

# 上海本地人源流主成分分析

李 辉<sup>1</sup>, 周振鹤<sup>2</sup>, 徐立群<sup>3</sup>, 张凌君<sup>4</sup>, 金建中<sup>1</sup>, 金 力<sup>1</sup>, 卢大儒<sup>1</sup>

(1. 复旦大学现代人类学研究中心, 上海 200433; 2. 复旦大学历史地理研究中心, 上海 200433;  
3. 复旦大学管理学院统计与运筹学系, 上海 200433; 4. 上海水产大学生物技术系, 上海 200090.)

**摘要:** 本文对分布于上海市郊9个区县13个乡镇的本地人群的体质特征和姓氏进行了主成分分析, 发现上海本地人主要可以分为3个类群, 与其他研究结果相整合推测他们可能分别源于越人、吴人和历史上的北方移民。各主成分所占数据信息的比例显示, 本地人中最主要的差异是北方移民和南方土著的差异, 其次是南方土著中吴人和越人的差异, 再次是越人内部也可以分出两种来源。把各主成分值在上海地图上分别画成趋势图, 发现各趋势图分别代表了上海历史上不同的移民事件。对姓氏的地理分析也揭示了南方姓氏和北方姓氏的扩散趋势和性质是不同的。

**关键词:** 上海本地人; 人群源流探索; 主成分分析; 吴越

**中图分类号:** Q983, Q988 **文献标识码:** A **文章编号:** 1005-8036(2003)02-0167-07

上海虽然作为一个国际大都市的历史只有短短的百来年, 但是却有着很悠久的人类定居史。考古发现不断证明五六千年前就已经有大量的人口居住在这一区域内<sup>[1]</sup>, 在历史上, 许多事件也引发了一批批的移民进入上海, 成为历史学研究的重要对象。上海特殊的地理位置使得本地人群在中国人群中有着独特的地位。遗传学研究证明, 长江是中国南北方人群的大致分界线<sup>[2]</sup>, 而上海正是处于这条线的端点, 同时受到南北方人群的影响, 对中国人群的遗传结构研究很有意义。在历史上, 上海曾经处于吴越两国的边境上, 长期以来又一直是江苏和浙江的边界地区, 故而上海本地人的源流涉及到许多历史问题。此外, 上海作为一个重要的大城市, 其基础人口的来源对上海城市社会研究本身也是很有意义的。所以, 从人类学研究进入中国开始, 就有学者把目光投向了上海本地人群。

十九世纪中叶, 一些西洋传教士的著作中就指出了上海地区的人群有着不同的方言和服饰文化<sup>[3]</sup>。之后很长时期内对上海人群的认识没有很大的进展。二十世纪初(1928)赵元任对吴语的研究指出, 今吴语区域(包括上海)春秋时主要居住着百越民族, 秦汉以来不断有越人北徙、汉人南迁的移民举动, 形成了汉越杂居的局面<sup>[4]</sup>。这是对上海人来源认识的一个突破。二十世纪八十年代以来, 语言学和人类学的研究又得到了长足的发展, 出现了许多新的观点。1987年钱乃荣研究上海奉贤的方言发现奉贤东西部的方言差异极大<sup>[5]</sup>。1988年郑张尚芳分析了浙南和上海一些方言的特殊语音, 推测两地的古代居民不但都是越族, 也许还有更亲的关系, 而与吴人有区别, 所以相信当地现代居民中应有相当比例的越人后裔<sup>[6]</sup>。1993年许宝华等完成了对上海地区的方言分片工作<sup>[7]</sup>, 其所分的7个片的人群来源很有可能不一样。至此为止对上海本地人的研究都限于语言学和文化人类学的范畴。

