

Y 染色体、线粒体 DNA 多态性与 云南宁蒗摩梭人的族源研究*

文波^{①②③} 石宏^{①②} 任玲^{①②} 奚慧峰^{①③} 李开源^{①②}
张文逸^② 宿兵^⑤ 司世恒^⑥ 金力^{①③⑤**} 肖春杰^{①②**}

①云南大学-复旦大学人类群体遗传学联合实验室, 昆明 650091; ②云南大学云南省生物资源保护与利用重点实验室及人类遗传学研究中心, 昆明 650091; ③复旦大学生命科学学院遗传工程国家重点实验室及现代人类学研究中心, 上海 200433; ④中国科学院昆明动物研究所细胞与分子进化重点实验室, 昆明 650031; ⑤Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45267, USA; ⑥云南省宁蒗县永宁中心医院, 宁蒗 674300

摘要 居住于云南宁蒗县泸沽湖畔的摩梭人是中国大陆目前惟一的母系社会群体. 在第一次民族识别中被定为纳西族, 但是大部分摩梭人认为自己与纳西族有本质的区别, 后来通过云南省人民代表大会决议, 称之为“摩梭人”. 将遗传学的方法引入摩梭人的族源研究, 对摩梭人以及居住于云南的纳西族、藏族、白族、彝族和普米族 6 个群体线粒体 DNA 第一高变区、Y 染色体上的 13 个 SNP 和 8 个 STR 位点进行了基因分型, 结果显示摩梭人相对缺乏南方民族特异的 Y 染色体类型, 而 mtDNA 具有南北双重特征. 主成分分析和分子系统学分析进一步表明, 摩梭人的父系遗传结构与云南藏族最接近, 而母系遗传结构最接近丽江纳西族, 提示其父系和母系基因库具有不同的来源. 摩梭人特殊的母系社会结构可能是导致其母系、父系遗传结构存在明显差异的原因之一. 该结果为摩梭人的族源研究提供了遗传学方面的线索.

关键词 摩梭人 族源 线粒体 DNA Y 染色体 多态性

摩梭人自称“纳日”, 是中国目前仅存的母系社会, 是研究母系社会结构的重要模式群体^[1]. 他们主要居住于泸沽湖周围的云南省宁蒗县和四川省盐源县, 1990 年人口普查时, 宁蒗县境内共有 15096 个摩梭人. 关于摩梭人的起源, 一般认为是中国西北的古代氏羌族群的后裔, 与氏羌南迁的一支“牦牛羌”关系密切^[2,3]. 汉语文献中最早提及与“摩梭”相似的称谓在东汉时期, 称为“摩沙”^[2](《华阳国志·蜀志》“…县在郡西, 渡泸水, 宾刚徼, 曰摩沙夷, 有盐池, 汉末, 夷皆固之…”). 在 20 世纪 50 年代的第一次民族识别中, 宁蒗摩梭人被识别为纳西族, 但是他们自认为与纳西族存在本质区别, 应为单一民族, 如若不然, 也宁愿被定为藏族的一个支系^[4]. 根据他们的意愿和要求, 1986 年云南省人民代表大会通过决议, 称这一部分“纳日”为摩梭人. 尽管如此, 摩梭人的族源及他们与其他兄弟民族之间的遗传关系仍然悬而未决.

遗传学指标的最大优点是稳定性, 并按照一定的频率发生突变. 遗传标记可为研究群体的传

2002-10-10 收稿, 2003-04-14 收修改稿

* “863”项目(批准号: 2001AA22C041)和国家自然科学基金资助项目(批准号: 30160037)

** 联系人, E-mail: cjxiao@public.km.yn.cn; Jin_li@fudan.edu

承、迁徙和融合的历史提供线索,因此被广泛地应用到人类群体的进化和遗传关系研究中^[5~8]。根据其传代行为,遗传标记可划分为 3 个系统:母系遗传的线粒体 DNA(mtDNA)、父系遗传的 Y 染色体 DNA(Y-DNA) 以及双系遗传的常染色体和 X 染色体 DNA。常染色体 DNA 虽然包含最多的变异,但是由于在传代过程中会发生重组,容易干扰对祖先历史的追溯,大大减弱了其在进化研究中的应用。线粒体 DNA 不发生重组,可形成完整的单体型,还具有突变率高、容易检测等特点,是在进化研究中使用最广泛的遗传系统,其 D 环区的第一高变区域(HVS1)成为进化研究的一种经典标记。Y 染色体的非拟常染色体区域同样不发生重组,但由于该区域多态性标记的相对缺乏,导致 Y 染色体的研究一度停滞不前^[9]。近年来,随着越来越多的 Y 染色体上多态性标记被发现,使得 Y 染色体在人类进化研究中日益显示出其优越性^[9,10]。Y 染色体上的单核苷酸多态性标记(SNPs)组成的单体型可以认为是一种“单位点多等位基因”的遗传标记,大部分突变是按照一定的时间顺序依次发生,具有很强的群体特异性,能清晰地记录群体的历史,是研究群体父系祖先迁徙史最理想的工具^[11,12]。Su 等人^[13]用 19 个 Y-SNP 构建了 17 种单体型,其中有些单体型在东亚人群中具有很强的特异性,通过这些标记的研究,在东亚和太平洋地区人群的迁徙和遗传关系方面获得了很多进展^[13~19]。短重复序列(STR)是 Y 染色体上另一类重要的遗传标记,它们具有非常高的突变率,其精密程度可以用来进行法医学鉴定,因此非常适合于研究短期进化事件和较近人群之间的遗传关系^[20,21]。

