

DOI: 10.3724/SP.J.1005.2009.00273

我国 3 个民族 13 个 SNPs 位点多态性及遗传学关系的比较

王瑞恒^{1,2}, 刘利民¹, 赵金玲¹

1. 中国医科大学法医学院, 沈阳 110001;

2. 辽宁师范大学法学院, 大连 116029

摘要: 采用荧光标记复合扩增毛细管电泳技术, 基于等位基因特异性 PCR 原理, 通过正交实验法建立了荧光标记复合扩增片段长度差异等位基因特异性 SNPs 分型体系, 该体系可以根据产物长度和产物峰的数量一次完成 13 个 SNPs 分型, 纯合子为单一产物峰, 杂合子为长度相差 4 bp 的两个产物峰。采用该体系对我国辽宁地区汉族、内蒙古地区蒙古族和广西地区壮族 3 个民族 13 个 SNPs 位点多态性进行群体调查, 获得了 3 个民族 13 个 SNPs 等位基因分布频率, 比较了 3 个民族等位基因的差异, 并对其遗传学关系进行了研究。结果显示: 3 个民族 13 个 SNPs 的等位基因分布具有多态性, 多个 SNPs 等位基因分布具有显著性差异($P \leq 0.01$), 抽样调查结果符合 Hardy-Weinberg 平衡; 辽南地区汉族人群与内蒙古蒙古族人群的亲缘关系更为接近, 与广西壮族之间的亲缘关系相对较远。

关键词: 荧光标记复合扩增; 单核苷酸多态性; 遗传学关系

A comparative study on genetic polymorphism and genetic relationship of 13 SNPs in three Chinese populations

WANG Rui-Heng^{1,2}, LIU Li-Min¹, ZHAO Jin-Ling¹

1. Forensic Medical College of China Medical University, Shenyang 110001, China;

2. Law School of Liaoning Normal University, Dalian 116029, China

Abstract: Using the fluorescence labeled capillary electrophoresis of multi-PCR technique, the single nucleotide polymorphism (SNP) typing system of fragment length discrepant allele specific fluorescence labeled multi-PCR technique is established based on the principle of allele-specific PCR. The typing of the 13 SNP loci can be completed simultaneously according to the length of PCR products and the number of product peaks. It appears a single product peak when the SNP is homozygous, and two product peaks with 4 bp differences will appear when it is heterozygous. By using this system, we conducted population census about allele frequencies for 13 autosomal SNP loci in Southern Liaoning Han samples, Mongolian samples in Inner Mongolia and Zhuang samples in Guangxi area, and got the allele frequencies of the 13 SNP loci in the three populations, then preliminarily discussed their genetic relationship by comparing their differences in allelic polymorphism. The results indicate that the allelic distributions of the 13 SNP loci in the three populations are polymorphic, and the difference is significant in some SNP loci ($P \leq 0.01$). The sampling survey shows that the result is consistent with

收稿日期: 2008-08-08; 修回日期: 2008-10-16

基金项目: 辽宁省教育厅科学研究基金项目(编号: 20060976, 2008346)资助

作者简介: 王瑞恒(1968-), 男, 副教授, 博士, 研究方向: 法医遗传学。E-mail: dalianwrh@126.com

通讯作者: 刘利民(1955-), 女, 教授, 博士生导师, 研究方向: 法医遗传学。E-mail: lmliu@mail.cmu.edu.cn; Tel: 024-82001810

Hardy-Weinberg equilibrium, and Han population in southern Liaoning has relatively closer relationship with Mongolian in Inner Mongolia than with Zhuang population in Guangxi by origin.

