

广东省汉族人群的亚群分析

吕德坚,刘秋玲,陆惠玲

(中山大学中山医学院法医物证教研室,广州 510089)

摘要:为调查广东地区汉族人群中是否存在亚群结构,对来自广州、佛山、东莞、江门和中山—珠海 5 个不同地区的 471 随机个体血样本进行 15 个 STR 的基因分型,并作 Hardy-Weinberg 平衡检验,比较群体之间的等位基因频率,计算这个 5 个地区人群间的共祖系数。结果显示,广东汉族人群没有明显的 Hardy-Weinberg 不平衡,5 个地区人群的等位基因分布均没有差异,他们之间的共祖系数小于 0.01。因此,广东地区汉族人群中没有明显的亚群结构。

关键词:亚群;群体遗传学;共祖系数;DNA 统计学;短串联重复

中图分类号:R394.5

文献标识码:A

文章编号:0253-9772(2003)06-0645-04

Analysis of Subpopulation in Guangdong Han Population

LÜ De-Jian, LIU Qiu-Ling, LU Hui-Ling

(Department of Forensic Serology, Zhongshan Medical College of Sun Yat-sen University, Guangzhou 510089, China)

Abstract: To investigate the subpopulation structure within the Han Population in Guangdong area, a total of 471 DNA samples from five populations in Guangdong Province, including Guangzhou, Foshan, Dongguan, Jiangmen and Zhongshan-Zhuhai region, were genotyped at 15 STR (short tandem repeats) markers. Hardy-Weinberg tests were performed, allele frequencies were compared, and the genetic coancestry coefficient (F_{ST}) was estimated. The results did not show significant departure from Hardy-Weinberg equilibrium in the total population. Difference of allele frequencies among these populations was not observed, and the coancestry coefficient (F_{ST}) was less than 0.01. Subpopulation structure within Han Population in Guangdong Province could not be detected.

Key words: subpopulation; population-genetics; coancestry coefficient; DNA statistics; short tandem repeats (STR)

在法医学上,通常是用被检验者所在群体的等位基因频率计算个人识别率和亲权概率。但是由于地域、历史文化背景和社会生活习俗等原因,在一个大群体中往往存在着某种程度的隔离群体,形成人群的亚群结构^[1,2]。亚群结构对 DNA 图谱概率的计算有较大的影响,是法庭 DNA 群体遗传学中研究的热点。现在,法庭统计遗传学家已普遍认为在应用 DNA 进行个体识别和亲权鉴定时要考虑亚群结构^[3,4]。为此,美国国家研究所(National Research Council, NRC)还规定了亚群结构的计算标准^[5]。短串联重复(short tandem repeats, STR)是

目前法医学检验工作中最常用的遗传标记。我国人口众多,幅员辽阔,虽然已有不少地区群体的 STR 等位基因频率报道^[6~8],但对各地区人群内是否存在亚群结构的研究还很少。本研究用 15 个 STR 的群体资料对广东地区的汉族人群进行亚群分析。

1 材料与 方法

1.1 血液样本

471 个随机抽样个体的血样本来自珠江三角洲地区的人群,按行政区域分为不同的亚人群,具体分布如下:广州地区(含从化、番禺、增城和花都)132

人,佛山地区(含佛山市区、南海、顺德和三水)119人,东莞市64人,江门地区96人,中山和珠海共60人。这些个体的祖籍均在当地,属于广东粤语系人。

1.2 STR分型

采用Chelex-100法处理血样的DNA,用Promega公司的PowerPlex® 16 System试剂盒(含D3S1358、TH01、D21S11、D18S51、Penta E、D5S818、D13S317、D7S820、D16S539、CSF1PO、Penta D、vWA、D8S1179、TPOX、FGA等15个STR和一个Amel性别基因座)和PE公司的ABI 377测序仪进行15个STR位点的检测与分型。具体方法参阅该试剂盒说明书。

