Am. J. Hum. Genet. 40:527–536, 1987

Population Genetics of Coagulant Factor IX: Frequencies of Two DNA Polymorphisms in Five Ethnic Groups

DENNIS B. LUBAHN,* SUSAN T. LORD,* JOHN BOSCO,† JULIE KIRSHTEIN,* O. J. JEFFRIES,* NANCY PARKER,* CYNDY LEVTZOW,* LAWRENCE M. SILVERMAN,* AND JOHN B. GRAHAM*

*Department of Pathology, School of Medicine, University of North Carolina, Chapel Hill, NC 27514; and †Department of Medicine, University of Malaysia, Kuala Lumpur, Malaya

SUMMARY

Two frequently used restriction-enzyme polymorphisms (RFLPs) of coagulant F.IX, *Taq*I and *Xmn*I, have been examined in five ethnic groups: white Americans, black Americans, East Indians, Chinese, and Malays. There is a distinct "cline" in the frequencies of both polymorphisms, from white Americans to Malays. The rarer type 2 alleles of both polymorphisms, in which middle recognition sites are present—and which in our sample reach their highest frequencies in white Americans—are marginally higher in four groups of Europeans previously reported by others. The frequencies of the rarer alleles are significantly higher in Europeans than in black Americans and East Indians, and these alleles are essentially absent in Chinese and Malays. The frequency of heterozygosity diminishes in the same order, being zero in Malays for both polymorphisms. The polymorphisms are in strong linkage disequilibrium, and in all groups the type 1 allele for *Taq*I is disproportionately accompanied by the type 1 allele for *Xmn*I. The paucity of type 2 alleles and the low rate of heterozygosity in four non-European groups suggest that the polymorphisms will be of little diagnostic value south of Gibraltar and east of Suez. This prediction is confirmed by the observed haplotype frequencies in the black American and the Oriental groups.

Received November 7, 1986; revision received January 5, 1987.

Address for correspondence and reprints: Dr. John B. Graham, Department of Pathology, School of Medicine, University of North Carolina, Chapel Hill, NC 27514.

 0002-9297/87/4006-0005$02.00

INTRODUCTION

F.IX is a glycoprotein involved in the blood-coagulation cascade whose gene on the X chromosome has been isolated and characterized several times in recent years (Choo et al. 1982; Kurachi and Davie 1982; Jaye et al. 1983; Anson et al. 1984; Jagadeeswaran et al. 1984; McGraw et al. 1985; Yoshitake et al. 1985). A number of restriction-enzyme polymorphisms (RFLPs) have been observed within this 35-kbp gene with eight exons, including polymorphisms for *Taq*I (Camerino et al. 1984), *Xmn*I and *Hinf*I/*Dde*I (Winship et al. 1984), *Msp*I/*Hpa*II (Camerino et al. 1985), and *Bam*HI (Hay et al. 1986). They have been used clinically with hemophilia B kindred for detecting carriers and affected fetuses in utero (Camerino et al. 1983; Choo et al. 1984; Giannelli et al. 1984; Grunebaum et al. 1984; Hassan et al. 1985). The diagnostic value of the polymorphisms depends in large part on the relative frequencies of the alleles at the polymorphic sites, since heterozygosity of the mother for both hemophilia B and an RFLP is the sine qua non. All earlier studies providing population data have come from Europe or the United Kingdom and have, presumably, been carried out on Caucasians. The allelic ratios have been in the range 65:35 –79:21, and it occurred to us that non-European populations might have quite different gene frequencies.

We have obtained estimates of the frequencies of the polymorphic *Taq*I and *Xmn*I alleles in five ethnic groups: white Americans, black Americans, East Indians, Chinese, and Malays. They vary greatly between the ethnic groups. The allelic ratios and linkage disequilibria suggest that the polymorphisms may be of limited diagnostic value for a large part of the world's population.