2000年, 我们的课题小组开始对上海郊区的本地人作体质人类学和遗传学的调查研究。首先发现了奉贤东西部的人群在体质上也存在显著的差异<sup>[8]</sup>, 与语言学的现象一致。在马桥地区和奉贤西部的遗传学分析发现这一相连地区的居民的Y染色体单倍型不是一般汉族的结构, 而是与现代的其他百越民族相当一致, 有些特殊的单倍型甚至目前只能追溯到柬埔寨等地区<sup>[11]</sup>。这也证实了语言学上关于本地存在越人后裔的推测。之后我们对散布上海郊区的13个乡镇的人群体质特征和姓氏等进行了调查, 发现在体质和姓氏遗传上, 上海本地人都可以分成3各类群, 即越人后裔、吴人后裔和北方移民<sup>[9, 10]</sup>。通过通

收稿日期: 2003-01-01

作者简介: 李辉(1978—), 男(汉族), 上海奉贤人, 复旦大学遗传所博士研究生, 主要研究方向为南方民族遗传关系。

婚机制的分析,也指出上海人群的变动不是由通婚引起的,移民才是主要因素;因为迁入地不同,所以不同人群之间大致还是分片而居,没有明显的混血<sup>[10]</sup>.在这些研究基础上,我们深入分析上海本地人中究竟保存着哪些移民事件留下的痕迹,这些痕迹可能对应的历史事件,对我们的研究作一个阶段性的总结.本文利用已发表过的 13 个人群的相关数据,用主成分分析的统计方法<sup>[2,11]</sup>,在地图上画出人群差异的趋势,来解答我们的问题.

## 1 材料与方法

### 1.1 研究对象

上海市郊 9 个区县的 13 个乡镇(金汇、头桥、商榻、朱家角、天马山、施湾、书院、枫泾、廊下、堡镇、南翔、朱家桥、马桥)的本地籍居民,13 个乡镇的具体地理位置见图 1.图中还标出了我们观察到的当地最有人群特征的头巾式样,不同的头巾式样表明人群的文化传统有差异.研究的个体一般状况已在相关论文中表述<sup>[9,10]</sup>.

### 1.2 分析项目

在相关论文中列出了 13 个人群的面貌观察性特征、头部和体部测量数据及标准化指数、肤纹参数<sup>[9]</sup>.考虑到各种参数对人群差异的体现程度不同,并且分析要求数据的标准化,我们选取了具有定性意义的所有的面貌观察性特征和体质测量的标准化指数作进一步的分析.13 个人群的姓氏在前文中列出<sup>[10]</sup>.

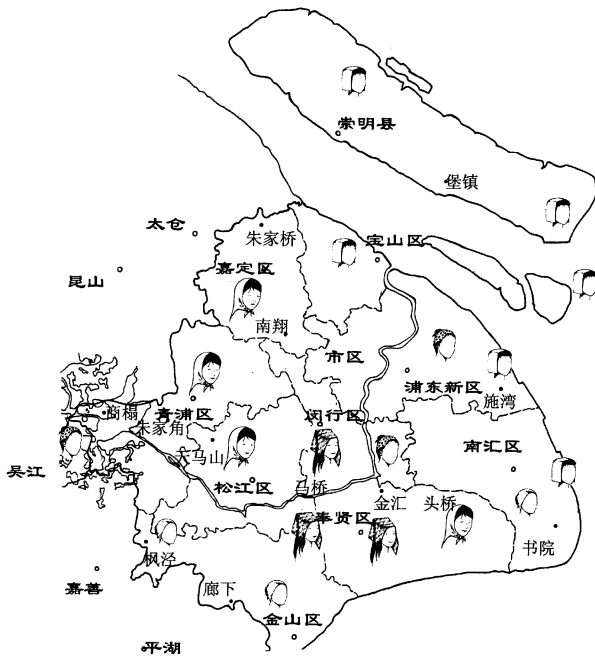


图 1 调查人群的地理分布及文化差异(以头巾式样的不同表示)

Fig. 1 Location and culture (shown with headkerchiefs such as babushka, coif and kerchief) of the populations we studies