1.3 统计分析

用计数法统计等位基因频率,分别用正合检验

(exact test)和杂合子缺少检验法(heterozygote deficiency test)进行Hardy-Weinberg平衡检验,并比较各亚群体之间的等位基因频率,计算共祖系数(F_{ST})。所有这些分析均用软件GDA^[9]或GenePop^[10]来完成。

2 结 果

通过检测5个地区471个样本的15个STR的基因型,分别建立相应地区人群的STR数据库。利用这些数据进行下列测验:

2.1 Hardy-Weinberg平衡检验

对5个地区的样本及其合并的总群体的471个样本进行正合检验,其P值见表1。可见在5%的显著性水平上,少数基因座偏离Hardy-Weinberg平衡。

表1 广东地区5个群体样本15个STR的Hardy-Weinberg平衡检验P值

Table 1 The P-value of Hardy-Weinberg equilibrium test of 15 STR markers in 5 populations in Guangdong area

基因座 Loci	东莞 Dongguan		佛山 Foshan		广州 Guangzhou		江门 Jiangmen		中山-珠海 Zhong-Zhu		总群体 Total	
	P_E	P_D	P_E	P_D	P_E	P_D	P_E	P_D	P_E	P_D	P_E	P_D
D3S1358	0.2766	0.0133	0.3132	0.9564	0.1971	0.8470	0.7321	0.6777	0.7069	0.8685	0.0428	0.7786
TH01	0.3282	0.8424	0.7404	0.5152	0.3738	0.8971	0.8913	0.9152	0.1571	0.5014	0.2275	0.9700
D21S11	0.7755	0.6874	0.3827	0.2857	0.7301	0.4706	0.5290	0.2810	0.5858	0.1756	0.6022	0.2613
D18S51	0.1147	0.3604	0.9858	0.6734	0.7605	0.4035	0.3183	0.7756	0.4824	0.2625	0.7569	0.7303
Penta E	0.1431	0.1604	0.2347	0.7054	0.2885	0.6283	0.3742	0.8016	0.0830	0.1200	0.1350	0.2778
D5S818	0.8151	0.5029	0.9383	0.2584	0.9064	0.4704	0.2058	0.2928	0.4737	0.4774	0.9747	0.3007
D13S317	0.5742	0.9423	0.4394	0.0431	0.3680	0.1676	0.2824	0.1206	0.2065	0.2736	0.6109	0.0824
D7S820	0.3459	0.5415	0.0427	0.3428	0.3318	0.9122	0.4352	0.9713	0.2824	0.0456	0.0100	0.7561
D16S539	0.1433	0.4815	0.6611	0.6522	0.4124	0.4581	0.9027	0.6596	0.7164	0.7218	0.4288	0.7019
CSF1PO	0.4295	0.4276	0.3438	0.0008	0.4925	0.4987	0.0632	0.7616	0.7129	0.4514	0.0556	0.0125
Penta D	0.2394	0.0292	0.6922	0.7213	0.8773	0.6617	0.5073	0.9018	0.0438	0.0180	0.8656	0.3474
vWA	0.6770	0.4969	0.2610	0.6273	0.2777	0.0117	0.1023	0.8149	0.2017	0.0949	0.7478	0.0720
D8S1179	0.7728	0.6963	0.5193	0.9476	0.1756	0.0700	0.0559	0.8597	0.0384	0.6408	0.0019	0.8384
TPOX	0.4393	0.8755	0.8097	0.4891	0.2871	0.8187	0.6372	0.4843	0.4585	0.5696	0.7784	0.8105
FGA	0.2403	0.7786	0.7329	0.1122	0.9358	0.5903	0.4758	0.7091	0.1032	0.1781	0.1381	0.3176