摩梭人具有悠久的历史和丰富的民俗文化,在其发展过程中先后融汇了藏文化、纳西文化、普米文化和彝文化的成分^[4]。本文对摩梭人以及 5 个对照群体(普米族、纳西族、藏族、彝族和白族)的线粒体 DNA 第一高变区的核苷酸序列, Y 染色体上的 13 个 SNP, 8 个 STR 位点进行了基因分型,建立 Y 染色体单体型和 mtDNA 单体型组,并对以上数据进行了主成分分析和分子系统学分析,通过综合利用多种遗传标记,以期从母系遗传和父系遗传两个角度为摩梭人的族源研究提供线索。

1 材料与方法

1.1 样本

所研究的 6 个群体均采自云南省,在知情同意原则下,随机选取身体健康、相互之间无亲源关系,3 代以内无外族通婚且居住当地的个体,其中摩梭人和普米族采自宁蒗县,白族、彝族、纳西族和藏族分别采自大理、双柏、丽江和中甸,每一个体抽取外周血 5 mL, ACD 抗凝,用常规酚-氯仿法抽提 DNA,各群体样本量见表 1。在样本量大小的确定上参照了同类型研究,以避免由于样本量不足导致的实验误差。对所有男性样本均进行了 Y 染色体分析。随机选取了

表 1 Y-SNP 单体型频率

| 群体 | 样本量 | 特征突变/单体型频率 | | | | | | | | | | | | |
|------|-----|------------|-----------|-----------|-----------|----------|------------|----------|------------|------------|-------------|------------|------------|------------|
| | | M130 H1 | YAP H2 | M15 H3 | M89 H4 | M9 H5 | M122 H6 | M7 H7 | M134 H8 | M119 H9 | M110 H10 | M95 H11 | M88 H12 | M45 H13 |
| 白族 | 50 | 6.0 | 2.0 | 4.0 | | 18.0 | 8.0 | 4.0 | 34.0 | 6.0 | | 12.0 | 2.0 | 4.0 |
| 彝族 | 50 | 8.0 | 2.0 | 10.0 | 38.0 | 16.0 | 10.0 | | 10.0 | 2.0 | | 4.0 | | |
| 摩梭人 | 47 | 6.4 | 23.4 | 4.3 | 2.1 | 31.9 | 4.3 | 6.4 | 12.8 | 8.5 | | | | |
| 纳西族 | 40 | 2.5 | 37.5 | | | 10.0 | | | 2.5 | | | 47.5 | | |
| 普米族 | 47 | 2.1 | 70.2 | 2.1 | 4.3 | 6.4 | 2.1 | | 6.4 | 4.3 | | | | 2.1 |
| 云南藏族 | 50 | 4.0 | 28.0 | 8.0 | | 12.0 | 6.0 | 4.0 | 34.0 | | | 4.0 | | |

大部分样品进行 mtDNA 分析, 由于男性同时拥有 Y 染色体和 mtDNA, 所以分析的样品大部分为男性(在纳西族中随机选取 5 名女性).

1.2 DNA 分析

线粒体 DNA 第一高变区用引物 L15996 和 H16401 进行 PCR 扩增^[22], PCR 产物与 Resin (Promega)混合后经 96 孔纯化柱纯化, Big-Dye 末端荧光标记试剂盒(PE 公司)进行测序反应, 测序反应产物经乙醇沉淀后在 ABI377 自动测序仪(PE 公司)上进行核苷酸序列测定, 每一个体用以上 2 条引物双向测序, 以保证数据的可靠性. 用 ABI Sequence Analysis 3.3 软件分析并读取序列. 梭梭人序列已提交 GenBank 数据库(登记号待定).

Su 等人^[13]用 19 个 Y-SNP 构建了 17 种单体型, 这里选择了在东亚地区具有丰富多态性的 YAP, M7, M9, M15, M45, M88, M89, M95, M110, M119, M122, M130, M134 共 13 个位点进行了基因分型. YAP 和 M15 两个位点的 PCR 产物分别通过 3% 和 4% 琼脂糖凝胶电泳直接观察结果. 用 PCR-RFLP 方法对其他 11 个 Y-SNP 进行了分型, 若多态不能形成酶切位点, 则在设计引物时引入. 每个位点的引物序列、限制性内切酶、PCR 扩增条件和多态的判定如表 2 所示, 3% 琼脂糖凝胶电泳观察结果. 引物在上海生工生物工程公司合成, 限制性内切酶购自 New England Biolab.

表 2 Y-SNP 分型位点 PCR 引物序列、限制性内切酶及多态判定

| 位点 | 退火温度/°C | PCR 引物序列(5'~3') | 内切酶 | 多态判定(电泳片段大小/bp) | |
|------|---------|--|----------------|------------------|-----------------|
| YAP | 53 | CAGGGGAAGATAAAGAAATA ACTGCTAAAAGGGGATGGAT | / | 无 Alu (150) | Alu 插入 (400) |
| M15 | 56 | ACA AATCCTGAACAATCGC GTCTGGGAAGAGTAGAGA AAAG | / | 9 bp 插入 (151) | 非插入 (142) |
| M89 | 52 | GAAAGTGGGGCCACAGAAGGA GCAAATCAGGCAAAGTGAGACAT | <i>Nla</i> III | C (79+21) | T (100) |
| M9 | 55 | GAAACGGCCTAAGATGGTTGGAT AAACTGAATCTTTTTCCTCATTTTTG | <i>Bam</i> H I | C (190+20) | G (210) |
| M119 | 55 | AGGTAAATGACTCACCTAAGGAAG GGGTTATTCCAATTCAGCATAACGC | <i>Bst</i> u I | A (161) | C (135+26) |
| M95 | 55 | ATAAGGAAAGACTACCATATTAGCG TTTGAAGCCCCAGTTGTGAG | <i>Hha</i> I | C (178+24) | T (202) |
| M122 | 55 | TAGAAAAGCAATTGAGATACTAATTCA GCGATGCTGATATGCTAGTTCAG | <i>Nla</i> III | C (100+22) | T (122) |
| M134 | 55 | AAGGACCAGGAAAAGTATGATCG TTTGATGATTCTCTTTGGGCTTC | <i>Nla</i> III | 非缺失 (100+22) | 1bp 缺失 (122) |
| M7 | 56 | TGTACCCTTGACCAATGCCTT TTGTAGTTGAGTTACTGTTCTCTTA | <i>Bfa</i> I | C (103+23) | G (126) |
| M130 | 56 | TATCTCTCTCTATTGCAG CCACAAGGGGGAAAAACAC | <i>Bst</i> I I | T (205) | C (162+43) |
| M110 | 55 | AACATTCTCTGTAGACTCACTGG ATTTAGCACTTCTTTTCCCC | <i>Nla</i> III | T (200) | C (88+122) |
| M88 | 55 | TCTTATTCTGCTTCTTCCGC CATGTGATGGTTCAGTAGGTGTGA | <i>Bst</i> u I | A (146) | G (125+21) |
| M45 | 55 | ATTGGCAGTGAAAAATTATAGCTA TGCCTTTGCTACAACCTCTCCTA | <i>Bfa</i> I | G (140+22) | A (162) |

