

广西仫佬族线粒体 DNA 高变 I 区多态性分析

王晓庆¹ 王传超² 邓琼英^{1*} 李辉²

(1. 广西医科大学基础医学院人体解剖学教研室, 南宁 530021;

2. 复旦大学生命科学学院现代人类学教育部重点实验室, 上海 200433)

[摘要] 目的 探讨广西仫佬族人群线粒体 DNA 高变 I 区的多态性特征, 了解仫佬族群体的母系遗传结构。方法 收集 91 例仫佬族无关男性个体的外周血样本, 目标序列用引物 L15947 和 R16488 进行 PCR 扩增, 用 ABI 3730 测序仪正反向测序; 计算多态性位点、核苷酸多态性、平均配对差异数目等多态性指标, 以及仫佬族与各民族之间的遗传距离, ME 法构建遗传进化树。结果 与修正后的剑桥标准序列 (rCRS) 比对, 91 例样本的线粒体高变 I 区序列共界定了 74 种单倍型, 核苷酸多态性为 0.0188 ± 0.010 , 平均核苷酸差异为 6.618 ± 3.154 ; 遗传距离显示仫佬族与南方各少数民族及南方汉族的亲缘关系较近, 与北方汉族的亲缘关系较远, 进化树中仫佬族与南方少数民族聚为一类。结论 仫佬族在母系遗传上属于典型的南方侗傣族群, 可能经历过群体扩张或选择效应; 高变 I 区序列具有较高的多态性, 可用于法医个体识别、民族起源等方面的研究。

[关键词] 广西; 线粒体 DNA; 高变 I 区; 测序; 遗传多态性; 聚合酶链反应; 仫佬族

[中图分类号] R394.5 **[文献标志码]** A **[DOI]** 10.3969/j.issn.0529-4356.2013.04.022

Genetic polymorphism of mitochondrial DNA hypervariable segment I of Mulam ethnic group in Guangxi

WANG Xiao-qing¹, WANG Chuan-chao², DENG Qiong-ying^{1*}, LI Hui²

(1. Department of Anatomy, Guangxi Medical University, Nanning 530021, China;

2. MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China)

[Abstract] **Objective** To demonstrate the polymorphism of the mitochondrial DNA (mtDNA) hypervariable segment I (HVS I) and to clearly understand the maternal genetic structure of Mulam ethnic group in Guangxi. **Methods** Venous blood samples from 91 unrelated Mulam individuals were collected. HVS I of mtDNA was amplified using primers L15947 and R16488. PCR amplicons were sequenced in both forward and reverse directions using the ABI 3730. Polymorphisms were scored relative to the revised Cambridge reference sequence (rCRS). The number of polymorphic sites, number of haplotypes, the mean number of pairwise differences and genetic distances between populations were calculated. Polygenetic tree was drawn based on dA distance. **Results** MtDNA HVS I sequences showed high levels of diversity. Seventy-four haplotypes were identified among the 91 samples. The nucleotide diversity was 0.0188 ± 0.010 and the mean number of pairwise differences was 6.618 ± 3.154 . Genetic distances between populations and the cluster map show that Mulam falls together with southern groups, but away from northern populations. Among southern populations, Mulam clusters were most closely with the Zhuang and Sui. **Conclusion** Mulam is a typical southern Kam-Tai population, and may experience population expansion or selection. MtDNA HVS-I reveals highly genetic diversities in Mulam, which may have potential use in population genetics and forensic practice.

[Key words] Guangxi; mtDNA; HVS I; Sequencing; Genetic polymorphism; PCR; Mulam

仫佬族是我国历史悠久的一个山地民族, 绝大多数居住在广西罗城仫佬族自治县。以往有关仫佬族起源问题的研究主要停留在历史学、语言学和体

质特征等方面的研究^[1-3], 而这些材料或研究的因素往往容易受到客观条件的限制, 难以获得直接证据。人体中的 DNA, 具有稳定的世代遗传的特点, 近年来随着 DNA 分析技术的发展, 特别是母系线粒体 DNA (mitochondrial DNA, mtDNA) 和父系 Y 染色体非重组区 (non-recombining region of Y chromosome, NRY) 的研究技术日趋完善, 这两组遗传标记系统已经成为世界公认的解读人群起源、迁徙和演化的最重要工具。迄今, 有关仫佬族 Y 染色

