

# 宋朝中国人的姓氏分布与群体结构分化

袁义达 金 锋 张 诚

(中国科学院遗传研究所 北京 100101)

斋藤成也

(国立遗传学研究所 日本国三岛市)

**摘要** 姓氏是一种十分有用的文化遗传因子,它的传递方式类似于Y染色体的表现。1000年前宋朝(公元960~1179年)中国人姓氏频率的分布,是一组十分吻合Karlin-McGregor的中性等位基因分布理论的随机数据。16个省区的姓氏种类丰度的相对参数 $\alpha$ 和迁移率的相对参数 $\nu$ 的分析,反映了这一时期的中国人群的迁移和人群间混合的程度。姓氏遗传距离和树状聚类图的分析,表明在1000年前的宋朝中国人群已经形成了南北两大区域的群体,并提示了南北两大群体的异源性,其明显的地域分界线在中国南部的武夷山和南岭。而且,树状聚类图的结构还表明1000年前的宋朝和当代的姓氏分布基本一致。宋朝和现代分别占总人口86%和87%的100个频率较高的姓氏分布曲线非常一致,反映了中国人姓氏在这1000年中的传递是比较稳定的,并证明中国人姓氏分布频率是研究和分析中国人群体结构、人群迁移和亲缘关系的一种非常有用的“遗传学”资源。

**关键词** 文化遗传因子,姓氏频率,群体结构,群体遗传学

**分类号** Q347

1875年G. Darwin根据人群中的姓氏分布频率,计算和分析了英国人群的近亲结婚系数<sup>[1]</sup>,从遗传学、文化和历史的角度,首次把姓氏作为生物学的标记应用于人类学研究之中。作为文明文化产物的姓氏,在人类社会中的传递类似于人类Y染色体的表现。姓氏相当于Y染色体上某个遗传座位,每一种姓氏相当于这个座位上的不同的等位基因<sup>[2,3]</sup>。在人类学<sup>[4]</sup>、近亲结婚<sup>[5]</sup>、人群间的生物相关性<sup>[6]</sup>、迁移<sup>[7]</sup>、群体结构<sup>[8,9]</sup>等一系列研究中,人们对姓氏作为生物学的标记有了更多的了解和发现。尤其姓氏作为一种文化“遗传因子”,在人类群体遗传学研究中是一个非常有用的标记,是一种廉价的和含有高度信息量的“遗传学”资源。

亚洲的姓氏群体遗传学的研究,最早见报于日本人的研究<sup>[10~12]</sup>。中国人的姓氏起源有记载的年代已至少4000年历史,而欧洲不足1000年,日本也仅仅几百年的历史。从姓氏历史和人口分布分析,中国人姓氏群体遗传学研究是很有前景的。80年代初,开始出现

---

本文于1998-03-04收到,1998-06-15修回

本文由国家自然科学基金资助(批准号39870406)

有关中国人姓氏群体遗传学的研究<sup>[13~15]</sup>。这些研究结果与许多遗传标记在中国人中分布是非常一致。本研究将对宋朝(公元960~1179年)的中国人姓氏分布进行分析,探讨中国人姓氏传递的稳定性,研究历史上中国人群的迁移和融合。

## 1 材料

本研究所用的宋朝姓氏样本全部取于《宋人传记资料索引》<sup>[16]</sup>,此书从490种文献和许多其他的资料中收集了宋朝320年间的有关的人物,是迄今为止所发现的收集宋朝人物最全的一部文献。我们根据每一人物的姓名和籍贯,或者居住地、工作地区、文献表明 的地点进行统计。为了便于与现代的姓氏研究的结果比较,宋朝地名的省份的归属按目前的行政区进行归并。东北地区、台湾省和西藏地区缺乏姓氏资料。在统计中,一些省份的资料由于太少,我们只好把它们与其相邻的省份合并,其中北京和天津的资料归于河北省,上海的资料归于江苏省,内蒙古和辽宁的数据合并到河北省,新疆、青海的数据合并到甘肃省,宁夏的数据合并到陕西省,云南和贵州的数据合并到四川省。总共得到23411个人物和436个姓氏的数据(表1)。