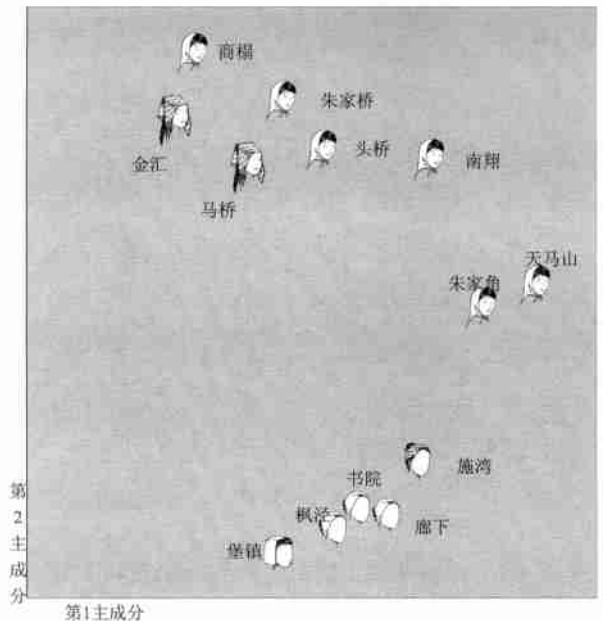


图 2 调查人群的主成分分析散点图

Fig. 2 Distances among populations shown by principle component analysis of physical characters

### 1.3 主成分分析

用美国 SPSS 公司的 SPSS10.0 统计分析通用软件对体质数据进行主成份分析<sup>[11]</sup>.用第一主成分和

第二主成分构建直角坐标系画出各人群的散点图观察群体间差距。根据人群的地理分布在上海地图上绘制等高线的原理画出各主成分的梯度分布图<sup>[2]</sup>。姓氏数据在前文中已做过了主成分分析,画出了散点图,在本文中再画出其地理梯度分布图。

## 2 结果与讨论

### 2.1 体质特征的主成分分析相关数值

对选取的体质特征数值作主成分分析共得到 13 个主成分,前 10 个主成分所占数据的方差较大,比较有意义,其占方差比例数值见表 1。

表 1 根据上海郊区 13 个人群的体质特征分析得的主成分所占的方差比例(%)

Tab. 1 Variance percentage of components from principle component analysis of physical characters of 13 populations in Shanghai suburb

主成分	1	2	3	4	5	6	7	8	9	10
方差比例	84.172	9.251	2.014	1.070	0.934	0.599	0.529	0.376	0.353	0.272

可见,第 1 主成分和第 2 主成分占了数据信息的绝大部分,代表着上海本地人群的最主要的 2 种差异,可以划分 3 种来源。这与之前的研究作出的本地人群可能存在越人、吴人和北方移民的 3 种来源的推测是吻合的。其余主成分也包含了有历史上其他移民事件留下的微弱痕迹。

### 2.2 主成分值揭示的上海本地人 3 大来源

利用各人群的前 2 主成分构建直角坐标系,各人群在坐标系中的分布见图 2。

图中可以明显地看出,具有相同或相似文化特征的人群相对聚在了一起,说明他们在体质上也相对相近。所以,这 13 个人群大致可以分为 3 类。金汇和马桥有相同的文化和相近的体质,我们已经从分子遗传学上证实了他们最可能是百越民族的遗民<sup>[1]</sup>,为了叙述简便,下文仍直接称他们为越人。堡镇、枫泾、廊下、施湾、书院代表的崇明、浦东、南汇、金山也相当接近,他们的文化特征与吴江、吴县等地居民相近,可能是吴人来源,下文直接称他们为吴人。其余群体虽然有相近的文化,但体质特征相对分散,可能是由于他们是来源较复杂的北方移民。这 3 类人群虽然有相互影响,但也保持了相对的独立性,是可以区分的。下文我们通过主成分的地理分析,研究 3 类人群形成现在的分布格局的几种过程。