MATERIAL AND METHODS

Blood samples for DNA studies were obtained in one of two ways. The American groups (white and black; female and male) consisted of sequential, unrelated specimens received by the clinical laboratories of the North Carolina Memorial Hospital for other purposes and destined to be discarded. They arrived consecutively during a period of several months during the winter and spring of 1985. All samples were <10 ml, and, since DNA was sparse in some instances, only one polymorphism was assessed in some individuals. The only selection in the sampling process consisted of rejection of hemolyzed specimens or tubes whose contents clearly appeared to be insufficient. DNA was prepared from the cell mass by standard methods (Maniatis et al. 1982).

Blood samples for the three Oriental groups were obtained in Kuala Lumpur, Malaysia, in February 1986 from subjects whose ancestry was regarded locally as solely East Indian, Chinese, or Malay. The subjects were either students or staff of the university and were unrelated as far as was known. One of us (D.B.L.) supervised the preparation of cell pellets under sterile conditions from ~50 ml of blood mixed with 0.4 ml of 0.5 M potassium ethylenediaminetetraacetate. The anticoagulated samples remained at room temperature for 1–24 h before addition of sucrose-Triton buffer. Aliquots of 12.5 ml each were diluted to 50 ml by addition of sucrose-Triton at 4 C, the final concentrations in the mixtures being 0.32 M sucrose, 10 mM Tris, 5 mM magnesium chloride, and

1% Triton X-100 at pH 7.6. After 10 min incubation at 4 C, samples were spun at 1,000 *g* at 4 C and the supernatants were decanted. The nuclear pellets were frozen at −70 C and stored until all had been collected. Pellets were transferred to 2-ml tubes that were then filled with a mixture of saline (75 mM), potassium ethylenediaminetetraacetate (24 mM at pH 8.0), 50 μg of proteinase K/ml, and 1% sodium dodecyl sulfate. The tightly capped tubes were transported to the United States at room temperature over a several day period and frozen at −70 C on arrival. A tube was thawed at 37 C for 2 h, and DNA was prepared by standard methods (Maniatis et al. 1982). Although the yields varied somewhat, the quality of the DNA was excellent; it spooled well and cut cleanly with the enzymes used; no unusual fragments of DNA were encountered. (More recently, cell pellets have been prepared in Malmö, Sweden, by means of the same protocol and shipped successfully to us by ordinary air mail.) All DNA samples were examined by Southern blotting (Southern 1975). Filters were probed for the DNA polymorphisms with pAla36, a genomic fragment from a patient with F.IX Alabama, a point mutation (Davis et al. 1987). The probe is the 2.444-kbp fragment of genomic DNA between the sixth *Hin*dIII and fourth *Eco*RI sites of genomic F.IX DNA (Yoshitake et al. 1985) and spans exon D.

RESULTS

Genotypes and Gene Frequencies

A total of 509 chromosomes were assessed for *Taq*I, 472 for *Xmn*I. The genotypes observed in the five ethnic groups are displayed separately by sex and polymorphism in table 1. The 1,1 and 1,2 genotypes are seen in the females of all ethnic groups except the Malay, which contained no heterozygotes. With

TABLE 1

FREQUENCIES OF THE GENOTYPES OBSERVED IN THE FEMALES AND MALES OF FIVE ETHNIC GROUPS FOR TWO F.IX POLYMORPHISMS

GENDER AND ETHNIC GROUP	FREQUENCY FOR POLYMORPHISM/GENOTYPE					
	*Taq*I			*Xmn*I		
Females	1,1	1,2	2,2	1,1	1,2	2,2
White American	29	19	5	33	17	3
Black American	38	14	0	37	11	0
East Indian	21	5	0	23	3	0
Chinese	17	1	0	17	1	0
Malay	23	0	0	19	0	0
Males	1	2		1	2	
White American	35	9		29	9	
Black American	28	9		31	5	
East Indian	26	2		26	2	
Chinese	32	0		32	0	
Malay	22	1		23	0	

one exception, no type 2 alleles were seen in Chinese or Malay males, and the 2,2 genotype was seen only in white Americans. In short, allele 1, the designation for the larger fragment that results from absence of the middle restriction-enzyme recognition site, is more common than allele 2 for both polymorphisms in all populations.

