

# 基于 STR 基因频率探究我国 32 个 行政区域汉族亚群的遗传特征\*

张 蒙<sup>1</sup>, 饶健安<sup>1</sup>, 赵艳超<sup>1</sup>, 窦浩宇<sup>1</sup>, 徐 暘<sup>2</sup>, 何 森<sup>1</sup>

(1. 中山大学生命科学学院, 广东 广州 510275;

2. 南方医科大学生物技术系, 广东 广州 510515)

**摘 要:** 探究我国不同行政区域汉族亚群间的分子遗传学关系一直是广受学术界关注的问题。短串联重复序列 (STR) 常应用于分子遗传学研究。综合采用了聚类分析、主成分分析和 MCOA 分析等统计学方法, 对我国 32 个行政区域汉族亚群的 9 个常见 STR 基因座 (D8S1179、D21S11、D7S820、D3S1358、D13S317、vWA、D18S51、D5S818、FGA) 的等位基因频率数据进行了系统计算分析, 初步探究了我国汉族亚群间的分子遗传关系、空间分布特征及分布格局的成因。研究发现, 以长江为界, 汉族可清晰地划分为南方汉族和北方汉族两大群体。北方汉族群体中, 山东、天津与其它汉族亚群的遗传距离较大; 南方汉族群体中, 香港、海南和广西汉族亚群亲缘关系密切, 但与南方其它汉族亚群遗传距离较大。厦门汉族亚群与北方汉族亚群的亲缘关系相对较近。主成分二维散点图从整体上体现出我国汉族亚群广分布、小聚集的空间分布格局。MCOA 分析发现, 形成我国汉族亚群目前分布格局的 3 个主要原因是: 长江的地理隔离、历史上的洪涝灾害、战争或人祸引发的人口迁徙等。

**关键词:** STR; MCOA; 汉族; *R*

中图分类号: Q348 文献标志码: A 文章编号: 0529-6579 (2014) 01-0106-07

## Genetic Characteristics of Han Subpopulations of 32 Administrative Area in China Based on Allele Frequencies of Short Tandem Repeat Loci

ZHANG Meng<sup>1</sup>, RAO Jianan<sup>1</sup>, ZHAO Yanchao<sup>1</sup>, DOU Haoyu<sup>1</sup>, XU Yang<sup>2</sup>, HE Miao<sup>1</sup>

(1. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China;

2. Department of Biotechnology Southern Medical University, Guangzhou 510515, China)

**Abstract:** Exploring the molecular genetics relationships between Han subpopulations in different administrative regions of China are always noticed by academics. Short tandem repeat (STR) has been often used in molecular genetics studies. The cluster analysis, principal component analysis and MCOA analysis had been used in this paper, we focused on the nine common STR loci (D8S1179, D21S11, D7S820, D3S1358, D13S317, vWA, D18S51, D5S818, FGA) allele frequency data of Han subpopulations from 32 administrative regions in China, in order to initially explore the molecular genetics relationships, the spatial distribution characteristics and distribution pattern of the Chinese Han subpopulations. The studies revealed that the Han nationality could be segregated into two groups which are the Southern and the Northern Han by the Yangtze River. In the Northern areas of China, the Han subpopulations from Shandong and Tianjin have large genetic distance with others; In the Southern areas, the genetic relationships within the Han subpopulations of Hong Kong, Hainan and Guangxi are much closed and greatly different from the others. The Han subpopulation of Xiamen is comparatively closed to Northern Han. Principal component scatter plot reflects that the spatial distribution pattern of Han subpopulations is wide distribution and gathered tightly in small areas in China. The three main factors contributing to the distribu-

\* 收稿日期: 2013-07-05

基金项目: 国家自然科学基金重大研究计划培育资助项目 (91130009); 广东省科技计划重大专项资助项目 (2003A3080503)

作者简介: 张蒙 (1990 年生), 男; 研究方向: 生物信息学; 通讯作者: 何森; E-mail: lsshem@mail.sysu.edu.cn

tion characteristics of Han subpopulations in China are geographic isolation by the Yangtze River, floods in history and the population migration caused by wars or disasters using MCOA analysis method.