Keywords: fluorescence labeled multi-PCR; single nucleotide polymorphism; genetic relationship

单核苷酸多态性是一种具有高度稳定性的遗传学标记^[1]。在基因多态性研究、基因图谱构建、疾病诊断、预防、药物筛选和人类起源与亲缘关系的研究中具有重要作用^[2]

。SNPs还是比较理想的研究人类起源、进化和迁移的遗传标记^[3-8]。SNPs的分型检测方法有很多种,基于PCR技术的常用方法主要有限制性酶切片段长度多态性(Restriction fragment length polymorphism, RFLP)^[9]、单链构象多态性(Single-strand conformation polymorphism, SSCP)^[10]、异源双链分析(Heteroduplex analysis, HA)^[11], 这些技术一次实验只能检测出单个SNPs位点的等位基因,检测过程较为繁琐,效率较低,较难适应大规模的群体调查要求。近几年发展起来的变性高效液相色谱(Denaturing high performance liquid chromatography, DHPLC)^[12]、基质辅助激光解析飞行时间质谱(Matrix-assisted laser desorption/ionization time of flight mass spectrometry, MALDI-TOF MS)^[13]、动态等位基因特异性杂交技术(Dynamic allele specific hybridization, DASH)^[14]、基因芯片法^[15]、TaqMan探针技术(TaqMan probes)^[16]和焦磷酸测序技术(Pyrosequencing)^[17]等新技术具有通量高、易于自动化等优点,但仪器设备昂贵,检验成本高,很难在普通实验室推广应用。

我们在等位基因特异性PCR的基础上,基于末端碱基错配技术,结合荧光标记复合扩增和毛细管电泳技术,建立了一个能够同时对多个SNPs位点进行复合扩增的快速检测系统,为SNPs基因分型提供一个高效率、大通量和低成本、且较为实用的SNPs分型新方法^[18]。采用该分型体系,本文对我国3个民族13个常染色体SNP位点进行了群体调查,获得各等位基因的频率,并以此为依据对我国3个民族间的遗传学关系进行了探讨。

1 材料和方法

1.1 材料

辽宁南部地区320名无血缘关系的汉族健康人

群静脉抗凝血样。内蒙古自治区锡林浩特市东乌旗地区242名无血缘关系的蒙古族健康人群静脉抗凝血样。广西壮族自治区隆安县314名无血缘关系的壮族健康人群静脉抗凝血样。

血液样品分别由大连市第五人民医院、锡林浩特市东乌旗人民医院、广西隆安县人民医院在当事人知情同意的前提下采集。

1.2 方法

1.2.1 引物序列

13个常染色体SNPs的选择原则如下:距离尽可能远,位于不同染色体上最佳;双等位基因;根据文献报道,频率较低的等位基因在东亚人群中的频率不低于0.1,大于0.3最佳。根据<http://www.ncbi.nlm.nih.gov/SNP>公布的参考序列自行设计引物,在两条上游引物3末端的第3或第4位碱基引入人为错配。错配的依据参照Little^[19]的原则:如果3端是“强”错配(A/G或C/T)则需要引入“弱”的错配(C/A或G/T)或无需引入错配,反之亦然;当3端是“中度”错配(A/A, C/C, G/G或T/T)则需要引入“中度”的错配。所有引物均由宝生物工程(大连)有限公司合成并标记(13个SNP位点的引物序列及荧光标记见参考文献[18])。

1.2.2 DNA模板制备及鉴定

使用天根公司生产的基因组DNA提取试剂盒(天根生物科技公司,北京)制备血液样品的扩增模板,提取后,再用Gene Quant Pro型紫外分光光度计(Amersham Biosciences公司,美国)检测DNA纯度和浓度,以去离子灭菌水调整终浓度为10 ng/ μ L。

1.2.3 PCR扩增及SNPs分型

20 μ L反应体系中,内含100 μ mol/L dNTPs, 50 mmol/L KCl, 10 mmol/L Tris-HCl (pH 8.3), 1.5 mmol/L $MgCl_2$, 0.5 U TaKaRa Ex TaqTM Hotstar DNA聚合酶, 20 ng模板DNA, 事先调制的引物混合物4 μ L。TaKaRa PCR Thermal Cycle Dice型梯度PCR

扩增仪[宝生物工程(大连)有限公司]扩增, PCR 循环参数为: 94 1 min, 62 1 min, 72 1.5 min, 共30个循环, 循环结束后72 延伸60 min, 4 保存(聚合酶及其他PCR反应试剂均购自宝生物工程(大连)有限公司)。

PCR产物采用310型DNA毛细管电泳测序仪分离(ABI公司, 美国)。GeneScan™ Size Standards LIZ-500(ABI公司, 美国)作为内标, POP4凝胶, 47 cm毛细管, 1.5 KV电泳25 min, GeneScan™3.0分析结果, 根据产物长度和产物峰的数量进行分型。

