

广西融水苗族 3 个 STR 基因座的 群体遗传学研究

周丽宁¹, 徐林¹, 李松锋¹, 龚继春¹, 魏博源¹, 邓祥发¹, 玉洪荣

(1. 广西医科大学解剖教研室人类学研究室, 南宁 530021)

摘要: 为了解广西融水苗族人群无关个体的 3 个短串联重复序列(short tandem repeat, STR): *HUMCSF1PO*, *HUMTPOX*, *HUMTH01* 遗传多态性分布情况, 本文用枸橼酸钠抗凝法采集血样, 酚-氯仿抽提法提取 DNA, 应用复合扩增技术对血样 DNA 的 3 个 STR 基因座进行扩增和检测。结果显示: 在三个 STR 位点共检测出 19 种等位基因, 48 种基因型, 频率分布分别在 0.0024—0.4663 和 0.0048—0.3173 之间; 基因型的分布符合 Hardy-Weinberg 平衡定律($P > 0.05$)。计算种族、民族之间的遗传距离, 并对之进行比较得出: 广西融水苗族与美国高加索人及美国非洲人存在显著差异, 且与美国非洲人之间的差异大于与美国高加索人之间的差异; 广西融水苗族与广西侗族的关系近于与其他少数民族的关系。

关键词: STR; 基因频率; 遗传距离; 苗族

中图分类号: Q987 文献标识码: A 文章编号: 1000-3193 (2005) 04-0307-08

短串联重复序列(Short Tandem Repeat, STR)是广泛存在于人类基因组中的一类具有片段长度多态性的 DNA 序列。其多态性主要由 2-7bp 组成的核心序列拷贝数目的变化而产生^[1]。STR 基因座具有高度的多态性和遗传稳定性, 符合孟德尔遗传规律^[2], 因其分布广、密度大、多态信息量高和易标准化等优点而成为人类基因组测定、致病基因和相关基因的寻找、法医鉴定、群体遗传学研究和生物考古学等方面理想的研究对象^[3]。四联体基因座 *HUMCSF1PO*, *HUMTPOX*, *HUMTH01* 的扩增条件及等位基因片段大小相近, 可同时扩增分析, 合称 *CTT* 位点, 是目前各国进行人类学、法医学研究常用的位点。美国 Promega 公司 1996 年首先推出了商品化的 *CTT* 复合扩增试剂盒^[4] 和银染试剂盒, 并调查了美国部分人群, 取得了满意的结果。随后, 国内学者邹浪萍^[5-6]、邹平^[7]、程宝文^[8] 等相继在不同民族群体中对 *CTT* 位点进行了检测, 并将之应用于群体遗传学研究和法医学鉴定, 均取得了满意的结果。本文随机抽样 208 名广西融水苗族健康的无关个体, 采用 STR 基因扫描技术对 *CTT* 位点的遗传结构和变化规律进行研究, 为建立广西融水苗族 STR 基因座等位基因和基因型频率数据库、了解民族起源、开展民族的群体遗传学研究和法医学鉴定提供基础数据。

收稿日期: 2004-06-28; 定稿日期: 2005-01-12

基金项目: 国家自然科学基金资助项目(编号:30260044)

作者简介: 周丽宁(1970-), 女, 汉族, 广西藤县人, 广西医科大学解剖教研室讲师, 硕士, 从事人体解剖学教学和体质人类学、分子生物学研究。E-mail: liningzhgx@163.com

1 材料与方法

1.1 研究对象

根据“知情同意”原则, 208 份血样采自广西融水苗族自治县民族中学的苗族学生(男 102 例, 女 106 例), 平均年龄 10—15 岁, 均身体健康, 彼此间无亲缘关系, 追溯三代皆为苗族。