**Key words:** STR; MCOA; Han nationality; *R*

汉族是目前世界上人数最多的民族群体。汉族起源于公元前 21 世纪至公元前 8 世纪居住于黄河流域中上游的华夏部落。随着历史的变迁，部分汉族人群逐渐南迁。在迁徙过程中，汉族不断与其他民族通婚，基因相互交融<sup>[1-2]</sup>，逐渐形成当前的分布格局。多年来，有关我国不同地区汉族亚群分子遗传学关系的研究受到国内外学者的广泛关注。

作为第二代遗传标记，短串联重复序列 (Short Tandem Repeats, STR) 通常由长度为 2~6 bp 的核心序列串联重复组成。目前，与线粒体 DNA 以及 Y 染色体位点一样，STR 已被广泛用于遗传图谱绘制、法医学个体识别<sup>[3]</sup>；同时，STR 数据也常应用于分子遗传学研究，例如，计算遗传距离等。

我国各民族间亲缘关系的相关研究已获得了丰硕成果。杜若甫等<sup>[4]</sup>在研究相关民族 STR 基因座等位基因频率数据时发现，我国汉族与少数民族分别归属于南方蒙古人种或北方蒙古人种，并且，各地汉族与当地少数民族亲缘关系相对较近。高雅等<sup>[5]</sup>对我国 19 个不同地区汉族亚群的 9 个 STR 基因座等位基因频率数据分析发现，不同汉族亚群等位基因存在多样性特点。Chu J 等<sup>[6]</sup>利用 30 个常染色体 STR 基因座数据分析了我国 28 个民族群体的遗传结构，发现我国南北汉族亚群间的遗传结构存在显著差异。Zhang H G 等<sup>[7]</sup>利用人类指纹图谱对我国少数民族间的亲缘关系开展了研究，发现我国少数民族之间存在明显的南北差异。但是，迄今为止，针对我国各地汉族亚群分子遗传关系的研究尚未涵盖全国所有行政区域。

我们较完整地收集了公开途径可以获取的我国现有 32 个行政区域汉族亚群的相关 STR 基因座等位基因频率数据，利用聚类分析、主成分分析和 MCOA 分析方法，尝试系统性探究我国各地汉族亚群间的分子遗传关系、空间分布特征以及分布格局成因。

## 1 数据与方法

### 1.1 数据来源

本研究采用的 STR 数据均来源于国内外权威学术期刊或专业数据库。数据筛选和验证的标准为：① STR 基因座等位基因频率之和为 1；② 符

合 Hardy-Weinberg 平衡；③ 样本采集数均大于 100 例 (澳门和深圳除外)。研究的 32 个行政区域包括：北京、天津、河北、内蒙古、辽宁、吉林、上海、江苏、浙江、安徽、福建、江西、山东、厦门、青岛、河南、湖北、湖南、广东、广西、海南、深圳、重庆、四川、云南、陕西、甘肃、青海、新疆、香港、澳门和台湾等。通过公开渠道均可获取这些行政区域汉族亚群的 D8S1179、D21S11、D7S820、D3S1358、D13S317、vWA、D18S51、D5S818 和 FGA 等 9 个 STR 基因座等位基因频率数据。

### 1.2 计算方法

1.2.1 聚类分析 聚类分析是鉴别人类族群间亲缘关系的有效方法之一<sup>[8]</sup>。本文采用奈氏 (Nei's) 标准计算遗传距离<sup>[9]</sup>，公式如下：

$$D_a = \ln\left(\frac{\sum \sum P_{1mi}P_{2mi}}{\sqrt{\sum \sum P_{1mi}^2 \times \sum \sum P_{2mi}^2}}\right)$$

