Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1045–1049, February 1986
Genetics

Evidence for increased recombination near the human insulin gene: Implication for disease association studies

(restriction fragment length polymorphism/recombination hot spot/linkage disequilibrium/diabetes mellitus)

ARAVINDA CHAKRAVARTI*†, STEVEN C. ELBEIN‡, AND M. ALAN PERMUTT‡

*Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261; and ‡Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110

Communicated by James V. Neel, October 11, 1985

ABSTRACT **Haplotypes for four new restriction site polymorphisms (detected by *Rsa* I, *Taq* I, *Hin*cII, and *Sac* I) and a previously identified DNA length polymorphism (5′ FP), all at the insulin locus, have been studied in U.S. Blacks, African Blacks, Caucasians, and Pima Indians. Black populations are polymorphic for all five markers, whereas the other groups are polymorphic for *Rsa* I, *Taq* I, and 5′ FP only. The data suggest that ≈1 in 550 base pairs is variant in this region. The polymorphisms, even though located within 20 kilobases, display low levels of nonrandom association. Population genetic analysis suggests that recombination within this 20-kilobase segment occurs 24 times more frequently than expected if crossing-over occurred uniformly throughout the human genome. These findings suggest that population associations between DNA polymorphisms and disease susceptibility genes near the insulin gene or structural mutations in the insulin gene will be weak. Thus, population studies would probably require large sample sizes to detect associations. However, the low levels of nonrandom association increase the information content of the locus for linkage studies, which is the best alternative for discovering disease susceptibility genes.**

The human insulin gene, located on chromosome 11 band p15 (1), is associated with a region of length heterogeneity ≈350 base pairs (bp) 5′ to the insulin mRNA initiation site (2). This DNA polymorphism, designated 5′ FP, is caused by variation in the number of tandem repeats of a 14-bp consensus sequence (3), and this variation is probably due to unequal crossing-over between homologous chromosomes containing variable numbers of repeats (3, 4). The potential importance of insulin in the pathogenesis of diabetes and the location of the 5′ FP close to putative promoter or enhancer sequences (5) have led several investigators to study the association of this region with diabetes mellitus. While extensive allelic heterogeneity exists at this locus, the alleles naturally fall into three general size classes. For purposes of association studies, these size classes have been called class 1 for the smallest alleles (570-bp repeats), class 3 for the largest (2400-bp repeats), and class 2 for alleles intermediate in size (6). Some groups have reported the association of the larger class 3 alleles with non-insulin-dependent diabetes mellitus (NIDDM) in initial studies (7–9), but other investigators have not confirmed these associations (6). More recently, studies with larger sample sizes in Caucasian (6), U.S. Black (10), and Pima Indian (11) populations failed to find any association of class 3 alleles with NIDDM. Bell *et al.* (6) reported an association of class 1 alleles with NIDDM in Caucasians, but combined data (12) from all published studies of Caucasians failed to demonstrate an association of any allele with NIDDM. In contrast, the increased frequency of class 1 alleles in Caucasians with insulin-dependent diabetes mellitus (IDDM) has been noted by all investigators (6–9), and the combined data show a significant association.

Current evidence suggests no functional role for the 5′ FP in insulin secretion (13) or insulin gene expression (5). Thus, any significance of 5′ FP in diabetes would most likely result from its close linkage to a disease susceptibility gene (insulin). Several factors might account for the inconsistent associations between NIDDM and 5′ FP—namely, (*i*) earlier associations may have been spurious; (*ii*) association studies have grouped diverse alleles into single classes, and the definition of subsets of these classes could permit identification of stronger associations; and (*iii*) increased recombination around the 5′ FP could significantly reduce association of this marker with potential disease susceptibility genes. Markedly increased rates of recombination of similar regions of tandem repeats have been noted in bacterial plasmids (14). Such increased recombination at the insulin locus would suggest caution in interpretation of the noted associations of the 5′ FP with IDDM (6), atherosclerosis (15, 16), and hypertriglyceridemia (17).

