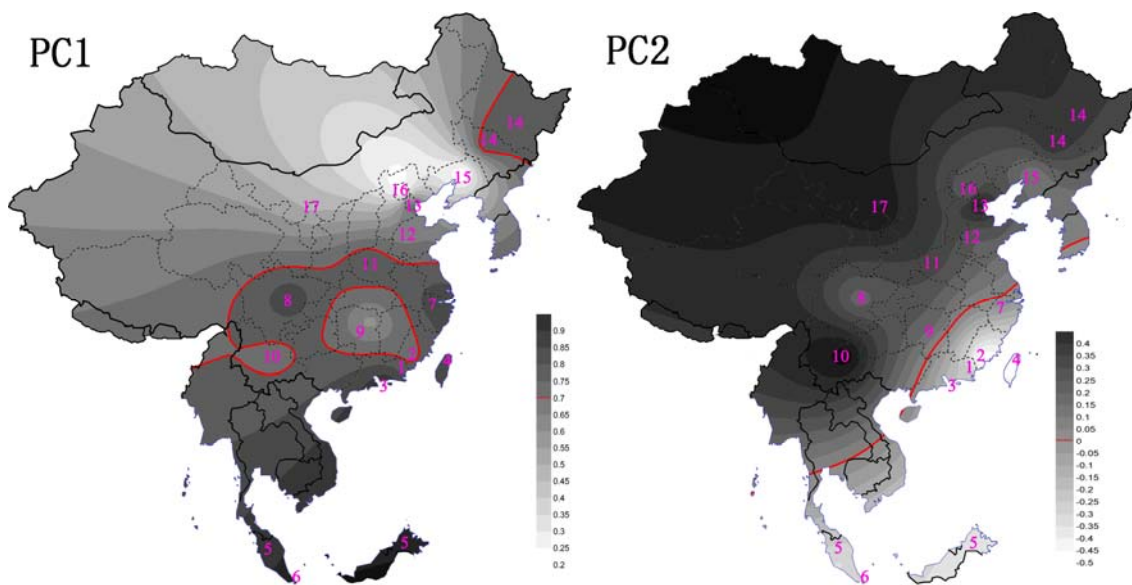


图 4 汉族人群 Y-STR 主成分的地理分布图 地图上的数字表示人群地点：1、潮汕，2、闽南，3、香港，4、台湾，5、马来西亚，6、新加坡，7、浙江，8、四川，9、湖南，10、云南，11、河南，12、山东，13、天津，14、东北，15、辽宁，16、北京，17、宁夏。

Fig.4. The geographic map of the principal components of Y-STR from Chinese Han populations. The numbers in the maps marked the population locations as follows: 1, Chaoshanese; 2, Minnamese; 3, Hong Kong Chinese; 4, Taiwanese; 5, Malaysia Chinese; 6, Singapore Chinese; 7, Zhejiang; 8, Sichuan; 9, Hunan; 10, Yunnan; 11, Henan; 12, Shandong; 13, Tianjin; 14, Northeast; 15, Liaoning; 16, Beijing; 17, Ningxia.

遗传标记的分析，我们[7,8]和其他学者[16,54,68-71]已经观察到南方和北方人群跨越各自语系家族界线，以北纬 30 度(大致沿长江一线)为界，分为南北两部分[54]。与之相反，本研究中 Y 染色体多态性的聚类分析(图 1)和主成分分析(图 2)显示，不管它们的地理分布和遗传距离如何，所有研究人群均按照语系分组聚类。父系遗传方面，中国人群往往与同一语系组内人群聚类，而非与同一地理组内人群聚类。这在南北分布的语系人群中尤其明显，例如藏缅语族和汉族。我们还进一步注意到在同一语系内部的语族基础上的聚类更明显，藏缅语族和汉族虽同属于汉藏语系，但是它们先是语族内部聚类然后再是两个语族聚类(图 1)。潮汕、闽南、台湾、马来西亚和新加坡汉族，尽管他们地理上距离较远，但是在南方汉中他们距离更近(图 1 和 4)。这五个人群均属于汉语十大方言之一的闽语人群[72]，并拥有共同的祖先[6,7,8,73-75]。本研究结果显示中国人群的父系遗传与语言相关性比与其地理位置的相关性强，这和世界人群研究中所观察到的语言与 Y 染色体高度相关[76-78]是一致的。众所

