

# 贵州壮侗语族7个少数民族21个Y-SNP多态性\*

张秀秀<sup>1,2</sup>, 喻艳琴<sup>1,2</sup>, 田薇<sup>1,2</sup>, 张婷<sup>1,2</sup>, 王婵娟<sup>1,2</sup>, 单可人<sup>1,2</sup>, 何燕<sup>1,2</sup>  
(1. 地方病与少数民族性疾病教育部重点实验室/贵州医科大学, 贵州 贵阳 550004;  
2. 贵州省医学分子生物学重点实验室, 贵州 贵阳 550004)

**摘要:** 调查贵州壮侗语族7个少数民族(水族、布依族、侗族、仡佬族、壮族、毛南族、仫佬族)人群Y染色体上21个单核苷酸多态性(single nucleotide polymorphism, SNP)位点的遗传多态性,并探讨该人群与其他民族以及其他语族之间的遗传关系。采用SNapShot法对445例贵州壮侗语族7个少数民族男性个体21个Y-SNP位点进行复合扩增检测,用直接计数法计算等位基因频率、单倍型频率与单倍群频率,运用SPSS 24软件进行主成分分析。在贵州水族、布依族、侗族、仡佬族、壮族、毛南族和仫佬族中,21个Y-SNP组成的单倍型多态性值(haplotype diversity, HD)分别为0.597 7、0.945 6、0.885 6、0.886 9、0.768 6、0.358 7和0.707 8;与已报道的其他9个民族人群单倍群分布频率进行主成分分析,壮侗语族与苗瑶语族聚在一起。从父系遗传角度分析,壮侗语族与苗瑶语族亲缘关系较近。

**关键词:** 壮侗语族; Y染色体; 单核苷酸多态性; 遗传多态性; 贵州

**中图分类号:** Q7 **文献标志码:** A **文章编号:** 0529-6579(2020)01-0125-08

## Single nucleotide polymorphism of 21 Y-chromosomal loci in seven ethnic minorities of Zhuang-Dong Language in the Guizhou Province of China

ZHANG Xiuxiu<sup>1,2</sup>, YU Yanqin<sup>1,2</sup>, TIAN Wei<sup>1,2</sup>, ZHANG Ting<sup>1,2</sup>,  
WANG Chanjuan<sup>1,2</sup>, SHAN Keren<sup>1,2</sup>, HE Yan<sup>1,2</sup>

(1. Key Laboratory of Endemic and Ethnic Diseases, Ministry of Education/Guizhou Medical University, Guiyang 550004, China;

2. Key Laboratory of Molecular Biology, Guizhou Medical University, Guiyang 550004, China)

**Abstract:** The study aimed to determine single nucleotide polymorphism of 21 Y-chromosomal loci in seven ethnic minorities, including Shui, Bouyei, Dong, Gelo, Zhuang, Maonan and Mulam, of Zhuang-Dong language in the Guizhou province of China, and explore differences in the polymorphism from other Language ethnics. The SNapShot method was used to detect the complex amplicons of the 21 Y-SNPs loci in 445 males from seven Guizhou Zhuang-Dong language ethnics. Frequencies of alleles, haplotypes, and haploid populations were calculated using a direct counting method. SPSS 24 was used for principal component analyses. The haplotype diversity values were 0.597 7, 0.945 6, 0.885 6, 0.886 9, 0.768 6, 0.358 7 and 0.707 8, respectively, in the seven ethnics. We performed principal component analyses based on the frequencies of haplogroups by comparing our data with those reported for other nine ethnic

\* 收稿日期: 2019-03-25

基金项目: 国家自然科学基金(31560306); 贵阳市科技计划项目([2017]5-9)

作者简介: 张秀秀(1988年生),女;研究方向:少数民族基因多态性;E-mail: 891991395@qq.com

通信作者: 何燕(1972年生),女;研究方向:少数民族基因多态性;E-mail: annieheyan@gmc.edu.cn

groups in China. The results showed that Zhuang-Dong Language and Hmong-Mien Language were clustered together. From the perspective of patrilineal genetics, the relationship of Zhuang-Dong Language to Hmong-Mien Language is relatively closer than those of other ethnics.

**Key words:** Zhuang-Dong Language; Y-chromosome; SNP; genetic polymorphism; Guizhou Province

Y 染色体的非重组区域单核苷多态性 (Y-SNP) 已被广泛用于研究人类种群的起源<sup>[1]</sup> 和迁徙<sup>[2-4]</sup>。Y-SNP 单倍群的分布频率有地区特异性<sup>[5]</sup>, 且和语言分类高度相关<sup>[6]</sup>, 研究人员可以通过单倍群频率来推断群体可能的地理起源<sup>[7]</sup> 和种族起源<sup>[8]</sup>。壮侗语族是一个历史非常悠久的民族群体, 与古代南方百越族群有渊源关系, 在距今 4 000 年前的新石器时代文化遗址中存在他们曾经生活过的迹象, 百越民族在千百年的分化融合过程中逐渐发展成黎族、侗族、水族、仡佬族、仡佬族、壮族等<sup>[9]</sup>。本文利用 Y-SNP 遗传多态性分析贵州壮侗语族 7 个民族 (水族、布依族、侗族、仡佬族、壮族、毛南族、仡佬族) 男性群体的遗传结构, 探讨与其他民族以及其他语族的遗传关系。

## 1 材料和方法

### 1.1 样本收集及 DNA 标化

从课题组根据知情同意原则建立的贵州世居少数民族 DNA 样本库中, 采用整群随机抽样方法, 从本民族聚集地采集样本, 3 代内无族外通婚史, 个体间无亲缘关系, 筛选出语言学分类隶属壮侗语族的贵州 7 个世居少数民族 (水族、布依族、侗族、仡佬族、壮族、毛南族和仡佬族) 男性 DNA 样本, 样本例数及采样地点等信息见表 1。

表 1 贵州省壮侗语族 7 个少数民族种类、样本例数和采样地点一览表

Table 1 List of the 7 ethnic minorities, sampled ethnic groups and sample numbers of Zhuang-Dong Language in Guizhou Province