在这份资料中,河北省的赵姓数据很大,总共515个,约占河北省全部数据(1231个)的42%。而且,我们发现河北涿州地区的数据是463个,其中赵姓的数据是453个,河北涿州地区的数据约占河北省数据的38%。按宋朝现在推测的人口数据<sup>[17]</sup>,当时河北涿州地区的人口仅仅占河北省人口的3%。为什么会出现河北省赵姓人数异常的情况?我们认为这与“皇族效应”有关,这是典型的“任人唯亲”的现象。宋朝的开国皇帝赵匡胤,其籍贯和出身地均在河北涿州,因此当时涿州赵姓容易做官和出名,由此造成了赵姓人数在文献中的异常。在历史上,因“皇族效应”而造成在文献中皇族姓氏的人数异常是非常明显的,如汉朝的刘姓、唐朝的李姓、明朝的朱姓等。为了获得更接近随机的数据,河北涿州地区的赵姓数据需要作必要的推算:

$$\begin{aligned} \text{河北省推算人数} &= (\text{河北省观察人数} - \text{涿州地区观察人数}) \div 97\% \\ &= (1231 - 463) \div 97\% = 792 \end{aligned}$$

$$\text{其多余的人数} = 1231 - 792 = 439$$

因此,河北省的数据和赵姓人数需要减少439个。表1中河北省和全国的合计人数中括号内的数据为赵姓未作推算前的原观察数。

## 2 姓氏频率的分布符合中性等位基因分布的理论

二千年以来,中国人姓氏的传递一般是以父系相传,子女继承父亲的姓,这与人类Y染色体的传递相似,但有所区别。我们注意到中国人中存在着改姓的社会现象,带有随机性;以及同姓并不同祖的问题,具有固定性。由于某种原因而改姓的现象时有发生,比如政治原因、过继改姓、随母姓等等,但是这类改姓的人数在整个人群中占有极小的比率,并不十分影响群体内姓氏分布的频率。这类改姓的现象十分相似于遗传学中等位基因的突变。而同姓并不同祖的问题在常见姓氏中普遍存在,但是,在世代传递过程中同样也很少有改姓的现象。目前要区分其祖宗是十分困难的。这相当于在常规下要鉴别某A型血型人是属于哪种A亚型血型一样困难。所以,同姓并不同祖的问题十分相似于等位基因中的亚型区分问题。姓氏“遗传座位”上姓氏种类已远远超过任何一个人类遗传座位上的等位基因的数目,所以,本文所需要的姓氏数目是足够了。在传递过程中,姓氏不受疾病、气

表1 宋朝(公元960~1279年)中国人姓氏分布和人口数

Table 1 The distribution of Chinese surnames and population density in the Song dynasty (AD. 960~1179)

省名 Province	宋朝中国人姓氏分布的参数 The parameters for Chinese surnames in the Song dynasty						宋朝历年人口数* The population density of difference times in the Song dynasty*							
	样本数 Individual	姓氏数 Surname	$v$	$\alpha$	$d.f$	$\chi^2$	$P$	*980年 980 y.①	*1079年 1079 y.②	*1102年 1102 y.③	*1210年 1210 y.④	加权平均值 W.average	占全国比重 Per cent(%)	④/①比率 Rate
河南 Henan	4011	262	0.015	62.79	31	27.43	0.650	269.5	517.9	518.4	1499.0	694.9	8.99	5.56
甘肃 Gansu	293	75	0.100	32.58	11	17.20	0.102	83.8	196.6	184.8	259.3	183.7	2.38	3.09
陕西 Shaanxi	477	98	0.073	37.38	16	14.92	0.531	168.2	446.5	484.7	419.8	391.2	5.06	2.50
河北 Hebei	(1231)792	122	0.048	40.29	22	27.08	0.208	266.3	481.4	505.4	1107.3	588.9	7.62	4.16
山东 Shandong	821	134	0.052	45.46	22	19.69	0.603	245.3	633.3	584.6	1009.8	624.3	8.08	4.12
湖北 Hubei	373	108	0.120	50.99	11	20.58	0.038	95.1	369.6	336.0	156.6	250.9	3.25	1.65
四川 Sichuan	1800	182	0.027	50.56	27	34.13	0.162	489.2	738.8	773.7	915.1	736.4	9.53	1.87
山西 Shanxi	522	111	0.076	43.15	17	32.64	0.013	131.9	310.5	339.2	718.2	376.0	4.86	5.45
湖南 Hunan	483	108	0.082	43.20	15	21.37	0.126	79.2	440.0	494.8	534.0	400.0	5.17	6.74
广西 Guangxi	182	70	0.186	41.64	6	5.53	0.478	33.9	117.8	131.0	214.0	126.1	1.63	6.31
安徽 Anhui	897	146	0.052	49.47	23	20.57	0.607	219.0	528.3	502.3	320.9	405.6	5.25	1.47
江西 Jiangxi	2467	187	0.019	46.99	29	35.75	0.181	280.6	734.5	856.1	1025.2	739.7	9.57	3.65
江苏 Jiangsu	1816	218	0.034	64.70	26	34.71	0.118	167.5	457.4	446.1	348.5	366.2	4.74	2.08
浙江 Zhejiang	4226	240	0.013	55.15	32	18.19	0.976	141.1	549.3	625.5	670.4	511.6	6.62	4.75
福建 Fujian	3431	155	0.010	33.39	33	26.78	0.769	191.8	428.0	461.0	625.5	432.9	5.60	4.75
广东 Guangdong	381	99	0.102	43.43	12	9.82	0.632	40.3	239.5	238.5	184.8	183.5	2.37	4.59
其他 Others								623.2	706.8	716.3	819.0	718.0	9.28	1.31
合计 Total	(23411)22972	436	0.003	76.31	41	50.21	0.153	3525.9	7896.2	8198.4	10827.4	7729.9	100.00	3.07

