

# 苗瑶族群父系遗传结构初探

文波<sup>1</sup> 张有隽<sup>2</sup> 李峰<sup>1</sup> 高扬<sup>1</sup> 李辉<sup>1</sup> 宋秀峰<sup>1</sup> 徐晓云<sup>1</sup> 肖春杰<sup>3</sup> 黄薇<sup>4</sup> 金力<sup>1</sup>

(1.复旦大学现代人类学研究中心, 上海 200433; 2.广西民族学院民族学人类学研究所, 南宁 530063;

3.云南大学生物系, 昆明 650091; 4.国家基因组南方研究中心, 上海 201203)

## 摘 要

近年来分子人类学在技术和方法上的进步以及大量新遗传标记的发现, 极大的推动了人类起源以及人群的迁徙、分化和融合的研究。特别是 Y 染色体非重组区域单核苷酸多态系统(Y-SNP)的引入, 使得从父系遗传的角度探讨群体遗传结构成为可能。虽然一些文献比较系统的讨论了东亚人群的遗传关系及群体结构, 但对苗瑶系统人群的研究还远不够深入。本研究利用 PCR-RFLP 技术, 对来自于 23 个苗瑶系人群的 684 个样品进行了 13 个 Y-SNP 的基因分型。在东亚及亚太地区存在 17 种 Y 染色体单倍型, 我们在苗瑶人群中观察到 9 种。其中带有 M122 突变的三种单倍型 (H6-H8) 的频率为 45%, 带有 M95 突变的 2 种单倍型 (H11-12) 的频率为 21.5%。前者在汉藏系统人群中的频率最高 (47.1%), 而后者则广泛分布于侗台系统民族 (平均频率 35.7%)。这一频率分布格局显示苗瑶系统人群的父系遗传结构具有双重特质。单倍型频率的主成分分析支持以上的观察: 苗瑶系统人群在主成分分布图上连接了汉藏和侗台人群, 并有部分群体与其交错分布。该结果提示苗瑶系统人群在父系遗传结构上与汉藏和侗台两个系统均有比较密切的联系, 这和历史记载相一致。

分子数据的方差分析 (AMOVA) 表明, 苗瑶与藏缅系统人群之间的差异最小 (1.2%), 汉族次之 (2.3%), 而与侗台 (4.5%) 和阿尔泰 (5.3%) 则差异较大 ( $p$  值均小于 0.01)。当单独考察苗族和瑶族, 苗族与藏缅的差异不显著 (1023 次非参数列阵检验,  $p=0.13$ ), 可能与三苗集团与炎黄集团的接触有关。在苗瑶族群内部, 布努与苗族群体的差异相当小 (-0.9%,  $p=0.58$ ), 而与瑶族群体的差异则相对较大 (2.7%,  $p=0.07$ ), 我们的遗传学观察与语言学的分类是一致的。

表 1. 各族群的 M122、M95 和 M7 频率 (%)

| 族群  | 样本量  | M122 | M95  | M7  |
|-----|------|------|------|-----|
| 阿尔泰 | 388  | 25.9 | 0.5  | 0.0 |
| 侗台  | 520  | 26.9 | 35.7 | 0.2 |
| 汉藏  | 1309 | 47.1 | 9.8  | 1.8 |
| 苗瑶  | 684  | 45.0 | 21.5 | 6.4 |

带有 M7 突变的 H7 单倍型在大部分苗瑶人群中均有分布 (平均频率 6%), 而在其它人群中则极其罕见, 显示出强烈的群体特异性。我们进一步对 H7 的个体进行了 5 个 Y 染色体上短重复序列 (Y-STR) 的分型。利用 Y-STR 单倍型构建的网络表明苗瑶系统的 H7 位于中心, 而客家人和土家族等人群的 H7 则位于旁支。提示苗瑶人群的 H7 更为古老, 是发散中心。用 5 个 Y-STR 重复数的方差估计 M7 发生的年代至少在 10300-11800 年之前, 这一估计可为苗瑶系统祖先群体的分化时间提供参照。

**关键词:** 苗瑶, Y 染色体, 遗传结构 .