对 DYS19, 388, 389-1, 389-2, 390, 391, 392 和 393 共 8 个 Y 染色体上的 STR 位点进行基因分型, 引物及片段大小信息参见 GDB 数据库(<http://www.gdb.org>), 上游引物分别用 6-FAM(388,

389), NED(390, 392, 393)和 HEX(19, 392)荧光标记. 8 个位点分为 2 个组(Panel)进行多重 PCR 扩增^[21]. 各取 0.5 μL PCR 产物与 0.2 μL ROX-500 荧光分子量内标(PE 公司), 1.8 μL 甲酰胺上样液混合, 95 $^{\circ}\text{C}$ 加热变性 2 min, 冰水骤冷, 在 ABI 377 测序仪上用 4%变性聚丙烯酰胺胶电泳 2 h. 通过 GenescanTM2.0(PE 公司)软件分析 DNA 片段大小, 应用 GenotypeTM1.1 软件(PE 公司)进行基因分型. 在每次电泳中加入同一个已知片段大小的标准样品, 以矫正不同电泳之间的误差.

1.3 数据分析

360 bp (16024~16383)的 mtDNA 第一高变区序列用 ClustalW 软件比对, 参照线粒体剑桥序列^[23], 用 ARLEQUIN 软件^[24]统计单体型频率, 计算群体间的 F_{st} 和净遗传距离 d_A ^[25], F_{st} 的显著性经 3000 次非参数列阵检验, 假设群体之间没有差异. 根据高变区 1 序列的变异位点, 依照 Yao 等人^[26]所述方法推断每一条序列所属的 mtDNA 单体型组, 同一单体型组内部的单体型具有比较近的分子系统关系. Y-SNP 单体型的构建和命名参照以前报道^[13]. 对于 Y-STR 数据, 用 ARLEQUIN 软件计算 R_{st} 距离^[27], 同样经过 3000 次非参数排列检验其显著性. Y-SNP 单体型频率、mtDNA 单体型组频率和遗传距离矩阵的主成分分析, 遗传距离矩阵的多维尺度分析(MDS), 遗传距离之间、主成分系数与单体型(群)频率之间的相关系数计算及显著性检验在 SPSS 软件上进行. 基于遗传距离的邻接法(NJ)系统树的构建在 MEGA2^[28]软件上进行.

Y-SNP 单体型频率的主成分分析中加入了西藏藏族^[15]、云南汉族^[18]、四川汉族^[18]、壮族^[14]、蒙古族^[14]、布依族^[19]及本实验室未发表的维吾尔族数据. mtDNA HVS1 的分析中引用了云南汉族^[26]、广东汉族^[26]、青岛汉族^[26]、新疆汉族^[26]、维吾尔族^[29]、哈萨克族^[29]、土族^[30]、蒙古族^[30]、傣族^[30]、壮族^[30]和佤族^[30]的序列数据.

2 结果

2.1 Y-SNP 单体型和 mtDNA 单体型组频率

表 1 列出了这 6 个群体的 Y-SNP 单体型频率. 在分析的 6 个群体中共发现 10 种单体型, 其中白族、彝族和摩梭人分别观察到 9 种, 云南藏族和普米族发现 8 种, 纳西族仅观察到 5 种. 摩梭人和藏族中频率最高的是 H5, H2 和 H8 三种单体型, 分别占总个体数的 68% 和 74%. 普米族频率最高的是 H2(70.2%), 纳西族的 H11 频率最高(47.5%). 白族和彝族的频率分布比较分散. 值得注意的是, 在纳西族中频率最高的 H11 单体型没有在摩梭人中观察到.

在分析的 239 个个体中, 共观察到 177 种 mtDNA HVS1 单体型, 其中摩梭人与纳西族共享 5 种, 与普米族共享 3 种, 与白族、藏族和彝族分别共享 2 种单体型. 根据高变区的某些特征变异位点, 可以将单体型归入不同的单体型组, 同一单体型组的单倍型具有比较近的系统演化关系. 结果有 87% 的序列能划分到特定的单体型组中(表 3). B, D, F, M7 是比较大的单体型组, 可划分为若干亚群, 表 1 同时统计出了它们的频率总计. 摩梭人主要的单体型组为 B, D, F, 占总个体数的 73.8%, 纳西族的 B, F 频率与摩梭人非常接近. 白族频率最高的是 F 和 G, 而 A, D 是云南藏族、普米族和彝族中主要的单体型组(C 在普米族种的频率也非常高).

2.2 遗传距离

Y-STR 的 R_{st} 矩阵表明(表 4), 摩梭人与云南藏族的遗传距离最近, 其 $R_{st} P$ 值大于 0.05, 不能否定它们之间无遗传差异的假设, 其后依次是大理白族、彝族、纳西族, 云南蒙古族和普米族.