[收稿日期] 2012-12-14 **[修回日期]** 2013-01-09

[基金项目] 广西自然科学基金资助项目(2011GXNSFA018240)

[作者简介] 王晓庆(1980—), 女(汉族), 山东省泰安市人, 在读硕士研究生。

* 通讯作者(To whom correspondence should be addressed)

E-mail: yingzideng@sina.com Tel: (0771) 5358854

体非重组区域的研究已经完成^[4] , 而从 mtDNA 的角度探讨仫佬族母系起源的研究尚未见报道。

线粒体 DNA 是人类细胞中唯一存在于细胞核外的基因组 , 由约 16 569 个碱基对组成双链闭合分子。在线粒体 DNA 中有一段约为 1100bp 的具有较高多态性的非编码区 , 又称控制区 (control region , CR) 或 D 环区 (D-loop) , 具有高变 I 区 (hypervariable segment I , HVS I) 和高变 II 区两个区 , 严格遵循母系遗传 , 同时具有多拷贝、突变率高及缺乏重组等特点 , 使其成为研究法医个体识别及群体遗传等方面的最理想遗传标记。

本研究从分子生物学角度对广西仫佬族群体的线粒体 DNA 高变 I 区进行测序 , 对所获得的 DNA 序列进行多态性分析并探讨仫佬族与其他民族的亲缘关系 , 获得了该民族的母系遗传特征。

材料和方法

1. 样本及 DNA 制备

91 例样本均采自广西罗城仫佬族自治县 , 并追溯其 3 代以上的家族史及居住史均为仫佬族的无亲缘关系的健康男性个体 , 采集其静脉血 2 ml , 酚氯仿法提取基因组 DNA。

2. 线粒体高变 I 区的扩增及测序

选用引物 L15947 和 R16488 对目的片段进行扩增 , 反应体系为 15 μ l , 其中包括 DNA 模板 1 μ l、Taq DNA 聚合酶 0.1 μ l、10X PCR buffer 1.5 μ l、1.2 μ l MgCl₂、dNTPs 1.5 μ l、正反向引物 0.2 μ l 及灭菌双蒸水 9.5 μ l。循环条件为 94 $^{\circ}$ C 预变性 3min; 然后 94 $^{\circ}$ C 变性 30s、62 $^{\circ}$ C 退火 30s、72 $^{\circ}$ C 延伸 1min , 循环 30 次后 72 $^{\circ}$ C 延伸 6min 结束。

扩增产物纯化后作为测序模板 , 由于高变 I 区有一段 poly C 处可能酶促反应无法进行 , 故用 PE 公司的 Big Dye Terminator Sequencing 试剂盒进行双向测序 , 测序反应体系 3 μ l 包含 5X sequence buffer 1.2 μ l、单向引物 0.06 μ l、Bigdye 0.4 μ l 及灭菌双蒸水 1.34 μ l。测序循环条件: 96 $^{\circ}$ C 预变性 1min , 96 $^{\circ}$ C 60s、50 $^{\circ}$ C 5s、60 $^{\circ}$ C 4min 25 个循环。测序产物纯化后在 ABI 3730 测序仪 (Appliedbiosystems , Carlsbad , CA) 上电泳。

3. 数据分析

高变 I 区序列用 DNASTAR-Seqman 软件将正反向测序结果拼接完整 , 并参照修正后的剑桥标准序列 (revised Cambridge Reference Sequence , rCRS) ^[5] 进行比对 , Bioedit 编辑比对结果。由于已经发表的一些其他民族的数据长度只有 360bp , 所以我们截取仫佬族序列中的相应片段 (rCRS: 16024 ~ 16383) , 利用 Arlequin 3.5 软件计算各种多态性指

标 , 包括多态性位点 (A) 、核苷酸多态性 (nucleotide diversity) 、平均配对差异数目 (average number of pairwise difference) 等 , 进行 Tajima's *D* 和 Fu's *F_s* 中性检验。