### 2.3 各体质特征主成分的地理分析

从图 3 到图 5 画出了各个主成分的地理分布梯度,每种地理分布代表着一个移民事件。我们对前 3 个主成分分布进行分析解释。

第 1 主成分是一个典型的扩散趋势,从松江向周边地区扩散。松江是上海历史上第一个官府驻地,中央政府在此设置行政机构,势必带来一大批的中原移民。官府的治所有一系列经济、政治、文化、军事的保证也吸引了北方移民在这一地区的聚集。相对而言,北方移民要在其他南方土著的地区定居难度较大,故而地图中的扩散明显地绕开了马桥等越人地区。松江作为上海的政治中心持续的时间最长,所以这一趋势在所有趋势中占的比例最大。这一主成分除了可以看作松江府居民的聚居和扩散之外,从另一方面也可看作是北方移民和南方吴越土著的差异,所以在人群差异中最主要的还是南北来源的差异。

第 2 主成分是典型的两种人群梯度式地交融的趋势,其中一个人群分布在中间,另一个分布在两侧。中间的人群处在西北-东南走向贯穿上海地区的冈身地带。冈身是形成于四千多年前的一系列贝壳沙堤,是上海平均地势最高,人类居住较早的地带。据考古发现和历史记载,吴人的湖熟文化进入太湖平原是三千多年前<sup>[12]</sup>。所以这一地带的许多居民应该是越人,从图上看是南多北少的分布,与文化的分布也相吻合。而两侧的居民则是吴人,与越人之间略呈渐进式的过渡。而且这种过渡是南部陡峭、北部平