P_E : P-value of exact test. 正合检验 P 值。

P_D : P-value of heterozygote deficit. 杂合子缺少检验 P 值。

另一方面也可以从是否有杂合子缺少(heterozygote deficiency)考察Hardy-Weinberg平衡。对5个地区的样本及其合并的总样本进行杂合子缺少的Hardy-Weinberg平衡检验,所得各基因座的P值见表1,相应的杂合率见表2。

2.2 群体间的等位基因频率比较

从总体上分析5个地区样本间各基因座的等位基因频率的分布,各基因座的等位基因频率均没有

显著性的差异。进一步对五个地区进行两两比较,也没有发现明显的不同(数据未显示)。

2.3 F_{ST} 值

从遗传学的角度来考虑,若群体中含有亚群结构时,可以用一个遗传学参数 F_{ST} 来估计其遗传差异^[12]。假定所有的总样本是由具有亚群结构的5个地区的样本所组成,经过计算得到这5个亚群体各位点的 F_{ST} 值(见表3),这些 F_{ST} 值均很小。

表 2 广东地区 5 个群体样本 15 个 STR 的杂合率

Table 2 The heterozygosity of 15 STR loci of 5 populations in Guangdong area

基因座 Loci	东莞		佛山		广州		江门		中山-珠海		总群体	
	Dongguan		Foshan		Guangzhou		Jiangmen		Zhong-Zhu		Total	
	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O
D3S1358	0.738	0.734	0.729	0.790	0.717	0.750	0.738	0.740	0.697	0.750	0.724	0.756
TH01	0.695	0.734	0.725	0.714	0.720	0.758	0.693	0.750	0.729	0.750	0.713	0.741
D21S11	0.836	0.828	0.817	0.840	0.832	0.818	0.824	0.792	0.807	0.767	0.823	0.813
D18S51	0.867	0.891	0.881	0.891	0.870	0.871	0.866	0.875	0.830	0.817	0.867	0.873
Penta E	0.922	0.844	0.905	0.916	0.915	0.917	0.899	0.938	0.921	0.900	0.911	0.909
D5S818	0.783	0.781	0.796	0.773	0.786	0.811	0.784	0.729	0.802	0.800	0.789	0.779
D13S317	0.813	0.875	0.781	0.697	0.795	0.750	0.800	0.760	0.804	0.767	0.795	0.758
D7S820	0.792	0.781	0.768	0.731	0.774	0.818	0.769	0.833	0.776	0.700	0.774	0.779
D16S539	0.781	0.766	0.758	0.782	0.776	0.795	0.778	0.771	0.794	0.800	0.776	0.783
CSF1PO	0.737	0.703	0.734	0.681	0.722	0.705	0.725	0.771	0.739	0.733	0.728	0.715
Penta D	0.783	0.734	0.812	0.824	0.816	0.811	0.786	0.823	0.769	0.717	0.798	0.794
vWA	0.798	0.813	0.792	0.790	0.807	0.758	0.759	0.823	0.802	0.700	0.792	0.779
D8S1179	0.850	0.875	0.854	0.899	0.852	0.818	0.859	0.875	0.851	0.883	0.855	0.866
TPOX	0.578	0.609	0.582	0.580	0.604	0.629	0.564	0.531	0.604	0.617	0.585	0.592
FGA	0.848	0.891	0.872	0.874	0.861	0.886	0.876	0.885	0.864	0.817	0.866	0.875

H_E : Expected heterozygosity. 杂合率期望值。

H_O : Observed heterozygosity. 杂合率观察值。

表 3 广东省 5 个地区人群之间 15 个 STR 基因座的 F_{ST} 值Table 3 F_{ST} value at 15 STR loci among 5 populations in the Guangdong Province

Loci	F_{ST}	Loci	F_{ST}
D3S1358	0.0017	D16S539	0.0019
TH01	0.0015	CSF1PO	0.0031
D21S11	0.0014	Penta D	0.0008
D18S51	0.0013	vWA	0.0001
Penta E	0.0006	D8S1179	0.0025
D5S818	0.0015	TPOX	0.0031
D13S317	0.0017	FGA	0.0006
D7S820	0.0002	Average	0.0004