其中， $m$  表示位点数量， $i$  表示第  $m$  位点的第  $i$  个等位基因， $P_{nmi}$  表示群体  $n$  中 ( $n=1, 2$ ) 第  $m$  位点第  $i$  个等位基因的频率。

基于奈氏距离计算结果，我们进一步采用邻接法 (Neibor-Joining, NJ) 构建聚类树。

1.2.2 MCOA 分析 MCOA 是基于主成分分析的一种多变量分析方法。该方法可揭示单个基因座之间的共有属性。该方法主要通过建立一个参考结构，使研究者采用图形方式，可比对单个基因座的结构和其参考结构间的差异<sup>[10-11]</sup>。Chessel 等<sup>[12]</sup>首先将该方法应用于生态学问题的研究。

MCOA 法建立参考结构的最优化标准：

$$\sum_{k=1}^K \omega_k \text{cov}^2(I_k, V^r) = \sum_{i=1}^K \omega_k \text{var}(I_k) \text{var}(V^r) \text{corr}^2(I_k, V^r)$$

上式中， $I_k$  表示在一个基因座  $k$  的频率数据在主成分分析中群体  $r^{\text{th}}$  的得分； $V^r$  表示  $r^{\text{th}}$  的参考得分。

本文采用 *R* 统计软件 *ade4* 子程序，对构建的数据集进行了 MCOA 分析<sup>[11,13]</sup>。通过计算 9 个 STR 基因座的前 3 个特征值，可以得到各地区汉族亚群在这 3 个主成分的得分，并绘制主成分二维散点图及 MCOA 分析得分行政区域分布图<sup>[14]</sup>。

## 2 结果与分析

### 2.1 聚类树构建

利用聚类分析, 我们构建了 32 个地区汉族亚群的聚类树 (见图 1)。分析初步发现, 不同地区汉族亚群可聚为 4 支, 包含两个大支和两个小支。在两个大支中, 一支是由香港、海南、广西、广东、澳门、台湾、四川、湖南、重庆、湖北、深圳、浙江、云南、江西等地区汉族亚群构成的南方族群分支; 另一支是由天津、江苏、吉林、北京、新疆、内蒙古、河北、山东、甘肃、辽宁等地区汉族亚群构成的北方族群分支; 在两个小支中, 一支是由安徽、河南、上海等地区的汉族亚群构成的族群分支, 另一支是由厦门、福建、青岛、陕西等地区汉族亚群构成的族群分支。分析同时发现, 山东、天津、厦门、澳门、深圳等地区汉族亚群与其所在族群分支中其它地区汉族亚群的遗传距离较远。

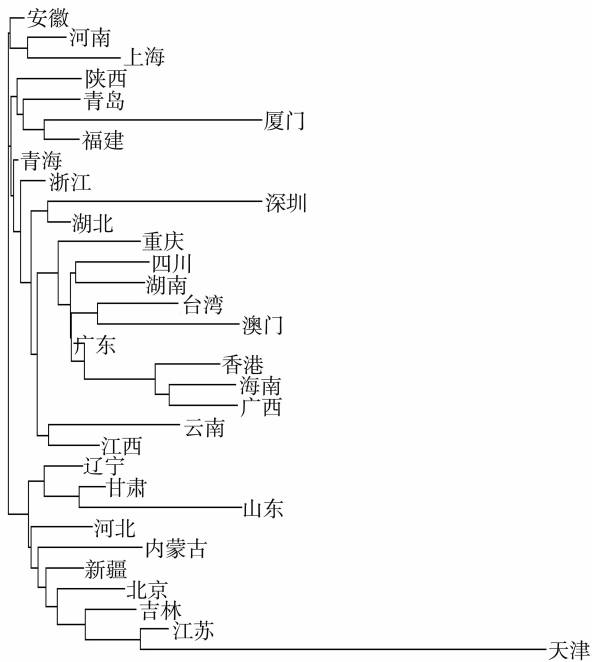


图 1 我国 32 个行政区域汉族亚群聚类树

Fig. 1 The clustering tree of Han subpopulations of 32 administrative areas in China

