

# RFLP 与三大人种间和人种内 分化演变的关系

吕宝忠 陈捷

(上海市肿瘤研究所, 上海 200032)

曾溢滔

(上海医学遗传研究所, 上海 200040)

**关键词** 限制性片段长度多态性; 人种; 遗传距离; 进化

## 内 容 提 要

限制性片段长度多态性(RFLP)指用限制性内切酶切割特定基因后产生的一类多态性, 不仅可广泛地应用于产前诊断, 且对于人类学研究和民族迁移的探讨也有重要意义。本文以 $\beta$ 珠蛋白、PAH和DMD基因连锁序列的RFLP频率为依据, 分析三大人种间和人种内的分化演变。研究表明, 本文以RFLP频率为依据构建的系统树在拓扑上与Nei等以基因座位频率为依据构建的树是一致的, 但尼格罗人种和其他人种间的遗传距离却比Nei等以基因座位频率数据为据报道的大一倍左右, 看来这是一个有待进一步探讨的问题。研究结果还指出, 高加索人种内的分化演变约发生在距今9000年前, 从而在分子水平上可靠地印证了人群地理迁移及历史语言学论断。本文还初步分析了遗传距离与遗传病发病率的关系, 经双对数变换处理后, 两者似呈现负相关, 或许提示DNA水平上的多态性维持机制可用中性突变学说来解释。

限制性片段长度多态性(restriction fragment length polymorphism, RFLP)不仅可用于疾病基因的产前诊断, 而且还可用于人类学研究。Nei(1987)曾以线粒体DNA的RFLP频率为据, 探讨了人类三大种族间和种族内的演变, 但迄今至少在国内尚未查见以核基因DNA的RFLP频率为据对人类三大种族间和种族内的分化演变的研究报道, 本文以 $\beta$ 珠蛋白基因簇、PAH(苯丙氨酸羟化酶)基因以及DMD(Duchenne肌营养不良症)基因相连锁的DNA序列的诸RFLP频率数据为

收稿日期: 1991-06-24

材料, 应用 Nei 的标准遗传距离(简称遗传距离)公式, 并以 UPGMA(不加权成对比较法)计算机程序构建系统树(吕宝忠等, 1989), 分析人类三大种族间和种族内不同群体的分化演变, 从而为人类学研究提供可靠依据。同时又从遗传距离期望值同其相应基因的遗传病发病率间的关系初步探讨 DNA 水平上的多态性维持机制。

### 一、材料与方法

对正常的  $\beta$  珠蛋白基因簇(酶切图谱见图 1)来说, 本文汇集了三大种族共有的 RFLP 频率数据, 资料来自曾溢滔等(1987a、1988b)(见表 1)。

表 1  $\beta$  珠蛋白基因簇、PAH 基因和 DMD 连锁序列的 RFLP 频率在人种间和人种内的分布

The RFLP frequencies of  $\beta$ -globin gene cluster, PAH gene and DMD-linked sequence among and within human races

基 因 Gene	RFLP	尼格罗人种 Negroid	高加索人种 Caucasoid		蒙古人种 Mongoloid
$\beta$ 珠蛋白基因簇		(美国黑人) (American Black)	(意大利人) (Italian)	(印度人) (Indian)	(中国人) (Chinese)
$\beta$ -globin gene cluster	Hinc II-5'c	0.10	0.76	0.78	0.75
	Hind III-G $\gamma$	0.41	0.26	0.30	0.19
	Hind III-A $\gamma$	0.16	0.06	0.06	0.28
	Hinc II- $\psi\beta 1$	0.15	0.20	0.17	0.07
	Hinc II-3 $\psi\beta 1$	0.76	0.28	0.27	0.15
	Hinf I-5 $\beta$	0.70	0.95	1.00	0.91
	Hgi AI-5 $\beta$	0.96	0.86	0.82	0.47
	Ava II- $\beta$	0.96	0.86	0.78	0.49
	Hpa I-3 $\beta$	0.93	1.00	1.00	0.89
	Bam HI-3 $\beta$	0.90	0.74	0.82	0.62
总人数 Total		(大) (Large)	(大) (Large)	(大) (Large)	(大) (Large)
PAH 基因及其侧翼 PAH gene and is untranslated regions			(法国人)	(丹麦人)	(中国人)
	Bgl II		0.32	0.41	0.13
	Pvu IIa		0.69	0.56	0.70
	Puv IIb		0.12	0.31	0.10
	Eco RI		0.56	0.41	0.83
	Msp I		0.60	0.62	0.12
	Hind III		0.29	0.39	0.81
	Eco RV		0.47	0.53	0.77
总人数 Total			(68)	(66)	(80)
DMD 连锁序列			(美国白人) (American)	(英国人) (British)	(中国人) (Chinese)
DMD-linked sequence	Xmn I(pERT87.1)		0.34	0.31	0.56
	Xmn I(pERT87.15)		0.68	0.68	0.48
	Taq I(pERT87.15)		0.67	0.67	0.75
	Taq I(pERT87.8)		0.58	0.74	0.31
总人数 Total			(25)	(79)	(70)