表 3 由 mtDNA 高变区 1 变异位点推断的单体型组频率

| mtDNA 单体型组 | 云南汉族 ^{a)} (43) | 青岛汉族 ^{a)} (50) | 新疆汉族 ^{a)} (47) | 广东汉族 ^{a)} (69) | 白族 (37) | 摩梭人 (46) | 纳西族 (45) | 普米族 (35) | 云南藏族 (36) | 彝族 (40) |
|------------------|----------------------------|----------------------------|----------------------------|----------------------------|------------|-------------|-------------|-------------|--------------|------------|
| A | 4.7 | 4.0 | 10.6 | | 5.4 | 4.3 | 8.9 | 14.3 | 13.9 | 19.5 |
| B* | 2.3 | 2.0 | | | | | | | | |
| B4 | 11.7 | 10.0 | 2.1 | 29.0 | | 13.0 | 17.8 | 2.9 | 5.6 | 12.2 |
| B5* | | | | | 5.4 | | | | | |
| B5a | 4.7 | | 4.3 | | | 17.4 | 6.6 | | 11.1 | |
| B5b | 2.3 | | 2.1 | 1.4 | | | | | | |
| B 总计 | 21.0 | 12.0 | 8.5 | 30.4 | 5.4 | 30.4 | 24.4 | 2.9 | 16.7 | 12.2 |
| C | 4.7 | | 6.4 | | | 13.0 | 8.9 | 22.9 | 8.3 | 2.4 |
| D4k | 9.3 | 26.0 | 19.1 | 10.1 | 10.8 | 13.0 | 6.7 | 17.1 | 19.4 | 14.6 |
| D5* | 2.3 | 3.9 | 2.1 | 5.8 | | | | | | |
| D5a | 2.3 | 6.0 | 4.3 | | | 8.7 | | 5.7 | 5.6 | 2.4 |
| D 总计 | 13.9 | 35.9 | 25.5 | 15.9 | 10.8 | 21.7 | 6.7 | 22.8 | 25.0 | 17.0 |
| F* | 2.3 | | 2.1 | 1.4 | | | | | | |
| F1a | 11.6 | 4.0 | 4.3 | 17.4 | 16.2 | | 17.8 | 2.9 | | 4.9 |
| F1b | 4.7 | 4.0 | 2.1 | 1.4 | 5.4 | 17.4 | 4.4 | 2.9 | | 4.9 |
| F1c | | 2.0 | 2.1 | 1.4 | 2.7 | | | | | |
| F2* | | | | 2.9 | | | | | | |
| F2a | 2.3 | 2.0 | 4.3 | 1.4 | 8.1 | 4.3 | 2.2 | 5.7 | 2.8 | 2.4 |
| F 总计 | 20.9 | 12.0 | 14.9 | 25.9 | 32.4 | 21.7 | 24.4 | 11.5 | 2.8 | 12.2 |
| G2 | | 6.0 | 2.1 | 1.4 | 10.8 | | 2.2 | | | 2.4 |
| M10 | 2.3 | 2.0 | | 2.9 | 8.1 | | 8.9 | 2.9 | 2.8 | 9.8 |
| M7* | 2.3 | | 2.1 | 1.4 | | | | | | |
| M7B | 16.3 | 4.0 | 6.4 | 8.7 | 5.4 | | 2.2 | | | 12.2 |
| M7c | | | 2.1 | 1.4 | | | | | | |
| M7 总计 | 18.6 | 4.0 | 10.6 | 11.5 | 5.4 | | 2.2 | | | 12.2 |
| M8a | | 8.0 | 4.3 | 2.9 | | | | | | |
| M9 | | 4.0 | 4.3 | 0.0 | 5.4 | | | 5.7 | 5.6 | 7.3 |
| N9a | 7.0 | 6.0 | | 1.4 | | | | | | |
| R9a | 2.3 | | 4.3 | 1.4 | | | 2.2 | | | |
| Y | | 2.0 | 2.1 | | | | | | | |
| Z | | | 2.1 | | | | | | 2.8 | |
| 其他 ^{b)} | 4.6 | 4.0 | 4.2 | 5.7 | 16.2 | 8.6 | 11.1 | 17.1 | 25 | 4.9 |

a) 数据引自文献[26], 括号内数字为样本量; b) 包含尚未确定的 M*, N*及 R*单体型组; * 示未确定的单体型组

表 4 Y-STR Rst 遗传距离^{a)}

| | 白族 | 彝族 | 摩梭人 | 纳西族 | 普米族 | 云南藏族 |
|------|-------|-------|-------|-------|-------|-------|
| 白族 | | 0.003 | 0.012 | 0.000 | 0.000 | 0.005 |
| 彝族 | 0.048 | | 0.001 | 0.000 | 0.000 | 0.009 |
| 摩梭人 | 0.043 | 0.076 | | 0.000 | 0.000 | 0.101 |
| 纳西族 | 0.260 | 0.191 | 0.138 | | 0.003 | 0.006 |
| 普米族 | 0.410 | 0.334 | 0.238 | 0.111 | | 0.000 |
| 云南藏族 | 0.065 | 0.057 | 0.021 | 0.084 | 0.193 | |

a) 左下角为 Rst, 右上角是无差异显著性检验 P 值

用 mtDNA 高变区 1 数据得到 Fst 和 d_A 两种遗传距离(表 5), 它们之间有相当好的相关性 ($r = 0.947, P < 0.01$). 总的看来, 各群体之间母系基因的差异小于父系基因. 摩梭人与纳西族

之间的遗传距离远小于与其他人群间的距离. 纳西族、云南汉族、壮族和傣族之间的遗传距离一般都比较小, 有的甚至估计出负值, 其 F_{st} 经检验后通常不能推翻没有差异的假设. 中国西北的土族、维吾尔族和哈萨克族之间的 F_{st} 也显示出类似的格局.