1.2 方法

1.2.1 采样与 DNA 提取

抽取肘正中静脉血 2ml, 枸橼酸钠抗凝, 酚- 氯仿抽提法提取全血 DNA, TE 缓冲液溶解, EP 管分装, 4 °C 冰箱保存备用。

1.2.2 PCR 扩增

用 AmpFSTR® Identifiler™ PCR Amplification kit 进行复合扩增, 在反应总体积为 10μl 反应体系中进行扩增: PCR Reaction Mix 4. 4μl, Primer Set 2. 2μl, AmpliAq Gold® DNA Polymerase (5U/μl) 0. 2μl, 模板 DNA (0. 5—1. 25ng) 1μl, 补水至 10μl。扩增使用 ABI9700 型扩增仪, 热循环条件: 96 °C 预变性 11min, 然后 (94 °C 1min → 59 °C 1min → 72 °C 1min) × 28 个循环 → 60 °C 60min → 4 °C 保温。扩增产物置于 4 °C 避光保存备用。

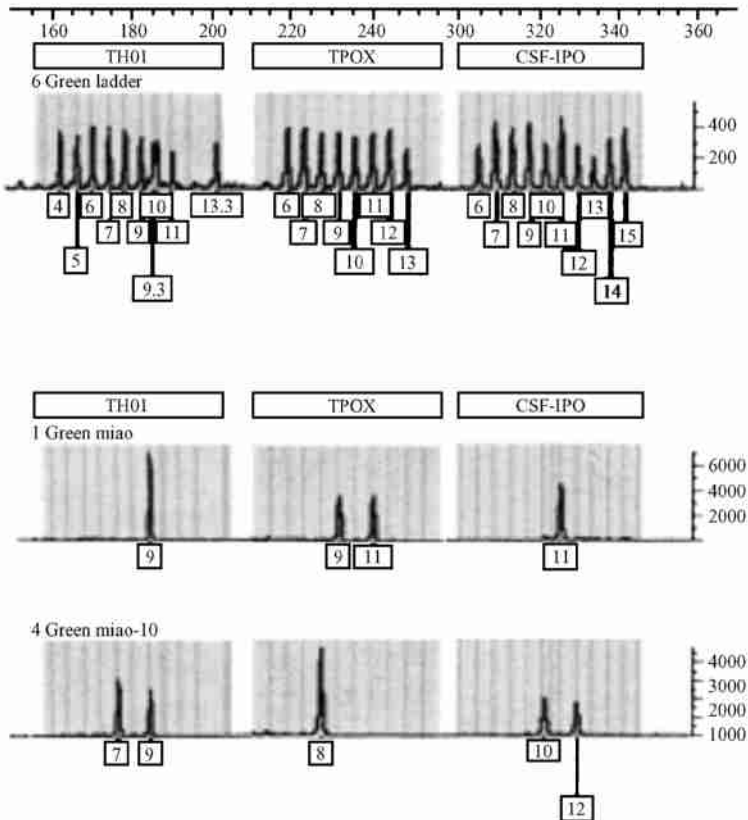


图 1 GeneScan 扫描图谱基因分型结果

The result of GeneScan

1.2.3 扩增产物分离

用 ABI Prism[®] 3100 型遗传分析仪进行毛细管电泳和检测。用 GeneScan Analysis 3.7 和 Genetyper 3.7 软件进行分析。

1.2.4 统计学分析

①用 Excel 和 SPSS 等统计软件对实验数据进行统计学分析处理。②按文献[9]对基因型分布进行 Hardy-Weinberg 平衡吻合性检验并计算杂合度、个体识别率、非父排除率、多态信息总量及累积非父排除率、累积个体识别率、累积多态信息量等。③用公式^[10] $I_k = \sum a_i b_j / (\sum a_i^2 \sum b_j^2)^{1/2}$ 和公式 $D = -\log_e I$ 或 $D = -\ln I$ 分别计算两群体间的遗传同一性及遗传距离 (其中 a_i, b_j 分别代表群体 A、群体 B 的等位基因)。

2 结果

2.1 三个 STR 基因座基因分型结果(图 1)

2.2 三个 STR 位点等位基因频率分布(表 1, 图 2—1、2—2、2—3)

2.3 三个 STR 位点的基因型频率分布(表 2)