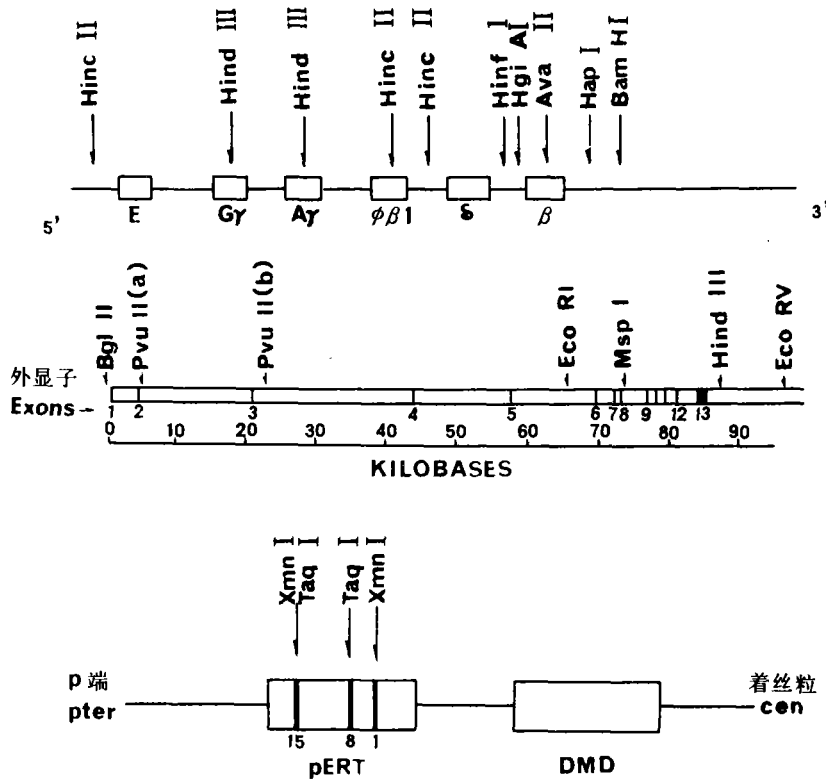


图 1  $\beta$  珠蛋白基因簇(上)、PAH 基因包括 5' 和 3' 非翻译区(中) 和 DMD 连锁序列(下)的 RFLP 位点分布图

The distribution map of RFLP sites for  $\beta$ -globin gene cluster(upper), PAH gene including its 5' and 3' untranslated regions(middle) and DMD-linked sequence(lower)

对正常的 PAH 基因来说, 本文汇集了两大种族内三个国家人群共有的 RFLP 频率(酶切图谱见图 1), 资料来自曾溢韬等(1988a c)、Rey 等(1988)和 Lidsky 等(1985)(见表 1)。对 DMD 基因连锁序列来说, 本文汇集了两大种族三个国家共有的 RFLP 频率数据(酶切图谱见图 1), 资料来自曾溢滔等(1987b)、Bertelson 等(1986)和 Norman 等(1989)(见表 1)。

Nei 的遗传距离观察值  $d$ 、方差  $V(d)$  以及遗传距离标准误 S.E. 的公式分别为:

$$d = -\ln I$$

而

$$I = \frac{\frac{\sum X_i Y_i}{n}}{\sqrt{\frac{\sum X_i^2}{n} \cdot \frac{\sum Y_i^2}{n}}} = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \cdot \sum Y_i^2}}$$