### 2.2 MCOA 分析

2.2.1 汉族 9 个基因座的特征值 分别计算我国汉族 9 个基因座的前 3 个特征值, 结果显示, 基因座 D3S1358、D13S317、D7S820、vWA、D8S1179、D5S818 的前 3 个特征值的累计贡献率均超过 80% (见表 1), 说明在汉族的各亚群中, 这 6 个基因座的差异性比较大; 基因座 FGA、D18S51 和 D21S11

的前 3 个特征值累计贡献率相对较低, 即在汉族的各亚群中, 这 3 个基因座的差异性较小。

表 1 我国汉族 9 个基因座的前 3 个特征值  
Table 1 The first three eigenvalues of nine STRs of Han nationality in China

STR 基因座	特征值 1	特征值 2	特征值 3	总和
D3S1358	0.569 90	0.253 12	0.116 90	0.939 92
D13S317	0.651 46	0.161 27	0.094 95	0.907 67
D7S820	0.578 04	0.188 25	0.140 58	0.906 87
vWA	0.470 95	0.211 35	0.162 62	0.844 91
D8S1179	0.501 86	0.206 91	0.125 15	0.833 92
D5S818	0.443 11	0.257 22	0.103 44	0.803 77
FGA	0.364 23	0.210 48	0.166 37	0.741 08
D18S51	0.433 20	0.175 54	0.114 39	0.723 13
D21S11	0.297 46	0.239 58	0.176 17	0.713 21

以前 3 个特征值累计贡献率超过 80% 的基因座等位基因频率数据为基础, 利用 *R* 统计软件 *ade4* 子程序, 可以计算出不同行政区域在这 3 个主成分的得分 (见表 2)。

在表 2 中, 广西、海南、香港和澳门的第一主成分得分较高, 均大于 1; 天津、山东、河北、辽宁和吉林的第一主成分分值较低, 均小于 -1; 表明在第一主成分中, 这些行政区域汉族亚群之间的遗传差异十分显著。同理, 在第二主成分中, 厦门、香港、山东、福建、甘肃、广西、湖南、深圳和澳门地区的汉族亚群之间的遗传差异十分显著。在第三主成分中, 江苏、吉林、深圳、湖南、云南、四川和山东地区的汉族亚群之间的遗传差异十分显著。

系统比较 3 个主成分的总得分发现, 山东、天津、澳门、厦门和江苏地区汉族亚群与其它区域汉族亚群间的遗传差异大, 说明这些区域汉族亚群在遗传结构上具有相对的独立性。

2.2.2 主成分二维散点图 依据各行政区域汉族亚群在 3 个主成分上的得分, 可绘制主成分二维散点图 (见图 2)。

在主成分二维散点图中, 山东、天津、澳门和江苏地区为离群值, 表明这些行政区域汉族亚群与其它区域汉族亚群的亲缘关系较远。图 2A 显示, 基于第一主成分和第二主成分, 可将相关汉族亚群分为四大支系: 香港、广西和海南的汉族亚群聚为一类; 四川、湖北、湖南和云南等南方汉族亚群聚为一类; 甘肃、河北、内蒙古、北京等北方汉族亚群聚为一类; 厦门、福建、台湾等区域汉族亚群聚

表 2 我国 32 个行政区域汉族亚群的 3 个主成分得分 1)

Table 2 The scores of 32 administrative areas in the first three principal components of Han subpopulations in China