表 1 广西融水苗族群体三个 STR 基因座等位基因频率分布

Allele frequency distribution of three STR loci in Miao

TPOX			TH01			CSF-IPO		
等位基因 Allele	计数 Count	基因频率 Frequency	等位基因 Allele	计数 Count	基因频率 Frequency	等位基因 Allele	计数 Count	基因频率 Frequency
6	0	0.000 0	4	0	0.000 0	6	0	0.000 0
7	0	0.000 0	5	0	0.000 0	7	8	0.019 2
8	192	0.461 5	6	59	0.141 8	8	0	0.000 0
9	59	0.141 8	7	103	0.247 6	9	9	0.021 6
10	9	0.021 6	8	18	0.043 3	10	91	0.218 8
11	146	0.351 0	9	194	0.466 3	11	107	0.257 2
12	10	0.024 0	9 3	15	0.036 1	12	171	0.411 1
13	0	0.000 0	10	26	0.062 5	13	28	0.067 3
			11	1	0.002 4	14	2	0.004 8
			13 3	0	0.000 0	15	0	0.000 0
合计 total	416	1.000 0		416	1.000 0		416	1.000 00

表 2 广西融水苗族群体三个 STR 基因座的基因型频率分布与 H-W 平衡检验

Genotypes distribution of three STR loci in Miao

STR 位点 Locus	基因型 Genotype	观察值(Observed)		期望值(Expected)		Hardy-Weinberg 平衡检验
		计数	频率	计数	频率	
		Count	Frequency	Count	Frequency	
TPOX	8 \ 8	48	0.230 8	44.307 7	0.214 9	$\chi^2 = 4.06$ df= 6 $P > 0.5$
	8 \ 9	22	0.105 8	27.230 8	0.132 1	
	8 \ 10	4	0.019 2	4.153 8	0.020 1	
	8 \ 11	66	0.317 3	67.384 6	0.326 9	
	8 \ 12	5	0.024 0	4.615 4	0.022 4	
	9 \ 9	4	0.019 2	4.183 9	0.020 3	
	9 \ 10	2	0.009 6	1.276 4	0.006 2	
	9 \ 11	26	0.125 0	20.706 7	0.100 4	
	10 \ 11	3	0.014 4	3.158 7	0.015 3	
	11 \ 11	23	0.110 6	25.620 2	0.124 3	
	11 \ 12	5	0.024 0	3.509 6	0.017 0	

续表 2

STR 位点 Locus	基因型 Genotype	观察值(Observed)		期望值(Expected)		Hardy-Weinberg 平衡检验
		计数	频率	计数	频率	
		Count	Frequency	Count	Frequency	
TH01	6\ 6	7	0.0337	4.1839	0.0203	$\chi^2 = 15.03$ df= 13 P > 0.25
	6\ 7	10	0.0481	14.6082	0.0709	
	6\ 8	2	0.0096	2.5529	0.0124	
	6\ 9	23	0.1106	27.5144	0.1336	
	6\ 9.3	4	0.0192	2.1274	0.0103	
	6\ 10	5	0.0240	3.6875	0.0179	
	6\ 11	1	0.0048	0.1418	0.0007	
	7\ 7	11	0.0529	12.7512	0.0619	
	7\ 8	4	0.0192	4.4567	0.0216	
	7\ 9	57	0.2740	48.0337	0.2332	
	7\ 9.3	4	0.0192	3.7139	0.0180	
	7\ 10	6	0.0288	6.4375	0.0313	
	8\ 8	1	0.0048	0.3894	0.0019	
	8\ 9	9	0.0433	8.3942	0.0408	
	8\ 9.3	1	0.0048	0.6490	0.0032	
	8\ 10	1	0.0048	1.1250	0.0055	
	9\ 9	43	0.2067	45.2356	0.2197	
9\ 9.3	6	0.0288	6.9952	0.0340		
9\ 10	12	0.0577	12.1250	0.0589		
10\ 10	1	0.0048	0.8125	0.0039		
CSF1PO	7\ 10	2	0.0096	1.7500	0.0086	$\chi^2 = 12.47$ df= 10 P > 0.25
	7\ 11	2	0.0096	2.0577	0.0101	
	7\ 12	4	0.0192	3.2885	0.0161	
	9\ 10	3	0.0144	1.9688	0.0096	
	9\ 11	1	0.0048	2.3149	0.0113	
	9\ 12	5	0.0240	3.6995	0.0181	
	10\ 10	6	0.0288	9.9531	0.0488	
	10\ 11	23	0.1106	23.4063	0.1147	
	10\ 12	43	0.2067	37.4063	0.1832	
	10\ 13	6	0.0288	6.1250	0.0300	
	10\ 14	1	0.0048	0.4375	0.0021	
	11\ 11	14	0.0673	13.7608	0.0674	
	11\ 12	47	0.2260	43.9832	0.2155	
	11\ 13	7	0.0337	7.2019	0.0353	
	12\ 12	29	0.1394	35.1454	0.1722	
12\ 13	14	0.0673	11.5096	0.0564		
13\ 14	1	0.0048	0.1346	0.0007		