In an attempt to define the amount of recombination at the insulin locus, we have examined the degree to which the 5′ FP is associated with *Rsa* I, *Taq* I, *Hin*cII, and *Sac* I restriction fragment length polymorphisms (RFLPs) near the human insulin gene (ref. 18; Fig. 1). Analysis of DNA haplotype data from U.S. Black, Caucasian, and Pima Indian populations enables a study of nucleotide variability and nonrandom association between these markers. A population genetic analysis of our data suggests increased meiotic recombination around the insulin gene and that for this region 1% recombination corresponds to 42 kilobases (kb) of DNA, whereas we had expected a distance of 1000 kb for 1% recombination.

MATERIALS AND METHODS

Subjects. All individuals in our study belonged to U.S. Black, African Black, Caucasian, or Pima Indian populations. The subjects were normal individuals except for a few carriers of the sickle hemoglobin gene (U.S. Black) and a few NIDDM patients (Caucasian). Since no relationship can be demonstrated between insulin alleles and NIDDM (10–12), and since the insulin gene is not sufficiently close to the β-globin gene (14% recombination; ref. 4) to display linkage disequilibrium, our sample of chromosomes represents a random sample.

Haplotype data were obtained from nuclear families, some

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kb, kilobase(s); RFLP, restriction fragment length polymorphism; NIDDM, non-insulin-dependent diabetes mellitus.

†To whom reprint requests should be addressed at: Room A310, Crabtree Hall, Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261.

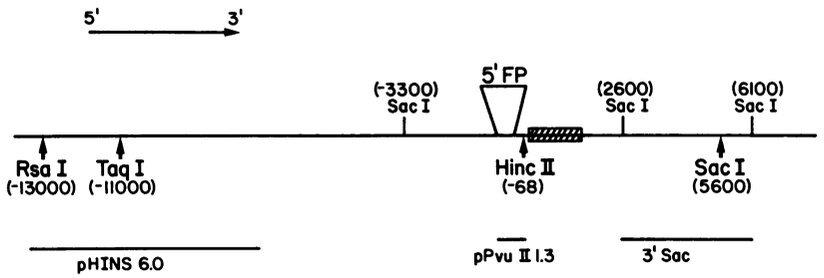

FIG. 1. Restriction map of 20 kb of DNA segment containing the human insulin gene (hatched). DNA polymorphisms are indicated by an arrow and by the symbol 5′ FP.

pedigrees, and population samples. In family data, haplotypes in parents can be determined from offspring phenotypes on the assumption that there are no recombinants in the families. Haplotypes from randomly chosen individuals could be determined only when individuals were heterozygous for at most one marker locus.

Nuclear DNA Preparation. High molecular weight DNA was isolated from the leukocytes of 10 ml of peripheral blood as described (2).

Probes and Restriction Digests. In our studies three probes were used as indicated in Fig. 1. A detailed description of these probes and the methods for restriction analysis are provided in refs. 18 and 19.

DNA Polymorphisms. The five DNA polymorphisms studied were detected by *Rsa* I, *Taq* I, *Pvu* II, *Hin*cII, and *Sac* I. The fragment sizes of the alleles are *Rsa* I (+, 1.25 kb; −, 1.45 kb), *Taq* I (+, 2.0–2.8 kb; −, 4.8 kb), *Hin*cII (+, 3.2 kb; −, 6.0–7.6 kb), and *Sac* I (+, 2.5 kb; −, 3.3 kb) (18), where +/− represents the presence/absence of the restriction site. The length polymorphism (5′ FP) can best be detected by using *Pvu* II, and it was originally described by Bell *et al.* (2), using *Sac* I. They also described a variant *Hin*cII site but gave no further details; our *Hin*cII polymorphism is probably identical to that of Bell *et al.* (2). Lebo *et al.* (4) described another length polymorphism in U.S. Blacks that was 3′ to insulin and was detected by using *Sac* I. We detect a restriction site polymorphism 3′ to the gene with *Sac* I, but using a probe specific for this region we have been unable to demonstrate a length polymorphism (18).

5′ FP is polymorphic in all racial groups. In Caucasian populations, using a *Pvu* II digest, one usually detects fragments 800 ± 200 bp long [class 1 alleles of Bell *et al.* (6)], or 2200+ bp (class 3 alleles), and rarely alleles of intermediate size (class 2). Class 2 alleles are quite common (10%) in Black populations (4, 10, 12) but are not present in Pima Indians (11). While this classification is useful for some kinds of analyses, it underestimates the actual number of alleles in our samples. It is impossible to accurately quantify the number of alleles at 5′ FP, since small (25–50 bp) differences in length alleles cannot be compared on different gels.

