

用 HLA 基因频率计算人群间的遗传距离

赵桐茂 张工梁

(上海市中心血站 HLA 分型实验室)

袁义达 杜若甫

(中国科学院遗传研究所)

关键词 人类群体遗传学; 遗传距离; 人类白细胞抗原

内 容 提 要

本文用白细胞抗原 (HLA) 的基因频率按 Nei 氏法计算了中国 11 个民族和其他人种与国家的六个人群间的遗传距离，并按 Sneath 和 Sokal 氏聚类原则绘制了这些群体的系统树。得出的结果与用红细胞血型资料所计算的结果基本吻合，群体间遗传距离和群体分布的地理距离呈一定平行关系。

一、前 言

研究人群间的血缘关系以及遗传差异的起源与维持，是群体遗传学中的一个重要内容。随着人类血型研究的进展，积累了大量的血型基因频率的数据，这就有可能定量描述两个群体间由于选择、突变、迁移和漂变所引起的遗传歧异 (genetic divergence) 的程度，为此，引进了“遗传距离”的概念，并导出数种根据基因频率估计“距离”的方法。国外这方面工作开展甚多，国内还只有根据 ABO、MN、P、Rh 等四个红细胞血型系统的基因频率研究遗传距离的报告(袁义达等，1983)。

人类白细胞抗原 (HLA) 是迄今所知人类最复杂的一个遗传多态性系统，HLA 遗传区域至少存在 A、B、C、D 和 DR 等五个遗传座位，已检出的等位基因在 90 种以上，表型有上亿种之多(赵桐茂等，1984)。在不同人种和同一人种的不同氏族间，HLA 基因频率表现出明显的差异，因此利用 HLA 系统估计群体遗传距离有一定意义。本文单独用 HLA 系统的基因频率计算了国内外 17 个群体间的遗传距离，现报告如下。

二、材料与方法

1. HLA-A, B 基因频率 汉族取自上海(赵桐茂等，1984)、海南岛(孙逸平等，1982)、广西(孙逸平等，1982)、四川(袁必文等，1982)、湖北(程敦秀等，1983)、北京(丁

素秋等, 1983) 等六个地区人群基因频率的算术平均值。蒙古族取自内蒙古和新疆两地三个蒙古族群体的基因频率算术平均值(宋长兴等, 1983)。其他为新疆的回族(宋长兴等, 1984)、哈萨克族(程定珍等, 1984) 和维吾尔族(程定珍等, 1983), 西南地区藏族(周鸣生等, 1982), 东北地区的朝鲜族(宋芳吉等, 1983), 四川的彝族(周鸣生等, 1982), 广西的壮族(顾文娟等, 1984)、瑶族和侗族(孙逸平等, 1982)。除彝、藏、瑶和侗族以外的资料, 均为与美国红十字会李正道博士合作调查取得。高加索人, 尼格罗人和日本人资料均直接取自第八届国际组织相容性专题讨论会论文集(Baur *et al.*, 1980)。

基因频率 P 计算公式为 $P = 1 - \sqrt{1 - f}$, 式中 f 为抗原频率。空白基因频率 = 1—(该座位上其它等位基因频率之和)。

2. 遗传距离 D 按 Nei 氏公式计算 (Nei, 1978)。

$$D = -\ln(J_{XY}/\sqrt{J_X \cdot J_Y})$$

$$J_X = \frac{1}{n} \left(\sum_{i=1}^n \sum_{j=1}^k X_{ij}^2 \right)$$

$$J_Y = \frac{1}{n} \left(\sum_{i=1}^n \sum_{j=1}^k Y_{ij}^2 \right)$$

$$J_{XY} = \frac{1}{n} \left(\sum_{i=1}^n \sum_{j=1}^k X_{ij} \cdot Y_{ij} \right)$$

X_{ij} 和 Y_{ij} 分别代表群体 X 和群体 Y 中第 i 个座位上第 j 个等位基因的频率; n 代表座位数, k 代表每一个座位上的等位基因数。

3. 群体的系统树 (phylogenetic tree) 根据 Sneath 和 Sokal(1973) 的聚类原则绘制。遗传距离的计算和聚类分析均用 TRS-80 微型电子计算机进行。

三、结果与讨论

1. 本文比较的 17 个群体 HLA 基因频率和遗传距离分别列在表 1 和表 2。根据系统树(图 1), 可以看到:

(1) 取 Nei 氏双平行线长度在 0.2—0.3 范围, 可以得到三个集群 (cluster), 恰好对应三个主要人种, 而且尼格罗人种与高加索人种的距离要比与蒙古人种距离为近。这和国外多数作者用不同的遗传标记得到的结论一致 (Letter, 1973)。