结合已经报道的青岛汉族、辽宁汉族、广东汉族、云南汉族、武汉汉族^[6] 和壮族、傣族^[7] , 水族、毛南族^[8] 和本实验所得仫佬族共 10 个群体 , 用 Arlequin 3.5 软件计算群体间的 *F_{st}* 和 dA 遗传距离。利用 SPSS 18.0 统计学软件对这两种遗传距离进行相关性检验。用最小进化法 (minimum evolution , ME) 在 MEGA 4.0 平台上计算群体间的净遗传距离 dA 矩阵 , 并绘制遗传进化树。

结 果

1. 仫佬族 mtDNA HVS I 多态性指标

本研究中仫佬族 mtDNA HVS I 的测序产物共 480bp , 与剑桥标准序列比对统计变异位点。91 例样本共发现 74 种单倍型 , 75 个多态位点 (图 1) , 图中第 1 行为碱基序列号 , 第 2 行为剑桥标准序列碱基。利用 360bp 序列长度计算出的核苷酸多态性为 0.0188 \pm 0.010 , 平均核苷酸差异为 6.619 \pm 3.154。依据中性检验理论 , 应用 Tajima's *D* 检验和 Fu's *F_s* 检验 , 对 8 个群体的线粒体高变 I 区序列多样性信息进行检验 , 发现 *D* 值或者 *F_s* 值均为负值。仫佬族与其他 7 个群体的遗传多态性指标见表 1 , 用于比较的数据来源于已发表文献 [8 9]。

2. 仫佬族与其他民族的遗传距离及聚类分析

根据线粒体高变 I 区的数据计算得到了 10 群体的 *F_{st}* 和 dA 遗传距离矩阵 (表 2) 。通过相关性检验 , 两种矩阵具有很好的相关性 , Pearson 相关系数为 0.995 *P* < 0.01。从遗传距离矩阵来看 , 仫佬族与南方各少数民族及南方汉族的遗传距离较近 , 尤其与壮族、水族的距离最近 , 与北方汉族的距离较远。进化树结果显示仫佬族与南方少数民族聚为一类 , 各北方汉族聚为一类 , 由此可见 , 仫佬族与广西各少数民族间的亲缘关系较近 , 而与各地汉族的亲缘关系较远 (图 2) 。

讨 论

仫佬族是广西古老的土著民族 , 历年来 , 通过语言、体质特征、体型和 DNA 等探索仫佬族的源流问题一直是学术界经久不息的热点。在语言学上 , 仫佬语属于汉藏语系-壮侗语族-侗水语支。王均等^[1] 研究认为 , 仫佬语无论在词汇、语法还是语音方面都与侗语最为接近; 范宏贵等^[10] 也认为仫佬语是壮语、侗语过程中相当重要的中间环节。在体质人类学方面 , 诸多研究表明 , 仫佬族不管在皮下脂肪还