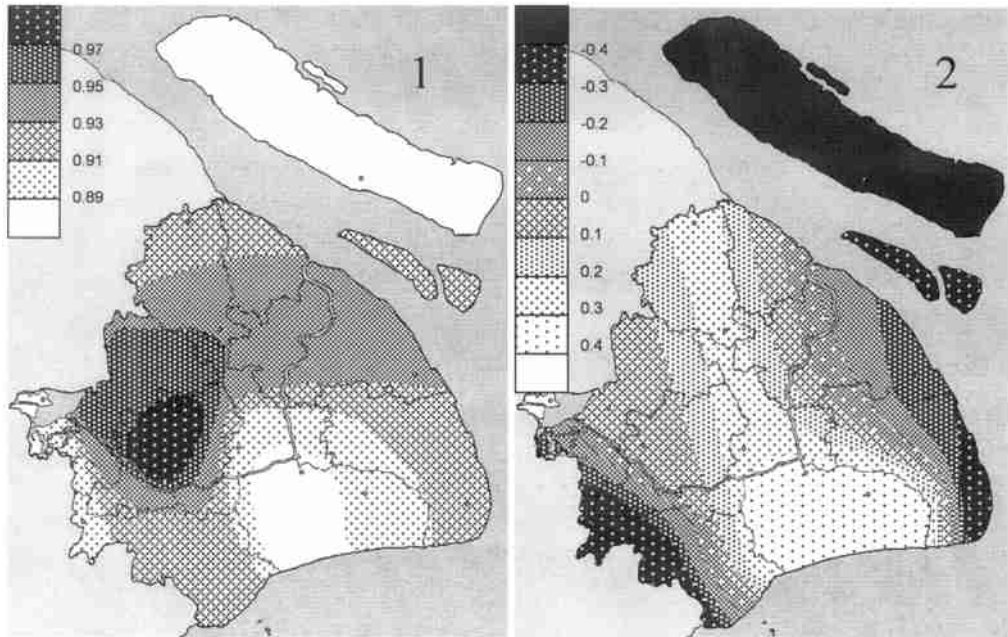


图 3 体质特征第 1、2 主成分的地理分析

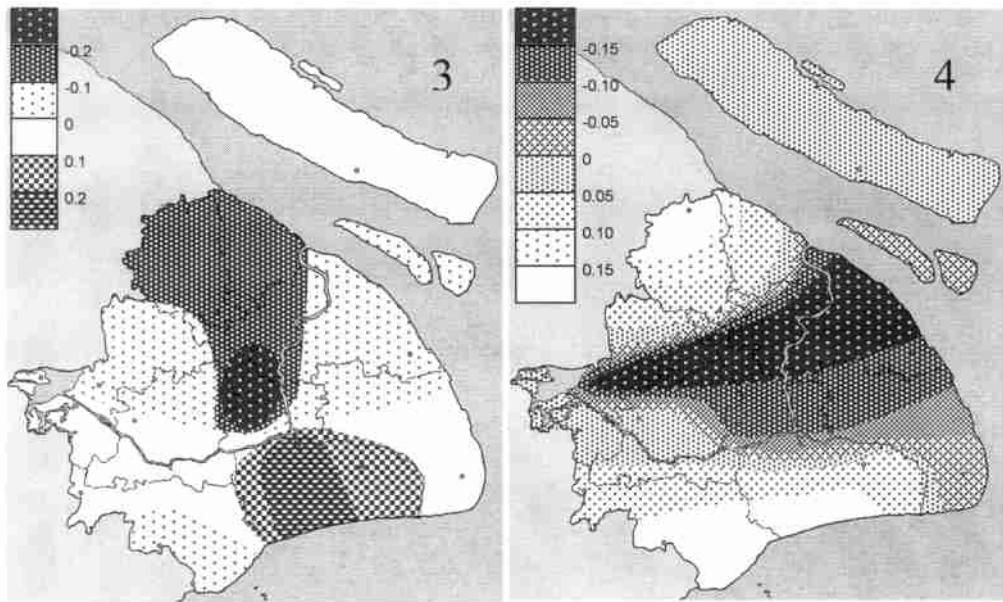
Fig. 3 Geographic analysis of the 1<sup>st</sup> and 2<sup>nd</sup> components of physical characters

图 4 体质特征第 3、4 主成分的地理分析

Fig. 4 Geographic analysis of the 3<sup>rd</sup> and 4<sup>th</sup> components of physical characters

缓,因为北部受北方移民的冲击影响,所以可见在南北差异之后的一种差异是南方的吴越之间的差异。

第 3 主成分也是两种人群相接触的趋势,马桥和金汇之间存在着区别,说明越人虽然文化相当一致,但是内部有两个来源。汉代浙南越族的东瓯国内迁江淮<sup>[6]</sup>,但是史书没有说明究竟迁到了哪里,很有可能他们迁到了上海,与上海更早的越族毗邻而居,因为同一民族内部的群体相对容易相处,马桥和金汇哪一个可能是东瓯内迁的越族后裔是很难定夺的,浙南和浙北的考古文化向来不一致<sup>[1]</sup>,所以这两地的越人之间有差异也是合理的,因此,上海本地人中的第 3 类差异是越人内部的差异。