Gene frequencies of alleles 1 and 2 are displayed separately for each polymorphism and ethnic group in table 2. Since the frequencies within ethnic groups were not significantly different between males and females, the data for each ethnic group were combined. Also shown are the frequencies of heterozygosity for each polymorphism in each ethnic type of female. Observed and expected heterozygote frequencies varied only slightly.

Table 3 shows, pairwise, the statistical tests of significance for the differences in gene frequencies between ethnic groups. The first row shows that the difference between white and black Americans is almost significant at the 5% level for *Taq*I and that it is significant at 5% for *Xmn*I. The second row shows that the differences between white Americans and East Indians is highly significant for both polymorphisms, while the third row shows that the differences between black Americans and East Indians are not significant.

For Europe and the British Isles, gene frequencies have been published for both *Taq*I and *Xmn*I. The mean allelic ratio was 66:34 for *Taq*I, based on the 263 chromosomes of four studies (Camerino et al. 1984; Choo et al. 1984; Giannelli et al. 1984; Hassan et al. 1985), while Winship et al. (1984) found a ratio of 71:29 for *Xmn*I in a sample of 72 chromosomes. Rows four and five of table 3 show comparisons of European frequencies with those of white and black Americans. The white American–European difference is almost significant for *Taq*I but not for *Xmn*I, while the black American–European differences are highly significant for both polymorphisms. (The latter result was similar for Europeans and white Americans vs. East Indians, Chinese, and Malays.)

It has been hypothesized that ~20% of the genes of black Americans originated from matings with white Americans (Reed 1969). When this estimate is used together with the American data, it is possible to estimate the average frequency of the rarer *Taq*I and *Xmn*I alleles in the parts of Africa from which black slaves were exported to America: let X be the African frequency sought; .8 = the proportion of African genes in American blacks; .2 = the proportion of white-American genes in American blacks; a = observed frequency of a polymorphism in American whites; and b = observed frequency of a polymorphism in American blacks. Then, $0.8\,(X) + 0.2\,(a) = b$. X for *Taq*I in Africans = 0.138, and X for *Xmn*I in Africans = 0.095.

Heterozygote Frequencies

The probability of encountering a female of an ethnic group who is heterozygous for either one or both of the polymorphisms can be calculated from the gene frequencies. The calculation assumes that the polymorphisms are distributed randomly with respect to the gene frequencies. The results are displayed in the first column of table 4, the expected frequencies being greatest in white

TABLE 2

FREQUENCIES OF GENES AND HETEROZYGOTES

ETHNIC GROUP	FREQUENCY							
	*Taq*I Polymorphism				*Xmn*I Polymorphism			
	No. of Chromosomes	F(Alleles)		Heterozygotes	No. of Chromosomes	F(Alleles)		Heterozygotes
		1	2	Observed/Expected		1	2	Observed/Expected
White American	150	.75	.25	.36/.38	144	.78	.22	.32/.34
Black American	142	.84	.16	.27/.27	132	.88	.12	.22/.21
East Indian	80	.91	.09	.19/.16	78	.94	.06	.12/.11
Chinese	68	.97	.03	.06/.06	56	.96	.04	.06/.08
Malay	69	.99	.01	.00/.02	62	1.00	.00	.00/.00

NOTE.—χ^2 Values for all comparisons of males and females within ethnic groups resulted in $P > .30$, with one exception: black American males and females, for whom $.10 > P > .05$.

TABLE 3

DIFFERENCES OF GENE FREQUENCIES BETWEEN ETHNIC GROUPS

COMPARISONS	DIFFERENCES			
	*Taq*I Polymorphism		*Xmn*I Polymorphism	
	χ^2	*P*	χ^2	*P*
White American vs. black American	3.68	.10 > .05	4.89	<.05
White American vs. East Indian	9.12	<.005	9.11	<.005
Black American vs. East Indian	2.42	>.10	1.81	>.10
European vs. white American	3.25	.10 > .05	1.25	>.20
European vs. black American	21.24	<.0001	9.11	<.0001

Americans and least in Malays. The second column shows that the observed order parallels the expected order but is less than expected in each instance. Column three shows that observed values ranged from 71% to 27% of expected values, the difference being due to linkage disequilibrium.