周知，婚姻模式严重影响着一个人群的遗传结构。在父系社会中，男性在文化实践中占据着与在政治领域中同等重要的地位，这使他们很少主动地改变他们的生活方式，而是将习俗代代传下去。女性婚后居住在男方家中，父系为主是中国的传统文化，并存在于当今中国，因此在中国人群中同样存在 Y 染色体和语言的共进化。

总之，本研究中我们首次分析了潮汕汉族的父系遗传结构及与其他中国人群的关系。我们的结果再次验证了潮汕人的中原起源，并进一步揭示了潮汕汉族的父系遗传成分主要来自中原汉族，而仅一小部分来自南方土著。而且，在父系遗传方面，中国人群倾向于按照语系进行聚类而非按照地理分布聚类。另外，从 Y 染色体遗传标记分析观察到，南方和北方人群的界线并不明显。

致谢

本项目得到以下基金支持：香港李嘉诚基金会剑桥大学国际合作项目基金；英国剑桥大学基金；广东省自然科学基金(000819)；汕头大学研究与发展基金(L00007)；汕头市重点科技计划项目(汕府科[2005]116号)。

参考文献

1. 韦东超,王瑞莲(2000)中国民族关系流变史. 武汉:湖北人民出版社.
2. 葛剑雄,吴松弟,曹树基(1997)中国移民史.福州:福建人民出版社.
3. 李辉(2007)分子人类学所见历史上闽越族群的消失.广西民族大学学报(哲学社会科学版) 29(2):42-47.
4. Li H, Cai XY, Winograd-Cort ER, Wen B, Cheng X, Qin ZD, Liu WH, Liu YF, Pan SL, Qian J, Tan CC, Jin L (2007) Mitochondrial DNA diversity and population differentiation in Southern East Asia. *Am J Phys Anthropol* 184:481-488.
5. Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu YF, Pan SL, Qin ZD, Liu WH, Cheng X, Yang NN, Li X, Tran D, Lu D, Hsu MT, Deka RJ, Marzuki S, Tan CC, Jin L (2008) Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol* 8:146.
6. 黄挺(1997)潮汕文化源流(第1版).广州:广东高等教育出版社.1-74.
7. Hu SP, Luan JA, Li B, Chen JX, Cai KL, Huang LQ, Xu XY (2007) Genetic link between Chaoshan and other Chinese Han populations: evidence from HLA-A and HLA-B allele frequency distribution. *Am J Phys Anthropol* 132:140-150.
8. 许丽娜,冯国英,周广彪,胡盛平(2008)常染色体 STR 的证据支持潮汕人的中原汉族起源.现代人类学通讯 2:11-16.
9. Carvajal-Carmona LG, Soto LD, Pineda N, Ortiz-Barrientos D, Duque C, Qspina-Duque J, McCarthy M, Montoya P, Alvarez VM, Bedoya G, Ruiz-Linares A (2000) Strong Amerind/white sex bias and a possible sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* 67:287-295.
10. Benedetto GD, Ergüven A, Stenico M, Castrifi L, Bertorelle G, Togan L, Barbujani G (2001) DNA diversity and population admixture in Anatolia. *Am J Phys Anthropol* 115:144-156.
11. Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen PD, Tournev L, Pablo RD, Kučinskis V, Perez-Leazun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (gypsies). *Am J Hum Genet* 69:1314-1331.
12. Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68:723-737.
13. Sans M, Welmer TA, Franco MHL, Salzano FM, Bentancor N, Alvarez L, Blanchi NO, Chakraborty R (2002) Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* 118:33-44.
14. Bosch E, Calafell F, Rosser ZH, Norby S, Lynnerup N, Hurler ME, Jobling MA (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum Genet* 112:353-63.
15. Kayser M, Lao O, Saar K, Brauer S, Wang XY, Nürnberg P, Trent RJ, Stoneking M (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am J Hum Genet* 82:194-198.
16. Wen B, Li H, Lu D, Song XF, Zhang F, He YG, Li F, Gao Y, Mao XY, Zhang L, Qian J, Tan JZ, Jin JZ, Huang WH, Deka RJ, Su B, Chakraborty R, Jin L (2004) Genetic evidence supports deep diffusion of Han culture. *Nature* 431:302-305.
17. Wen B, Xie XH, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, Lu D, Chakraborty R, Jin L (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in Southern Tibeto-Burmans. *Am J Hum Genet* 74:856-865.
18. 刘静文,胡盛平,李旸(2006)潮汕地区汉族人群 12 个 Y 短串联重复序列基因座遗传多态性分析.中华医学遗传学杂志 23(3):352-353.
19. 李旸,胡盛平,刘静文(2006)闽南地区汉族人群 12 个 Y-STR 基因座遗传多态性.中国法医学杂志 21(3):172-174.