\* 此表中宋朝历年的人口基本数据引自文献[17] \* The data in the Song dynasty is quoted from reference document [17]

候、居住的地区和环境的影响,每一种姓氏相当于Y染色体上某个遗传座位上的一个中性等位基因。Karlín-McGregor的中性等位基因分布的理论可以分析姓氏在群体中分布的特征<sup>[18]</sup>。在具有 $k$ 个等位基因的 $N$ 个单倍体的群体中,每一个体携带有一个等位基因,在我们的研究中“等位基因”是姓氏。在一个群体内,姓氏的数目是固定的。群体内姓氏种类和每种姓氏频率,因群体迁移并与周围人群的融合而发生变化,这类变化相当于遗传等位基因频率的变化,或者称为基因流动。携带 $k$ 个体的姓氏数目的期望值可以根据以下公式获得<sup>[19,20]</sup>:

$$E(S_k) = \alpha(1 - v)^k / k \quad (1)$$

$$N = \alpha(1 - v) / v \quad (2)$$

$$S = -\alpha \log v \quad (3)$$

$$S / N = [v / (1 - v)] \log v \quad (4)$$

在以上公式中, $N$ 表示人群的样本大小, $S$ 表示人群中的姓氏种类数, $\alpha$ 和 $v$ 分别表示群体内姓氏种类丰度(surname abundance)和人群迁移率(migration rate)的相关参数, $E(S_k)$ 表示携带 $k$ 个体的姓氏数目的期望值。 $\alpha$ 和 $v$ 是分析姓氏频率分布的两个重要参数。从同一个群体中或者同一个省区内,随机获得的不同样本的 $\alpha$ 值是一个常数,与样本的大小( $N$ )不存在函数关系,但与姓氏种类的多少( $S$ )有函数关系,图1表示宋朝16个省区的 $\alpha$ 值与姓氏种类 $S$ 的关系,其相关系数为 $r = 0.78$ ,具有显著的相关。 $v$ 表示人群的迁移率, $v$ 值的大小直接反映了人群迁移程度。

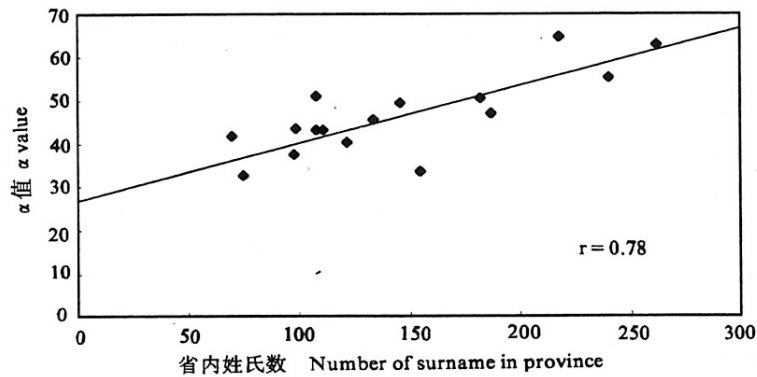


图1  $\alpha$  值与姓氏数的相关

Fig.1 The relativity between  $\alpha$  value and number of surname

### 3 结果和讨论

#### 3.1 宋朝中国人姓氏分布和人群迁移

各省姓氏分布的观察值和期望值的吻合程度见表1。其中山西省和湖北省的 $\chi^2$ 值有显著差异,其余各省的姓氏分布吻合 Karlín-McGregor理论的中性等位基因分布的模式。甘肃、山西、河北、湖北、湖南、广东和广西等7省的迁移的相关参数 $v$ 值均很高,表明在这些省区人群的迁移十分频繁。在宋朝河北、山西和甘肃等北方省区不属于宋朝的版图,而在唐朝时汉人已经逐渐失去了对这些省区的控制。唐宋时期北部省区前后由西夏、辽、金