民族	样本例数	采样地点
水族	64	三都县三洞乡
布依族	58	平塘县四寨乡、荔波县钟山乡
侗族	65	从江县丙妹镇和高增乡
仡佬族	67	黔西县沙井乡
壮族	66	从江县刚边乡
毛南族	60	平塘县卡浦乡
仡佬族	65	麻江县宣威乡
合计	445	

每份 DNA 样本用 Thermo Scientific™ NanoDrop Lite 分光光度计定量后, 取少量标化为 20 ng/ $\mu$ L 作为实验的模板,  $-40\text{ }^{\circ}\text{C}$  保存备用。

### 1.2 Y 染色体基因分型

**1.2.1 21 个 Y-SNP 多重 PCR 扩增及纯化** 多重 PCR 扩增: 在 Y 染色体进化树选取东亚主要的单倍群<sup>[10]</sup> 上的 M145、RPS4Y711、M89、M9、M214、M175、M119、P31、M95、SRY465、47Z、M122、M324、P201、M159、M7、M134、M133、M217、M48、M407 21 个 Y-SNP 为研究靶点, 依据文献 [11] 分成 4 组 (I、II、III、IV 组) 进行 PCR 扩增 (引物序列及分组情况见表 2)。体系包括: 20 ng/ $\mu$ L 的模板 DNA 1.5  $\mu$ L、引物 MIX 15  $\mu$ L、10 nmol/L dNTP 3.0  $\mu$ L、10  $\times$  Buffer 2.5  $\mu$ L、Taq DNA 聚合酶 0.5  $\mu$ L、1 mmol/L 甜菜碱 1.0  $\mu$ L (其作用在于: ① 富含 GC 模板的 PCR 扩增; ② 提高 Taq DNA 聚合酶的稳定性)、5 mmol/L MgCl<sub>2</sub> 1.0  $\mu$ L、500  $\mu$ g/mL 牛血清蛋白 (BSA) 0.5  $\mu$ L。循环条件: 95  $^{\circ}\text{C}$  10 min; 95  $^{\circ}\text{C}$  30 s, 58  $^{\circ}\text{C}$  30 s, 72  $^{\circ}\text{C}$  30 s, 循环 35 次; 72  $^{\circ}\text{C}$  7 min, 产物置 4  $^{\circ}\text{C}$  保存。

纯化: 第 I、II 组 PCR 产物各取 1  $\mu$ L 混合, 加入 1 U/ $\mu$ L 虾碱酶 (shrimp alkaline phosphatase, SAP) 1  $\mu$ L 和 1 U/ $\mu$ L 大肠杆菌核酸外切酶 I (exonuclease, Exo I) 1  $\mu$ L, 37  $^{\circ}\text{C}$  保温 70 min 后 75  $^{\circ}\text{C}$  15 min 灭活酶, 即得纯化后的多重 PCR 产物, 4  $^{\circ}\text{C}$  保存, 充当单碱基扩增时 A 组的模板。第 III、IV 组扩增产物也如法纯化, 充当单碱基扩增时 B 组的模板。

**1.2.2 SNaPShot 单碱基扩增及纯化** 分 A、B 两组进行单碱基扩增 (分组情况及引物信息见表 2)。体系包括: 模板 0.75  $\mu$ L、SNaPShot Mix 1.25  $\mu$ L、单碱基扩增引物 MIX 0.5  $\mu$ L。循环条件: 96  $^{\circ}\text{C}$  10 s, 50  $^{\circ}\text{C}$  5 s, 60  $^{\circ}\text{C}$  30 s, 循环 28 次, 产物 4  $^{\circ}\text{C}$  保存。

纯化: 单碱基扩增产物加入 1 U/ $\mu$ L 的 SAP 0.5  $\mu$ L, 混匀, 瞬时离心, 37  $^{\circ}\text{C}$  保温 70 min 后 75  $^{\circ}\text{C}$  15 min 灭活酶, 即得纯化后的 SNaPShot 单碱基延伸产物, 4  $^{\circ}\text{C}$  保存。

**1.2.3 ABI 3130 毛细管电泳检测** 纯化的单碱基

延伸产物 0.5 μL、GeneScan – 120LIZ Size Standard 0.05 μL 和 Hi – Di™ 甲酰胺 9.45 μL, 混匀, 离心并用 ABI 3130 遗传分析仪 (Applied Biosystems) 进行毛细管电泳分析, ABI 3130 Genetic Analyzer Data Collection Software v3.0 进行数据收集。

### 1.3 数据统计分析

用直接计数法计算 21 个 Y-SNP 等位基因频率、单倍型频率与单倍群频率。单倍型多样性 (HD) 和基因多样性 (genetic diversity, GD) 根据公式  $HD$  或  $GD = n(1 - \sum P_i^2) / (n - 1)$  ( $P_i$  为单倍型频率或等位基因频率,  $n$  为样本数) 计算。运用 SPSS 24 软件进行主成分分析 (principle component analysis, PCA)。

## 2 结 果

### 2.1 贵州 7 个世居少数民族男性样本 21 个 Y-SNP 的遗传多态性分析

本文采用 SNaPShot 法对贵州省壮侗语族 7 个世居少数民族 445 例男性样本的 21 个 Y-SNP 位点进行基因分型, 等位基因频率见图 1, 基因多态性见图 2。21 个 Y-SNP 位点中 7 个世居少数民族男性样本的 M407、47Z 的突变频率均为 0, 水族群体的 M175、M214、M9、M89 的突变频率均为 1, 上述位点均没有多态性 ( $GD = 0.0000$ ); M324、M122、M119、M95、P31 在 7 个世居少数民族男性样本中均具有多态性, M48 位点仅侗族有多态性

表 2 21 个 Y-SNP 位点的多重 PCR 引物和 SNaPShot 单碱基扩增引物及分组情况  
Table 2 The sequences of Multiplex PCR primers and SNaPShot microsequencing primers and grouping for 21 SNPs on Y – chromosome