表 5 mtDNA 高变区 1 估计的群体间 F_{st} 和净遗传距离(d_A)^{a)}

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---------|-------|-------|--------------|-------|-------|--------------|---------------|-------|--------------|--------------|-------|--------|-------|-------|
| 1 白族 | | 0.845 | 0.313 | 0.451 | 0.499 | 0.196 | 0.264 | 0.199 | 0.269 | 0.349 | 0.497 | 0.508 | 0.188 | 0.529 |
| 2 摩梭人 | 0.079 | | 0.211 | 0.779 | 0.515 | 0.492 | 0.441 | 1.329 | 0.517 | 0.504 | 1.045 | 0.312 | 0.687 | 1.131 |
| 3 纳西族 | 0.031 | 0.019 | | 0.569 | 0.353 | 0.203 | 0.108 | 0.630 | 0.201 | 0.214 | 0.572 | 0.137 | 0.263 | 0.572 |
| 4 普米族 | 0.052 | 0.071 | 0.054 | | 0.438 | 0.305 | 0.661 | 0.939 | 0.680 | 0.445 | 0.660 | 0.927 | 0.376 | 0.504 |
| 5 云南藏族 | 0.057 | 0.048 | 0.034 | 0.048 | | 0.273 | 0.514 | 1.162 | 0.481 | 0.271 | 0.642 | 0.540 | 0.394 | 0.624 |
| 6 彝族 | 0.022 | 0.046 | 0.020 | 0.033 | 0.030 | | 0.215 | 0.702 | 0.282 | 0.049 | 0.250 | 0.400 | 0.140 | 0.579 |
| 7 云南汉族 | 0.032 | 0.043 | <u>0.011</u> | 0.073 | 0.058 | 0.024 | | 0.486 | -0.019 | 0.145 | 0.545 | -0.012 | 0.245 | 0.685 |
| 8 佤族 | 0.028 | 0.130 | 0.068 | 0.117 | 0.141 | 0.086 | 0.064 | | 0.541 | 0.722 | 0.976 | 0.783 | 0.584 | 0.965 |
| 9 壮族 | 0.035 | 0.058 | 0.025 | 0.082 | 0.060 | 0.036 | <u>-0.002</u> | 0.072 | | 0.217 | 0.573 | 0.019 | 0.229 | 0.580 |
| 10 土族 | 0.045 | 0.050 | 0.022 | 0.055 | 0.034 | <u>0.006</u> | 0.018 | 0.105 | 0.029 | | 0.275 | 0.242 | 0.098 | 0.426 |
| 11 蒙古族 | 0.059 | 0.085 | 0.046 | 0.072 | 0.070 | 0.022 | 0.061 | 0.147 | 0.070 | 0.039 | | 0.766 | 0.211 | 0.605 |
| 12 傣族 | 0.062 | 0.031 | <u>0.014</u> | 0.104 | 0.063 | 0.046 | <u>-0.002</u> | 0.106 | <u>0.003</u> | 0.033 | 0.093 | | 0.380 | 0.874 |
| 13 维吾尔族 | 0.026 | 0.071 | 0.030 | 0.049 | 0.050 | 0.018 | 0.032 | 0.086 | 0.031 | <u>0.014</u> | 0.029 | 0.051 | | 0.195 |
| 14 哈萨克族 | 0.069 | 0.108 | 0.059 | 0.063 | 0.076 | 0.068 | 0.084 | 0.141 | 0.075 | 0.062 | 0.089 | 0.111 | 0.029 | |

a) 对角线左下为 F_{st} 遗传距离, 对角线右上为净遗传距离, 下画线的 $F_{st} P > 0.05$

用 Y-STR 和 mtDNA 估计出的遗传距离之间相关性很差($r = 0.298, P = 0.28$), 提示摩梭人与各群体之间父系和母系遗传关系有比较大的差别.

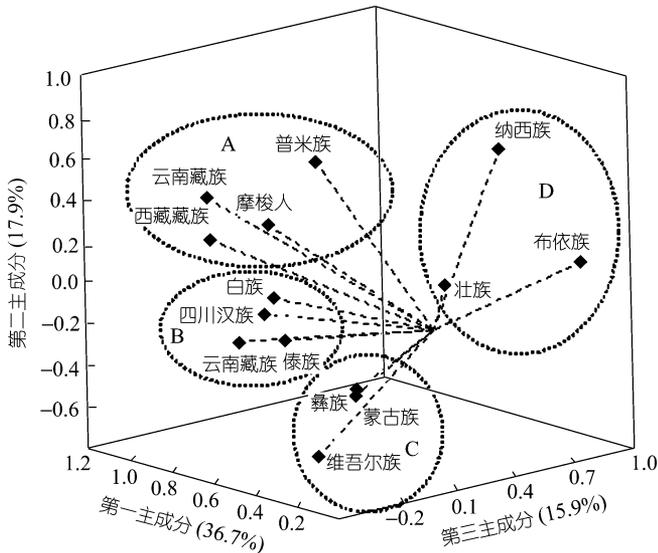


图 1 Y-SNP 单体型频率主成分坐标图(前 3 个主成分共占总体方差的 70.4%)

综合起来, 将摩梭人与藏族聚在一起主要是由于 H2 和 H8 的贡献, 而把纳西族与摩梭人分开的主要原因是他们在 H8 和 H11 频率上的差异.