2.4 群体遗传学资料

广西融水苗族 3 个 STR 基因座杂合度、个体识别力、非父排除率、多态信息总量等群体遗传学资料见表 3。

表 3 广西融水苗族群体三个 STR 基因座的遗传学指标
 Statistical parameters of three STR loci in GuangXi Miao group

STR 位点 Locus	观测杂合度 (Hobs)	期望杂合度 (Hexp)	个体识别力 (DP)	非父排除率 (EPP)	多态信息总量 (PIC)
TPOX	0.6394	0.6442	0.8048	0.4619	0.5809
TH01	0.6971	0.6957	0.8549	0.5796	0.6519
CSF1PO	0.7644	0.7133	0.8613	0.5770	0.6637
平均值	0.7003 > 0.7		0.8403 > 0.8	0.5395 > 0.5	0.6322 > 0.5
累积效力			0.9961	0.9043	0.9509

2.5 遗传距离

种族、民族间遗传距离计算结果见表 4。

表 4 遗传距离计算结果

The result of figuring genetic distance

	广西汉族 Guangxi Han	云南苗族 Yunnan Miao	非洲裔 美国人 Black	高加索裔 美国人 Caucasia	广西侗族 Guangxi Dong	蒙古族 Mongol	藏族 Zang	广西苗族 Guangxi Miao
广西壮族	0.027 7	0.029 2	0.101 8	0.142 4	0.029 1	0.036 0	0.048 6	0.039 1
广西汉族		0.031 9	0.117 4	0.143 5	0.039 4	0.051 7	0.046 5	0.046 6
云南苗族			0.147 2	0.139 7	0.017 0	0.011 8	0.015 4	0.012 8
非洲裔美国人				0.134 4	0.163 9	0.108 8	0.176 4	0.155 1
高加索裔美国人					0.164 2	0.099 9	0.144 1	0.137 2
广西侗族						0.025 9	0.024 9	0.010 4
蒙古族							0.013 4	0.012 2
藏族								0.011 8

2.6 种族、民族间的聚类分析 见图 3。

3 讨 论

3.1 广西融水苗族群体的遗传结构

3.1.1 等位基因与基因型

广西融水苗族,其聚居地多为山区,交通极为不便,且很少与外界通婚,可视之为隔离人群,他们在长期的生活中形成了自己的遗传特征。我们本次对他们的 *HUMTH01*、*HUMCSFIPO*、*HUMTPOX* 三个 STR 位点研究结果显示:*HUMTPOX* 检出 5 种等位基因,11 种基因型,最常见是等位基因 *HUMTPOX** 8 (频率为 0.4615);*HUMTH01* 检出 7 种等位基因,20 种基因型,最常见的是 *HUMTH01** 9 (频率为 0.4663);

HUMCSFIPO 检测出 7 种等位基因,17 种基因型,等位基因 *HUMCSFIPO** 12 为最常见(频率为 0.4111)。对三个 STR 位点基因型进行 χ^2 检验,均符合 Hardy-Weinberg 平衡定律 ($P > 0.05$)。表明该 3 个 STR 基因座在广西融水苗族人群中的分布较好,是一组较有应用价值的遗传标记,可为建立广西融水苗族 STR 位点等位基因和基因型频率数据库及开展其民族的群体遗传学研究和法医学鉴定提供基础数据。