位点	多重 PCR 引物上游	多重 PCR 引物下游	多重 PCR 分组情况	单碱基扩增引物	单碱基扩增引物分组
M145	F: gcctccacgacttctctagac	R: aggttctcccactect	I	(t) <sub>25</sub> ctagacaccagaagaaagc	A
RPS4Y <sub>711</sub>	F: cagggaataaaccttggat	R: gtggccagcctcttatctctc		(t) <sub>47</sub> agggaataaaccttggatttc	
M89	F: :agcttctggattcagctctc	R: caggatcaccagcaaaggtag		(t) <sub>51</sub> caactcaggcaaaagttagat	
M9	F: ggacctgaaatacagaactgc	R: cgttgaacatgtctaaattaaagaaaa		(t) <sub>19</sub> aacggcctaagatgggtgaat	
M214	F: cactggaagaaaaaagtctg	R: agcctgggagacagtgtag		(t) <sub>25</sub> tggtacttctgcttatttttc	
M175	F: acccaaatcaactcaactcca	R: tgatacctttgttcttcttctt		(t) <sub>38</sub> gcacatgccttctcacttctc	
M119	F: caaacccagtgctatgtgt	R: atgggtattccaattcagca		(t) <sub>23</sub> ggttattccaattcagcatacagcc	
P31	F: tgggaacaggtaggtggta	R: gtgtgagactccatcgcaaa	(t) <sub>12</sub> aggttacataaaataaggttttttttgggtg	II	
M95	F: gggatcaaatggagttcctg	R: gcctacaggttgaaaggcta	(t) <sub>7</sub> tgaggataaggaaagactaccatattagtg		
SRY <sub>465</sub>	F: atcccgtctcggactctctg	R: tcttgagtgtgtgcttctct	(t) <sub>56</sub> cctgttgctcagttgcacttc		
47Z	F: tctcctgacctgtgaltcg	R: tcattgacatgggctggact	(t) <sub>34</sub> tgggctggacttgggtggctca		
M122	F: cttagttgccttttggaaatgaa	R: gctttattcagattttcccctga	(t) <sub>44</sub> ttcagattttcccctgagagc		
M324	F: tgattgatctacctgcccttt	R: aagggaacaaattgatttcag	(t) <sub>33</sub> (a)2tgatctacctgccctttcct	III	
P201	F: tgtgctgtcaagttgtgtg	R: tgggtgcagttaaagcaatga	(t) <sub>36</sub> agatcttggttaagtcatttgatctcag		
M159	F: ttcagccttctctgtacttttta	R: tcctctggagtcgaaagagtg	(t) <sub>12</sub> agtttattattgatgcaagccctaa		
M7	F: caaagggcatgtaatcattct	R: tgatccaattatttccattgtgtt	(t) <sub>9</sub> attaaattttagttgagttactgttctt		
M134	F: atcaaacccagaagggtaaaga	R: gagatactttgatccccacca	(t) <sub>31</sub> agaaaaggcccaggaaagat		
M133	F: aaggtgggcttctgaag	R: gattgtctggttggggaa	(a) <sub>34</sub> ctttctgaagcaaataccagcttaaaaaaaa		
M217	F: ggagaatgaaaaagttgggtg	R: aagctgctgtggcttctc	(t) <sub>22</sub> agaatgaaaaagttgggtgacac		
M48	F: tccttccactcttagcttgac	R: ctgagggaactattaaggca	(t) <sub>37</sub> gacaattaggattaagaatgat	IV	
M407	F: ctgaaagttggggacagctcat	R: tggcactaaatcaacttctctt	(t) <sub>11</sub> gcactaaatcaacttctcttgg		

( $GD = 0.0308$ ), SRY465 和 M159 仅仡佬族有多态性,  $GD$  值均为  $0.0298$ 。

经单倍型多态性公式计算水族、布依族、侗族、仡佬族、壮族、毛南族、侗族单倍型多态性分别为  $0.5977$ 、 $0.9456$ 、 $0.8856$ 、 $0.8869$ 、 $0.7686$ 、 $0.3587$ 、 $0.7078$ ; 依照国际系谱遗传 (International Society of Genealogy, ISOGG) 网站 <https://isogg.org/tree/index.html> 上发布的 Y 单倍群系统进化树进行单倍群的划分, 通过直接计数法获得单倍群频率 (见表 3); 通过 Excel 绘制贵州省 7 个世居少数民族 Y 染色体单倍群频率热图, 该热图对不同单倍群在 7 个世居少数民族的分布情况进行直观观测, 绿色→蓝色→红色单倍群频率逐渐增加, 显而易见, 在 7 个世居少数民族中主要单倍群为 O1b1a1a - M95, 该单倍群在毛南族、仡佬族、水族人群中呈高频分布 ( $0.8167$ 、 $0.5692$ 、 $0.5000$ ), 在侗族人群中分布频率较低 ( $0.0923$ )。

## 2.2 贵州 7 个世居少数民族与其他 9 个少数民族的群体遗传学分析

通过 Excel 绘制贵州省壮侗语族的 7 个世居少数民族和其他 9 个少数民族 Y 染色体单倍群频率热图, 见表 4, 该热图对不同单倍群在 16 个民族的分布情况进行直观观测, 绿色→蓝色→红色单倍群频率逐渐增加, 可以看到单倍群 O1b1、O2a2 在壮侗语族和苗瑶语族的分布频率较高; O1a 仅在壮

侗语族存在高频分布; 单倍群 C 在北方民族群体中的分布频率明显高于南方民族群体。运用 SPSS 24 软件对表 4 进行主成分分析 (见图 3)。如图 3 所示, 北方民族与南方民族各自聚在一起, 其中归属壮侗语族的毛南族、壮族、水族、仡佬族紧密相聚后再与仡佬族相聚, 而同属壮侗语族的侗族和布依族则与苗瑶语族的苗族、瑶族、畲族较近, 提示上述民族之间可能发生了基因交融。