2.3 主成分分析

图 1 是 Y-SNP 单体型频率主成分分析三维坐标图, 前 3 个主成分解释了 70.3% 的总体方差. 图上的 14 个人群可以分为 4 个聚类: A. 云南藏族、西藏藏族、摩梭人、普米族; B. 白族、傣族、四川汉族和云南汉族; C. 蒙古族、维吾尔族和彝族; D. 壮族、布依族和纳西族. 第二主成分将滇西北的 5 个人群(聚类 A 加上纳西族)与其他人群分开, 相关分析表明, 第二主成分与 H2 的频率极显著相关($r = -0.844, P < 0.01$), 提示将它们分开的主要原因是 H2 频率的差异. 第一、三主成分共同将聚类 A 和 D 分开, 第一主成分

与 H8 极显著相关($r = 0.810, P < 0.01$), 第三主成分与 H11 极显著相关($r = 0.898, P < 0.01$).

在 mtDNA 单体型组的主成分分析中, 把 B, D, F, M7 下的亚群各自合并一起分析, 因为同一单体型组下的亚群之间有比较近的亲缘关系, 若将各亚群作为独立的变量分析, 将消灭亚群之间实际存在的紧密关系而导致分析结果的偏差甚至错误. 图 2 为前两个主成分的坐标图, 占总体方差的 76.5%. 前两个主成分共同将各人群分为 2 个聚类. 两个南方汉族人群与佤族、白族、纳西族共同组成南方人群的聚类, 而两个北方汉族人群其他人群聚在坐标图的左上部分. 相关分析表明, 第一主成分与 F 极显著相关($r = 0.799, P < 0.01$), 与 B 单体型组显著相关($r = 0.654, P < 0.05$), 与 A 显著负相关($r = 0.667, P < 0.05$); 第二主成分与 D 单体型组显著相关($r = 0.721, P < 0.05$), 与 F 极显著负相关($r = -0.876, P < 0.01$), 提示将南北人群分开主要是这 4 种单体型组的贡献. 摩梭人的第一主成分与南方类群靠近, 而第二主成分则更接近北方类群, 其整体位置居于南北人群之间, 但是最靠近纳西族. 结合相关分析的结果, 高频率的 B, F 单体型组频率把摩梭人的第一主成分拉到南方类群, 高频率的 D 单体型组使其第二主成分靠近北方类群.

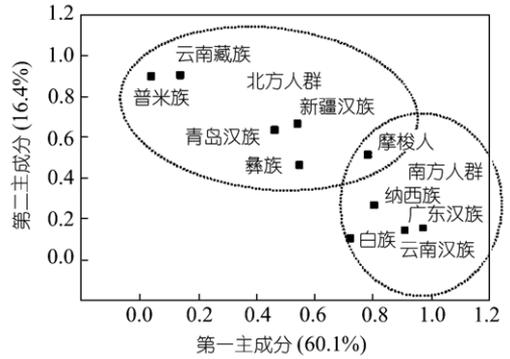


图 2 mtDNA 单体型组频率主成分坐标图

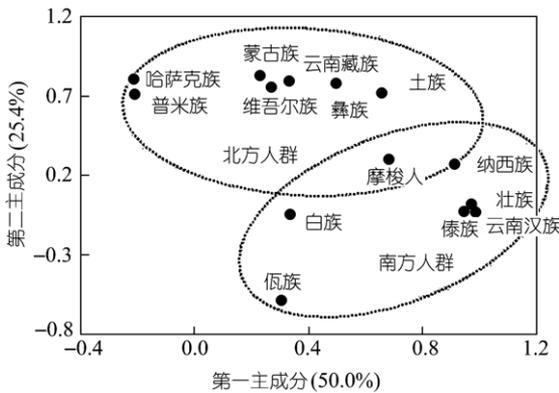


图 3 mtDNA Fst 遗传距离主成分坐标图

用表 5 中的 Fst 矩阵进行主成分分析(图 3), 该坐标图解释了 75.4% 的总体方差. 结果与图 2 非常一致. 图中的第二主成分将南北方人群分得很清楚. 土族、维吾尔族和哈萨克族 3 个典型北方民族与彝族、普米族和云南藏族聚在主成分图的上方, 而其他人群与傣族、壮族、佤族等典型南方民族聚在坐标图的下面, 摩梭人仍然与纳西族的位置最接近, 居于北方与南方类群之间.

2.4 分子系统树

图 4 为 Y-STR Rst 距离矩阵构建的邻接法系统树和多维尺度分析图谱. 两种分析方法的结果完全一致, 摩梭人、云南藏族、彝族和白族形成一个聚类, 纳西族和普米族形成一个聚类, 其中摩梭人与云南藏族最为接近.

基于 mtDNA Fst 距离矩阵的邻接法系统树也与主成分分析的结果相一致. 所有人群分为南北两大支, 土族、维吾尔族、哈萨克族、彝族、普米族和云南藏族形成北支, 其他人群为南支, 摩梭人与纳西族相距最近. 在南支中, 纳西族和摩梭人共同的分支与北支最接近.

3 讨论

从父系遗传角度看, 摩梭人的 Y 染色体类型主要为 H2, H5 和 H8, 同时具有高频率的 YAP+ 突变(H2, H3)和 H8 是藏族的特征, 被认为是藏族双重起源的遗传学证据^[15], YAP+ 为中亚起源, 而 H8 发源于东亚. 摩梭人与云南藏族在 Y 染色体类型上极高的相似性($r = 0.642, P < 0.05$)表明他们在父系遗传结构上非常相似. M95 突变产生的 H11 以及 M119 突变导致的 H9 单

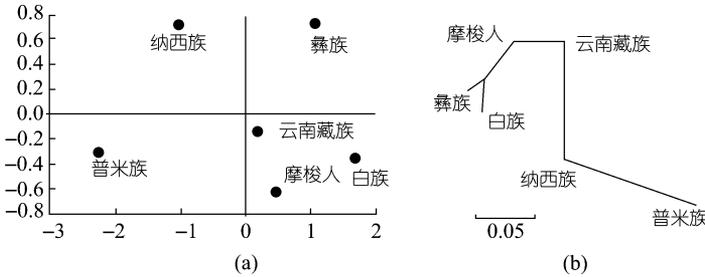


图 4 基于 YSTR R_{st} 距离矩阵的((a))多维尺度分析(MDS)坐标图和无根邻接(NJ)系统树((b))