1.2.4 测序验证

每个SNPs各随机挑选已经分型的样品6例, 测序验证。测序引物序列及PCR片段长度见参考文献[18]。

1.2.5 统计学分析

直接计数法计算等位基因与基因型频率; 基因型 Hardy-Weinberg 平衡检验、基因分布差异性检验及聚类分析用SPSS 11.5统计软件完成, 聚类方法用离差平方和法(Ward' method), 遗传距离测量用欧氏距离平方方法(squared Euclidean distance)。

2 结果与分析

2.1 荧光标记长度差异等位基因特异性复合扩增体系进行多位点SNPs分型

本文建立的荧光标记片段长度差异等位基因特异性复合扩增体系, 能够一次对13个SNPs分型。不同的位点, 在设计引物时, 两条特异性引物的长度相差4 bp, 由此使扩增的片段长度不同, 可以利用毛细管电泳根据片段长度差异直接检测3种基因型。纯合子为单一产物峰, 杂合子则为长度相差4 bp的两个产物峰, 尽管有的产物长度有所重叠, 但使用不同的荧光物质标记, 可以正确区分并判定基因型(图1)。经测序验证, 本方法分型结果与测序分型结果一致。

2.2 13个SNPs等位基因频率

对辽宁南部地区320名汉族个体、内蒙古地区242名蒙古族个体和广西地区314名壮族个体进行了群体调查, 获得了3个民族13个SNPs等位基因分布频率(表1)。结果显示: 3个民族13个SNPs等位基因频率有较大差异; 3个民族13个SNPs基因型

观察值与期望值之间吻合度检验结果显示: 辽南地区汉族群体 χ^2 值位于0.026~1.241之间; 在内蒙古地区蒙古族群体中 χ^2 值位于0.529~1.842之间; 广西地区壮族群体 χ^2 值在0.281~1.536之间。

2.3 3个民族13个SNPs基因型频率分布的比较

本文对我国辽南地区汉族、内蒙古地区蒙古族、广西地区壮族3个民族的13个SNPs的基因型分布状态进行了比较(表2)。由结果可见: 3个民族的基因型分布有较大的差异。比较结果显示, 在汉族和蒙古族之间, 有3个位点 *rs2073498*、*GSTM4(T2517C)* 和 *rs1056836* 呈极显著差异($P<0.001$), 1个位点 *rs17050550* 有差异($0.01 < P<0.05$), 其余9个位点没有差异($P=0.05$); 在汉族和壮族之间, 4个位点 *rs140693*、*rs1136410*、*GSTM4(T2517C)* 和 *rs17050550* 有极显著差异($P<0.001$), 2个位点 *rs3087386* 和 *rs8305* 有显著差异($0.001 < P<0.01$), 1个位点 *rs2073498* 有差异($0.01 < P<0.05$), 其余6个位点没有差异($P=0.05$); 在蒙古族和壮族之间, 2个位点 *rs140693* 和 *rs3087386* 有极显著差异($P<0.001$), 2个位点 *rs1136410* 和 *rs1800566* 有显著差异($0.001 < P<0.01$), 3个位点 *rs1056836*、*rs1799782* 和 *GSTM4(T2517C)* 有差异($0.01 < P<0.05$), 其余6个位点没有差异($P=0.05$)。

2.4 3个民族13个SNPs的聚类分析

根据3个民族13个SNPs等位基因频率, 利用SPSS软件对我国3个民族进行聚类分析, 得到欧氏不相似系数平方矩阵表(表3)和聚类分析树状关系图(图2)。由表3的结果可见, 辽南汉族和内蒙古蒙古族之间的欧氏不相似系数最小, 为0.295; 蒙古族与壮族之间的欧氏不相似系数次之, 为0.432; 辽南汉族和广西壮族之间的欧氏不相似系数最大, 为0.475。由图2的结果可见, 在聚类的过程中, 汉族首先与蒙古族合并后, 再共同与壮族合并。