## 2.3 贵州省壮侗语族与其他语族之间的群体遗传学分析

为了验证贵州省壮侗语族和其他语族之间的关系, 通过直接计数法获得贵州省壮侗语族的单倍群频率, 通过文献报道的民族单倍群频率计算语族的单倍群频率, 通过 Excel 绘制贵州省壮侗语族和其他语族 Y 染色体单倍群频率热图, 见表 5, 该热图对不同单倍群在 7 个语族间的分布情况进行直观观测, 绿色→蓝色→红色单倍群频率逐渐增加, 很明显单倍群 C 在北方民族群体分布频率较高; O2a2 在南方群体分布频率较高, O1b1 在壮侗语族群体存在较高频率分布, 该单倍群可能与壮侗语族群体相关; O2 在汉族语族群体中分布频率较高。根据表 5 的单倍群频率采用 SPSS 24 软件进行主成分分析 (见图 4)。图 4 中, 前三个主成了解释了  $84.79\%$  的总方差, 图中归属汉藏语系的壮侗语族、苗瑶语

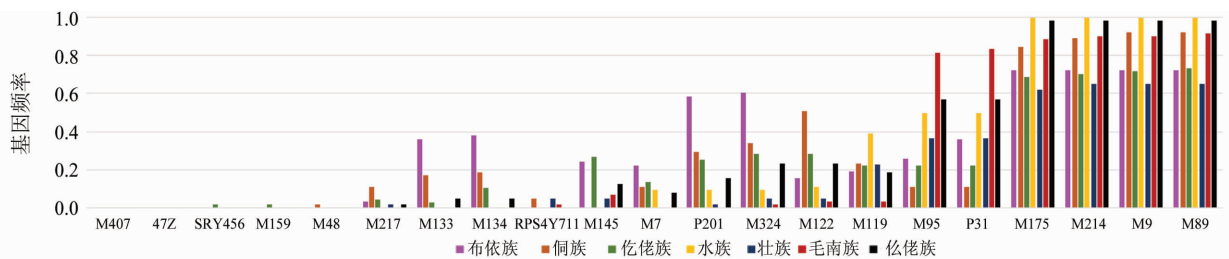


图 1 贵州壮侗语族 7 个民族人群 21 个 Y-SNP 的基因频率

Fig. 1 Frequency of 21 Y-SNPs loci 7 ethnic groups in Guizhou Zhuang-Dong Language

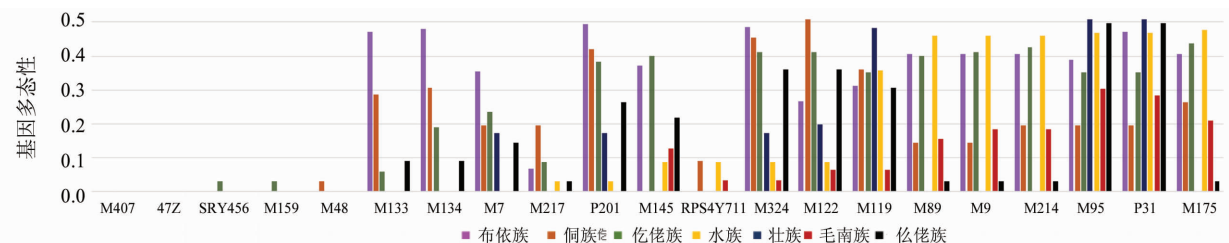


图 2 贵州壮侗语族人群 7 个民族 21 个 Y-SNP 的基因多态性

Fig. 2 GD value of 21 Y-SNPs loci 7 ethnic groups in Guizhou Zhuang-Dong Language

表 3 贵州壮侗语族的 Y 染色体单倍群频率<sup>1)</sup>

Table 3 Y-SNP haplotype frequency of Y chromosome of Guizhou Zhuang-Dong Language

民族	例数	C-RPS4Y711	C2-M217	D-M145	F-M89	K-M9	NO1-M214	O-M175	O1a-M119	O1b-P31
水族	64	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.390 6	0.000 0
布依族	58	0.000 0	0.000 0	0.103 4	0.000 0	0.000 0	0.000 0	0.017 2	0.051 7	0.017 2
侗族	65	0.000 0	0.046 2	0.000 0	0.000 0	0.030 8	0.030 8	0.000 0	<b>0.184 6</b>	0.000 0
仡佬族	67	0.000 0	0.044 8	<b>0.194 0</b>	0.014 9	0.000 0	0.014 9	0.000 0	<b>0.209 0</b>	0.000 0
壮族	66	0.045 5	0.015 2	0.000 0	0.000 0	0.000 0	0.015 2	0.000 0	<b>0.227 3</b>	0.000 0
毛南族	60	0.016 7	0.000 0	0.066 7	0.000 0	0.000 0	0.016 7	0.000 0	0.033 3	0.016 7
仡佬族	65	0.000 0	0.015 4	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	<b>0.184 6</b>	0.000 0

民族	例数	O1b1a1a-M95	O1b2-SRY465	O2-M122	O2a-M324	O2a2-P201
水族	64	<b>0.500 0</b>	0.000 0	0.015 6	0.000 0	0.000 0
布依族	58	0.172 4	0.000 0	0.034 5	0.017 2	0.000 0
侗族	65	0.092 3	0.000 0	<b>0.276 9</b>	0.046 2	0.000 0
仡佬族	67	<b>0.223 9</b>	0.014 9	0.000 0	0.029 9	0.000 0
壮族	66	<b>0.363 6</b>	0.000 0	0.000 0	0.030 3	0.015 2
毛南族	60	<b>0.816 7</b>	0.000 0	0.016 7	0.016 7	0.000 0
仡佬族	65	<b>0.569 2</b>	0.000 0	0.000 0	0.076 9	0.030 8

民族	例数	O2a2a1a1a-M159	O2a2a1a2-M7	O2a2b1-M134	O2a2b1a1a-M133	Other
水族	64	0.000 0	0.093 8	0.000 0	0.000 0	0.000 0
布依族	58	0.000 0	<b>0.206 9</b>	0.017 2	<b>0.362 1</b>	0.000 0
侗族	65	0.000 0	0.107 7	0.015 4	0.169 2	0.000 0
仡佬族	67	0.014 9	0.134 3	0.074 6	0.029 9	0.000 0
壮族	66	0.000 0	0.000 0	0.000 0	0.000 0	<b>0.287 9</b>
毛南族	60	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0
仡佬族	65	0.000 0	0.076 9	0.000 0	0.046 2	0.000 0