行政区域	样本数	PC1 得分	PC2 得分	PC3 得分
广西	308	2.275 95	1.068 01	0.288 20
海南	200	2.001 47	0.066 89	-0.163 88
香港	284	1.961 99	1.627 28	0.476 83
澳门	87	1.160 09	-2.673 78	0.310 45
湖南	270	0.953 20	-1.253 03	-1.666 27
四川	260	0.882 27	-0.246 29	-2.032 40
广东	1 000	0.819 31	-0.012 81	-0.516 33
重庆	225	0.644 71	0.067 85	0.290 03
江西	200	0.601 76	0.145 36	0.197 76
湖北	305	0.493 38	-0.304 54	0.387 32
深圳	69	0.346 71	-1.291 37	1.202 00
云南	109	0.244 71	-1.099 24	-1.825 86
台湾	189	0.174 14	0.923 41	0.835 63
浙江	598	0.042 66	0.165 46	0.042 20
上海	106	0.018 01	0.168 30	-0.126 77
福建	364	-0.064 97	1.206 72	1.084 21
厦门	122	-0.094 63	2.747 47	0.145 30
青海	206	-0.229 05	-0.261 74	0.082 53
安徽	1 000	-0.264 69	-0.131 49	0.736 81
陕西	203	-0.423 77	-0.161 26	0.706 79
河南	220	-0.500 33	0.014 93	0.078 85
北京	236	-0.711 37	-0.120 51	-0.891 51
青岛	387	-0.732 07	-0.227 01	0.115 14
甘肃	215	-0.790 09	1.152 45	0.816 67
江苏	120	-0.879 40	-1.039 83	2.354 20
新疆	200	-0.909 47	0.659 55	-0.723 46
内蒙古	129	-0.951 93	-0.355 46	-1.269 40
辽宁	316	-1.001 15	-0.855 53	0.095 66
吉林	200	-1.018 69	0.474 24	1.245 05
河北	345	-1.176 81	-0.033 90	0.115 09
山东	100	-1.625 56	1.501 16	-2.409 10
天津	260	-1.692 97	-1.005 80	0.127 99

1) PC1、PC2 和 PC3 分别为第一、第二和第三主成分

为一类。图 2B 显示, 基于第二主成分和第三主成分, 可将汉族亚群划分为福建、广西和香港等东南地区族群, 其中, 甘肃和吉林为特例; 四川、云南、内蒙古和新疆等西南地区族群, 其中, 广东和北京为特例; 湖北、河南、重庆、天津和上海等中东部地区族群, 其中, 辽宁、海南和天津为特例。图 2C 显示, 基于第一主成分和第三主成分, 在南方汉族中, 海南、四川和云南等汉族亚群聚为一类, 广西、香港、澳门、湖北和江西等汉族亚群聚为一类; 在北方汉族中, 北京、内蒙古和新疆等汉

族亚群聚为一类; 安徽、陕西、甘肃和吉林等汉族亚群可聚为一类; 而上海、浙江、辽宁、厦门、河南、青岛和河北等汉族亚群聚为一类。

主成分散点图从整体上体现出了我国汉族亚群广分布、小聚集的空间分布格局。比较分析发现, 主成分二维散点图的聚集效果与聚类树吻合度很高, 可用于交叉验证结果。

2.2.3 主成分区域分布图 依据表 2 各个区域汉族亚群的 3 个主成分得分, 采用 MCOA 分析法, 分别制作我国 32 个行政区域汉族亚群的 3 个主成分得分区域分布图 (见图 3)。结合历史文献记载, 分析 3 个主成分得分区域分布图的汉族亚群空间分布特征, 可初步推测我国目前行政区划条件下汉族亚群分布现状的可能成因。

图 3A 显示, 汉族清晰地被长江分隔为北方汉族和南方汉族。但是, 位于南方的福建汉族亚群与北方汉族的遗传距离较近。第一主成分得分区域分布图可能揭示了长江对汉族亚群的地理分隔作用。