(其中 $\sum$ 表示*i*从第1至第*n*个的RFLP频率的总和)

$$V(d) = \frac{1-I}{In} \quad S.E. = \sqrt{\frac{V(d)}{n}}$$

(上述公式中,  $x_i$  和  $y_i$  分别表示相比较的一对  $x$  和  $y$  群体中某基因中第  $i$  个 RFLP 的频率,  $n$  为该基因的不同 RFLP 的种类数)。当得到诸遗传距离观察值后, 即将其排成矩阵输入 UPGMA 程序以构建系统树。

通过高加索人种内不同人群间的遗传距离同人群地理迁移和历史语言学的资料进行比较, 则可以可靠地估计该种族内不同人群间分化起源的大致时间。

## 二、结 果

### 1. $\beta$ 珠蛋白基因簇的 RFLP 构建的系统树

表 2 从  $\beta$  珠蛋白基因簇的 RFLP 比较了三大种族间和高加索人种内不同人群间的遗传距离观察值; 然后以矩阵形式输入 UPGMA 程序, 以获得遗传距离期望值数据; 再根据聚类指标(由程序运算后得知)构建成图 2 的系统树。

表 2 人种间和人种内的  $\beta$  珠蛋白基因簇的多态性遗传距离期望值和观察值

Expected and observed genetic distances among and within the human races

for  $\beta$ -globin gene cluster polymorphism

		1	2	3	4
1. 中国人	Chinese		0.030	0.030	0.121
2. 意大利人	Italian	0.034(0.018)		0.002	0.021
3. 印度人	Indian	0.025(0.016)	0.002(0.005)		0.021
4. 美国黑人	U.S.Black	0.179(0.044)	0.089(0.031)	0.096(0.032)	

注: 右上方的数值为遗传距离的期望值, 左下方的数值为遗传距离的观察值。左下方括号内的数值为观察值的标准误。

Expected and observed genetic distances are showed in the upper fight half and in the lower left half, respectively. Standard errors of observed genetic distances in parenthesis.

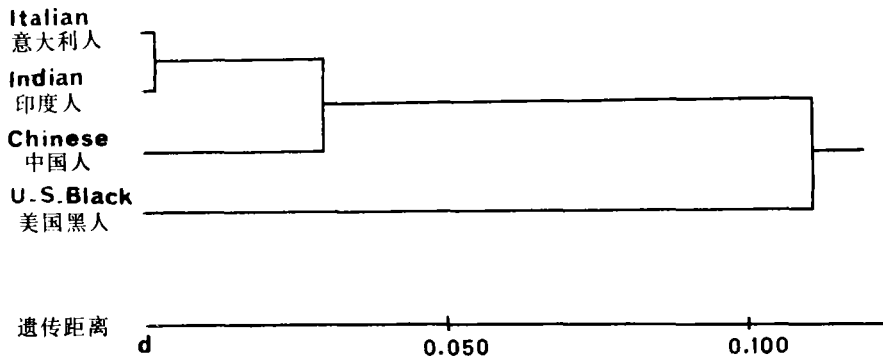


图 2 关于人类三大种族间和种族内  $\beta$  珠蛋白基因簇多态性数据的系统树  
Phylogenetic tree for  $\beta$ -globin gene cluster polymorphism among  
and within the three main races of *Homo sapiens*

由表 2 和图 2 可见, 尼格罗人种与其他人种间的遗传距离最远, 其次为蒙古人种与高加索人种间, 最近的理所当然的为人种内不同人群的距离。此结果与 Nei 和 Roychoudhury(1974)所获得的基本一致, 但尼格罗人种与其他人种间的遗传距离比本文报道的比 Nei 和 Roychoudhury(1974)的要大一倍左右。

## 2. PAH 基因的 RFLP 构建的系统树

表 3 从 PAH 基因的 RFLP 比较了两大种族间和高加索人种内不同人群间的遗传距离。根据同上处理, 获得遗传距离期望值数据和构建成图 3 的系统树。该树再一次指出, 人种间的遗传距离比人种内不同人群间的要大。

表 3 人种间和人种内的 PAH 基因多态性的遗传距离期望值和观察值

Expected and observed genetic distances between and within the

human races for PAH gene polymorphism

		1	2	3
1. 中国人	Chinese		0.181	0.181
2. 丹麦人	Dana	0.196(0.050)		0.032
3. 法国人	French	0.165(0.060)	0.032(0.026)	

注: 见表 2(see table 2).