1) 表中数字(≤1)代表单倍群在该民族中所占比例;绿色:0.000 0 ~ 0.010 0(不含);蓝色:0.010 0 ~ 0.180 0(不含);蓝色加粗:0.180 0 ~ 0.300 0(不含);红色:0.300 0 ~ 0.600 0(不含);红色加粗:≥0.600 0

表 4 贵州省壮侗语族的 7 个世居少数民族和其他 9 个少数民族 Y 染色体单倍群频率<sup>1)</sup>

Table 4 Y-SNP haplotype frequency of Y chromosome of Guizhou seven ethnic groups and 9 ethnic minority population

人群	例数	语系	语族	C	D	F	K	O	O1a	O1b	O1b1	O1b2	O2	O2a	O2a2	Other
水族	64	汉藏	壮侗	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.390 6	0.000 0	<b>0.500 0</b>	0.000 0	0.015 6	0.000 0	0.093 8	0.000 0
布依族	58	汉藏	壮侗	0.000 0	0.103 4	0.000 0	0.000 0	0.017 2	0.051 7	0.017 2	0.172 4	0.000 0	0.034 5	0.017 2	<b>0.586 2</b>	0.000 0
侗族	65	汉藏	壮侗	0.046 2	0.000 0	0.000 0	0.061 6	0.000 0	<b>0.184 6</b>	0.000 0	0.092 3	0.000 0	<b>0.276 9</b>	0.046 2	<b>0.292 3</b>	0.000 0
仡佬族	67	汉藏	壮侗	0.044 8	<b>0.194 0</b>	0.014 9	0.014 9	0.000 0	<b>0.209 0</b>	0.000 0	<b>0.223 9</b>	0.014 9	0.000 0	0.029 9	<b>0.253 7</b>	0.000 0
壮族	66	汉藏	壮侗	0.060 7	0.000 0	0.000 0	0.015 2	0.000 0	<b>0.227 3</b>	0.000 0	<b>0.363 6</b>	0.000 0	0.000 0	0.030 3	0.015 2	<b>0.287 9</b>
毛南族	60	汉藏	壮侗	0.016 7	0.066 7	0.000 0	0.016 7	0.000 0	0.033 3	0.016 7	<b>0.816 7</b>	0.000 0	0.016 7	0.016 7	0.000 0	0.000 0
仡佬族	65	汉藏	壮侗	0.015 4	0.000 0	0.000 0	0.000 0	0.000 0	<b>0.184 6</b>	0.000 0	<b>0.569 2</b>	0.000 0	0.000 0	0.076 9	0.153 9	0.000 0
瑶族 <sup>[11]</sup>	41	汉藏	苗瑶	0.024 4	0.073 2	0.024 4	0.048 8	0.048 8	0.122 0	0.000 0	<b>0.317 1</b>	0.000 0	0.146 3	0.000 0	<b>0.195 2</b>	0.000 0
苗族 <sup>[11]</sup>	49	汉藏	苗瑶	0.061 2	0.061 2	0.000 0	0.020 4	0.122 4	0.061 2	0.000 0	<b>0.306 1</b>	0.000 0	<b>0.183 7</b>	0.000 0	<b>0.183 6</b>	0.000 0
畲族 <sup>[11]</sup>	56	汉藏	苗瑶	0.035 7	0.000 0	0.017 9	0.000 0	0.035 7	0.089 3	0.000 0	0.035 7	0.000 0	<b>0.303 6</b>	0.000 0	<b>0.482 1</b>	0.000 0
蒙古族 <sup>[12]</sup>	50	阿尔泰	蒙古	<b>0.400 0</b>	0.020 0	0.000 0	0.040 0	0.040 0	0.000 0	0.000 0	0.020 0	0.000 0	0.160 0	0.000 0	0.160 0	0.160 0
土族 <sup>[12]</sup>	50	阿尔泰	蒙古	<b>0.280 0</b>	0.060 0	0.020 0	0.020 0	0.060 0	0.060 0	0.000 0	0.020 0	0.000 0	<b>0.240 0</b>	0.000 0	0.020 0	<b>0.220 0</b>
锡伯族 <sup>[12]</sup>	32	阿尔泰	满- 通古斯	0.125 0	0.000 0	0.000 0	0.093 8	<b>0.218 8</b>	0.125 0	0.000 0	0.000 0	0.000 0	0.062 5	0.000 0	0.156 3	<b>0.218 8</b>
维吾尔族 <sup>[12]</sup>	50	阿尔泰	突厥	0.040 0	0.020 0	0.000 0	<b>0.180 0</b>	0.060 0	0.000 0	0.000 0	0.000 0	0.000 0	0.040 0	0.000 0	0.000 0	<b>0.660 0</b>
哈萨克 <sup>[12]</sup>	41	阿尔泰	突厥	<b>0.585 4</b>	0.024 4	0.048 8	0.024 4	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.073 2	<b>0.243 9</b>
俄罗斯族 <sup>[12]</sup>	19	印欧	斯拉夫	0.105 3	0.052 6	0.000 0	0.000 0	0.105 3	0.052 6	0.000 0	0.052 6	0.000 0	<b>0.368 4</b>	0.000 0	0.105 3	0.157 9

1) 表中数字(≤1)代表单倍群在该民族中所占比例;绿色:0.000 0 ~ 0.010 0(不含);蓝色:0.010 0 ~ 0.180 0(不含);蓝色加粗:0.180 0 ~ 0.300 0(不含);红色:0.300 0 ~ 0.600 0(不含);红色加粗:≥0.600 0

表 5 贵州省壮侗语族和其他语族 Y 染色体单倍群频率<sup>1)</sup>

Table 5 Y-SNP haplotype frequency of Y chromosome of Guizhou Zhuang-Dong Language and other language population