Linkage Disequilibrium and the Haplotypes

The haplotypes for two diallelic genes that are linked are the four combinations in which they may occur on a single chromosome. Since the exact haplotypes of double heterozygotes cannot be determined without family studies yet comprised a large part of our data, we devised the following scheme to utilize them. The number of doubly heterozygous chromosomes in each ethnic group was multiplied by the probability of the occurrence of each haplotype (from the gene frequencies), and each result was converted to the nearest integer. There were 12 white-American, 10 black-American, 3 East-Indian, 1 Chinese, and 0 Malay double heterozygotes who were haplotyped by this method. The method is conservative, since it will tend to underestimate slightly

TABLE 4

FREQUENCY OF HETEROZYGOSITY FOR EITHER ONE OR BOTH POLYMORPHISMS

Ethnic Group (*N*)	Expected	Observed	Observed/Expected
White American (39)	.607	.436 (17/39)	.71
Black American (36)	.401	.278 (10/36)	.69
East Indian (26)	.269	.192 (5/26)	.71
Chinese (17)	.219	.060 (1/17)	.27
Malay (19)	.000	.000 (0/19)	.00

TABLE 5

HAPLOTYPES FOR TWO F.IX POLYMORPHISMS

ETHNIC GROUP (*N*)	HAPLOTYPES *TaqI*/*XmnI* 1/1	2/1	1/2	2/2	NO OF. CHROMOSOMES	χ^2	*P*	D
White American:								
Males (38)	25	4	5	4	38			
Females (52) ..	75	13	6	10	104			
Total (90) ...	100	17	11	14	142	20.76	<.0001	.0602
Black American:								
Males (36)	26	5	0	5	36			
Females (43) ..	78	6	2	0	86			
Total (79) ...	104	11	2	5	122	...	.000364[a]	.0199
East Indian:								
Males (28)	26	0	0	2	28			
Females (26) ...	49	2	1	0	52			
Total (54) ...	75	2	1	2	80	...	.0056[a]	.0220
Chinese:								
Males (32)	32	0	0	0	32			
Females (18) ..	34	1	1	0	36			
Total (50) ...	66	1	1	0	68	...	...	...
Malay:								
Males (23)	22	1	0	0	23			
Females (19) ..	38	0	0	0	38			
Total (42) ...	60	1	0	0	61	...	...	...

[a] By Fisher's exact method.

the true value of the disequilibrium. The scheme was used only for the females, since the haplotype is identical with the genotype in males.

Table 5 shows the haplotypes for the *Taq*I and *Xmn*I polymorphisms, both separately by sex and combined, for each of the five ethnic groups. When possible, statistical tests were performed on the combined data for each group, i.e., χ^2-test on the white-American data and Fisher's exact test on the black-American and East-Indian data. The Chinese and Malay data were not tested because of the absence of one or more haplotypes. The large χ^2-value for white Americans and the very small probability imply that linkage disequilibrium exists (Ott 1985, p. 151 ff.). Exact probabilities calculated for black Americans and East Indians are of the same order and have the same implication. The parameter *D*, a measure of disequilibrium obtained from the haplotype frequencies as the difference between cross-products in a 2 × 2 matrix (Ott 1985, p. 151 ff.), was greatest in white Americans and much lower in black Americans and East Indians. It has no meaning in Chinese and Malays, in whom there appears to be fixation of type 1 alleles.

DISCUSSION

The *Taq*I polymorphism was first described by Camerino et al. (1984), the *Xmn*I polymorphism by Winship et al. (1984). They are present in the 5′ half of the F.IX gene, the polymorphic sites being ~3.8 kbp apart. Because of their proximity, the polymorphisms are expected to be in strong linkage disequilibrium (Graham et al. 1985). Our data show this particularly well in white Americans, but in all groups there is an obviously strong association between the type 1 allele of *Taq*I and the type 1 allele of *Xmn*I. The predominance of type 1 alleles means that absence of the middle recognition site is much more common than its presence. Our data show that the association between type 1 alleles is greatest in Africans and Orientals. Since Africa and China are considered the likeliest sites of origin(s) of *Homo sapiens*, absence of the middle sites may have been the original condition in humans. Polymorphism at the middle sites may have resulted more recently from mutations—perhaps in Europe, from where they then spread outward. It will be of interest to see whether any well-defined European population has particularly high frequencies of the rarer type 2 alleles.