图 3B 显示, 在北方汉族中, 新疆、甘肃、河南、山东和吉林等汉族亚群与其他北方汉族亚群的遗传距离较远; 在南方汉族中, 重庆、广西、海南、江西、福建、浙江和上海等汉族亚群与其他南方汉族亚群的遗传距离较远。南、北方的这些特殊的汉族亚群主要位于黄河或长江流域的上、下游, 或位于水系丰富、河水泛滥及自然灾害频发的地区。史料记载, 宋代黄河和长江的水灾十分严重, 有记录的水灾达到 465 次<sup>[15]</sup>; 其中, 黄河水灾尤为突出, 仅北宋时期, 发生在河北境内的决溢次数就达 32 次之多<sup>[16]</sup>。因此, 推测第二主成分得分区域分布图可能揭示了洪涝等自然灾害对人口迁移的影响。

图 3C 显示, 在北方汉族中, 新疆、内蒙古、山东和北京等汉族亚群与其他北方汉族亚群遗传距离较大; 在南方汉族中, 四川、云南、湖南、广东和海南等汉族亚群与其他南方汉族亚群的遗传距离较大, 而与中原地区汉族亚群之间的遗传距离相对较小。这可能与历史上朝代更迭、战乱与人祸所导致的人口迁徙和民族融合有关。例如, 公元 316 年, 西晋政权被匈奴、鲜卑、羯、羌、氐即 (五胡) 肢解后, 中原汉人无以安身, 远离故土, 迁徙到南越之地。因此, 推测第三主成分得分区域分布图可能揭示了历史上战乱和人祸对人口迁移的影响。