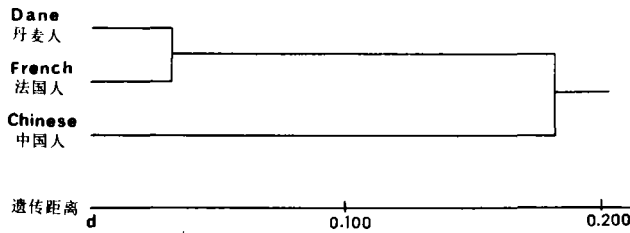


图 3 关于两大人种间和人种内 PAH 基因多态性数据的系统树

Phylogenetic tree for PAH gene polymorphism between and within two races of *Homo sapiens*

### 3. DMD 基因连锁序列的 RFLP 构建的系统树

表 4 从 DMD 基因连锁序列的 RFLP 比较了两大种族间和高加索人种内不同人群间的遗传距离。根据同上处理，获得遗传距离期望值数据和构建成图 4 的系统树。该树再次指出，人种间的遗传距离比人种内不同人群间的遗传距离要大。

表 4 人种间和人种内的 DMD 连锁序列多态性的遗传距离期望值和观察值

Expected and observed genetic distances between and within  
the human races for DMD-linked sequence polymorphism

		1	2	3
1. 中国人	Chinese		0.086	0.086
2. 英国人	British	0.104(0.084)		0.007
3. 美国人	American	0.067(0.066)	0.007(0.021)	

注: 见表 2(see table 2).

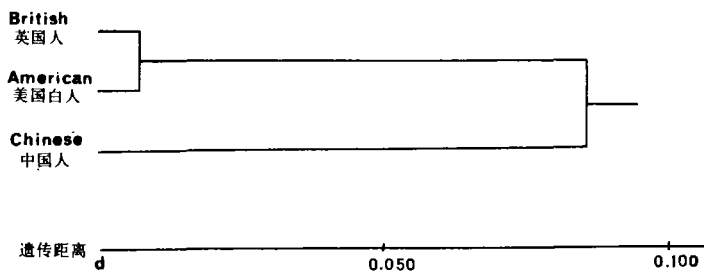


图 4 关于两大人种间和人种内 DMD 连锁序列多态性数据的系统树

Phylogenetic tree for DMD-linked sequence polymorphism between  
and within two races of *Homo sapiens*

#### 4. RFLP 的遗传距离期望值与该基因发病率的关系

虽然遗传病发病率的研究尚处于数据积累阶段, 但较大样本或较长时间积累的数据应该说是较可信的, 据此本文分别采用中国人的  $\beta$  地中海贫血症和苯酮尿症的发病率  $1/152$  和  $1/16000$ , 以及欧洲人 DMD 的发病率  $1/7000$ (因欧洲白人男性发病率约为  $1/3500$ , 而女性发病率显然要低得多, 故人群总发病可考虑为  $1/7000$ )作为参数。如以该参数的对数值同其相应的种族间(高加索人种与蒙古人种间)和种族内(高加索人种的不同人群间)的 RFLP 遗传距离期望值的对数值分别作相关分析, 则可获得两大人种间的 RFLP 遗传距离期望值与其相应遗传病发病率的相关系数为  $-0.943$ , 以及高加索人种内不同人群的 RFLP 遗传距离期望值与其相应遗传病发病率的相关数为  $-0.935$ 。如以  $t$  值检验法作显著性测验, 则可用 Fisher 的小样本  $r$  检验公式, 从而获得  $t_{\text{种族间}} \approx 2.8$  以及  $t_{\text{种族内}} \approx 2.6$ 。查  $t$  值单侧界限表当自由度为 1 时,  $P \approx 0.10$ 。两个相关系数均接近可能显著程度, 似表明可能存在负相关趋势。