和元蒙古等北方民族政权统治,西南和西北地区则由吐蕃和回鹘等民族支配,这些地区一直受到战争的蹂躏,加上北方地区地形环境的特点,人群迁移十分频繁,外来民族与当地人群的混居也十分普遍。随着战争的间隙及战后的生产的恢复和人民的安居乐业,这些地区外来民族的汉化进程加快。从汉晋、南北朝、隋唐以来,北方地区少数民族的姓氏的汉化一直没有间断过<sup>[21,22]</sup>,而且,姓氏的进化一直伴随着民族的融合和汉化进程。在宋朝的320年间,湖北、湖南、广东和广西等省区的人口增长非常之快,其增长速度比东南省区还要快,这与人群迁移的相关系数 $\nu$ 值和表1中1210年(南宋)和980年(北宋初期)之间的人口比率所反映的现象基本相符。湖北的人口比率值较低,而迁移的相关参数 $\nu$ 值又很高,这与湖北在宋朝320年中人口激增锐减有关,说明人群迁入和迁出的频繁程度。由于北方民族的不断强大和南移,中原地区的大批居民向南部和东南省区迁移,到北宋末期,中国人口比重在历史上第一次出现了南方大于北方的局面。从此结束了我国北方人口多于南方的历史,政治重心和经济中心也开始由北方移向了南方。

### 3.2 树状聚类分析

图2是宋朝16个省区人群根据姓氏遗传距离数值 $D_1$ (the negative logarithms of relative isonymies,表2),采取平均连接法获得的树状聚类结果。姓氏遗传距离数值 $D_1$ 计算公式:

$$I = \sum_{i=1}^s p_i q_i \quad (5)$$

$$D_1 = -\log \left( I \sqrt{\left( \sum_{i=1}^s p_i^2 \sum_{i=1}^s q_i^2 \right)} \right) \quad (6)$$

在以上公式中, $I$ 表示群体 $p$ 和群体 $q$ 之间的同姓率(isonymy); $p_i$ 和 $q_i$ 分别表示 $p$ 和 $q$ 群体的第 $i$ 种姓氏的频率; $s$ 表示姓氏种类数;全部16个省区间的姓氏遗传距离数值列于表2。

聚类图(图2)十分明显地把16个省区划分为2大区域:南部的广东和福建(A区)与其余14个省区(B区)。这两区域间的遗传距离十分大,提示了南北两大人群的异源性,其明显的地理区分地域应在中国南部福建省与江西省交界的武夷山和广东省与湖南省交界的南岭。B区又明显的划分为C区和D区。C区包括浙江、江苏、江西和安徽4省,形成了东南地区人群,其聚类的顺序也很合理,反映了古吴越民族与中原地区南下人群混合的历史痕迹。D区进一步可划分成E区和F区。E区含湖南和广西二省区,在宋朝该地区是一块很荒凉的“蛮”地,从表1的宋朝320年的人口增长率分析,人口增长十分迅速,这主要包含

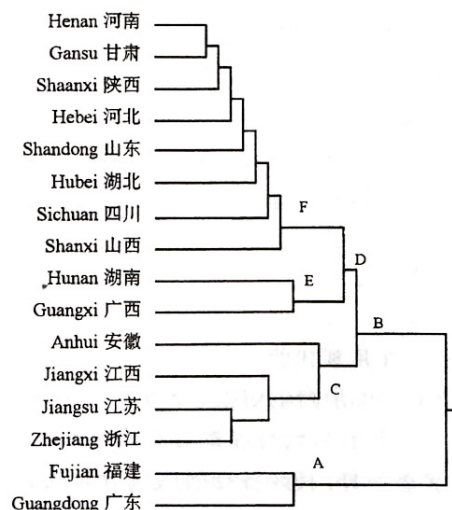


图2 宋朝16省区姓氏频率分布聚类图  
Fig.2 The surname phylogenetic dendrogram of sixteen provinces in the Song dynasty