All polymorphisms are located within a 20-kb region, with locations as indicated in Fig. 1. Restriction mapping has determined the physical distance between the RFLPs (18). Since 5′ FP is a length polymorphism, a unique location for the locus cannot be obtained. We have taken the coordinate as the midpoint of the harmonic mean of the several *Pvu* II fragment lengths. The adjacent distances between the RFLPs *Rsa* I, *Taq* I, 5′ FP, *Hin*cII, and *Sac* I are then 2.0, 11.9, 0.9, and 5.3 kb, respectively.

Statistical Methods. The methods used to calculate the nucleotide diversity (π) and the standardized nonrandom association (Δ) are provided in Chakravarti *et al.* (20). The regression analysis used to estimate the recombination frequency per kb (ϕ) is that described by Chakravarti *et al.* (21).

RESULTS

Population Screening. DNA samples were screened by using 13 different restriction endonucleases and four different probes covering 20 kb of the insulin region. However, not all probe–enzyme combinations were tested. Our screening results are shown in Table 1. The number of samples examined was usually >36, and of the 140 restriction sites consistently screened in the U.S. Black sample, four RFLPs were discovered. The screening of Caucasian samples was not as extensive, but each polymorphic site identified in Blacks was investigated in Caucasians as well and only two RFLPs were discovered. However, the proportion of sites polymorphic in U.S. Blacks (0.029) and Caucasians (0.032) were not significantly different and were similar to that observed for the human growth hormone cluster (20).

Nucleotide Variability. On the assumption that the RFLPs are selectively neutral or that selection is weak, we may estimate the nucleotide diversity π. This estimates the heterozygosity per nucleotide site over the 20-kb segment from the site polymorphism data by using the method of Ewens *et al.* (20, 22). The data necessary are the number of bp in the recognition sequence for an endonuclease (b), the number of restriction sites screened (s), the number of alleles (DNAs) screened (n), and the number of RFLPs discovered (κ), as provided in Table 1. π is then estimated as the ratio of the number of RFLPs to the estimate of the number of bp screened (22).

We estimate $\hat{\pi} = 0.0017$ for the U.S. Black sample and $\hat{\pi} = 0.0020$ for the Caucasian sample. The Caucasian value is slightly larger than the U.S. Black estimate because four fewer enzymes were used, but screening included those endonucleases known to yield RFLPs in the U.S. Black sample. These π values are similar to those obtained for the

Table 1. Restriction enzymes used for polymorphism survey at the insulin locus

Restriction enzyme	No. of bases	U.S. Black			Caucasian		
		s	n	κ	s	n	κ
*Bam*HI	6	5	48	—	—	—	—
Bcl I	6	3	36	—	2	36	—
Bgl I	6	8	48	—	8	54	—
Bgl II	6	4	48	—	2	48	—
*Eco*RI	6	2	56	—	—	—	—
*Hin*cII	5	4	116	1	1	60	—
*Hin*dIII	6	3	44	—	—	—	—
*Hin*fI	4	16	38	—	5	50	—
Pst I	6	35	108	—	16	44	—
Pvu II	6	28	64	—	12	52	—
Rsa I	4	16	57	1	7	36	1
Sac I	6	5	117	1	5	64	—
Taq I	4	11	47	1	5	40	1
Totals	—	140	—	4	63	—	2

s, Number of restriction sites screened; n, number of DNAs screened; κ, number of polymorphisms discovered.

Table 2. Estimates of frequencies, SEM, and heterozygosities of RFLPs at the insulin locus