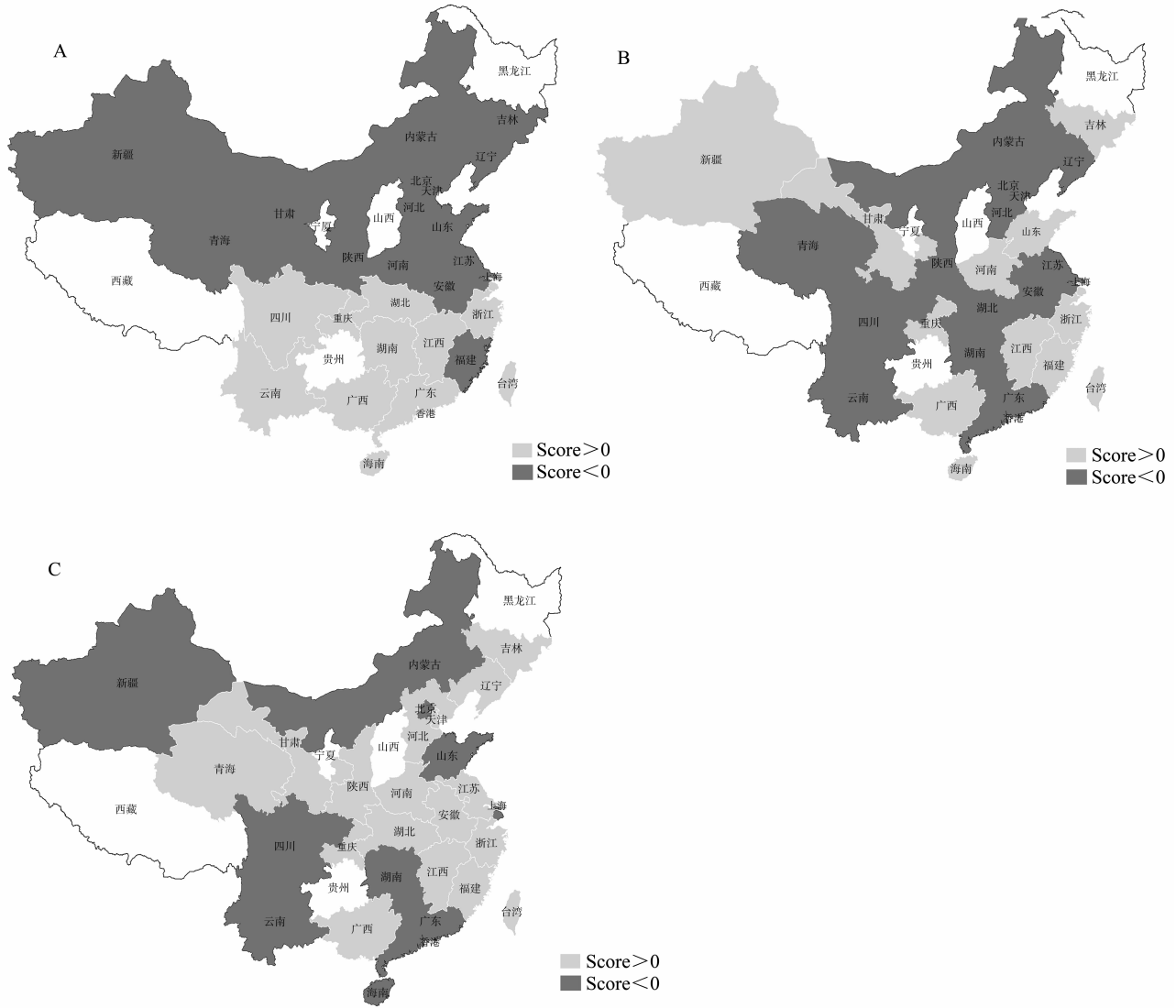


图 3 MCOA 分析得分区域分布图

Fig. 3 Spatial distribution of the MCOA score in different regions

(A: 第一主成分得分区域分布图; B: 第二主成分得分区域分布图; C: 第三主成分得分区域分布图)

### 3 结论与讨论

本文基于我国 32 个行政区域汉族亚群的 9 个 STR 基因座等位基因频率数据, 首次综合利用聚类分析、主成分分析和 MCOA 分析等统计学方法, 从基因水平上系统性探究了我国各地汉族亚群的分子遗传关系、空间分布特征以及分布格局成因。

分析表明, 我国汉族可显著地划分为南方汉族和北方汉族两大群体, 支持了我国民族有南北之分的理论<sup>[7]</sup>。北方汉族可划分为两大支, 说明北方地区汉族亚群之间的基因交流较频繁。南方汉族群体可划分为 3~4 个大支, 并且, 各区域汉族亚群可聚为不同的小支, 例如, 广西、香港和海南地区小支等, 说明南方汉族与当地的少数民族之间的可

能存在较频繁的基因交流; 该结论与相关研究结论一致<sup>[17-18]</sup>。在南方地区, 福建和厦门汉族亚群与其它南方汉族亚群的遗传距离较远, 但是, 与北方汉族的亚群能够聚在一起, 该结论支持了高雅<sup>[5]</sup>等人的观点。山东、天津、澳门和江苏等汉族亚群与其它区域汉族亚群间的遗传距离较远, 说明这些区域汉族亚群在遗传结构上保持了相对的独立性。

本文研究发现, 我国汉族亚群中差异性比较大的 6 个 STR 基因座分别是 D8S1179、D13S317、D5S818、FGA、D3S1358 和 D18S51。主成分二维散点图从整体上体现出我国汉族亚群广分布、小聚集的空间分布格局特征。通常聚类分析会丢失基因频率数据中的部分信息, 无疑会对分析汉族亚群空间分布特征的成因产生不利影响。但是, MCOA 分

析方法可以最大限度地保留数据中的关联信息。通过计算各区域汉族亚群 6 个 STR 基因座的 3 个主成分得分, 我们有效绘制了 MCOA 分析得分区域分布图。分析初步发现, 导致我国目前汉族亚群空间分布特征及分布格局的 3 个主要原因是: 长江的地理隔离、历史上的洪涝灾害和改朝换代的战争或人祸引发的人口迁徙等。