语系	语族	例数	C	D	F	K	O	O1a	O1b	O1b1	O1b2	O2	O2a	O2a2	Other
汉藏	壮侗语族	445	0.026 3	0.052 0	0.002 1	0.015 5	0.002 5	<b>0.183 0</b>	0.004 8	<b>0.391 2</b>	0.002 1	0.049 1	0.031 0	<b>0.199 3</b>	0.041 1
汉藏	苗瑶语族 <sup>[11]</sup>	754	0.085 8	0.039 1	0.014 2	0.034 6	0.108 0	0.057 1	<b>0.000 0</b>	<b>0.198 8</b>	<b>0.000 0</b>	0.163 8	<b>0.000 0</b>	<b>0.297 7</b>	<b>0.001 2</b>
汉藏	藏缅语族 <sup>[13]</sup>	154 7	0.029 1	0.067 9	0.014 9	0.079 5	<b>0.000 0</b>	0.014 2	<b>0.000 0</b>	0.016 8	<b>0.000 0</b>	0.079 5	<b>0.000 0</b>	<b>0.336 8</b>	<b>0.361 3</b>
汉藏	汉语族 <sup>[14]</sup>	170 6	0.075 0	0.018 8	0.062 1	0.161 2	<b>0.000 0</b>	0.079 7	<b>0.000 0</b>	0.035 2	<b>0.000 0</b>	<b>0.284 9</b>	<b>0.000 0</b>	<b>0.242 7</b>	0.040 4
阿尔泰	突厥语族 <sup>[15]</sup>	102 4	<b>0.368 7</b>	0.030 2	0.001 4	<b>0.000 0</b>	<b>0.000 0</b>	<b>0.000 0</b>	0.002 3	0.003 3	<b>0.000 0</b>	0.001 9	0.006 9	0.171 7	<b>0.413 6</b>
阿尔泰	蒙古语族 <sup>[15]</sup>	250	<b>0.201 8</b>	<b>0.224 3</b>	<b>0.000 0</b>	<b>0.000 0</b>	<b>0.000 0</b>	<b>0.000 0</b>	0.049 4	<b>0.000 0</b>	<b>0.000 0</b>	0.012 4	0.023 3	0.109 6	<b>0.379 5</b>
阿尔泰	满-通古斯语族 <sup>[14]</sup>	124	<b>0.543 8</b>	<b>0.000 0</b>	<b>0.006 4</b>	0.140 3	<b>0.000 0</b>	0.019 6	<b>0.000 0</b>	0.006 8	<b>0.000 0</b>	0.110 9	<b>0.000 0</b>	0.105 7	0.066 5

1) 表中数字 ( $\leq 1$ ) 代表单倍群在该民族中所占比例, 绿色: 0.000 0 ~ 0.010 0 (不含); 蓝色: 0.010 0 ~ 0.180 0 (不含); 蓝色加粗: 0.180 0 ~ 0.300 0 (不含); 红色: 0.300 0 ~ 0.600 0 (不含); 红色加粗:  $\geq 0.600 0$

族、汉语族聚在一起, 归属阿尔泰语系的突厥语族、蒙古语族相聚后再与满通古斯语族聚在一起, 藏缅语族位于两组之间。

### 3 讨论

#### 3.1 贵州省壮侗语族 7 个世居少数民族的 Y 染色体遗传结构

Y-SNP 多态性分布具有明显的民族特异性<sup>[16]</sup>, 各个民族之间具有其独特的遗传结构<sup>[17-18]</sup>。本研究对贵州省壮侗语族 7 个世居少数民族 445 例无关男性个体进行基因分型, 并对结果进行相关统计分析。首先对表 2 进行单倍群相关性分析, O1b1a1 - M95 与 O2 - M122 的皮尔逊相关性  $R = -0.526 0$ , 呈负相关, 但显著性 (双尾)  $P = 0.225 0$ , 说明负相关未达到显著水平; O1a - M119 与 O1b - P31 为显著负相关 ( $R = -0.804 0$ ,  $P = 0.029 0$ )。通过热图绘制对不同单倍群在 7 个民族分布情况进行直观观测 (见表 3), 单倍群 D、F、O1b2、O2a 的频率在各个语族的分布频率均很低; O2 - M122 在各语族之间无显著差异; O1a 在壮侗语族中分布频率较高; O1b1、O2a2 在壮侗语族和苗瑶语族中分布频率均较高。水族人群中, 主要单倍群为 O1b1 (0.500 0) 和 O1a (0.390 6); O2a2 (0.586 2) 和 O1b1 (0.172 4) 在布依族人群中具有代表性; O1a (0.184 6)、O2 (0.276 9)、O2a2 (0.292 3) 在侗族人群中具有代表性; 仡佬族的主要单倍群有 D (0.194 0)、O1a (0.209 0)、O1b1 (0.223 9)、O2a2 (0.253 7); 壮族除了主要单倍群 O1a (0.227 3)、O1b1 (0.363 6) 之外, 存在完全野生型 SNP 的频率达 0.287 9, 可能提示部分壮族人群携带的是最古老的遗传标记, 也可能是本研究选择的 Y-SNP 位点有限, 未包含突变的基因座; 毛南族的单倍群 O1b1 高达 0.816 7; 仡佬族的主要单倍群有 O1a (0.184 6)、O1b1 (0.596 1)、O2a2 (0.153 9)。通过以上数据分析

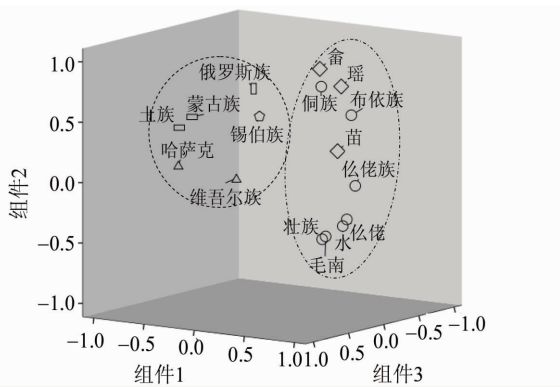


图 3 贵州 7 个世居少数民族和 9 个少数民族人群 Y 染色体主成分分析三维图  
Fig. 3 The principal component analysis of Y chromosome of Guizhou seven ethnic groups and 9 ethnic minority population