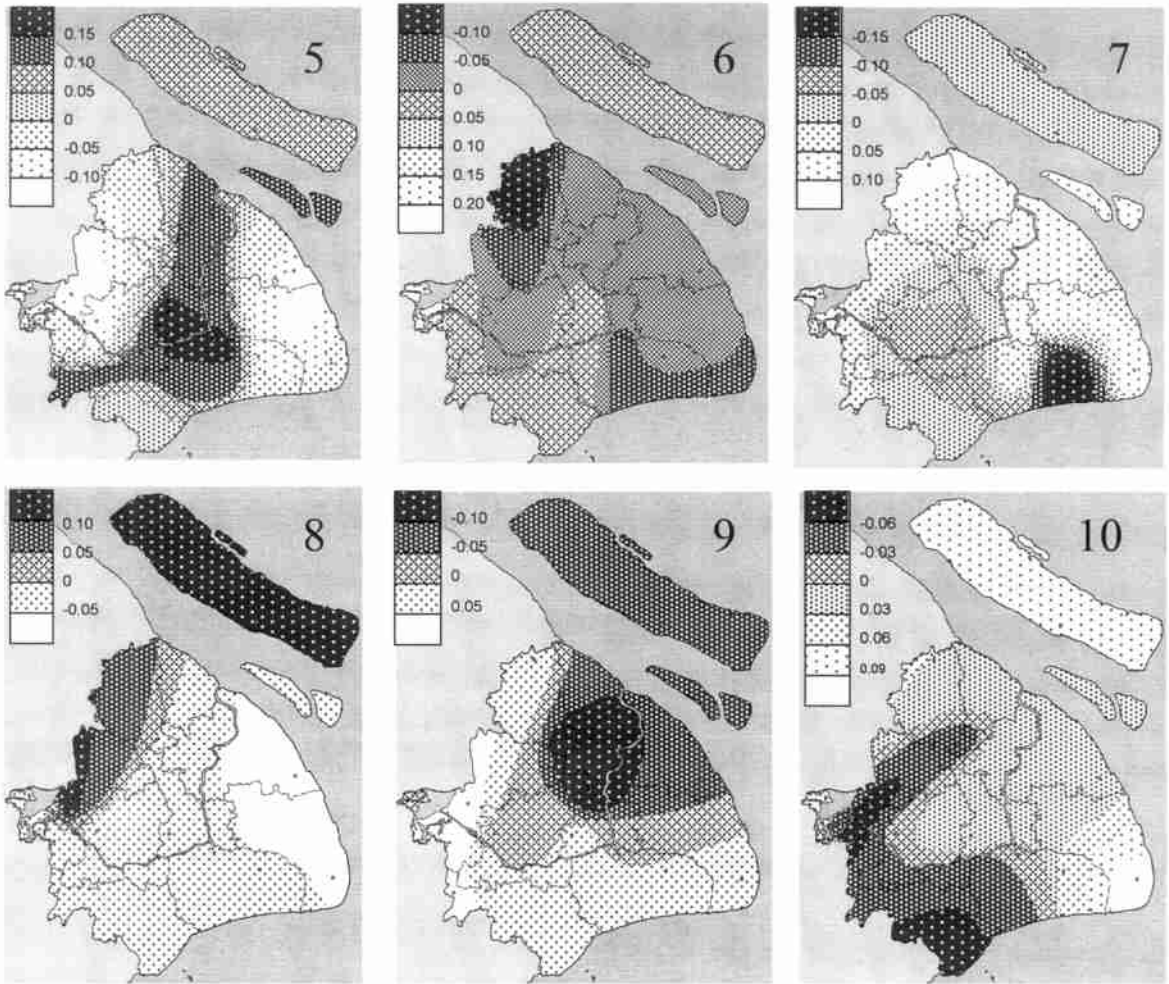


图 5 体质特征第 5~10 主成分的地理分析

Fig. 5 Geographic analysis of the 5<sup>th</sup> ~ 10<sup>th</sup> components of physical characters

第 4 主成分之后的各主成分也可能有一些具体意义,但是这些信息涉及的数据很少,意义是不明确的。

#### 2.4 姓氏主成分分析数据和地理分析

对 13 个人群地姓氏构成的研究也可以提示各家族的渊源.我们对姓氏进行的主成分分析也析出了 13 种主成分,各主成分所占的信息比例比体质分析结果接近.由于姓氏并不是自然的遗传特征,所以不像体质特征那样能反映许多现象,我们只观察前 4 个主成分,其所占数据方差比例见表 2.

表 2 根据上海郊区 13 个人群的姓氏分析得的主成分所占的方差比例 (%)

Tab. 2 Variance percentage of components from principle component analysis of surnames of 13 populations in Shanghai suburb

主成分	1	2	3	4
方差比例	51.35	8.343	6.865	6.35

人群的主成分散点图已在相关论文中画出<sup>[10]</sup>.由于越人原先没有汉姓,现有的姓氏是历史上借鉴于周边人群的,所以散点图中没有析出越人的类群.