对 Y-STR 的遗传距离估计和分子系统学分析均支持以上观察。

从母系遗传角度看, B, F, D 单体型组是摩梭人主要的 mtDNA 类型. B, F 单体型组在东亚的频率有从南往北递减的趋势^[26,31], 因此高频率的 B, F 单体型组是南方人群的特征. 摩梭人与纳西族这 2 种单体型组的频率非常相近, 而且它们总的单体型组频率分布也极显著正相关 ($r = 0.827, P < 0.01$), 说明它们在母系遗传结构上有高度相似性. 需要注意的是, D 单体型组的频率分布与 F 单体型组呈显著负相关 ($r = -0.785, P < 0.05$), 在汉族中也观察到从北向南递减的趋势(参见文献[26]表 4 的数据), 提示该高频率的 D 单体型组可能是北方人群的特征. 摩梭人的 mtDNA 单体型频率分布格局提示其在母系遗传结构上具有南方和北方双重特征. 这一点在主成分分析和系统树分析的结果中均显示出来.

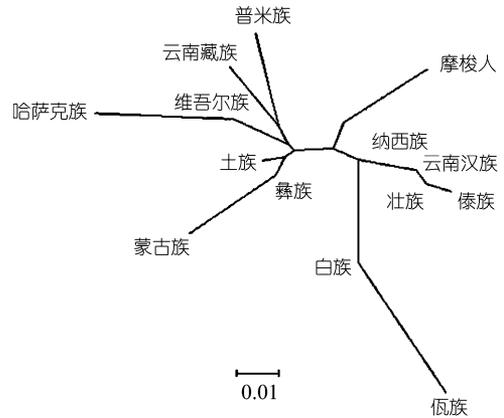


图 5 基于 mtDNA 高变区 1 Fst 遗传距离构建的邻接法系统树

综合以上多种分析的结果, 可以得出以下结论: 摩梭人的父系遗传结构缺乏南方民族的特征, 与云南藏族之间的关系最接近, 而母系遗传结构具有南北并蓄的特征, 最接近丽江纳西族. 摩梭人的父系、母系遗传结构存在比较大的差异, 无论是丽江纳西族还是藏族均不能包括其遗传结构的所有特点.

单体型(群)实质上是将亲源关系近的 Y 染色体(mtDNA)进行分类的方法, 其可靠性不仅取决于分类标准正确与否, 还与分类标准的分辨率有很大关系, 若分辨率不够, 可能导致信息的丧失. 为了进一步验证结果, 我们用更精细的 Y-STR 和 mtDNA HVSI 的序列数据进行了分子系统学分析和主成分分析, 考虑到了每一个体的数据, 结果完全一致. 用不同类型的遗传标记、不同的分析方法得到一致的结果, 大大加强了本结论的可靠性和完整性.

遗传学所揭示的人群演化史在母系和父系之间存在一定程度的差异^[32,33], 但像摩梭人这样巨大的反差却比较罕见, 基因交流是形成摩梭人复杂遗传结构最可能的原因. 据史书记载, 居住于云南包括摩梭人在内的藏缅语族人群是源于中国西北的氐羌族群的后裔, 氐羌的大规模南迁开始于公元前 7 世纪^[34]. 在其漫长的演化过程中, 不同的迁徙支系、居住环境以及受其

体型在南方人群中分布相当广泛, 而在北方人群中非常罕见, 是南方特有的染色体类型. 摩梭人中只观察到 8.5% 的 H9, 相对缺乏南方特有的单体型; 反之, 纳西族则具有相当高频率的 H11 单体型(47.5%), 因此摩梭人与纳西族在父系遗传结构上有着相当大的差异. 主成分分析以及

他人影响的程度不同,导致了现在不同人群之间的差异。摩梭人和丽江纳西族有相似的族称,均源自“牦牛羌”,秦汉以后被称为“摩沙夷”,其历史沿革过程大致分布于今四川盐源、云南宁蒞、丽江一带^[35],其语言也同属于纳西语,但属不同方言区。他们在母系遗传结构上的相似性与语言和历史的相似性是一致的。同为北方起源的纳西族和摩梭人母系基因库中均含有比较多的南方成分,可能是由于在他们迁徙和居住过程中融合了部分当地兄弟民族的血缘,在“摩沙夷”形成的早期阶段可能就已经开始发生融合。造成摩梭人与藏族父系遗传结构相近有两种可能的原因:(i) 他们有比较近的共同起源;(ii) 由基因混合导致。虽然摩梭人和藏族同属于古羌人的后裔,但源于不同的分支,藏族与“发羌”关系比较密切^[34],其语言也分别属于彝语支和藏语支。历史和语言的证据并不支持摩梭人与藏族之间共同起源的时间晚于纳西族与摩梭人之间的共同起源时间。摩梭人是中国目前仅存的母系社会,至今仍保持着“男不娶,女不嫁”的走婚制度,据徐胜^[4]的调查,在泸沽湖周围的婚龄人口中仍有 75% 保持“阿夏偶居婚(走婚)”,这种以女性为中心的社会结构和婚配方式增加了父系基因来源的复杂程度。关于摩梭人走婚制度形成及延续的年代,学术界尚没有定论,一般认为是对古羌人遗风的继承,因此摩梭人父系基因混合可能经历了相当长的时间。藏文化对摩梭文化的影响极其深远,600 年前由西藏传入的喇嘛教在其宗教信仰中占统治地位,在宁蒞地区甚至出现了类似于藏区的政教合一组织^[36],而喇嘛公开结交“阿夏”已经形成了一种特殊的宗教仪式^[4]。经过几百年的基因交流,形成现今摩梭人与藏族父系遗传结构的相似性是很有可能的。综上所述,与异族的基因融合可能是形成摩梭人复杂遗传结构的主要原因,较早开始的融合引入了摩梭人母系基因库中的南方成分,而比较晚近的与藏族等兄弟民族的融合改变了摩梭人的父系基因库,导致其父系遗传结构与藏族相近。