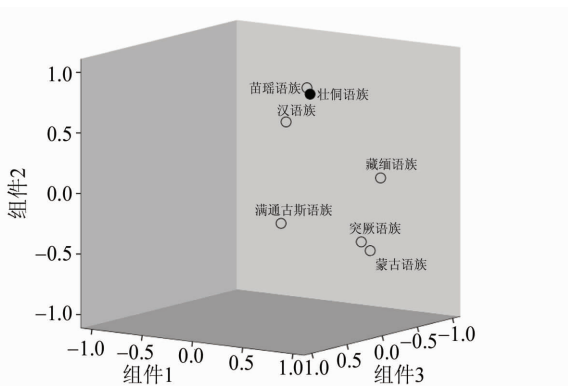


图 4 壮侗语族和其他语族人群 Y 染色体主成分分析三维图  
Fig. 4 The principal component analysis of Y chromosome of Guizhou Zhuang-Dong Language and other language population

很明显可以看出,除了毛南族,本研究的其他 6 个民族均具有复杂的父系遗传结构,提示毛南族的父系遗传结构比较单一,与毛南族土著说<sup>[19]</sup>相符。单倍群 O\* 是东亚现今人群中分布最广泛的特异单倍群,综合频率约占 50% 以上<sup>[20]</sup>。在本研究中,水族、布依族、侗族、仡佬族、壮族、毛南族、仡佬族单倍群 O\* 的分布频率分别为 1.000 0、0.896 4、0.892 3、0.731 4、0.636 4、0.900 1、0.984 6,平均频率高达 86.30%,与上述文献相符。O1 单倍群占大多数壮侗语族的 60.00% 以上<sup>[6]</sup>,经直接计数法计算这 7 个民族水族 (0.890 6)、布依族 (0.241 3)、侗族 (0.276 9)、仡佬族 (0.447 8)、壮族 (0.590 9)、毛南族 (0.866 7)、仡佬族 (0.753 8) 中约一半的人群在 O1 单倍群所占比重与之相符。O1b1a1a - M95 是高度分化的支系,在七个壮侗语民族群体中是最常见的单倍群,在东南亚频率较高,其次是中国南部<sup>[6]</sup>,本研究中,在水族 (0.500 0)、毛南族 (0.816 7)、仡佬族 (0.569 2) 呈高频分布 (均大于 50%),提示水族、毛南族、仡佬族遗传背景较为单一,与土著说的观点相符,尤其是毛南族高达 81.67%,提示毛南族经历了非常强烈的瓶颈效应。O2a2a1a2 - M7 与苗瑶语族、孟 - 高棉语族的人群分布有关<sup>[21]</sup>,贵州省壮侗语族长期与苗瑶语族混居,故 O2a2a1a2 - M7 在壮侗语族也占有一定比例,但是壮族与毛南族人群的分布频率却为 0,壮族有就近嫁娶,民族内婚的习俗<sup>[22]</sup>,但是毛南族婚姻比较复杂<sup>[23]</sup>,为何 O2a2a1a2 - M7 为 0,还不得而知。壮侗语族在我国历史上起源于百越族群,前期有学者研究发现 O1a - M119 是百越群体的特征性单倍群<sup>[24]</sup>,我们的研究展示了这个结果——壮侗语族 7 个民族群体

均含有单倍群 O1a - M119 的频率分布与上述观点吻合,进一步印证了壮侗语族与百越族群的关系:在民族学中,壮侗语族族群来源于百越民族。

### 3.2 贵州壮侗语族 7 个世居少数民族与其他民族以及其他语族之间的遗传关系

壮侗语族 7 个民族的单倍群频率与已有文献报道的 9 个少数民族人群的单倍群频率 (见表 4) 运用 SPSS 24 软件进行主成分分析。结果如图 3 所示,归属汉藏语系的壮侗语族 (水族、布依族、侗族、仡佬族、壮族、毛南族、仡佬族) 与苗瑶语族 (苗族、瑶族、畲族) 聚在一起;提示壮侗语族与苗瑶语族之间的遗传关系较近,这与梁祚仁<sup>[25]</sup>在广西 10 个少数民族 Y 染色体 17 个 STR 基因座的遗传多态性研究中根据 14 个民族人群之间的遗传距离绘制的 N - J 系统进化树中壮侗语族民族群体与苗瑶语族民族群体聚为一类的研究结果相一致,也与何燕等<sup>[26]</sup>的研究结果吻合。

为了验证贵州省壮侗语族和其他语族之间的关系,将贵州省壮侗语族 7 个民族的数据通过直接计数法获得贵州省壮侗语族的单倍群频率,通过文献报道的民族单倍群频率计算语族的单倍群频率 (见表 5),然后进行主成分分析,得到图 4,结果显示汉藏语系 (壮侗语族、苗瑶语族、汉语族) 聚在一起,归属阿尔泰语系的突厥语族、蒙古语族聚在一起后再与满通古斯语族相聚,而归属汉藏语系的藏缅语族位于上述两聚群之间,这与各族民族的地理分布、历史源流等相对应。其中归属汉藏语系的壮侗语族与苗瑶语族发生了部分重叠,进一步提示了两语族民族由于长期毗邻而居,可能发生了基因交融,显示出较近的亲缘关系。

### 参考文献:

- [1] BRIÓN M, SANCHEZ J J, BALOGH K, et al. Analysis of 29 Y-chromosome SNPs in a single multiplex useful to predict the geographic origin of male lineages [J]. International Congress, 2006, 1288: 13 - 15.
- [2] WETTON J H, TSANG K W, KHAN. Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs [J]. Forensic Science International, 2005, 152(1): 45 - 53.
- [3] INÁCIO A, COSTA H A, SILVA C V D, et al. Study of Y-SNPs genetic markers with forensic interest and ancestry informative power in PALOP's immigrant populations in Lisboa [J]. Forensic Science International (Genetics Supplement Series), 2015, 5: e3 - e4.
- [4] GAUDÊNCIO S, COSTA H A, AMORIM A, et al. Y-SNPs genetic variation in the Caucasian population living in southern Portugal [J]. Forensic Science International (Genetics Supplement Series), 2015, 5: e344 - e345.
- [5] VILLAESCUSA P, ILLESCAS M J, VALVERDE L, et al. Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain [J]. Forensic Science International Genetics, 2016, 27: 142 - 148.
- [6] 张雯. Y 染色体 O1 单倍群 SNP 位点在广西六个民族中遗传多态性 [D]. 昆明: 昆明医科大学, 2017.