图 6 画出了前 4 个主成分的地理梯度分布.

第 1 主成分从西部的苏州地区向上海全境扩散,扩散是单一性的主成分值的降低.这应该是代表了

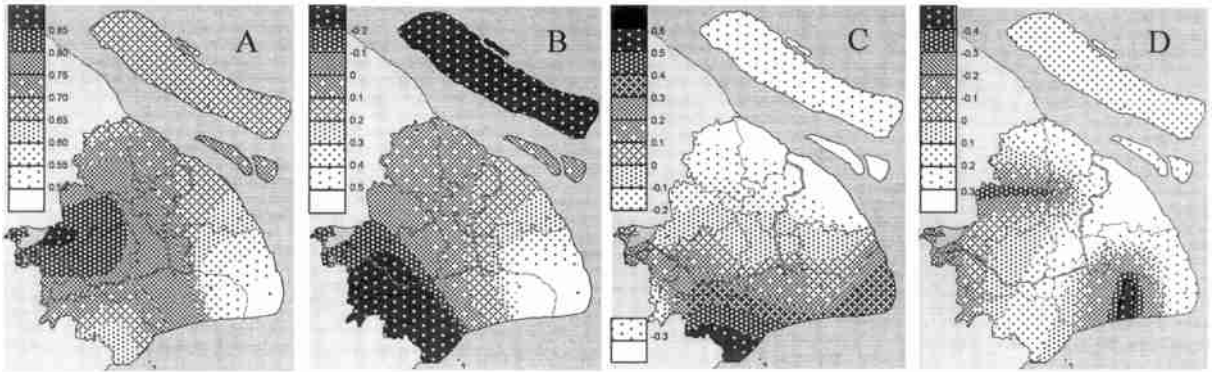


图 6 姓氏第 1~4 主成分的地理分析

Fig. 6 Geographic analysis of the 1<sup>st</sup> ~ 4<sup>th</sup> components of surnames

南方家族及其姓氏的扩散. 由于南方的家族和姓氏人口多而种类相对北方移民少, 这些家族从西部向 3 个方向在本地人中的比例逐次下降, 而姓氏种类基本没有减少. 这是本地家族的一个来源.

第 2 主成分是从上海中北部向西南和东南呈两个不同的趋势变化. 这应该使代表了北方家族和姓氏的扩散. 来自北方的一批批移民带来了大量的姓氏种类, 具有姓氏种类多而各姓氏的人口较少的特点. 所以从北部进入之后向两个方向的移民带有不同的姓氏, 在主成分上体现为从 0 值向正负两个方向的变化.

第 3 主成分展示的是来自浙江和江苏的家族地理分布上相过渡的趋势. 其间的分界线 (0 度线) 相对于古代江浙界线偏向于浙江一侧, 说明江苏的移民多于浙江.

第 4 主成分可能是到西部湖沼地带和东部海岸滩涂垦殖的家族迁入的现象.

## 2.5 对人群各种特征参数的主成分分析在移民史研究中的意义

对上海本地人的体质特征的主成分分析不但划分出了吴、越、北方移民 3 种来源, 还勾勒出了历史上的许多移民事件留下的痕迹. 这对移民史的研究应该是很好的启示作用的. 在如此复杂的上海地区尚且能分析出各种现象, 其他地区应该可以更清晰地看到移民的痕迹, 解决许多文献中不详的悬疑问题.

各主成分占数据的方差比例也有很重要的意义, 在一定程度上反映了各种移民在当地人口中占的比例. 比如本文中体质特征主成分分析, 得到的与越人有关的主成分 (2, 3, 5) 的方差总和约为 12.2%, 这与马桥等地越人风俗的 40 多万人口在上海全部本地人中占的比例大致相等.

相信这种数理统计方法结合体质人类学调查的研究形式必将为移民史研究注入新的动力.