RFLP	Allele	Value of	U.S. Black	African Black	Caucasian	Pima Indian
Rsa I	+	*p*	0.67 ± 0.06	—	0.61 ± 0.05	0.47 ± 0.09
		n	57	—	80	34
		h	0.44	—	0.47	0.50
Taq I	+	*p*	0.94 ± 0.04	—	0.89 ± 0.04	0.87 ± 0.06
		n	47	—	64	38
		h	0.12	—	0.19	0.23
5′ FP	1	*p*	0.54 ± 0.04	0.23 ± 0.08	0.74 ± 0.04	0.76 ± 0.04
	2	*p*	0.13 ± 0.03	0.35 ± 0.09	0.02 ± 0.01	.00
	3	*p*	0.33 ± 0.04	0.42 ± 0.10	0.24 ± 0.04	0.24 ± 0.04
		n	125	26	108	114
		h	0.58	0.65	0.39	0.36
*Hin*cII	+	*p*	0.50 ± 0.05	0.54 ± 0.10	1.00	1.00
		n	119	26	56	30
		h	0.50	0.50	—	—
Sac I	+	*p*	0.03 ± 0.02	—	1.00	1.00
		n	117	—	56	30
		h	0.07	—	—	—
		$\bar{h}$ (above sites)	0.342	0.573	0.354	0.363
		$\bar{h}$ (all 140 sites)	0.012	—	0.008	0.008
	Proportion of sites polymorphic		0.036	—	0.021	0.021

For each RFLP, *p* is the frequency of the indicated allele ± SEM, *n* is the number of chromosomes sampled, *h* is heterozygosity, and $\bar{h}$ is average heterozygosity.

human growth hormone, β-globin gene cluster, and the α_1-antitrypsin gene (20, 23, 24), and they suggest that, on average, 1 in 550 bp are different between two randomly chosen chromosomes.

Variation at Single RFLPs. In a sample of *n* chromosomes, of which r_i are of allelic type i ($\Sigma r_i = n$), the maximum likelihood estimate of the frequency of allele i is $\hat{p}_i = r_i/n$, with standard error $[p_i(1 - p_i)/n]^{1/2}$. The extent of polymorphism at individual RFLPs can then be evaluated from the heterozygosity as $\hat{h} = 1 - \Sigma\hat{p}_i^2$, with an average $\bar{h} = \Sigma\hat{h}_i/m$ for *m* restriction sites. Our calculations are presented in Table 2.

Table 3. Complete haplotypes for insulin polymorphisms in the U.S. Black, Caucasian, and Pima Indian populations

Rsa I	*Taq* I	5′ FP	*Hin*cII	*Sac* I	U.S. Black	Caucasian	Pima Indian
+	+	1	+	−	4	24	10
−	+	1	+	−		12	13
+	−	1	+	−	1	4	2
+	+	1	−	−	10		
−	+	1	−	−	4		
+	+	2	+	−	2		
+	−	2	+	−	1		
+	+	2	+	+	2		
−	+	2	+	+	1		
+	+	3	+	−		4	2
+	−	3	+	−		2	
−	+	3	+	−	4	10	3
+	+	3	−	−	4		
−	+	3	−	−	1		
+	−	3	−	−	1		
				Total	35	56	30

The expected number of haplotypes under linkage equilibrium was calculated from the allele frequencies in Table 2. Thus, the expected frequency of (+ + 1 + −) in U.S. Blacks is 0.67 × 0.94 × 0.54 × 0.50 × 0.97 × 35 = 5.8, etc. For each population, the data were pooled into three or four classes to yield expected values of 5 or greater. This analysis shows the haplotypes to be in linkage equilibrium: U.S. Black (χ^2 = 6.51, df = 3, *P* = 0.09), Caucasian (χ^2 = 1.67, df = 3, *P* = 0.65), Pima Indian (χ^2 = 1.65, df = 2, *P* = 0.44).

Nonrandom Association Between RFLPs. For studying linkage disequilibrium in this region, it would be ideal to construct haplotypes for the five RFLPs. A tabulation of complete haplotypes in three populations is given in Table 3. It should be emphasized that the diversity of haplotypes is even greater than Table 3 implies, since the three classes of alleles at 5′ FP greatly underestimate the number of alleles at this site (see *Materials and Methods*). Note that in Caucasians and Pima Indians, *Hin*cII is always + and *Sac* I is always −. The overall level of linkage disequilibrium can be simply gauged by a χ^2 test on the observed and expected (under linkage equilibrium) haplotype frequencies. These frequencies did not differ significantly, thus providing prima facie evidence for low levels of linkage disequilibrium (Table 3).