- haplogroup in six ethnic groups of Guangxi, Southwest of China[D]. Kunming; Kunming Medical University, 2017.
- [7] MEDINA L S J, MUZZIO M, SCHWAB M, et al. Human Y-chromosome SNP characterization by multiplex amplified product-length polymorphism analysis [J]. Electrophoresis, 2015, 35(17): 2524 - 2527.
- [8] OCHIAI E, MINAGUCHI K, NAMBIAR P, et al. Evaluation of Y chromosomal SNP haplogrouping in the HID-Ion AmpliSeq identity panel[J]. Legal Medicine, 2016, 22: 58 - 61.
- [9] 王文光, 李晓斌. 百越民族发展演变史: 从越、僚到壮侗语族各民族[M]. 北京: 民族出版社, 2007.
- [10] PARK M J, LEE H Y, KIM NY, et al. Y-SNP mini-plexes for East Asian Y-chromosomal haplogroup determination; in degraded DNA[J]. Forensic Sci Int Genet, 2013, 7(1): 75 - 81.
- [11] CAI X, QIN Z, WEN B, et al. Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes [J]. PLoS One, 2011, 6(8): e24282.
- [12] SHOU W H, QIAO E F, WEI C Y, et al. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians [J]. Journal of Human Genetics, 2010, 5(55): 314 - 322.
- [13] 谢选华. 从Y染色体遗传结构看藏缅语族人群的起源[D]. 上海: 复旦大学, 2004.
- XIE X H. The origin of Tibeto-Burman speaking group from the perspective of Y chromosome genetic structure [D]. Shanghai: Fudan University, 2004.
- [14] 文波. Y染色体、mtDNA多态性与东亚人群的遗传结构[D]. 上海: 复旦大学, 2004.
- WEN B. Y chromosome, mtdna polymorphism and genetic structure of East Asian population [D]. Shanghai: Fudan University, 2004.
- [15] 陆艳. 中国西部人群的遗传混合[D]. 上海: 复旦大学, 2011.
- LU Y. Genetic admixture in western China [D]. Shanghai: Fudan University, 2011.
- [16] ZHOU D, YAN J, TANG D Z, et al. Validation of a novel Y-SNPs multiplex system for forensic application [J]. Forensic Science International Genetics Supplement, 2013, 4(1): e314 - e315.
- [17] 黄玉静. 云南佤族、白族、傣族 Y-SNP 与 Y-STR 多态性研究与法医学意义[D]. 昆明: 昆明医学院, 2011.
- HUANG Y J. Genetic polymorphism study of Y-SNP and Y-STR in the Wa, Bai, Dai Ethnic Groups of Yunnan and their forensic implications [D]. Kunming: Kunming Medical University, 2011.
- [18] MEZZAVILLA M, GEPPERT M, TYLER-SMITH C, et al. Insights into the origin of rare haplogroup C3 \* Y chromosomes in South America from high-density autosomal SNP genotyping [J]. Forensic Science International: Genetics, 2015, 15: 115 - 120.
- [19] 刘世彬. 贵州毛南族(佯僮人)族源探析[J]. 黔南民族师范学院学报, 2009, 29(2): 33 - 36.
- LIU S B. On the ethnic source of the Maonans (People of Yangguang) in Guizhou Province [J]. Journal of the Qiannan Normal College for Nationalities, 2009, 29(2): 33 - 36.
- [20] 刘树虎, 热比亚木·巴克, 伊力哈木·乃扎木, 等. 和田维吾尔族人群 Y 染色体遗传多样性分析[J]. 生物技术, 2017, 27(1): 38 - 44.
- LIU S H, RABIYAMU B, BUPATIMA A, et al. Y-chromosome genetic diversity of the uygur populations in Hedian [J]. Biotechnology, 2017, 27(1): 38 - 44.
- [21] XIAOYUN C, ZHENDONG Q, BO W, et al. Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes [J]. PLoS One, 2011, 6(8): e24282.
- [22] 韦小鹏. 壮族婚姻圈的变迁[D]. 桂林: 广西民族大学, 2008.
- WEI X P. The changes of the Zhuang's marriage circles [D]. Guilin: Guangxi University for Nationalities, 2008.
- [23] 周颖虹. 毛南族族群认同研究[D]. 桂林: 广西师范大学, 2007.
- ZHOU Y H. The study of Maonan Ethnic group identity [D]. Guilin: Guangxi Normal University, 2007.
- [24] 杨波. 黎族三个支系 Y-SNP 多态性研究[D]. 西安: 第三军医大学, 2007.
- YANG B. Populations genetic analysis of Y-chromosomal single nucleotide polymorphism in three branches of Hlai [D]. Xi'an: Third Military Medical University, 2007.
- [25] 梁祚仁. 广西 10 个少数民族 Y 染色体 17 个 STR 基因座的遗传多态性[D]. 广州: 广州医学院, 2009.
- LIANG Z R. Polymorphism of 17 Y-chromosome short tandem repeats loci in 10 minority populations in Guangxi of China [D]. Guangzhou: Guangzhou Medical University, 2009.
- [26] 何燕, 文波, 单可人, 等. 贵州三都水族 Y 染色体单倍型频率分析[J]. 遗传, 2003, 25(3): 249 - 252.
- HE Y, WEN B, SHAN K R B, et al. The distribution of Y-chromosome haplotypes of Shui Ethnic in Sandu, Guizhou [J]. Hereditas, 2003, 25(3): 249 - 252.