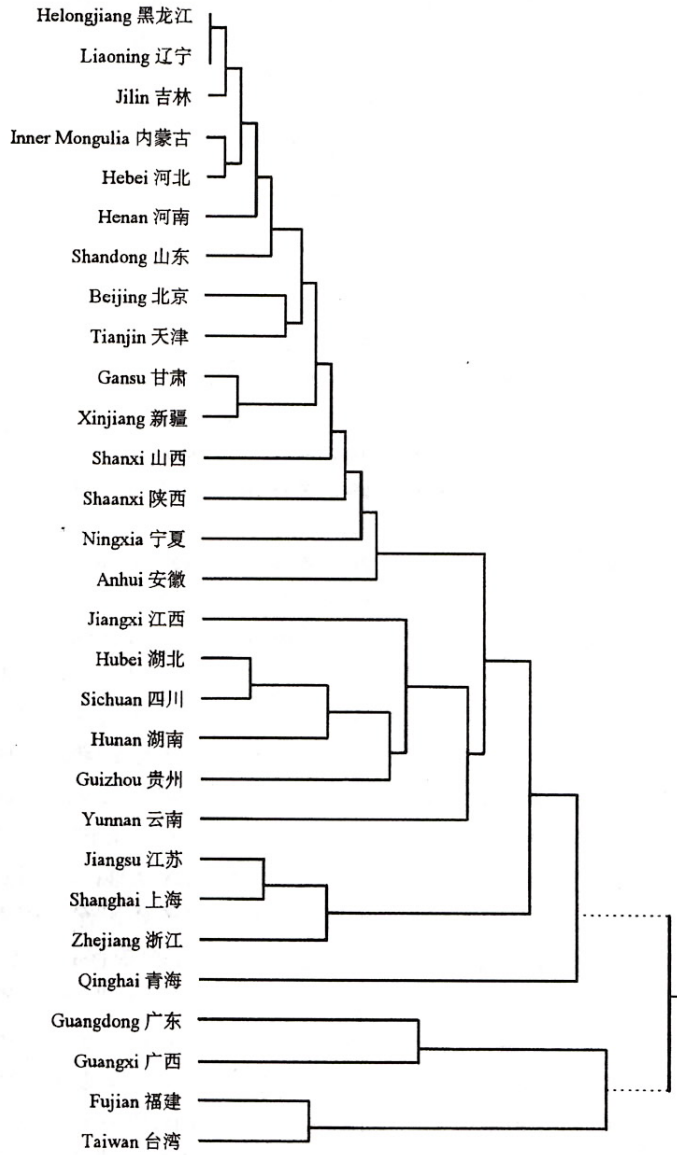


图3 现代29省区的姓氏频率分布聚类图

Fig.3 The surname phylogenetic dendrogram of 29 provinces in the present

方政权频繁交战的边界省份的样本数相对较少。在经济文化发达和社会稳定的省区,做官的人士多,出现在文献上的人士也多;反之则少。这就是造成了在文献中各省区之间的样本数不随机平衡的主要原因。为了了解宋朝时期全中国的姓氏分布的情况,我们依据宋朝4次人口的统计值(表1),以年份间隔为加权数计算宋朝的全国和各省平均人口数和比重(表1),从而推算出宋朝时期全国样本中出现的姓氏的分布频率,可以认为这436个姓氏分布的频率基本上反映了1000年前中国人的主要姓氏分布的情况。表3列出的是宋

表3 宋朝和现代100个常见姓氏的分布

Table 3 The distribution of 100 common surnames in the Song dynasty and the present