3 讨论

3.1 3个民族13个SNPs基因型多态性

本文经群体调查, 获得了3个民族13个SNPs等位基因频率。3个民族13个SNPs基因型观察值与期望值之间吻合度检验结果显示: 辽南汉族群体

表 2 3 个民族 13 个 SNPs 基因型分布比较结果

位点	基因型	汉族	蒙古族	壮族	汉族 vs 蒙古族		汉族 vs 壮族		蒙古族 vs 壮族	
					χ^2	$P(df=2)$	χ^2	$P(df=2)$	χ^2	$P(df=2)$
rs1800566	C/C	94	81	73	2.3529	0.308	3.9480	0.139	10.9732	0.004
	C/T	170	129	172						
	T/T	56	32	69						
rs2073498	C/C	248	151	212	15.4441	0.000	9.1043	0.011	4.2502	0.119
	C/A	66	85	88						
	A/A	6	6	14						
rs140693	A/A	58	35	116	1.7307	0.421	52.0640	0.000	52.5053	0.000
	A/G	152	126	158						
	G/G	110	81	40						
rs13181	T/T	226	152	220	5.7902	0.055	0.2299	0.891	4.1578	0.125
	T/G	84	74	82						
	G/G	10	16	12						
rs1799782	C/C	188	121	186	4.9988	0.082	4.4115	0.110	6.2319	0.044
	C/T	120	106	105						
	T/T	12	15	23						
GSTM4(T2517C)	C/C	121	110	172	14.4695	0.001	23.1804	0.000	6.2161	0.045
	C/T	140	114	115						
	T/T	59	18	27						
rs1136410	C/C	67	36	26	4.6309	0.099	35.4428	0.000	12.4610	0.002
	C/T	171	129	147						
	T/T	82	77	141						
rs17050550	G/G	261	176	215	6.9235	0.031	14.6471	0.001	1.4934	0.474
	T/G	53	56	87						
	T/T	6	10	12						
rs1056836	C/C	252	156	237	15.3004	0.000	1.0333	0.597	8.7184	0.013
	C/G	61	72	68						
	G/G	7	14	9						
rs3087386	C/C	38	22	22	3.4341	0.180	12.7631	0.002	18.6736	0.000
	C/T	132	118	102						
	T/T	150	102	190						
rs2228001	A/A	138	88	135	4.2959	0.117	0.4057	0.816	2.9551	0.228
	A/C	154	122	147						
	C/C	28	32	32						
rs8305	A/A	26	8	18	5.6524	0.059	2.5982	0.273	2.5824	0.275
	A/G	120	97	134						
	G/G	174	137	162						
	G/G	29	20	16	1.6091	0.447	11.1342	0.004	3.9398	0.139
	A/G	119	79	90						
	A/A	172	143	208						

表 3 欧氏不相似系数平方矩阵

民 族	不相似系数平方矩阵		
	汉族	蒙古族	壮族
汉族	0.000	0.295	0.475
蒙古族	0.295	0.000	0.432
壮族	0.475	0.432	0.000

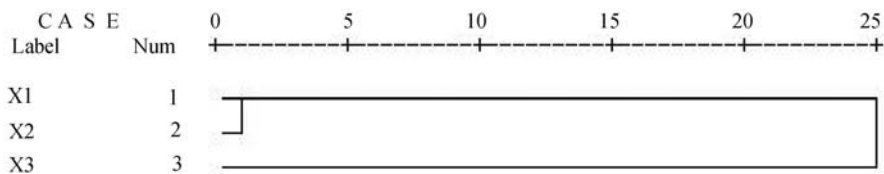


图 2 3 个民族聚类分析树状关系图

X1: 汉族; X2: 蒙古族; X3: 壮族。

χ^2 值位于 0.026~1.241 之间; 内蒙古蒙古族群体 χ^2 值位于 0.529~1.842 之间; 广西壮族群体 χ^2 值在 0.281~1.536 之间, 3 个民族的 χ^2 值均小于自由度为 1, 且 $P > 0.05$ 时的临界值 3.84, 由此可见, 各个位点的调查结果其观察值与期望值之间均符合 Hardy-Weinberg 平衡, 说明本次抽样调查符合统计学的要求, 样本具有代表性。