3 讨 论

Hardy-Weinberg 平衡检验的结果是判断有无亚群结构的指标之一^[1]。但是正合检验与杂合子缺少检验两者的检验假设是不一样的,前者检验的是所有基因型之间的关系,后者则是观察群体中的杂合子的比例。在人类群体中,亚群主要是由于婚配大多局限于亚群内个体之间而形成的,实质上是相当于某种程度的近亲婚配^[5]。当群体之内存在着近亲婚配时,就会出现某些等位基因的频率较高,部分基因型所占的比例较大,群体中的纯合子增加,杂合子缺少,从而偏离 Hardy-Weinberg 平衡。

当群体之中存在着亚群结构时,各亚群之间的等位基因频率分布可能并不相同,若用总群体的等位基因频率来计算某一 STR 分型在亚群中出现的概率,显然是有误差的,将影响到法医鉴定的准确性^[4]。本文比较不同地区人群之间的等位基因频率,并没有发现明显的差异,表明可用总群体的等位基因频率计算相关的法医学数据。

共祖系数 F_{ST} 也是衡量群体内有无亚群结构的遗传学参数之一^[11]。本研究所有基因座的 F_{ST} 值均较小,没有明显的亚群结构存在。

亚群结构与遗传标记是无关系的。当群体有亚群结构时,应当是表现在所有的或大多数的基因座,不应是只在少数基因座体现出来。以上分析表明,少数基因座的 Hardy-Weinberg 不平衡结果,可能是由于样本数量不足,或样本在群体不具代表性,是样本抽样误差或分型不准确引起的,不能表明有明显的亚群存在。

在法医的案件中,当 DNA 的检验结果不能下排除的结论时,通常要计算 DNA 图谱的相同概率(probability of match, PM 值),以评估两个 DNA 样本系同一个体的可能性有多大。在计算 PM 值时,理想的情况是群体中的个体均是随机个体,但是受各种因素的影响,人类群体中可能存在着无法预

知的亚群结构,使等位基因频率之间、基因型频率之间表现出某种程度的相关性,因此要引入共祖系数 F_{ST} 以修正基因型频率^[4,12]。美国 NRC^[5] 推荐 F_{ST} 的取值在 0.01~0.03 之间,我国目前尚未有 F_{ST} 取值的建议。本文获得了广东粤语系的人群的 F_{ST} 值,15 个 STR 基因座的值均小于 0.01,说明若按 NRC 的标准来计算粤语地区人的 DNA 图谱概率,是一种保守的估计,对被告有利。

参考文献(References):

- [1] Zhivotovsky L, Ahmed S, Wang W, Bittles A H. The forensic DNA implications of genetic differentiation between endogamous communities[J]. *Forensic Sci Int*, 2001, 119:269~272.
- [2] YU Jian-Kun, CHU Jia-You, QIAN Ya-Ping, SUN Hao, SHI Lei, SHI Li, CHU Zheng-Tao, YANG Zhao-Qing, LIN Ke-Qing, TAO Yu-Fen, HUANG Wei, JIN Li. Genetic relationships among six Chinese populations revealed by analysis of 30 autosomal STRs[J]. *Acta Genetica Sinica*, 2001, 28(8): 699~706.
俞建昆, 褚嘉佑, 钱亚屏, 孙浩, 史磊, 史荔, 初正韬, 杨昭庆, 林克勤, 陶玉芬, 黄薇, 金力. 应用 30 个常染色体 STR 位点研究中国 6 个民族群体的遗传关系[J]. *遗传学报*. 2001, 28(8): 699~706.
- [3] Ayres K L. Relatedness testing in subdivided populations[J]. *Forensic Sci. Int.*, 2000, 114(2): 107~115.
- [4] Lee J W, Lee H S, Park M, Hwang J J. Evaluation of DNA match probability in criminal case[J]. *Forensic Sci Int*, 2001, 116(2-3): 139~148.
- [5] National Research Council. The evaluation of forensic DNA evidence[M]. Washington D C: National Academy Press, 1996.
- [6] LI Sheng-Bin, LAI Jiang-Hua, GAO Shu-Hui, ZHENG Hai-