族源研究是一项系统工程,必须综合历史学、民族学、语言学、人类学和遗传学等多方面证据,本研究的主要目的是为揭开“中国原始社会最末一支玫瑰^[37]”的神秘面纱提供遗传学方面的证据,以供不同学科的工作者共同商榷和探讨。

参 考 文 献

- 1 Gottner-Abendroth H. The structure of matriarchal societies. *ReVision*, 1999, 21(3): 31~35
- 2 尤 中. 云南民族史. 昆明: 云南大学出版社, 1997. 19~29
- 3 和家修. 简论摩梭风情的实质和存在价值. *学术探索*, 1999, (5): 60~62
- 4 徐 旌. 摩梭人民俗文化探析. *云南地理环境研究*, 1998, 10(1): 89~95
- 5 Cann R L, Stoneking M, Willson A C. Mitochondrial DNA and human evolution. *Nature*, 1987, 325: 31~36
- 6 Cavalli-Sforza L, Menozzi P, Piazza A. The history and geography of human genes. Princeton: Princeton University Press, 1994. 373~374
- 7 肖春杰, 杜若甫, Cavalli-Sforza L, et al. 中国人群基因频率主成分分析. *中国科学, C 辑*, 2000, 30(4): 434~442
- 8 杜若甫, 肖春杰, Cavalli-Sforza L, et al. 用 38 个基因座的基因频率计算中国人群间的遗传距离. *中国科学, C 辑*, 1998, 28(1): 83~89
- 9 Jobling M A, Tyler-Smith C. Father and sons: the Y chromosome and human evolution. *Trends Genet*, 1995, 11: 449~456
- 10 Underhill P A, Passarino G, Lin A A, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet*, 2001, 65(1): 43~62
- 11 Underhill P A, Shen P, Lin A A, et al. Y chromosome sequence variation and the history of human populations. *Nature Genetics*, 2000, 26: 358~361

- 12 Su B, Xiao C, Deka R, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics*, 2000, 107: 582~590
- 13 Su B, Xiao J, Underhill P, et al. Y chromosome Evidence for a northward migration of modern humans in East Asia during the last Ice Age. *American Journal of Human Genetics*, 1999, (3)65: 1718~1724
- 14 Li Jin, Bing Su. Natives or immigrants: modern human origin in East Asia. *Nature Genetics Reviews*, 2000, (1): 126~132
- 15 Qian Y, Qian B, Su B, et al. Multiple origins of Tibetan Y chromosomes. *Human Genetics*, 2000, 106: 453~454
- 16 Su B, Jin L, Underhill P, et al. Polynesian origins: new insights from the Y-chromosome. *Proceedings of National Academy of Sciences USA*, 2000, 97: 8225~8228
- 17 Ke Y H, Su B, Song X F, et al. African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes. *Science*, 2001, 292: 1151~1153
- 18 柯越海, 宿兵, 肖君华, 等. Y染色体单体型在中国汉族人群中的多态性分布与中国人群的起源与迁移. *中国科学, C辑*, 2000, 30(6): 614~620
- 19 李永念, 左丽, 文波, 等. 中国布依族人的起源及迁移初探: 来自Y染色体和线粒体的线索. *遗传学报*, 2002, 29(2): 196~200
- 20 Kayser M, Caglia A, Corach D, et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med*, 1997, 110(3): 125~133
- 21 Kayser M, Krawczak M, Excoffier L, et al. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet*, 2001, 68(4): 990~1018
- 22 Lum J K, Cann R L, Martinson J J, et al. Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am J Hum Genet*, 1998, 63(2): 613~624
- 23 Anderson S, Bankier A T, Barrell B G, et al. Sequence and organization of the human mitochondrial genome. *Nature*, 1981, 290: 457~465
- 24 Schneider S, Roessli D, Excoffier L. *Arlequin ver 2.0: A Software for Population Genetic Data Analysis*. Switzerland: University of Geneva, 2000
- 25 Nei M, Kumar S. *Molecular evolution and phylogenetics*. New York : Oxford University Press, 2000. 256
- 26 Yao Y G, Kong Q P, Bandelt H J, et al. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet*, 2002, 70(3): 635~651
- 27 Slatkin M. A new measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 1995, 139: 457~462
- 28 Sudhir Kumar, Koichiro Tamura, Ingrid B, et al. *MEGA2: Molecular Evolutionary Genetics Analysis Software*. Arizona: Arizona State University, Tempe, 2001
- 29 Yao Y G, Lu X M, Luo H R, et al. Gene admixture in the silk road region of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet Syst*, 2000, 75(4): 173~178
- 30 Yao Y G, Nie L, Harpending H, et al. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol*, 2002, 118(1): 63~76
- 31 Wallace D C, Brown M D, Lott M T. Mitochondrial DNA variation in human evolution and disease. *Gene*, 1999, 238(1): 211~230
- 32 Gibbons A. The women's movement. *Science*, 1997, 278: 805
- 33 Elizabeth Pennisi. Tracking the sexes by their genes. *Science*, 2001, 291: 1733~1734
- 34 苍 铭. 云南民族迁徙文化研究. 昆明: 云南民族出版社, 1997. 10~24
- 35 王钟翰. 中国民族史. 北京: 中国社会科学出版社, 1994. 289~291
- 36 和少英. 浅论藏文化于纳西文化之交汇. *民族研究*, 1995, (1): 27~32
- 37 余洪涛. 中国原始社会最末一支玫瑰. 北京: 中国青年出版社, 1996. 315~320