3.1.2 群体遗传学分析

反映 DNA 多态性的遗传标记成千上万,不同标记的多态性、法医学应用价值及特点都

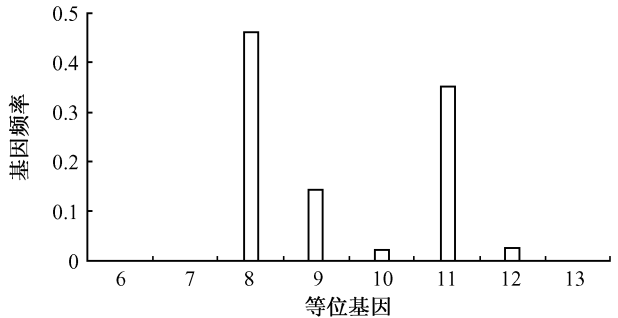


图 2-1 TPOX 基因座等位基因频率分布直方图

The Allele frequency distribution of TPOX locus

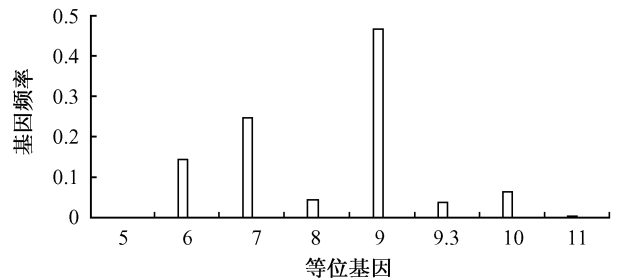


图 2-2 TH01 基因座等位基因频率分布直方图

The Allele frequency distribution of TH01 locus

可能不同。评价遗传标记的多态性及其应用价值的参数一般常用杂合度(H)、个体识别率(DP)、多态信息总量(PIC)、非父排除率(EPP)等来衡量:其中杂合度是群体在某一基因座上遗传变异程度的测度,能客观地反映出群体的遗传变异水平,平均杂合度越大,表明群体内遗传差异越大^[11],所达到的遗传平衡程度越高,多态性也越高。 PIC 是评价一个多