20. Excoffier L, Guillaume L, Schneider S (2005) Arlequin (ver.3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50.
21. Yang BQ, Gu M, Wang G, Li X, Liu Y, Yang W (2006) Population data for 11 Y-chromosome STRs in northeast China Han. *Forensic Sci Int* 164:65-71.
22. 巴华杰,林子清,李树(2007)我国东北地区汉族群体11个Y染色体STR基因座的遗传多态性.法医学杂志 23(3):206-209.
23. Wang XL, Sawaguchi T (2006) Analysis of Y-STR Loci in a Population Sample from Northeast China. *J Forensic Sci* 51:195-198.
24. Zhu BF, Deng YJ, Zhang FX, Wei WJ, Chen LP, Zhao J, He YF, Tian YF, Xu YC, Yu RJ, Fang JB, Liu Y (2006) Genetic Analysis for Y Chromosome Short Tandem Repeat Haplotypes of Chinese Han Population Residing in the Ningxia Province of China. *J Forensic Sci* 51:1417-1420.
25. 匡金枝,朱巍,聂同钢,刘毅,刘孟男,王玉健(2005)天津汉族人群 12 个 Y-STR 基因座的遗传多态性.刑事技术 175(1):19-26.
26. 冯常俊,相志德,申成斌(2005)河南汉族人群 12 个 Y-STR 基因座的遗传多态性.刑事技术 177(3):23-28.
27. Yan JW, Tang H, Liu YC, Jing YT, Jiao ZP, Zhang QX, Gao JW, Shang LP, Guo H, Yu J (2007) Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. *Leg Med* 9:196-202.
28. Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, Krawczak M, Tyler-Smith C, Kim W (2005) Y-chromosomal STR haplotypes and their applications to forensic and population studies in east Asia. *Int Legal Med* 119:195-201.
29. Guo H, Yan JW, Jiao ZP, Tang H, Zhang QX, Zhao L, Hu N, Li HF, Liu YC (2008) Genetic polymorphisms for 17 Y-chromosomal STRs haplotypes in Chinese Hui population. *Leg Med* 10:163-169.
30. Zhu YS, Huo ZH, Yu B, Zhang HB, Wang YJ, Zhao W, Jiao HY, Dang J, Li SB (2007) Genetic polymorphism of twelve Y chromosomal short tandem repeat loci in Chinese Hui ethnic group. *Chin J Med Genet* 24:594-597.
31. Zhang YJ, Zhang HJ, Cui Y, Cui H, Xu QS, Sun S, Sun LP, Lee JB (2007) Population genetics for Y-chromosomal STRs haplotypes of Chinese Korean ethnic group in northeastern China. *Forensic Sci Int* 173:197-203.
32. Zhu BF, Shen CM, Xun X, Yan JW, Deng YJ, Zhu J, Liu Y (2007) Population genetic polymorphisms for 17 Y-chromosomal STRs haplotypes of Chinese Salar ethnic minority group. *Leg Med* 9:203-209.
33. Zhu BF, Li XS, Wang ZY, Wu HY, He YF, Zhao J, Liu Y (2005) Y-STRs haplotypes of Chinese Mongol ethnic group using Y-PLEXTM 12. *Forensic Sci Int* 153:260-263.
34. Zhu BF, Wang ZY, Yang CH, Li XS, Zhu J, Yang G, Huang P, Liu Y (2005) Y-chromosomal STR haplotypes in Chinese Uigur ethnic group. *Int J Legal Med* 119:306-309.
35. 杨亚军,安黎哲,徐玖瑾,张伟华,周瑞霞,王勋陵,谢小冬(2005) Human Y-specific STR haplotypes in Dongxiang ethnic group from Northwest China.兰州大学学报(自然科学版) 41:34-38.
36. Zhu BF, Wu YM, Shen CM, Yang TH, Deng YJ, Xun X, Tian YF, Yan JC, Li T (2008) Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic group residing in Qinghai province of China. *Forensic Sci Int* 175:238-243.
37. 陈水琴,陈和军,曾晓冠,李琼,朱忠炼,聂绍祿(2005)湖南汉族群体 11 个 Y-STR 基因座的遗传多态性.中国法医学杂志 20(3):174-176.
38. Hidding M, Schmitt C (2000) Haplotype frequencies and population data of nine Y-chromosomal STR polymorphisms in a German and a Chinese population. *Forensic Sci Int* 113:47-53.
39. Zhang HJ, Yun LB, Li YB, Zhang J, Wu J, Yan J, Hou YP (2008) Haplotype of 12 Y-STR loci of the PowerPlex Y-system in Sichuan Han ethnic group in west China. *Forensic Sci Int* 175:244-249.