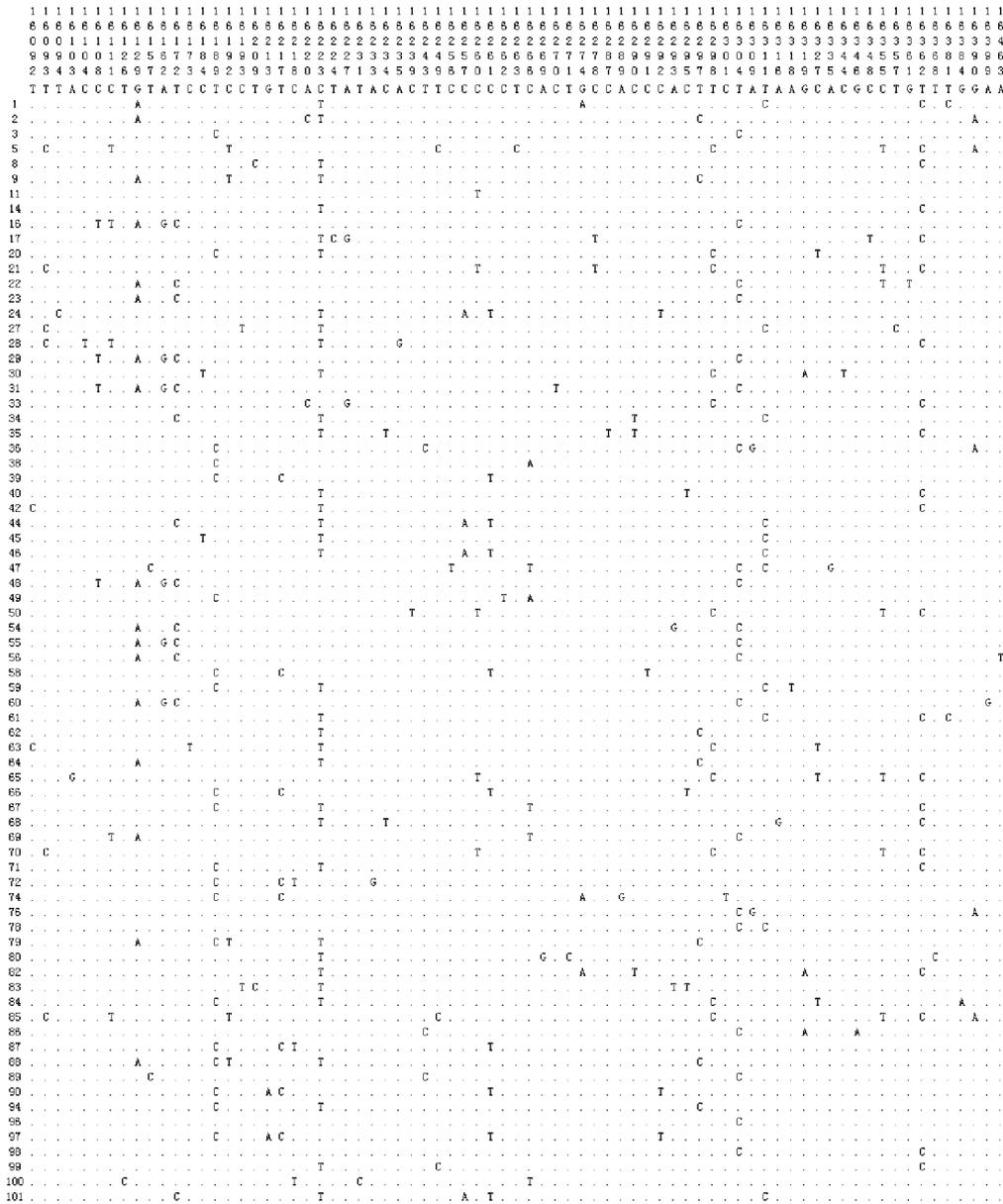


图1 仫佬族 mtDNA HVS-I 单倍型图和多态位点

Fig. 1 Polymorphic sites and haplotype map of mtDNA HVS I in Mulam population(n =91)

表1 仫佬族 mtDNA HVS I 多态性指标与其他民族群体的比较

Table 1 Comparison of genetic diversity of HVS I between Mulam and other populations

群体	样本数目	多态位点	平均核苷酸配对差异数目	核苷酸多态性	θ (s)	θ (pi)	F_s 值	D 值	P 值
Pop	N	S	P_i	π	Theta (S)	Theta (Pi)	Fu's F_s	Tajima's D	$P(D)$
仫佬族 (Mulam)	91	71	6.619(3.154)	0.018(0.010)	13.969(3.767)	6.618(3.494)	-25.046	-1.724	0.017
水族 (Shui)	30	38	6.906(3.341)	0.019(0.010)	9.592(3.280)	6.906(3.718)	-25.117	-1.028	0.151
壮族 (Zhuang)	83	65	7.328(3.464)	0.020(0.011)	13.026(3.590)	7.328(3.838)	-24.922	-1.44	0.035
毛南 (Maonan)	32	33	6.462(3.140)	0.018(0.010)	8.194(2.815)	6.462(3.493)	-25.194	-0.761	0.218
白族 (Bai)	55	78	7.343(3.488)	0.020(0.011)	17.048(4.941)	7.343(3.870)	-24.99	-1.983	0
苗族 (Miao)	142	87	6.482(3.085)	0.018(0.009)	15.734(3.910)	6.482(3.415)	-24.913	-1.863	0.004
彝族 (Yi)	87	91	8.057(3.778)	0.022(0.012)	18.065(4.807)	8.057(4.185)	-24.779	-1.843	0.011
傣族 (Dai)	38	59	8.283(3.924)	0.023(0.012)	14.042(4.429)	8.283(4.360)	-24.856	-1.484	0.061