The reduction in heterozygosity of the polymorphisms that were due to linkage disequilibrium averaged ~28%–30% in white and black Americans and East Indians and was essentially absent in Chinese and Malays. This means that clinical studies in Africa and the Orient will depend on the discovery and study of other polymorphisms. The most valuable polymorphisms for use in all groups might be those located at some distance from the region containing the polymorphic *Taq*I and *Xmn*I sites. Perhaps, in a variety of ethnic groups, the *Bam*HI polymorphism near exon I (Hay et al. 1986) should be explored with this in mind. Special attention might also be given to the 3′ end of the gene and the 5′ flanking DNA, since most of the polymorphisms reported to date have been in the gene's midportion.

The frequency of the rarer *Taq*I allele is higher in Europeans (reported by others) than in our group of white Americans. This is somewhat of a surprise, since the ancestors of the white Americans in the North Carolina sample came overwhelmingly from Scotland, Wales, Ireland, and northern England. This suggests that significant genetic drift may have occurred since the American migrations from Britain began in earnest about 1700. More striking are the lower values in black Americans than in white Americans and, especially, Europeans. When the black-American frequencies are adjusted for white-American admixture, values for the rarer alleles that might tentatively be regarded as truly African—i.e., .138 for *Taq*I and .095 for *Xmn*I—are fairly close to those for East Indians—.09 and .06, respectively. Perhaps the African:East Indian convergence is an expression of genetic similarity resulting from gene flow over the land bridge connecting Asia and Africa and around the edges of the rather small ocean that separates the Indian subcontinent from Africa.

Our results show very clearly that the frequencies of these two F.IX DNA polymorphisms vary greatly with ethnicity and that local frequencies are an important consideration when a laboratory is selecting polymorphisms for diag-

nostic purposes. The low frequency of the type 2 alleles in African and Oriental populations means that the *Taq*I and *Xmn*I polymorphisms will have little diagnostic utility in large areas of the world.

ACKNOWLEDGMENTS

We wish to thank the patients of the North Carolina Memorial Hospital of Chapel Hill; the students and staff of the School of Medicine of the University of Malaysia, Kuala Lumpur, Malaya, for providing the samples of blood that made this work possible; and Dr. Philip Green of Collaborative Research, Boston, for statistical advice. We especially wish to thank the following members of the staff of the medical school of the University of Malaysia for their many acts of kindness, including the use of their laboratory facilities: Dr. T. Pang, Dr. C. K. Sam, J. Dyck, K. Kalaimathee, S. Ponniah, S. Samuel, and S. Y. Wong. The study was supported in part by National Institutes of Health research grants HL-06350 and HL-35152 and by funds from the University of Malaysia.

REFERENCES

Anson, D. S., K. H. Choo, D. J. G. Rees, F. Gianelli, K. Gould, J. A. Huddleston, and G. G. Brownlee. 1984. The gene structure of human anti-haemophilic factor IX. EMBO J. **3**:1053–1060.

Camerino, G., K. H. Grzeschik, M. Jaye, H. de la Salle, P. Tolstoshev, J.-P. Lecocq, J. R. Heilig, and J. L. Mandel. 1984. Regional localization on the human X chromosome and polymorphism of the coagulation factor IX gene (hemophilia B locus). Proc. Natl. Acad. Sci. USA **81**:498–502.

Camerino, G., M. G. Mattei, J. F. Mattei, M. Jaye, and J. L. Mandel. 1983. Close linkage of fragile X-mental retardation to haemophilia B and transmission through a normal male. Nature **306**:701–705.