## 参考文献:

- [1] 宋建. 马桥[M]. 上海: 上海书画出版社, 2002.
- [2] XIAO C J, CAVALLI-FORZA L L, MINCH E, et al. Principle Component Analysis of Gene Frequencies of Chinese Populations[J]. *Sci China Ser C*, 2000, 43(5): 472 - 481.
- [3] 周同春. 十九世纪的上海语音[A]. 中国语言文学研究所吴语研究室. 吴语论丛[C]. 上海: 上海教育出版社, 1988.
- [4] 钱乃荣. 当代吴语研究[M]. 上海: 上海教育出版社, 1992.
- [5] 钱乃荣. 奉贤东西乡的语言同言线[A]. 李振麟. 语言研究集刊 I[C]. 上海: 复旦大学出版社, 1987.
- [6] 郑张尚芳. 浙南和上海方言中的紧喉浊塞音声母初探[A]. 中国语言文学研究所吴语研究室. 吴语论丛[C]. 上海: 上海教育出版社, 1988.
- [7] 许宝华, 汤珍珠, 陈忠敏. 上海地区方言的分片[J]. *方言*, 1993(1): 14 - 30.
- [8] 林凌, 李辉, 陈莉, 等. 奉贤县东西部两个人群的体质差异[J]. *复旦学报(自然科学版)*, 2002, 41(1): 97 - 101.
- [9] 林凌, 李辉, 陈莉, 等. 上海郊区人群的体质特征和遗传关系[J]. *人类学学报*, 2002, 21(4): 293 - 306.

- [10] 徐立群,李辉,奚慧峰,等.上海郊区姓氏和通婚分析[J].遗传学报,2002,29(8):666-673.
- [11] CAVALLI-SFORZA L L, MENOZZI P, PIAZZA A. The History and Geography of Human Genes[M]. Princeton: Princeton University Press, 1996.
- [12] 鲍明炜. 六朝金陵吴语辨[A]. 中国语言文学研究所吴语研究室. 吴语论丛[C]. 上海:上海教育出版社, 1988.

## Principle Component Analysis of Shanghai Suburbanites Headstream

LI Hui<sup>1</sup>, ZHOU Zhen-he<sup>2</sup>, XU Li-qun<sup>3</sup>, ZHANG Ling-jun<sup>4</sup>,  
JIN Jian-zhong<sup>1</sup>, JIN Li<sup>1</sup>, LU Da-ru<sup>1</sup>

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

2. Center for Historical Geographic Studies, Fudan University, Shanghai 200433, China;

3. Department of Statistics, Management College, Fudan University, Shanghai 200433, China;

4. Department of Biotechnology, Shanghai Fisheries University, Shanghai 200090, China.)

**Abstract:** Principle component analysis on physical and surnames divides Shanghai suburbanites from 13 townships of 9 counties into 3 clusters. Conferring the result of other research fields, the 3 clusters shall be Yue(Daic), Wu and historical immigrants from North China. The variance percentages of the components show that the primary difference in Shanghai suburbanites is between Northern Immigrants and Southern Natives. Secondary difference is between two kinds of Southern Natives, Wu and Yue, then Yue is divided into two parts with different origins. Marking the components into the contour on the Shanghai map, the historical immigration events are shown in different maps. Geographic analysis on surname components also tells that the diffusion features and patterns of Southern surnames and Northern surnames are entirely different.

**Key words:** Shanghai suburbanites; historical immigrations; principle component analysis; Wu Yue

### 简讯

2003年4月4日中央民族大学学报(自然科学版)编委会召开重组后的第一次会议,会议对2003年初改为季刊和扩大版面后的工作进行总结,为进一步办好学术刊物、提高学报的编辑质量,编委会提出不少中肯的意见和宝贵建议。进入核心期刊,是本届编委会始终努力奋斗的目标,为此,编委会诚恳希望学校对自然科学版学报予以政策倾斜,我校教员在本学报上发表的优秀论文,在职称评定和计算科研工作量中应有所优惠。会后,编委会还将该项建议写成报告,争取得到学校有关方面的批准与支持。