宋朝 Song dynasty				现代 The present		宋朝 Song dynasty				现代 The present	
排 序 No.	姓氏 Surname	观察值 Obs.(%)	推算值 Exp.(%)	姓氏 Surname	观察值 Obs.(%)	排 序 No.	姓氏 Surname	观察值 Obs.(%)	推算值 Exp.(%)	姓氏 Surname	观察值 Obs.(%)
1	王 Wang	6.67	7.61	李 Li	7.94	51	魏 Wei	0.43	0.44	余 Yu	0.41
2	李 Li	6.08	7.19	王 Wang	7.41	52	罗 Luo	0.43	0.44	潘 Pan	0.41
3	张 Zhang	5.24	6.24	张 Zhang	7.07	53	叶 Ye	0.80	0.44	杜 Du	0.40
4	赵 Zhao	5.07	5.64	刘 Liu	5.38	54	沈 Shen	0.67	0.43	戴 Dai	0.39
5	刘 Liu	4.49	4.67	陈 Chen	4.53	55	孟 Meng	0.26	0.42	夏 Xia	0.39
6	陈 Chen	5.42	3.87	杨 Yang	3.08	56	姚 Yao	0.41	0.41	钟 Zhong	0.38
7	杨 Yang	2.33	2.70	赵 Zhao	2.29	57	傅 Fu	0.44	0.38	汪 Wang	0.38
8	吴 Wu	2.39	2.26	黄 Huang	2.23	58	丁 Ding	0.41	0.37	田 Tian	0.38
9	黄 Huang	2.19	1.56	周 Zhou	2.12	59	章 Zhang	0.51	0.36	任 Ren	0.38
10	朱 Zhu	1.70	1.43	吴 Wu	2.05	60	萧 Xiao	0.34	0.35	姜 Jiang	0.37
11	孙 Sun	1.33	1.35	徐 Xu	1.66	61	蒋 Jiang	0.38	0.34	范 Fan	0.36
12	郭 Guo	1.14	1.30	孙 Sun	1.54	62	卢 Lu	0.36	0.34	方 Fang	0.36
13	胡 Hu	1.38	1.27	胡 Hu	1.31	63	陆 Lu	0.52	0.33	石 Shi	0.35
14	吕 Lu	1.25	1.21	朱 Zhu	1.26	64	袁 Yuan	0.40	0.33	姚 Yao	0.35
15	高 Gao	1.01	1.15	高 Gao	1.21	65	江 Jiang	0.47	0.32	谭 Tang	0.34
16	宋 Song	1.00	1.13	林 Lin	1.18	66	晁 Zhao	0.21	0.32	廖 Liao	0.34
17	徐 Xu	1.42	1.08	何 He	1.17	67	谭 Tan	0.16	0.30	邹 Zou	0.33
18	程 Cheng	0.93	1.08	郭 Guo	1.15	68	邵 Shao	0.34	0.29	熊 Xiong	0.32
19	林 Lin	2.03	0.98	马 Ma	1.05	69	欧阳OuYang	0.24	0.29	金 Jin	0.32
20	郑 Zheng	1.48	0.97	罗 Luo	0.86	70	孔 Kong	0.18	0.28	陆 Lu	0.31
21	范 Fan	0.85	0.90	梁 Liang	0.84	71	俞 Yu	0.38	0.27	郝 Hao	0.30
22	何 He	1.00	0.85	宋 Song	0.81	72	尹 Yin	0.24	0.27	孔 Kong	0.29
23	韩 Han	0.75	0.82	郑 Zheng	0.78	73	廖 Liao	0.21	0.26	白 Bei	0.29
24	曹 Cao	0.72	0.82	谢 Xie	0.72	74	阎 Yan	0.19	0.25	崔 Cui	0.28
25	马 Ma	0.63	0.80	韩 Han	0.68	75	洪 Hong	0.35	0.25	康 Kang	0.28
26	许 Xu	0.84	0.74	唐 Tang	0.65	76	夏 Xia	0.26	0.24	毛 Mao	0.27
27	田 Tian	0.51	0.74	冯 Feng	0.64	77	雷 Lei	0.15	0.24	邱 Qiu	0.27
28	冯 Feng	0.66	0.72	于 Yu	0.62	78	葛 Ge	0.24	0.23	秦 Qin	0.26
29	杜 Du	0.64	0.70	董 Dong	0.61	79	文 Wen	0.19	0.23	江 Jiang	0.26
30	周 Zhou	0.64	0.69	萧 Xiao	0.59	80	柳 Liu	0.15	0.22	史 Shi	0.25
31	曾 Zeng	0.78	0.66	程 Cheng	0.57	81	陶 Tao	0.15	0.22	顾 Gu	0.25
32	汪 Wang	0.60	0.60	曹 Cao	0.57	82	毛 Mao	0.29	0.21	侯 Hou	0.25
33	苏 Su	0.62	0.60	袁 Yuan	0.54	83	丘 Qiu	0.27	0.21	邵 Shao	0.24
34	董 Dong	0.50	0.60	邓 Deng	0.54	84	龚 Gong	0.21	0.20	孟 Meng	0.24
35	方 Fang	0.98	0.58	许 Xu	0.54	85	康 Kang	0.18	0.19	龙 Long	0.24
36	蔡 Cai	0.73	0.57	傅 Fu	0.51	86	蒲 Pu	0.14	0.19	万 Wan	0.24
37	梁 Liang	0.52	0.57	沈 Shen	0.50	87	邢 Xing	0.17	0.19	段 Duan	0.23
38	石 Shi	0.51	0.52	曾 Zeng	0.50	88	郝 Hao	0.12	0.19	雷 Lei	0.23
39	谢 Xie	0.66	0.51	彭 Peng	0.49	89	庞 Pang	0.10	0.19	钱 Qian	0.22
40	贾 Jia	0.40	0.51	吕 Lu	0.47	90	安 An	0.15	0.19	汤 Tang	0.19
41	薛 Xue	0.51	0.50	苏 Su	0.47	91	裴 Pei	0.13	0.19	尹 Yin	0.19
42	彭 Peng	0.44	0.50	卢 Lu	0.47	92	折 Zhe	0.09	0.19	黎 Li	0.18
43	崔 Cui	0.39	0.48	蒋 Jiang	0.47	93	施 Shi	0.21	0.18	易 Yi	0.19
44	唐 Tang	0.43	0.48	蔡 Cai	0.46	94	游 You	0.19	0.18	常 Chang	0.18
45	潘 Pan	0.63	0.47	贾 Jia	0.42	95	金 Jin	0.21	0.17	武 Wu	0.18
46	邓 Deng	0.39	0.46	丁 Ding	0.42	96	邹 Zou	0.20	0.17	乔 Qiao	0.18
47	任 Ren	0.42	0.46	魏 Wei	0.42	97	汤 Tang	0.20	0.16	贺 He	0.18
48	史 Shi	0.54	0.45	薛 Xue	0.42	98	虞 Yu	0.17	0.16	赖 Lai	0.18
49	钱 Qian	0.61	0.45	叶 Ye	0.42	99	严 Yan	0.15	0.16	龚 Gong	0.17
50	侯 Hou	0.33	0.45	阎 Yan	0.41	100	钟 Zhong	0.14	0.16	文 Wen	0.17