(2) 在属于蒙古人种的群体中, 明显地表现出不均一性。如取双平行线长度 0.17 的水平, 每个集群内的 D 值小于 0.15, 12 个群体相聚为三组: 汉、壮、瑶、侗、藏、彝为一组, 其中汉族首先依次和壮族、瑶族、侗族相聚, 然后与藏族相聚, 最后与彝族相聚。蒙古、哈萨克、回、维吾尔为一组, 蒙古族首先与哈萨克相聚, 再与回族相聚, 最后与维吾尔族相聚。朝鲜族和日本人为一组。这和用四个红细胞血型系统基因频率计算所得的结果基本吻合, 和民族学研究结果也是很一致的 (袁义达等, 1983)。值得注意的是这三组之间的遗传距离和群体分布的地理距离明显地呈平行关系, 提示这些群体间的迁移和混杂明显受到地理位置的限制。

表 1 HLA 基因频率 (%) ($\times 10^{-4}$)

基因	汉 1		壮 2		瑶 3		侗 4		藏 5		彝 6		蒙古 7		哈萨克 8		回 9		维吾尔 10		朝鲜 11		日本人 12		非洲尼格罗 13		美洲尼格罗 14		澳洲高加索 15		欧洲高加索 16		北美美洲高加索 17			
	汉	1	壮	2	瑶	3	侗	4	藏	5	彝	6	蒙古	7	哈萨克	8	回	9	维吾尔	10	朝鲜	11	日本人	12	非洲尼格罗	13	美洲尼格罗	14	澳洲高加索	15	欧洲高加索	16	北美美洲高加索	17		
A1	227	0	0	45	69	69	270	1127	742	1120	1313	101	53	245	273	1876	1478	1478	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379	1379
A2	3094	3072	3400	2742	4225	2881	2663	2504	2532	2109	2789	2377	1042	1605	2392	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492	2492		
A3	275	101	50	45	353	0	764	897	392	837	253	53	624	988	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159	1313	1159		
A9	1443	1743	2165	2160	1782	477	2015	1743	1408	1348	2909	3589	1470	1422	972	1422	972	1422	972	1422	972	1422	972	1422	972	1422	972	1422	972	1422	972	1422	972	1422		
A10	395	101	353	89	488	538	421	581	291	526	563	1061	1091	1164	722	610	597	597	597	597	597	597	597	597	597	597	597	597	597	597	597	597	597	597	597	
A11	2793	3836	2825	3661	1834	2975	1164	950	1508	888	1056	879	35	82	673	585	585	633	633	633	633	633	633	633	633	633	633	633	633	633	633	633	633	633	633	
A28	77	0	50	181	138	67	48	241	145	239	202	53	1108	633	494	391	391	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	
Aw19	906	504	250	409	703	805	443	482	733	1598	455	1491	3366	2751	1221	1493	1493	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	1530	
Aw36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Aw43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
A 空白	790	643	907	668	408	1987	1356	1860	1871	1142	1672	417	700	918	278	278	278	278	278	278	278	278	278	278	278	278	278	278	278	278	278	278	278			
B5	704	457	881	411	1438	1309	1005	980	1141	1581	1096	1922	502	1147	552	955	955	661	661	661	661	661	661	661	661	661	661	661	661	661	661	661	661	661	661	
B7	189	253	100	0	139	762	609	690	194	483	356	579	1158	738	1065	857	857	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	958	
B8	91	50	50	45	69	133	210	438	48	483	50	11	420	241	1288	817	817	905	905	905	905	905	905	905	905	905	905	905	905	905	905	905	905	905	905	
B12	189	101	50	0	210	408	508	433	743	731	1168	658	985	1255	1322	1202	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445	1445		
B13	1030	945	1383	2160	425	477	594	589	854	940	408	203	35	109	233	233	233	233	233	233	233	233	233	233	233	233	233	233	233	233	233	233	233			
B14	29	0	99	0	0	131	96	241	48	287	101	11	385	399	581	294	294	483	483	483	483	483	483	483	483	483	483	483	483	483	483	483	483	483	483	
B15	1131	945	1045	1238	1259	981	593	488	646	47	1225	889	35	109	634	588	588	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	
Bw16	780	925	0	646	1024	762	335	239	643	333	252	313	350	55	232	458	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507	507		
B17	820	945	561	551	210	270	948	845	543	633	461	84	1855	1183	478	425	425	477	477	477	477	477	477	477	477	477	477	477	477	477	477	477	477	477	477	
B18	62	50	150	45	0	66	84	241	145	95	0	0	0	0	420	349	349	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569	569
Bw21	26	0	0	0	0	0	0	0	151	241	392	336	0	32	70	464	445	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352	352
Bw22	563	771	881	791	425	762	418	290	286	922	1126	140	55	349	276	276	276	276	276	276	276	276	276	276	276	276	276	276	276	276	276	276	276			
B27	241	101	150	318	69	270	151	339	145	190	101	42	140	138	520	391	391	382	382	382	382	382	382	382	382	382	382	382	382	382	382	382	382	382	382	
Bw35	327	151	353	135	281	338	561	95	442	483	780	733	438	760	611	959	803	803	803	803	803	803	803	803	803	803	803	803	803	803	803	803	803	803	803	
B37	42	0	0	135	139	0	379	144	47	50	53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
B40	1868	1637	1735	2039	2500	1724	1618	1622	1091	673	1725	1541	35	273	875	507	507	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660	660	
Bw41	8	0	0	0	0	0	0	0	0	0	47	50	37	222	82	87	103	103	202	202	202	202	202	202	202	202	202	202	202	202	202	202	202	202	202	202
Bw42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Bw46	671	1574	1262	318	281	131	134	0	646	336	151	332	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
Bw47	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Bw48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
B 空白	1221	1095	1300	1168	1531	1476	1605	2085	1744	1989	1054	1124	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100		
样本数	930	100	101	112	72	75	286	105	104	106	100	950	143	184	172	172	172	172	172	172	172	172	172	172	172	172	172	172	172	172	172	172	172			