### 三、 讨 论

现代人(*Homo sapiens*)根据肤色、面型以及毛发质地等形态特征, 可以明显地分为三大人种: 蒙古人种、高加索人种和尼格罗人种。尽管在这三大人种间已经而且还在继续发生着基因混杂, 但它们间如此明显的形态差异表明, 它们间已经历了漫长时间的隔离和分化。关于现代人的起源问题, 主要可划分为两种不同的理论: 多区域模型(认为不同人种种族特征是在当今发现它们的那些区域中经过漫长的期间进化而成, 只有在相对较近的时间中出现过人群迁移的情况, 如亚洲的蒙古人种越过白令海峡陆桥进入美洲演化成印第安人才属于例外)和“源于非洲”的单源模型(认为现代人起源于非洲, 后经人群迁移演化成不同人种)(见 Stringer, 1990)。目前两种理论仍在学术界进行争辩。与现代人起源密切有关的一个问题是种族特征即人种分化的起源问题。古人类学发掘的化石资料, 经热释光法和电子自旋共振法新技术断定年代, 支持非洲人种首先作为现代人分化而来。但这个推论能否获得分子群体遗传学或分子人类学的的数据支持呢? Nei 和 Roychoudhury(1974)根据三大人种共有的 35 个组位基因频率数据估计的遗传距离表明: 高加索人种和蒙古人种间约在 55 000 年前分化, 而尼格罗人种则分别在 115 000 年前和 120 000 年前同高加索人种和蒙古人种间发生分化。Nei(1987)于后来又以线粒体 DNA 的 RFLP 频率数据分析该问题, 鉴于线粒体 DNA 的突变率远远高于核基因 DNA, 因而在种族内多态性与种族间多态性之间看不出存在显著差异, 难以回答上述问题。本文以  $\beta$  珠蛋白基因簇等的 RFLP 频率数据所绘制的系统树在拓扑上与 Nei 和

Roychoudhury(1974)的结果大体一致,但本文中尼格罗人种比其他两个人种间的遗传距离约大 4 倍,显著不同于 Nei 和 Roychoudhury(1974)的结果,而于 Vigilant 等(1991)的结果更类似。由于本文采用的 RFLP 数据中多态性的产生绝大多数来自中性突变(见后述),似更符合“分子钟”假说,且与 Vigilant 等(1991)应用线粒体 DNA 测序比较结果相类似;但 Nei 和 Roychoudhury(1974)的优点是使用的座位数目较多。看来这仍是有待进一步研究的一个课题。

本文中高加索人种内不同群体间的遗传距离约为蒙古人种同高加索人种间的遗传距离  $1/6$ — $1/15$  之间。如果采纳 Nei 和 Roychoudhury(1974)蒙古人种同高加索人种间的分化发生在 55 000 年前的观点,则高加索人种不同群体间的分化起源应发生在  $55\ 000/6$  即 9167 年前。这个推论与历史和历史语言学的证据能否相符吗?人类的早期农业似起源于小麦、大麦的栽培以及羊群的放牧,发生在距今 9 000 年前。已经查明,欧洲并非早期农业的发源地之一。它是从近东 Anatolia 中部平原(今土耳其境内)传进来的。当然,早期农业的出现并不局限于上述近东地区,事实上还有一些区域,但欧洲的早期农业无疑是上述近东地区的早期农民及其后裔迁移形成的(Renfrew, 1989; 吕宝忠, 1983)。与农业传入欧洲密切相关的是历史语言学的证据,已表明:高加索人种内的不同人群包括亚洲的印度人直至几乎所有欧洲人在约 7 000 年前都使用同一种原始的印欧语(Davies, 1979)。历史语言学的研究还提示:在 10 000 年前的近东地区,当地人群使用着一种更原始语言,该原始语言是印欧语系、亚非语系和 Dravidian 语言未分化前的祖先语言(Renfrew, 1989)。近东早期农民迁移到欧洲,不仅传入了早期农业,也带来了原始的印欧语。我们的分子群体遗传学有关高加索人种不同群体的分化时间推论基本上与上述历史与历史语言学提供的资料是相吻合的。由于 Davies(1979)的论断自认为是比较粗略的估计,因而综合各方面资料包括本文的估计似可把高加索人种内现有人群的分化起源推定在 9 000 年前。对有 4 000 多年文字可考证的由 56 个民族组成的中华民族,起源与何时这一问题,本文未予讨论,但相信只要获得充足的各民族的 RFLP 数据,完全可望予以满意解答。

本文结果的第 4 部分指出, RFLP 频率数据推论的遗传距离期望值与其相应基因突变后的遗传病发病率的相关系数,在种族间和种族内分别为  $-0.943$  和  $-0.935$ 。尽管  $t$  测验仅表明两个相关系数均接近可能显著程度,但已可看到似存在负相关的趋势。如再检阅图 1,则可清楚地看到,三个基因的 RFLP 酶切位点大多数位于有编码功能的外显子外(仅  $\beta$  珠蛋白基因簇中  $Ava$  II 酶切位点在具有编码功能的  $\beta$  珠蛋白外显子中),因此由酶切位点估计的遗传距离期望值绝大多数都是由中性突变造成的。对中性突变来说, Kimura(1983)指出,基因的突变率受到其功能的制约。据此,遗传病发病率的大小应与该基因替换率大小呈现反比关系,因为发病率愈大的基因,其中性突变比例愈小,而发病率愈小的基因,其中性突变比例