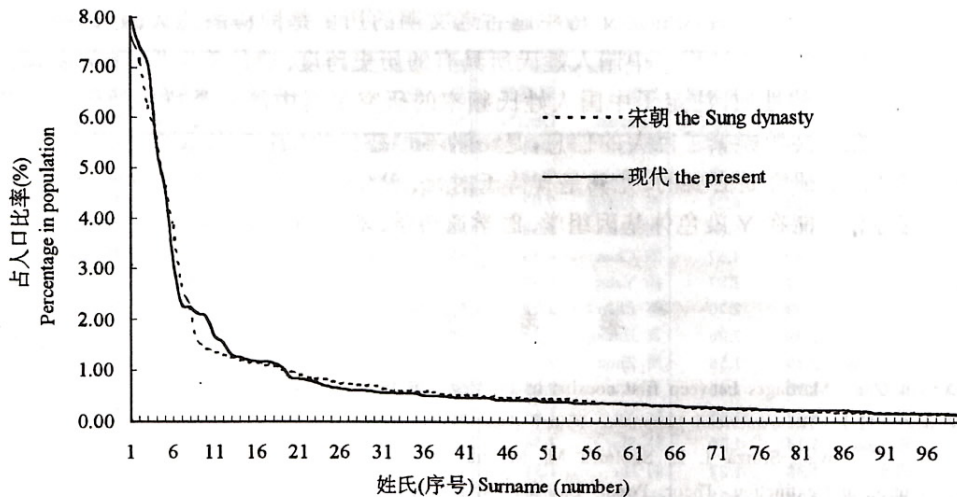


图4 宋朝和现代100个常见姓分布曲线

Fig.4 The curves for the distribution of 100 common surname in the Song dynasty and the present

朝和现代(1982年)的各100个常见姓氏的分布频率。现代(1982年)的100个常见姓氏的数据引自《中华姓氏大辞典》的“后记”<sup>[23]</sup>,与中国人现代姓氏频率分布一文<sup>[15]</sup>所用的是同一数据。

宋朝和现代各100个常见姓氏的分布频率十分一致。两个时期的100个常见姓氏分别覆盖了总人口的86.17%和87.19%,其中最常见姓氏(占总人口1%以上)分别为18个和19个,这些最常见的姓氏分别代表了总人口的52.73%和55.63%。中国现代的7大姓氏为李、王、张、刘、陈、杨和赵,宋朝的7大姓氏为王、李、张、赵、刘、陈和杨,姓氏完全一样,但在人口数的排列上略有区别。宋朝和现代各100个常见姓氏的分布频率曲线非常一致(见图4),说明中国人姓氏的传递在这1000年间是十分稳定的。

## 5 总结

宋朝(公元960~1179年)中国人姓氏的分布十分符合中性等位基因分布的特征,反映了1000年以来人群的结构变化和姓氏“基因”的稳定性和可靠性。宋朝的姓氏频率分布和姓氏遗传距离分析结果表明,在中国1000年前的宋朝已经形成了南北两大区域的人群,从姓氏遗传信息上明显地显示出南北两大区域人群结构的区别,提示南北两大区域人群的不同源性,其明显的地域分界线是在南方的武夷山和南岭这条东西走向的地带。中国人姓氏的历史至少已有4000年以上,主要起源于中国西北和中原地区的民族,融合并同化了各个时期周边各民族的姓氏,形成了中国几千年来一直在使用的汉字姓氏。根据1000年前宋朝的中国人姓氏分布和对人群迁移、混居和融合情况的分析,结果基本上反映了当时的民族联合和进化的历史,也说明了中原地区的人群是组成中华民族的主体,它融合了南北各个时期的民族成分,尤其是北方地区的民族成分,形成了互相依存不可分割的当今的中华民族整体。现代人群分布与宋朝人群分布具有很高相似度,它们之间的区别点正是在这1000年期间人群进一步迁移与分化的结果。

宋朝频率较高的姓氏分布与现代的姓氏分布的惊人一致,进一步说明了中国人姓氏的遗传稳定性。中国人姓氏的历史与中国古老文明的历史是同样的悠久,姓氏一直伴随着整个中华民族的进化过程。中国人姓氏所具有的历史跨度、遗传稳定性、种类多样和中性等位基因分布的性质,注定了中国人姓氏频率的研究是对中国人类群体遗传学研究的一种重要补充。姓氏凝聚了巨大的信息,是一种不可替代的“遗传学”资源。因此,对中国各个时期姓氏的研究是必要的,尤其是有待于对元、明、清3个时期的姓氏分布的研究。这一研究也将有可能在Y染色体基因组学、医学遗传学、地方病和遗传流行病等研究中发挥作用。