N . 样本数目; S . 多态性位点数; P_i . 平均核苷酸配对差异数目; π . 核苷酸多态性

N , Number of individuals; S , Number of variable nucleotide positions; P_i , Average number of pairwise differences; π , Nucleotide diversity

是体型方面都略逊于北方民族,究其原因除了遗传因素外,还与地理环境、生活习俗等有很大关系^[2,11]。近年来,随着分子生物学技术的发展,DNA在辨析仫佬族源流问题上也起到了重要作用。李辉

等^[4]通过研究广西罗城仫佬族 Y 染色体单倍群,认为仫佬族属百越民族系统的侗水族群。徐林等^[3]对仫佬族 15 个常染色体 STR 位点进行扩增,通过与其他民族进行主成分分析、遗传距离等多元的统

表 2 根据线粒体高变 I 区计算的 10 个群体的遗传距离

Table 2 The genetic distance among 10 populations based on mtDNA HVS I

	佤族 Mulam	广东汉 GD Han	辽宁汉 LN Han	青岛汉 QD Han	武汉汉 WH Han	云南汉 YN Han	水族 Shui	壮族 Zhuang	毛南族 Maonan	傣族 Dai
佤族(Mulam)										
广东汉(Guangdong Han)	0.020		0.175	0.175	0.192	0.154	0.053	0.105	0.244	0.147
辽宁汉(Liaoning Han)	0.025	0.007		0.031	0.045	0.003	0.321	0.084	0.299	0.283
青岛汉(Qingdao Han)	0.025	0.005	-0.044		0.053	0.162	0.433	0.184	0.470	0.384
武汉汉(Wuhan Han)	0.028	0.006	0.007	0.011		0.050	0.261	0.070	0.373	0.236
云南汉(Yunnan Han)	0.023	0.000	0.020	0.020	0.006		0.197	-0.018	0.243	0.216
水族(Shui)	0.008	0.042	0.054	0.061	0.034	0.024		0.084	0.255	0.068
壮族(Zhuang)	0.015	0.012	0.024	0.027	0.010	-0.002	0.011		0.204	0.149
毛南(Maonan)	0.036	0.041	0.060	0.064	0.049	0.031	0.037	0.027		0.248
傣族(Dai)	0.022	0.034	0.046	0.051	0.029	0.026	0.008	0.020	0.032	

对角线左下角是 *Fst* 遗传距离, 对角线右上角是 dA 遗传距离。

The data of the left bottom represent *Fst* distance, the data of the right upper represent dA distance.

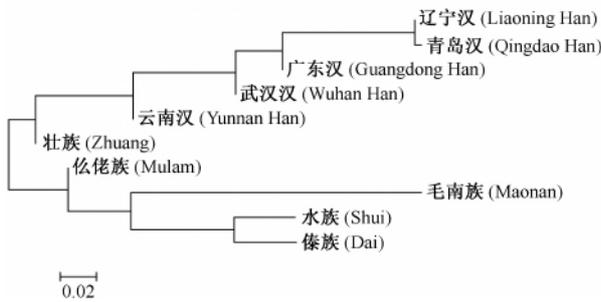


图 2 基于 dA 遗传距离的遗传进化树

Fig. 2 Polygenetic tree based on dA distance

计分析表明, 佤族与毛南、水、苗、瑶等百越族群遗传距离较近, 与各北方汉族遗传距离较远。综上所述, 目前学术界普遍认同佤族来自南方百越族群, 但尚未见有关佤族母系起源的报道。在本研究中, 我们首次对佤族线粒体 DNA 的高变 I 区进行测序, 并计算了 *Fst* 和 dA 两种遗传距离和构建遗传进化树, 结果显示, 佤族在母系遗传结构与地理位置较近的南方群体有较近的亲缘关系, 尤其与水族、壮族的关系最近, 而与北方群体的关系较远, 可见佤族在母系遗传上属典型的南方侗傣族群。