Bo, FENG Ji-Dong, ZHAO Jian-Min, LI Shuang-Ding, FENG Cheng-Bo, JIN Tian-Bo, WANG Jian, YANG Huan-Ming. STR polymorphisms in five Chinese ethnic groups(2)[J]. *Acta Genetica Sinica*, 2000, 27(12): 1035~1041.

李生斌, 赖江华, 高树辉, 郑海波, 冯继东, 赵建民, 李双顶, 冯成博, 金天博, 汪健, 杨焕明. 中国五个民族 STR 位点遗传多态性(2)[J]. *遗传学报*. 2000, 27(12): 1035~1041.

- [7] GAO Shu-Hui, WANG Xin-Huai, LAI Jiang-Hua, ZHENG Hai-Bo, LI Sheng-Bin. Polymorphism of nine STR locus in Nu population from Yunnan province [J]. *Hereditas (Beijing)*, 2002, 24(2): 125~130.

高树辉, 王新淮, 赖江华, 郑海波, 李生斌. 云南怒族 STR 基因座遗传多态性研究[J]. *遗传*, 2002, 24(2): 125~130.

- [8] FENG Ming-Liang, JI Yun, LU Qiong, Ma Jun, JI Yue-Hua, YANG Ying. The polymorphism distributions of nine STR loci in Shanghai Han population[J]. *Hereditas (Beijing)*, 2002, 24(4): 403~406.

冯明亮, 季芸, 陆琼, 马俊, 稽月华, 杨颖. 用多重 PCR 检测上海地区汉族人群 9 个 STR 基因座的多态性[J]. *遗传*, 2002, 24(4): 403~406.

- [9] Zaykin D, Zhivotovsky L, Weir B S. Exact tests for association between alleles at arbitrary numbers of loci [J]. *Genetica*, 1995, 96(1-2): 169~178.

- [10] Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenism. [J]. *J Heredity*, 1995, 86: 248~249.

- [11] Hammond H A, Jin L, Zhong Y, Caskey C T, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications[J]. *Am J Hum Genet*, 1994, 55(1): 175~189.

- [12] Graham J, Curran J, Weir B S. Conditional genotypic probabilities for microsatellite loci[J]. *Genetics*, 2000, 155(4): 1973~1980.

《中华医学遗传学杂志》被美国《工程索引》(EI) 收录为核心期刊

《中华医学遗传学杂志》在编委会和编辑部的辛勤耕耘和努力下,经过多年来与作者、读者和审稿老师的共同努力和合作,自 1998 年被收录 Ei 扩展版(Ei PageOne)以来,2003 年又被收录入 Ei 核心期刊(Ei Compendex)的网络版和光盘版。从 1998 年以来在本刊发表有英文文摘的文章,作者都可以通过教育网在 Ei CompendexWeb 数据库的主页 <http://www.engineeringvillage2.com.cn> 进行检索自己的文章。如不能进入该数据库的作者,可与编辑部联系(028-85501165)。目前《中华医学遗传学杂志》已被美国《生物学文摘》(BA)、《化学文摘》(CA)、《工程索引》(Ei)、《医学索引》(IM)及荷兰《医学文摘》(EM)和俄罗斯《文摘杂志》(AJ)六个著名检索系统收录。这些检索系统的收录使该刊的知名度和国际化程度进一步提高,国际交流范围进一步扩大,这对促进我国医学遗传学事业的发展和对外交流有着重大的历史意义。欢迎投稿,欢迎订阅(邮发代号:62-613)。