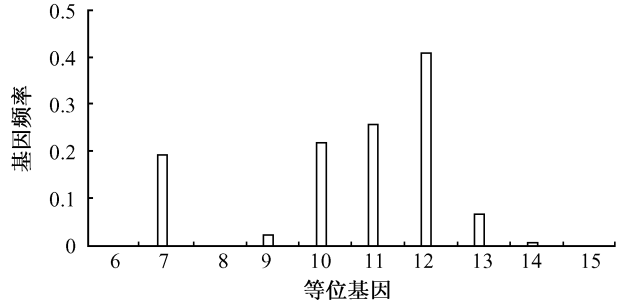


图 2—3 CSF1PO 基因座等位基因频率分布直方图
The Allele frequency distribution of CSF1PO locus

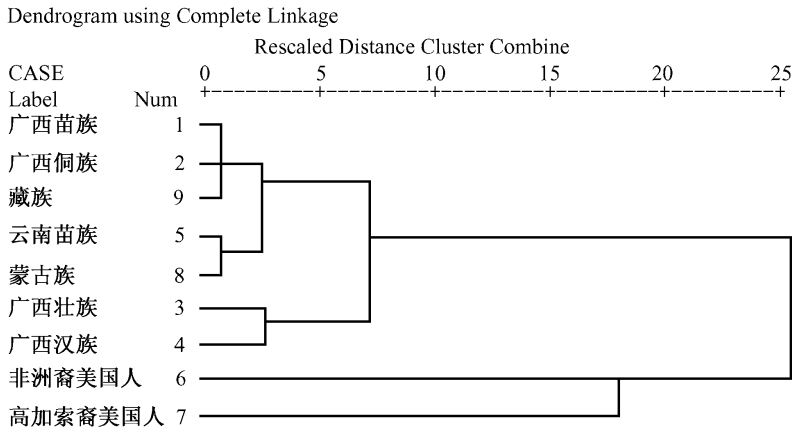


图 3 种族、民族间的聚类分析图
The result of clustering

态性基因座的使用价值的一个量化指标,直接反映一个遗传标记所包含或所能提供的遗传信息总量。 $PIC > 0.5$ 时,遗传标记具有高度的可提供信息性^[12]。 DP 和 EPP 则反映了该遗传标记在法医学个体识别及亲权鉴定中的能力。一般认为 $DP > 0.8$ 、 $EPP > 0.5$ 时,属于高度多态性遗传标记,具有较高的应用价值^[13]。本研究的广西融水苗族 3 个 STR 基因座多态性结果:平均 $H = 0.7003 > 0.7$;平均 $PIC = 0.6322 > 0.5$;平均 $DP = 0.8403 > 0.8$, $TDP = 0.9961$;平均 $EPP = 0.5395 > 0.5$, $CEP = 0.9043$ 。说明了 $TPOX$ 、 $TH01$ 、 $CSF1PO$ 位点在广西融水苗族人群中具有高度多态性和较好的应用价值;同时,也肯定了本研究群体调查资料和实验结果的可靠性。

但从单个基因显示的多态性来看, $TPOX$ 基因座的多态性略差(EPP 为 0.4619, PIC 为 0.5809, H 为 0.6394)证明不同的遗传标记在群体中的遗传多态性是不同的,提示在实际工作中,应该根据每个民族的遗传结构特点去选择最具价值的遗传标记进行研究,也进一步说明建立不同民族群体遗传学数据库的重要性和必要性。

3.2 种族、民族的比较

人类的不同种族、不同民族或不同群体,由于地理风俗习惯,宗教信仰,社会政治或传统上的原因,彼此隔离,随着时间的推移,它们就有可能积累不同的基因^[14]。而这些基因的分

化是在突变、自然选择以及随机漂变等因子的作用下产生的,其分化的程度可用遗传距离来衡量^[15]。本文通过群体间同一座位上相等等位基因的频率来计算广西融水苗族和其他人群的遗传距离,并对结果进行比较:广西融水苗族与非洲裔美国人、高加索裔美国人存在显著性差异,并与非洲裔美国人差异较大;从本文比较的我国这几个民族看,与非洲裔美国人、高加索裔美国人的差异均大于民族相互间的差异,这是否说明了种族之间差异大于民族之间差异,有待进一步调查研究。

在现有资料中,同一种族不同民族间相比,广西融水苗族与广西侗族、藏族的关系较近,而与广西汉族、壮族关系较远;同时也可得出蒙古族与云南苗族的遗传距离较近。而在广西区内的苗、侗、壮、汉四民族中,壮族与汉族的遗传距离小于苗、侗族与汉族的遗传距离,即广西汉族与壮族的关系近于与苗、侗族的关系。另外对遗传学资料进行聚类分析结果也显示了广西融水苗族与侗族、广西汉族与壮族各聚为一类,此可能是由于地理原因(融水苗族与侗族多集居于广西西北部,壮、汉族则大多居住在广西东部和东北部)及风俗习惯(不与外族通婚等)等原因造成。

总之,中华民族包含多个民族,各个民族的地理环境、宗教信仰和风俗习惯等各有不同,使其彼此间相对隔离,并聚居在相对稳定的地域里,就可能会积累下不同的祖系基因和群体间的遗传差异,形成遗传进化。用 STR 基因扫描技术,研究群体遗传结构,对基因研究、基因起源和进化都具有重要意义。

参考文献:

- [1] Edwards A, Civifello A, Hammond HA, *et al.* DNA typing and genetic mapping with trimeric and tetrameric tandem repeats [J]. *Am J Hum Genet*, 1991, 49 (4): 746—756.
- [2] Weber JL, May FE. Abundant class of human DNA polymorphism which can be typed using the polymorphisms chain reaction [J]. *Am J Hum Genet*, 1989, 44 (3): 388—396.
- [3] 李生斌. 人类 DNA 遗传标记 [M]. 北京: 人民卫生出版社, 2000, 212—217.
- [4] Promega. Gene print™ STR system technical manual [Z]. Printed in the USA Revised, 1996, 4—34.
- [5] 邹浪萍, 杨燕, 等. CSF IPO, TPOX 和 TH01 基因座复合扩增及其在苗族人群中的基因频率分布 [J]. *中国法医学杂志*, 1997, 12 (3): 148—151.
- [6] 邹浪萍, 杨燕, 诸嘉佑, 等. 多重 PCR 检测 CSF IPO, TPOX 和 TH01 基因座在中国汉族人群中的多态性 [J]. *遗传学报*, 1998, 25 (3): 199—204.
- [7] 邹平, 杨燕, 李德林, 等. 云南白族六个基因座的遗传多态性调查 [J]. *中国法医学杂志*, 1999, 16 (3): 160—163.
- [8] 程宝文, 辛军平, 邹浪萍, 等. 中国云南基诺族、佤族、布朗族 3 个 STR 基因座的遗传多态性研究及其在法医学中的应用 [J]. *华西医科大学学报*, 2000, 31 (4): 485—490.
- [9] 郑秀芬. 法医 DNA 分析 [M]. 北京: 中国人民公安大学出版社, 2002, 383—390.
- [10] 金天博. 广西壮族群体 STR 遗传结构研究 [D]. 西安: 西安交通大学, 2002, 15—16.
- [11] 刘雅诚, 霍振义, 唐晖. STR 三组复合扩增 (9 个基因座) 的法医学应用 [C]. 重庆: 第二次全国法医物证学学术交流会论文集, 1999, 76—79.
- [12] 赖江华, 张保华, 郑海波, 等. 中国纳西族 STR 遗传结构研究 [J]. *遗传学报*, 2002, 29 (7): 576—580.
- [13] Hammond HA, Li Jing, Zhong Y, *et al.* Evaluation of 13 short tandem repeat loci for use in personal identification application [J]. *Am J Hum Genet*, 1994, 15 (2): 175—182.
- [14] 杨电, 刘超, 彭汝标, 等. 南方汉族、黎族人群 15 个 STR 基因座频率调查 [J]. *法医学杂志*, 2002, 18 (4): 207—212.
- [15] 高树辉, 王新淮. 新疆乌孜别克 STR 基因频率分布及在法医学中的应用研究 [J]. *公安大学学报 (自然科学版)*, 2003, 3: 28—32.

A Study of the Polymorphisms of 3 STR Loci in the Miao Ethnic Group of Rongshui, Guangxi Province

ZHOU Li-ning¹, XU Lin¹, LI Song-feng¹, GONG Ji-chun¹,
WEI Be-yuan¹, DENG Xiang-fa¹, RU Hong-rong¹

(1. *Humanics Laboratory of Anatomy staff room of Guangxi Medical University, Nanning 530021*)

Abstract: To investigate the distribution of three short tandem repeat (STR) loci in the Miao ethnic group of the Guangxi Rongshui. We collected sodium-citrated blood specimens from 208 (male 102, female 106) healthy unrelated Miao individuals in Guangxi Rongshui, and used the phenol-chloroform method to extract DNA. By using the AmpFSTR Identifiler™ PCR Amplification Kit and 3100 Genetic Analyzer, 19 alleles and 48 genotypes of three STR were observed. The allele frequency and genotype frequency were 0.002 4—0.466 3 and 0.004 8—0.317 3, respectively. The expected and observed genotype distributions of three loci according to the Hardy-Weinberg equilibrium was given at $P > 0.05$. Genetic distance with 3 STR data showed that there were significant differences between Guangxi Rongshui Miao national minorities with American Caucasians and American Blacks. There was little difference between Guangxi Rongshui Miao national minorities and the Guangxi Dong.

Key words: Short tandem repeat; Gene frequency; Genetic distance; Miao ethnic group