A more detailed analysis of the linkage disequilibrium can, however, be made. Because the majority of our data are from U.S. Blacks, we restrict our detailed analysis to this sample. By considering both complete and partial haplotype data we can greatly increase the sample size for all pairs of RFLPs, as shown in Tables 4 and 5. The nonrandom associations (Δ) between any pair of RFLPs are presented in Table 6, together with the physical distance in kilobases between them. Δ is the correlation coefficient between the uniting gametes at two markers and takes values between −1 and 1. The method for calculating Δ and testing the hypothesis Δ = 0 (linkage

Table 4. Frequencies of pairwise haplotypes between the *Rsa* I, *Taq* I, *Hin*cII, and *Sac* I RFLPs in the U.S. Black population

Locus 1	Locus 2	Frequency of haplotype ++	+−	−+	−−	Sample size
Rsa I	*Taq* I	22	3	10	0	35
Rsa I	*Hin*cII	17	21	7	12	57
Rsa I	*Sac* I	2	32	1	14	49
Taq I	*Hin*cII	22	22	2	1	47
Taq I	*Sac* I	3	41	0	3	47
*Hin*cII	*Sac* I	4	50	0	53	107

Table 5. Frequencies of pairwise haplotypes between the 5′ FP and *Rsa* I, *Taq* I, *Hin*cII, *Sac* I RFLPs in the U.S. Black population

5′ FP allele	*Rsa* I		*Taq* I		*Hin*cII		*Sac* I	
	+	−	+	−	+	−	+	−
1	24	9	22	1	29	32	0	61
2	5	1	6	1	12	3	4	10
3	7	7	16	1	16	23	0	40
n	53		47		115		115	

equilibrium) is described by Chakravarti *et al.* (20) and in ref. 25. The values in Table 6 demonstrate significant associations only for the cluster 5′ FP/*Hin*cII/*Sac* I. Of course, only in the comparisons involving these RFLPs was the sample size large ($n > 107$).

Relationship Between Δ and Physical Distance. If nonrandom associations (Δ) were solely determined by recombination (the physical distance between RFLPs), then Δ would decline with increasing physical distance. Table 6 shows that this trend is true in general as one compares an RFLP to RFLPs located 3′ to it. Furthermore, for neutral RFLPs at genetic equilibrium, $\Delta^2 = 1/(1 + 4N_e c)$, where N_e is the effective population size and c is the recombination value between the markers (26–28). For small distances, $c = k \cdot d$, where k is recombination frequency per kb and d is the distance in kb. Therefore, for any pair of RFLPs, we can estimate the quantity $4N_e c = 4N_e kd = (\Delta^{-2} - 1)$, and by linear regression on known d values we can obtain an estimate of $\phi = 4N_e k$ (21) from the data in Table 6. ϕ is a standardized measure of recombination frequency per kb that allows us to compare values in different regions of the genome. Our analysis gives $\hat{\phi} = 8.75 \pm 1.49$. On a linear scale, the regression of $(\Delta^{-2} - 1)$ on d was significant and showed a good fit to the data ($r = 0.9$). Fig. 2 shows the plot of observed Δ and d values and the curve obtained by regression analysis.

The estimated $\hat{\phi}$ value was next compared to that expected (ϕ_e) under the hypothesis of uniform recombination. The calculation of ϕ_e requires an estimate of the quantity N_e. To obtain N_e, we next considered haplotype data on the *HLA-A, -C, -B, -DR*, and *-Bf* loci from the histocompatibility testing workshop (29) and calculated Δ values for all pairwise comparisons. Since the interlocus recombination frequencies for this gene cluster are known (30), we performed an identical regression analysis, but this time to compute N_e. Our analysis of haplotype data from European Caucasians, Japanese, African Blacks, and U.S. Blacks gave $N_e = 8950 \pm 4500$. If recombination occurs uniformly, $k = 10^{-5}$, so that $\phi_e = 0.358 \pm 0.180$.

The estimated ϕ value at the insulin locus (8.75 ± 1.49) is thus significantly different ($P < 0.005$) from the expected value (0.358 ± 0.180) and is 24 times greater than expected. Furthermore, since $\hat{\phi}$ has 95% confidence limits 5.83–11.67, the 95% confidence limits on the rate of recombination is 16–33 times higher than expected.