则愈大。本文估计的遗传距离期望值是基因替换率大小的测度, 其与该基因发病率呈现负相关关系(经双对数变换), 似表明 RFLP 多态性维持机制由中性突变所维持, 即可用中性突变学说来解释。

### 参 考 文 献

- 吕宝忠, 1983. 优生学与人种遗传学. 自然杂志, 7:483—486.
- 吕宝忠、M. Nei, 1989. 免疫球蛋白重链可变段和 T 细胞受体可变段的分子进化研究. 遗传学报, 16:140—149.
- 曾溢滔等, 1988(a). 苯酮尿症的产前诊断. 中华医学杂志, 68:61—64.
- 曾溢滔等, 1988(b). 用 RFLP 连锁分析产前诊断  $\beta$ -地中海贫血. 遗传与疾病, 5(3):129—132.
- 曾溢滔等, 1988(c). 中国人苯丙氨酸羟化酶的限制酶位点多态性研究. 中国科学 B 辑, (3):297—302.
- 曾溢滔等, 1987(a). Disorders of haemoglobin in China. *J. Med. Genet.*, 24:578—583.
- 曾溢滔等, 1987(b). Restriction fragment length polymorphisms on Xp21 region of the chromosome in patient with Duchenne muscular dystrophy. *Kexue Tongbao*, 33(11):965.
- Bertelson, C. J. *et al.*, 1986. Localisation of Xp21 meiotic exchange points in Duchenne muscular dystrophy families. *J. Med. Genet.*, 23:531—537.
- Davies, P.(ed.), 1979. *The American Heritage Dictionary of the English Language*. pp.361—363. Dell Publishing Co. Inc. New York.
- Kimura, M., 1983. *The Neutrality Theory of Molecular Evolution*. Cambridge Uni. Press, London.
- Lidsky, A. S. *et al.*, 1985. Prenatal diagnosis of phenylketonuria by DNA analyses. *Lancet*, 1:549—551.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia Uni. Press, New York.
- Nei, M. and A. K. Roychoudhury, 1974. Genic variation within and between the three major races of man, Caucasoids, Negroids and Mongoloids. *Am. J. Hum. Genet.*, 26:421—443.
- Norman, A. M. *et al.*, 1989. Duchenne muscular dystrophy in Wales: impact of DNA linkage analysis and cDNA detection screening. *J. Med. Genet.*, 26:565—571.
- Renfrew, C., 1989. The origins Indo-European languages. *Scientific American*, 261(4):82—90.
- Rey, F. *et al.*, 1988. Clinical and molecular heterogeneity of phenylalanine hydroxylase deficiencies in France. *Am. J. Hum. Genet.*, 43:914—921.
- Stringer, C. B., 1990. The emergence of modern humans. *Scientific American*, 263(6):68—75.
- Vigilant, L. *et al.*, 1991. African populations and the evolution of human mitochondrial DNA. *Science*, 253:1503—1507.

DIFFERENTIATION AND EVOLUTION OF THREE MAJOR RACES  
OF *HOMO SAPIENS* INFERRED FROM RFLP RATE

Lü Baozhong    Chen Jie

(Shanghai Cancer Institute, Shanghai 200032)

Zen Yitao

(Shanghai Institute of Medical Genetics, Shanghai 200040)

**Key words** RFLP; Human races; Genetic distance; Evolution

**Summary**

This paper studies the difference of RFLP rates(including  $\beta$ -globin gene cluster, PAH gene and DMD-linked sequences)among and within the three major races of *Homo sapiens* on one hand, and the differentiation and evolution in modern humans on the other hand. The phylogenetic trees we reconstructed with UPGMA are substantially similar to those by Nei and Roychoudhury, but the genetic distance estimated is twice as that estimated by them between Negroid and other races. The differentiation and evolution within Caucasoids occurred as late as about 9 000 years ago. This is essentially concordant(and probably more reliable)with the evidences collected from human migration and etymology. This paper provides a preliminary analysis of the relationship between genetic distances and corresponding incidences of genetic diseases and, shows the tendency of negative correlation between them through double logarithmic transformation. Therefore, the neutral theory seems to be more reasonable for the maintenance of RFLP polymorphism.