表 2 十七个群体间的遗传距离 $D(\times 10^{-4})$

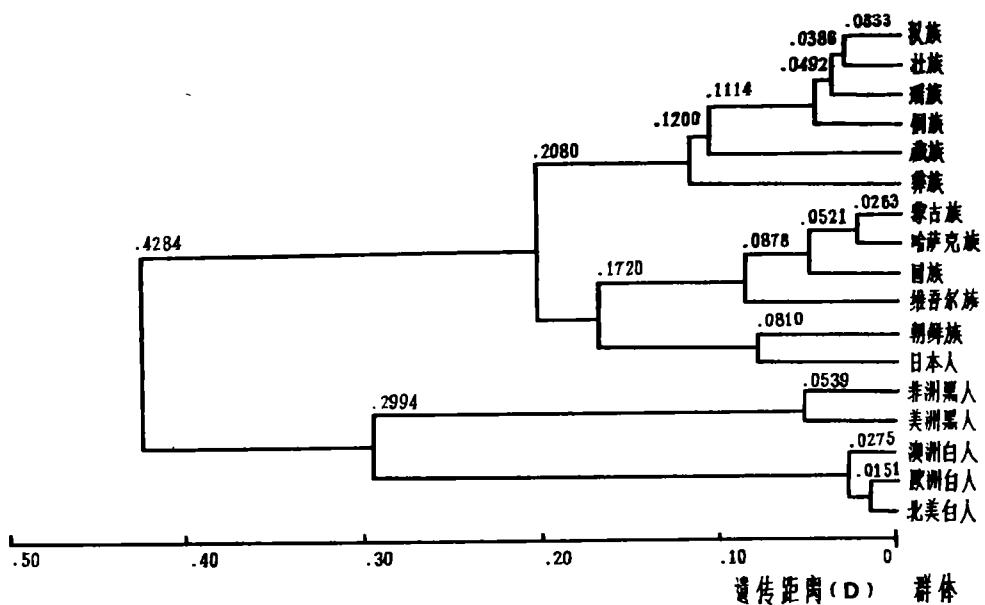


图 1 17 个群体的系统树 (根据 HLA 基因频率绘制)

Phylogenetic tree for 17 populations (Constructed on the basis of gene frequencies of HLA)

2. HLA-A1 和 Bw 21 等基因，在高加索人种中的频率比较高，而在东方人中比较少，新疆维吾尔族的这些基因频率比国内其他民族高，接近高加索人种的数值。对维吾尔族的免疫球蛋白 Gm 因子的调查（赵桐茂等，1983）也发现该群体中存在一些高加索人种的 Gm 单倍型，如 Gm^{3,5}。此外，Rh 阴性的比例，也高于国内其他民族，接近高加索人种。这些事实提示维吾尔族有高加索人种血缘。而被认为是蒙古人种特有的 HLA-Bw 46 抗原，在维吾尔人中仍占有相当比例（6%），基因频率约为汉族人的一半。而且根据遗传距离，维吾尔族也首先和蒙古族、哈萨克族、回族相聚，然后再和朝鲜族、日本人以及汉族、壮族等一大组相聚，最后才与尼格罗人及高加索人种相聚。这说明即使维吾尔族有高加索人种