40. Zhang XH, Wu WW, Tang JX, Qian GL, Zhang XM (2006) Polymorphisms of eleven Y-chromosome STR loci and forensic application in Yunnan Han population. *J Forensic Med* 22:291-294.
41. 吴微微, 郑小婷, 潘立鹏, 郝宏蕾, 傅汀 (2005) 浙江汉族人群 16 个 Y-STR 基因座遗传多态性调查. *刑事技术* 179(3):11-17.
42. Yeung SM, Wong LM, Cheung BKK, To KY (2006) Allele frequencies and haplotypes of 12 Y-STR loci for the local Chinese population in Hong Kong. *Forensic Sci Int* 162:55-63.
43. Huang TY, Hsu YT, Li JM, Chung JH, Shun CT (2008) Polymorphism of 17 Y-STR loci in Taiwan population. *Forensic Sci Int* 174:249-254.
44. Tang JSW, Wong HY, Syn CKC, Tan-Siew WF, Chow ST, Budowle B (2006) Population study of 11 Y-chromosomal STR loci in Singapore Chinese. *Forensic Sci Int* 158:65-71.
45. Yong RYY, Lee LKH, Yap EPH (2006) Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. *Forensic Sci Int* 159:244-257.
46. Chang YM, Perumal R, Keat PY, Kuehn DLC (2007) Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci Int* 167:70-76.
47. Zhang QX, Yan JW, Tang H, Jiao ZP, Liu YC (2006) Genetic polymorphisms of 17 Y-STRs haplotypes in Tibetan ethnic minority group of China. *Leg Med* 8:300-305.
48. Kang LL, Liu K, Ma YM (2007) Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. *Forensic Sci Int* 172:79-83.
49. Zhu BF, Liu SZ, Ci D, Huang JF, Wang YC, Chen LP, Zhu J, Xu YC, Zhao QZ, Li SB, Liu Y (2006c) Population genetics for Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group in Tibet. *Forensic Sci Int* 161:78-83.
50. Xin N, Chen T, Yu B, and Li SB (2008) 12 Y-STRs haplotypes in Chinese Naxi ethnic minority Group. *Forensic Sci Int* 174:244-248.
51. Zhu BF, Shen CM, Qian GL, Shi RY, Dang YH, Zhu J, Huang P, Xu YC, Zhao QZ, Ma J, Liu Y (2006) Genetic polymorphisms for 11 Y-STRs haplotypes of Chinese Yi ethnic minority group. *Forensic Sci Int* 158:229-233.
52. Shi MS, Bai R, Wan L, Yu X, Chang L (2008) Population genetics for Y-chromosomal STRs haplotypes of Chinese Tujia ethnic group. *Forensic Sci Int: Genetics* 2:e65-e68.
53. 刘莹 (2004) 广西都安瑶族线粒体 DNA 控制区及 Y-STR 多态性研究. 四川大学法医学硕士学位论文 36-54.
54. 肖春杰, 杜若甫, Cavalli-Sforza LL (2000) 中国人基因频率的主成分分析. *中国科学(C 辑)* 43:472-481.
55. Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) ver. 3.6. Univ Washington, Dept Genome Sciences, Seattle. (<http://evolution.genetics.washington.edu/phylip.html>)
56. Nei M (1987) Genetic variation within species. In: *Molecular evolutionary genetics*. New York: Columbia University Press, 220-221
57. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
58. Dupanloup I, Bertorelle G. (2001) Inferring admixture proportions from molecular data: extension to any number or parental populations. *Mol Biol Evol* 18:672-675.
59. Wang J (2003) Maximum-likelihood estimation of admixture proportions from genetic data. *Genetics* 164:747-765.
60. Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. *Mol Biol Evol* 15:1298-1311.
61. Roberts DF, Hiorns RW (1965) Methods of analysis of the genetic composition of a hybrid population. *Hum Biol* 37:38-43.
62. Li D, Li H, Ou CY, Lu Y, Sun YT, Yang B, Qin ZD, Zhou ZJ, Li SL, Jin L (2008) Paternal Genetic Structure of Hainan Aborigines Isolated at the Entrance to East Asia. *PLoS ONE* 3:e2168.
63. Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE (1997) Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol* 102:153-159.
64. Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR (1998) Female gene flow stratifies Hindu castes. *Nature* 395:651-652.
65. Seielstad MT, Minch E, Cavalli-Sforza L (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278-280.
66. 陈训先 (2006) 潮汕先民探源. 广州: 广东人民出版社.
67. 范玉春 (2000) 论中国古代军事移民对移居地的影响. *广西师范大学学报(哲学社会科学版)* 36:81-84
68. Zhao TM, Lee TD (1989) Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Hum Genet* 83:101-110.
69. Du RF, Yuan YD, Hwang J, Mountain J, Cavalli-Sforza LL (1992) Chinese surnames and the genetic differences between north and south China. *J Chin Ling, monograph series* no.5.
70. Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. *Proc Natl Acad Sci USA* 95:11763-11768.
71. Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635-651.
72. 甘瑞静, 潘尚领, 覃振东, 蔡晓云, 徐杰舜, 李辉 (2008) 桂北平话汉族的遗传成分主要源于当地少数民族. *现代人类学通讯* 2:1-10.
73. 汤漳平 (2002) 初唐中原移民入闽与闽台文化之形成. *许昌师专学报* 21(1):82-86.
74. 林国平, 邱季端 (2005) 福建移民史. 北京: 方志出版社.
75. 张应龙 (2006) 郑和下西洋与满刺加的中国移民史. *学术论坛* 3:187-189.
76. Cavalli-Sforza LL, Minch E, Mountain JL (1992) Coevolution of genes and languages revisited. *Proc Natl Acad Sci USA* 89:5620-5624.
77. Chen J, Sokal RR, Ruhlen M (1995) Worldwide analysis of genetic and linguistic relationships of human populations. *Hum Biol* 67:595-612.
78. Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup L, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome p49a, f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61:1015-1035.