### 参 考 文 献

- 1 Darwin G H. Marriages between first cousins in England and their effects. *J. Stat. Soc.*, 1875, 38: 153~184
- 2 Mencken H L. *The American Language*. Knopf, New York. 1936
- 3 Yasuda N, Cavalli-Sforza L L, Skolnick M, Moroni A. The evolution of surnames: An analysis of their distribution and extinction. *Theor. Popul. Biol.*, 1974, 5: 123~142
- 4 Fisher R A, Vaughan J. Surnames and blood groups. *Nature*, 1939, 144: 1047~1048
- 5 Crow J F, Mange A P. Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eug. Quart.*, 1965, 12: 199~203
- 6 Lasker G W. A coefficient of relationship by isonymy: A method for estimating the genetic relationship between populations. *Human Biol.*, 1977, 49: 489~493
- 7 Zei G, Guglielmino C R, Siri E, Moroni A, Cavalli-Sforza L L. Surnames as neutral alleles: Observations in Sardinia. *Human Biol.*, 1983, 55: 357~365
- 8 Roberts D F, Rawling C P. Secular trends in genetic structure: An isonymic analysis of northumberland parish records. *Ann. Hum. Biol.*, 1974, 1: 393~410
- 9 Lasker G W. *Surnames and genetic structure*. Cambridge University Press, Cambridge, 1985
- 10 Kamizaki M. Frequency of isonymous marriages. *Seibutsu Tokei-gaku Zasshi*, 1954, 2: 292~298 (in Japanese)
- 11 Yasuda N. Studies of isonymy and inbreeding in Japan. *Human Biol.*, 1983, 55: 263~276
- 12 Yasuda N, Saitou N. Random isonymy and surname distribution in Japan. *Biology and Society.*, 1984, 1: 75~84
- 13 Chen K H, Cavalli-Sforza L L. Surnames in Taiwan: Interpretations based on geography and history. *Human Biol.*, 1983, 55: 367~374
- 14 袁义达. 中国汉族常见姓氏分布. 载: 陈原主编, 现代汉语定量分析. 上海教育出版社, 1989, 226~232
- 15 Du R F, Yuan Y D, Hwang J, Mountain J, Cavalli-Sforza L L. Chinese surnames and the genetic differences between north and south China. *J. Chinese Linguistics*, 1992, monograph series no. 5
- 16 昌彼得, 王德毅, 程元敏, 侯俊德. 宋人传记资料索引. 台湾鼎文书局, 1974
- 17 赵文林, 谢淑君. 中国人口史. 北京: 人民出版社, 1988
- 18 Karlin S, McGregor, J. The number of mutant forms maintained in a population. *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, 1967, 4: 415~438
- 19 Zei G, Guglielmino Matessi, R, Siri E, Moroni A, Cavalli-Sforza L L. Surnames in Sardinia I. Fit of frequency distributions for neutral alleles and genetic population structure. *Ann. Hum. Genet.*, 1983, 47: 329~352
- 20 Piazza. A, Rendine S, Zei G, Moroni A, Cavalli-Sforza L L. Migration rate of human populations from surname distribution. *Nature*, 1987, 22: 714~716
- 21 姚薇元. 北朝胡姓考. 北京: 科学出版社, 1958

- 22 陈连庆. 中国古代少数民族姓氏研究. 长春: 吉林文史出版社, 1993  
23 袁义达等. 中华姓氏大辞典. 北京: 教育科学出版社, 1996, 1213~1215

## The Study of the Distribution of Chinese Surnames and the Diversity of Genetic Population Structure in the Song Dynasty

YUAN Yi-Da JIN Feng ZHANG Cheng

(*Institute of Genetics Chinese Academy of Sciences Beijing 100101*)

Naruya Saitou

(*National Institute of Genetics Mishima, Japan*)

### Abstract

Surname in China, as a cultural genetic factor, is a very huge and remarkable human population study resource. The transmission feature of surname has been proved to be similar to that of the genetic markers on Y chromosome of males. It, therefore, has been applied in the population genetic studies since 1965's. In the present study, we have analyzed a group of Chinese surname data of the Song dynasty (AD. 960~1179). The distribution of the present surname data is corresponded with the Karlin-McGregor's neutral allele model. The surname abundance  $\alpha$  and the migration rate parameter  $\nu$  in the Song dynasty in 16 provincial regions reflected degrees of migration and admixture at that time. Surname phylogenetic dendrogram analysis demonstrated that the Han Chinese had been segregated into North and South two parts in Song dynasty of 1000 years ago. Wuyi and Nanling Mountains divided Northern and Southern Han populations into the two major parts of Han Chinese. The results also showed that the surname distribution of Song dynasty 1000 years ago was quite similar to the present model. The 100 high frequency surnames in Song dynasty were well corresponded with the current surname distribution. This phenomenon suggests that Chinese surnames have been steadily transmitted from ancient to the present. The present study improved that the Chinese surname is a very important genetic resource for the studies of structure of the Chinese populations, migration, and affinities among different populations.

**Key words** Cultural genetic factor, Surname frequencies, Population structure, Population genetics

Received March 4, 1998, revision received June 15, 1998

This project was supported by National Natural Science Foundation of China (No. 39870406)