在人类漫长的历史发展进程中, 我国汉族经历了无数次的迁移、融合和分化, 从而形成当前不同区域亚群独特的分子遗传结构特征和地域文化特征, 如广东地区的客家文化等。由于我国目前各行政区域有关汉族亚群 STR 基因座数据积累的不一致性, 同时, 个别区域诸如黑龙江、山西、贵州、宁夏、西藏等汉族亚群的 STR 基因座数据还存在缺失现象; 均会对本文研究结论产生不利影响。本文结论将为开展我国汉族分子遗传学进一步的研究工作提供有益的借鉴。

#### 参考文献:

- [1] WEN B, LI H, LU D, et al. Genetic evidence supports demic diffusion of Han culture[J]. *Nature*, 2004, 431(7006): 302-305.
- [2] 于冰筠, 胡丽莉. 华南地区汉族人群 MDR1 基因单核苷酸多态性研究[J]. *中山大学学报: 自然科学版*, 2011, 50(2): 93-98.
- [3] 任凌雁, 何燕, 张婷, 等. 贵州 9 个少数民族线粒体 DNA Region V 及 Y 染色体 DYS287 位点多态性研究[J]. *中山大学学报: 自然科学版*, 2013, 52(4): 121-124.
- [4] 杜若甫, 肖春杰. 用 38 个基因座的基因频率计算中国人群间遗传距离[J]. *中国科学: C 辑*, 1998, 28(1): 83-89.
- [5] 高雅, 阎春霞, 金天博, 等. 中国 19 个不同地区汉族亚群的分子遗传学关系[J]. *西安交通大学学报: 医学版*, 2005, 26(2): 111-114.
- [6] CHU J, HUANG W, KUANG S, et al. Genetic relationship of populations in China[J]. *Proceedings of the National Academy of Sciences*, 1998, 95(20): 11763-11768.
- [7] ZHANG H G, CHEN Y F, DING M, et al. Dermato-glyphics from all Chinese ethnic groups reveal geographic patterning[J]. *PLoS One*, 2010, 5(1): e83-87.
- [8] LANGFELDER P, ZHANG BHORVATH S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R[J]. *Bioinformatics*, 2008, 24(5): 719-720.
- [9] NEI M. Estimation of average heterozygosity and genetic distance from a small number of individuals[J]. *Genetics*, 1978, 89(3): 583-590.
- [10] LALO D, JOMBART T, DUFOUR A B, et al. Consensus genetic structuring and typological value of markers using multiple co-inertia analysis[J]. *Genetics Selection Evolution*, 2007, 39(5): 545-567.
- [11] SCHILLING D A. Strategic facility planning: The analysis of options[J]. *Decision Sciences*, 1982, 13(1): 1-14.
- [12] CHESSEL DHANAFI M. Analyses de la co-inertie de K nuages de points[J]. *Revue de statistique appliquée*, 1996, 44(2): 35-60.
- [13] JOMBART T. adegenet: a R package for the multivariate analysis of genetic markers[J]. *Bioinformatics*, 2008, 24(11): 1403-1405.
- [14] 樊文平, 石忆邵, 车建仁, 等. 基于 GIS 与空间句法的道路网结构对城市商业中心布局的影响[J]. *中山大学学报: 自然科学版*, 2011, 50(3): 112-117.
- [15] 袁冬梅. 对宋代黄河水灾原因的分析[J]. *乐山师范学院学报*, 2004, 19(9): 83-99.
- [16] 郭志安, 李京龙. 略论北宋黄河水患下的河北水环境变迁[J]. *保定学院学报*, 2010, 23(002): 60-63.
- [17] 宋土生, 黄辰. 应用网络 STR 生物信息对中国 14 个人群遗传距离的研究[J]. *西安交通大学学报: 医学版*, 2003, 24(001): 1-4.
- [18] YANG L, ZHAO Y, LIU C, et al. Allele frequencies of 15 STRs in five ethnic groups (Han, Gelao, Jing, Shui and Zhuang) in South China[J]. *Forensic Science International Genetics*, 2013, 7(1): e9-14.