3 个民族 13 个 SNPs 组成的基因型分布的比较结果显示: 在汉族和蒙古族之间, 4 个位点有差异 ($P < 0.05$), 9 个位点没有差异 ($P > 0.05$); 在汉族和壮族之间, 7 个位点有差异 ($P < 0.05$), 6 个位点没有差异 ($P > 0.05$); 在蒙古族和壮族之间, 7 个位点有差异 ($P < 0.05$), 其余 6 个位点没有差异 ($P > 0.05$), 由此可见, 3 个民族在多个 SNPs 位点上的基因型分布具有显著性差异, 等位基因分布频率在我国不同地区、不同民族之间具有差异性, 在进行人类遗传学分析时数据不能通用。

3.2 3 个民族 13 个 SNPs 的聚类分析

单核苷酸多态性是研究人类遗传关系的重要标记。Kim 等^[4]通过对亚洲地区 1260 名男性 5 个 Y-SNPs (DYS287, DXYS5Y, SRY465, DYS19, DXYS 156Y) 单倍型的研究结果显示, 在亚洲人群中, 韩国人群和日本人群间的亲缘关系最为接近, 并且韩国人群并非单一起源于东北亚, 而是由东北亚和东亚共同起源。柯越海等^[5]对来自中国各地 9 988 例男性随机样本的 M89、M130 和 YAP 3 种 Y 染色体单倍型进行研究, 研究结果为中国人非洲起源假说提供了最新的

遗传学证据。刘烜等^[6]通过对我国贵州地区 5 个少数民族 Y-SNPs 分布状态的研究认为, 布依族、仡佬族、仡佬族、毛南族、壮族 5 个民族之间有密切联系, 且与国内其他民族有较大的遗传差异, 是相对独立的群体。杨波等^[7]通过对来自海南省白沙、保亭和乐东地区 3 个黎族支系的 206 个男性群体样本进行 Y 染色体非重组区 SNPs 的单倍型研究, 提出了海南黎族 3 个支系父系族源来源于东亚南部的观点。Mohyuddin 等^[8]通过对 95 例巴基斯坦人群相关的 5 个 SNPs 的分析, 认为巴基斯坦群体与印度系民族间具有很近的亲缘关系。

本文对 3 个民族 13 个 SNPs 共 78 个等位基因频率的数据, 采用欧氏距离平方法进行聚类分析, 得到了欧氏不相似系数平方矩阵。辽南地区汉族和内蒙古地区蒙古族的欧氏不相似系数最小, 为 0.295; 蒙古族与壮族之间的欧氏不相似系数次之, 为 0.432; 辽南地区汉族和广西地区壮族之间的欧氏不相似系数最大, 为 0.475。在聚类的过程中, 汉族首先与蒙古族合并后, 再共同与壮族合并, 由于在聚类分析时, 距离最近或最相似的聚为一类, 此结果表明: 辽南地区汉族与内蒙古地区蒙古族的亲缘关系比辽南地区汉族与广西壮族的亲缘关系更为接近。

参考文献(References):

- [1] Lander ES. The new genomics: global views of biology. *Science*, 1996, 274(5287): 536-539. [\[DOI\]](#)
- [2] 刘万清, 贺林. SNP——为人类基因组描绘新的蓝图. *遗传*, 1998, 20(6): 38-40.