不同群体或种族之间的差异主要是由 DNA 分子造成的, 研究这些 DNA 分子的差异可揭示民族的起源、进化及迁徙情况。比较常用的度量 DNA 分子多态性的指标是分离位点数及核苷酸多态性, Tajima 检验通过计算分离位点数和任意两条序列之间差异的平均数来评估是否遵循中性理论^[12]。从佤族与南方各民族的多态性指标对比来看, 各民族 Tajima' *D* 值均为负值, 均偏离中性理论的模型, 其中佤族、壮族、白族、苗族和彝族的 *D* 值差异显著, 具有统计学意义, 提示在这些民族群体的进化过程受到了基因随机漂变以外的影响, 有可能发生过大规模的群体扩张或者受到过选择作用; 而水族、毛南族和傣族的差异不显著, 可能是由于这 3 个群体所选的样本量太小造成的。

参 考 文 献

[1] Wang J, Zheng GQ. A Sketch of Mulam [M]. Beijing: Ethnic Publishing House, 1980: 1-5. (in Chinese)
王均, 郑国乔. 佤语简志 [M]. 北京: 民族出版社, 1980: 1-5.

[2] Zheng LB, Lu ShH, Ding B. A study on the physical characteristics of Mulam nationality [J]. Acta Anthropologica Sinica, 2006, 25 (3): 242-250. (in Chinese)
郑连彬, 陆舜华, 丁博. 佤族体质特征研究 [J]. 人类学学报, 2006 25 (3): 242-250.

[3] Xu L, Deng QY, Li SF, et al. Genetic analysis of Mulao nationality using 15 short tandem repeats [J]. Chinese Journal of Medical Genetics, 2008 25 (1): 96-100. (in Chinese)
徐林, 邓琼英, 李松峰, 等. 佤族 15 个短串联重复序列的遗传分析 [J]. 中华医学遗传学杂志, 2008 25 (1): 96-100.

[4] Li H. Preliminary analyses of genetic structure of daic populations [C]. Shanghai: Proceedings of The International Symposium of Anthropological Studies, 2002: 89-94. (in Chinese)
李辉. 百越族群遗传结构分析的初步思路 [C]. 上海: 现代人类学国际研讨会论文集, 2002: 89-94.

[5] Andrews RM, Kubacka I, Chinnery PF, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA [J]. Nat Genet, 1999, 23 (2): 147.

[6] Yao YG, Kong QP, Bandelt HJ, et al. Phylogeographic differentiation of mitochondrial DNA in Han Chinese [J]. Am J Hum Genet, 2002, 70 (3): 635-651.

[7] Yao YG, Nie L, Harpending H, et al. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity [J]. Am J Phys Anthropol, 2002, 118 (1): 63-76.

[8] Li H, Cai X, Winograd-Cort ER, et al. Mitochondrial DNA diversity and population differentiation in southern East Asia [J]. Am J Phys Anthropol, 2007, 134 (4): 481-488.

[9] Wen B, Li H, Gao S, et al. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages [J]. Mol Biol Evol, 2005, 22 (3): 725-734.

[10] Fan HG. Homologous Nationalities: the Origin and Cultural of Zhuang and Tai [M]. Beijing: Ethnic Publishing House, 2007: 127-128. (in Chinese)
范宏贵. 同根生的民族: 壮泰各族渊源与文化 [M]. 北京: 民族出版社, 2007: 127-128.

[11] Deng QY, Gong JG, Zhou LN, et al. A study on the developmental tendency of subcutaneous fat of Mulao adults in Guangxi [J]. Anatomical Research, 2011 33 (1): 49-51. (in Chinese)
邓琼英, 龚建古, 周丽宁, 等. 广西佤族成人皮下脂肪发育趋势的研究 [J]. 解剖学研究, 2011 33 (1): 49-51.

[12] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism [J]. Genetics, 1989, 123 (3): 585-595.

(编辑 张艳)