血缘，其程度也是有限的。

3. 根据 ABO、P、MN 和 Rh 血型计算的遗传距离进行聚类，汉族首先和中国北方的蒙古、回等民族相聚（袁义达等，1983）；而根据 HLA 数据，汉族首先和南方的一些民族相聚。这可能是由于取样差异所造成的。因为从 ABO 血型（陈稚勇等，1982）和 Gm 因子（赵桐茂等，1983）分布看，汉族本身并不是一个均一的群体，南方和北方汉族之间存在一定差异。本文 HLA 基因频率取自六个汉族群体，其中五个在长江以南或长江附近，北方的群体只有一个，因此本文汉族一般只代表南方汉族的情况。事实上，根据 HLA 基因频率，也可以看出不同汉族群体之间的遗传距离和地理分布也呈一定的平行关系（表 3）。

表 3 六个汉族群体间的遗传距离 ($\times 10^{-4}$)

海南岛 (1)					
广西 (2)		557			
四川 (3)		403	342		
上海 (4)		785	439	364	
武汉 (5)		704	882	532	630
北京 (6)		739	943	480	450 627
		(1)	(2)	(3)	(4) (5) (6)

4. 遗传距离是各种遗传标记的基因频率平均值的函数，反映特定群体间遗传歧异的程度。所用的遗传标记越多，基因频率越精确，越能反映出实际情况。本文单独使用 HLA 系统，得到的结果大体上是可信的。但为了确切估计这些群体间的遗传距离，最好是合并使用其他一些遗传标记，特别是红细胞酶型和血清蛋白型等。预料在不远的将来，国内将会积累这方面的资料。

（1983 年 10 月 10 日收稿）

参 考 文 献

- 丁素秋等，1983。HLA 抗原在北京人中的分布。中华器官移植杂志，4(3): 110。
- 孙逸平等，1982。我国部份少数民族白细胞抗原 (HLA) 多态性的调查。中华微生物和免疫学杂志，2(3): 133—137。
- 宋长兴等，1983。我国蒙古族 HLA-A、B、C 抗原分布的调查。中华器官移植杂志，4(3): 106—107。
- 宋长兴等，1984。新疆地区回族 HLA-A、B、C 抗原调查。中华器官移植杂志，5(1): 40。
- 宋芳吉等，1983。沈阳地区朝鲜族人群中 HLA 抗原的分布。中华器官移植杂志 4(3): 108—109。
- 周鸣生等，1982。西南地区藏族和彝族人群的 HLA-A、B 和 DR 抗原的分布调查。四川医学院学报，13: 372—377。
- 陈稚勇等，1982。中国人 ABO 血型分布。遗传，4(2): 4—7。
- 赵桐茂等，1983。人体免疫球蛋白 Gm、Km 因子在汉族和维吾尔族人群的分布。遗传学报，10: 311—316。
- 赵桐茂等，1984。HLA 分型原理与应用。上海科技出版社。
- 袁义达等，1983。中国十七个民族间的遗传距离的初步研究。遗传学报，10: 398—405。
- 袁必文等，1982。四川地区汉族人群 HLA-A、B 抗原的分布调查。四川医学院学报，13: 361—365。
- 顾文娟等，1984。HLA 抗原在壮族人中的分布。中华器官移植杂志，5(1): 38—39。
- 程定珍等，新疆哈萨克族人群 HLA-ABC 抗原分布调查。(待发表)。
- 程定珍等，1983。新疆维吾尔族人群 HLA 抗原分布的调查。中华器官移植杂志，4(3): 111—112。
- 程教秀等，1983。HLA 抗原在湖北汉族人群中的分布调查。中华器官移植杂志，4(3): 102—103。
- Baur, M. P. et al., 1980. Population analysis of HLA-A, B, C, DR and other genetic markers. in *Histo-compatibility Testing*, ed. Terasaki, P. I., UCLA, California, pp. 955.
- Latter, B. D. H., 1973. The estimation of genetic divergence between population based on gene frequency data. *Am. J. Hum. Genet.*, 25(3):247.

- Nei, M., 1978. The theory of genetic distance and evolution of human races. *Jap. J. Hum. Genet.*, **23**; 341—369.
Sneath, P. H. A. and Sokal, R. R., 1973. *Numerical taxonomy*. Freeman, San Francisco, pp. 201—213.

GENETIC DISTANCES ESTIMATED ON THE BASIS OF HLA

Zhao Tongmao, Zhang Gongliang

(*HLA typing Laboratory, Shanghai Blood Transfusion Center*)

Yuan Yida, Du Ruofu

(*Institute of Genetics, Academia Sinica*)

Key words Human population genetics; Genetic distance; Human leucocyte antigen

Abstract

The genetic distances between 17 populations were computed by Nei's method (1978) and a phylogenetic tree of these populations was constructed on the basis of genetic distances by the unweighted pair-group method of clustering (Sneath and Sokal, 1963). The results of present study agree with those obtained on the basis of gene frequencies of red cell blood group systems. There is a parallel relationship between genetic distances and the geographic distances of 17 populations.