- [3] Mansoor A, Mazhar K, Khaliq S, Hameed A, Rehman S, Siddiqi S, Papaioannou M, Cavalli-S L, Mehdi S, Ayub Q. Investigation of the Greek ancestry of populations from northern pakistan. *Hum Genet*, 2004, 114(5): 484–490. [\[DOI\]](#)
- [4] Kim W, Shin DJ, Harihara S, Kim YJ. Y-chromosomal DNA variation in east Asian populations and its potential for inferring the peopling of Korea. *J Hum Genet*, 2000, 45(2): 76–83. [\[DOI\]](#)
- [5] 柯越海, 宿兵, 李宏宇, 陈黎峰, 春建, 郭新军, 黄薇, 金建中, 卢大儒, 金力. Y 染色体遗传学证据支持现代中国人起源于非洲. *科学通报*, 2001, 46(5): 411–414.
- [6] 刘烜, 单可人, 齐晓岚, 何燕, 赵艳, 吴昌学, 李毅, 褚迅, 任锡麟. 贵州布依族、仡佬族、佤族、毛南族、壮族 Y-SNP 的初步研究. *遗传*, 2006, 28(11): 1350–1354.
- [7] 杨波, 李冬娜, 孙元田, 区采莹, 应大君. 海南黎族不同支系 Y-SNP 的多态性分析. *中国热带医学*, 2007, 7(3): 341–343.
- [8] Mohyuddin A, Ayub Q, Qamar R, Khaliq S, Mansoor A, Mehdi SQ. Detection of novel Y SNPs provides further insights into Y-chromosome variation in pakistan. *J Hum Genet*, 2006, 51(4): 375–378.
- [9] Sibley LD, Leblanc AJ, Pfeifferkorn ER, Boothroyd JC. Generation of a restriction fragment length polymorphism linkage map for *Toxoplasma gondii*. *Genetics*, 1992, 132(4): 1003–1015.
- [10] Glavac D, Dean M. Optimization of the single-strand conformation polymorphism(SSCP) technique for detection of point mutations. *Hum Mutat*, 1993, 2(5): 404–414. [\[DOI\]](#)
- [11] Nataraj AJ, Isabelle OG, Noriko K, Highsmith Jr WE. Single-strand conformation polymorphism and heteroduplex analysis for gel-based mutation detection. *Electrophoresis*, 1999, 20(6): 1177–1185. [\[DOI\]](#)
- [12] Nickerson ML, Weirich G, Zbar B, Schmidt LS. Signature-based analysis of MET proto-oncogene mutations using DHPLC. *Hum Mutat*, 2000, 16(1): 68–76. [\[DOI\]](#)
- [13] Fei Z, Smith LM. Analysis of single nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Rapid Commun Mass Spectrometry*, 2000, 14(11): 950–959. [\[DOI\]](#)
- [14] Jonathan AP, Lars F, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Research*, 2001, 11(1): 152–162. [\[DOI\]](#)
- [15] Schmalzing D, Belenky A, Novotny MA, Koutny L, Salas-Solano O, El-Difrawy S, Adourian A, Matsudaira P, Ehrlich D. Microchip electrophoresis: a method for high-speed SNP detection. *Nucleic Acids Res*, 2000, 28(9): e43. [\[DOI\]](#)
- [16] Shi MM, Myrand SP, Bleavins MR, De la Iglesia FA. High throughput genotyping for the detection of a single nucleotide polymorphism in SAD(P)H quinone oxidoreductase(DT diaphorase) using TaqMan probes. *Mol Pathol*, 1999, 52(5): 295–299. [\[DOI\]](#)
- [17] Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyren P. Single nucleotide polymorphisms analysis by pyrosequencing. *Anal Biochem*, 2000, 280(1): 103–110. [\[DOI\]](#)
- [18] 王瑞恒, 刘利民, 赵金玲, 孙学科, 孙琳琳, 刘奉君. 荧光标记复合扩增毛细管电泳法在 SNP 分型中的应用. *中国法医学杂志*, 2008, 23(5): 326–329.
- [19] Little S. In: Taylor G, ed. *Laboratory Methods for the Detection of Mutations and Polymorphisms in DNA*. CRC Press, Boca Raton, 1997, 45–51.

•遗传咨询•

糖原贮积症的产前检查

问: 已生过一胎患有糖原贮积症, 如生第二胎, 预防这种病的发生, 需要做些什么检查, 小孩得这种病的几率是多大? 有生育不患病的胎儿的可能吗?

答: 糖原贮积症 (Glycogen Storage Disease, GSD), 是由于糖原分解代谢中各种酶的缺陷所引起。GSD 分为 I~XI 型, 相应的致病基因均已阐明, 分别定位于多条常染色体和 X 染色体上。临床上一般分为两组, 一组主要以肝受累为主, 另一组主要以肾受累为主, 分别以 GSD Ia 型和 GSD II 型为代表。其中除 GSD 和 GSD b 为 X 连锁隐性遗传外, 其他各型均为常染色体隐性遗传。尽管遗传方式不同, 但仍有生育正常胎儿的几率。可以通过生化检查和基因诊断确定患儿 GSD 分型以及父母相应的基因型, 然后实施再怀孕胎儿的产前诊断。

(中国科学院遗传与发育生物学研究所 张喆、李巍)