Camerino, G., I. Oberle, D. Drayna, and J. L. Mandel. 1985. A new *Msp*I restriction fragment length polymorphism in the hemophilia B locus. Hum. Genet. **71**:79–81.

Choo, K. H., D. George, G. Filby, J. L. Halliday, M. Leversham, G. Webb, and D. M. Danks. 1984. Linkage analysis of X-linked mental retardation with and without fragile-X using factor IX gene probe. Lancet **2**:349.

Choo, K. H., K. G. Gould, D. J. G. Rees, and G. G. Brownlee. 1982. Molecular cloning of the gene for human anti-haemophilic factor IX. Nature **299**:178–180.

Davis, L. M., R. A. McGraw, J. Ware, H. R. Roberts, and D. W. Stafford. 1987. Factor IX Alabama: a point mutation in a clotting protein which results in hemophilia B. Blood **69**:140–143.

Giannelli, F., D. S. Anson, K. H. Choo, D. J. G. Rees, P. R. Winship, N. Ferrari, C. R. Rizza, and G. G. Brownlee 1984. Characterisation and use of an intragenic polymorphism for detection of carriers of haemophilia B (factor IX deficiency). Lancet **1**:239–241.

Graham, J. B., P. P. Green, R. A. McGraw, and L. M. Davis. 1985. Application of molecular genetics to prenatal diagnosis and carrier detection in the hemophilias: some limitations. Blood **66**:759–784.

Grunebaum, L., J.-P. Cazenave, G. Camerino, C. Kloepfer, J.-L. Mandel, P. Tolstoshev, M. Jaye, H. de la Salle, and J.-P. Lecocq. 1984. Carrier detection of hemophilia B by using a restriction site polymorphism associated with the coagulation factor IX gene. J. Clin. Invest. **73**:1491–1495.

Hassan, H. J., M. Orlando, A. Leonardi, C. Chelucci, R. Guerriero, P. M. Mannucci, G. Mariani, and C. Peschle. 1985. Intragenic factor IX restriction site polymorphism in hemophilia B variants. Blood **65**:441–443.

Hay, C. W., K. A. Robertson, S.-L. Yong, A. R. Thompson, G. H. Growe, and R.

MacGillivray. 1986. Use of a *Bam*HI polymorphism in the factor IX gene for the determination of hemophilia B carrier status. Blood **67**:1508–1511.

Jagadeeswaran, P., D. E. Lavelle, R. Kaul, T. Mohandas, and S. T. Warren. 1984. Isolation and characterization of human factor IX cDNA: identification of *Taq*I polymorphism and regional assignment. Somatic Cell Mol. Genet. **10**:465–473.

Jaye, M., H. de la Salle, F. Schamber, A. Balland, V. Kohli, A. Findeli, P. Tolstoshev, and J.-P. Lecocq. 1983. Isolation of human antihaemophilic factor IX cDNA using a unique 52-base synthetic oligonucleotide probe deduced from the amino acid sequence of bovine factor IX. Nucleic Acids Res. **11**:2325–2335.

Kurachi, K., and E. W. Davie. 1982. Isolation and characterization of a cDNA coding for human factor IX. Proc. Natl. Acad. Sci. USA **79**:6461–6464.

McGraw, R. A., L. M. Davis, C. M. Noyes, R. L. Lundblad, H. R. Roberts, J. B. Graham, and D. W. Stafford. 1985. Evidence for a prevalent dimorphism in the activation peptide of human coagulant factor IX. Proc. Natl. Acad. Sci. USA **82**:2847–2851.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Ott, J. 1985. Analysis of genetic linkage. The Johns Hopkins Press, Baltimore.

Reed, T. E. 1969. Caucasian genes in American Negroes. Science **165**:762–768.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

Winship, P. R., D. S. Anson, C. R. Rizza, and G. G. Brownlee. 1984. Carrier detection in haemophilia B using two further intragenic restriction fragment length polymorphisms. Nucleic Acids Res. **12**:8861–8872.

Yoshitake, S., B. G. Schach, D. C. Foster, E. W. Davie, and K. Kurachi. 1985. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). Biochemistry **24**:3736–3750.