Table 6. Standardized nonrandom associations (Δ) between insulin RFLPs in the U.S. Black population (upper right) and physical distance in kb (lower left) between the markers

	Rsa I	*Taq* I	5′ FP	*Hin*cII	*Sac* I
Rsa I	—	−0.194	0.170	0.075	−0.015
Taq I	2.0	—	0.098	−0.082	0.068
5′ FP	13.9	11.9	—	0.222*	0.361*
*Hin*cII	14.8	12.8	0.9	—	0.195*
Sac I	20.1	18.1	6.2	5.3	—

*Significantly different from 0 at the 5% level.

DISCUSSION

Polymorphism at the Insulin Locus. In this study, 140 restriction sites were consistently screened in the U.S. Black sample, and four RFLPs were identified. Since the majority of restriction sites were monomorphic, it is more realistic to compute the average heterozygosity ($\bar{h}$) over all sites. In the U.S. Black sample, $\bar{h} = 0.012$, as compared to 0.010 for β-globin (A.C. and H. H. Kazazian, unpublished data) and 0.016 for growth hormone (20). The $\bar{h}$ value of 0.008 in Caucasians and Pima Indians is similar to the value for U.S. Blacks, given that screening of these populations was not extensive. We also calculated that the heterozygosity per bp is 0.0017, as previously observed for the growth hormone (20), β-globin (23), and α_1-antitrypsin (24) loci. However, our current calculations do not include the length polymorphism 5′ FP, so that even greater variability exists. Since $\pi = 0.0017$, on average, 34 variants should exist in a 20-kb segment, but the probability that any of these would also alter a restriction site is small (20, 22).

Nonrandom Association Between RFLPs. The presence of extensive disequilibrium between physically close RFLPs reduces their usefulness in linkage studies or as markers for genetic counseling (31). Fortunately, disequilibrium is low at the insulin locus, so the majority of matings would be informative for linkage. In fact, these insulin RFLPs allow three or four different parental alleles to be distinguished in 40% and 60% of families, respectively (18).

We have provided evidence for low levels of nonrandom associations at the insulin locus. That the Δ values in Table 6 are small can also be seen in comparison to the growth hormone (20), β-globin (21), albumin (32), or D11S12 (33) loci. Specifically, at β-globin, RFLPs spanning a recombinational hot spot show $\bar{\Delta} = 0.13$, whereas RFLPs in neighboring DNA show $\bar{\Delta} = 0.28$, in U.S. Blacks. Table 6 gives $\bar{\Delta}$ as 0.15—a value closer to those at the β-globin hot spot. A more accurate comparison takes physical distance into account and thus we measured the standardized recombination frequency per kb, ϕ. The insulin value $\hat{\phi} = 8.75$ is even greater than the β-globin hot spot value ($\hat{\phi} = 6.35$) and greater still than either the DNA 5′ to the hot spot ($\hat{\phi} = 0.01$) or 3′ to the hot spot ($\hat{\phi} = 0.96$) (21).

Relationship of Δ to Physical Distance. In Fig. 2, we have plotted the observed Δ values against physical distance and we also plotted the fitted curve obtained by using $\hat{\phi} = 8.75$. The degree of nonrandom association is clearly related to physical distance, even though this is not true for growth hormone RFLPs (20) or RFLPs at the D11S12 locus (33). Our

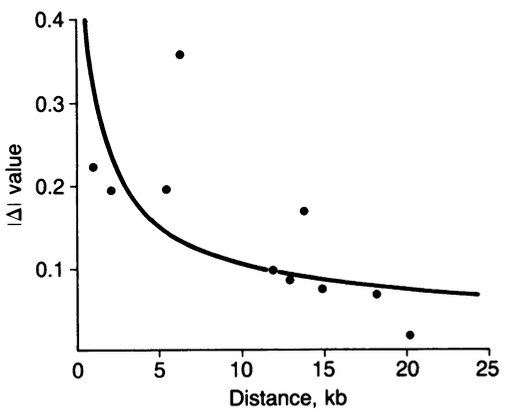

FIG. 2. Relationship between absolute value of the observed nonrandom association (Δ) and physical distance (d) in kb. The curve $\Delta = [1/(1 + \hat{\phi}d)]^{1/2}$ is also shown. $\hat{\phi}$ was obtained by linear regression of $\Delta^{-2} - 1$ on d. The regression line demonstrated acceptable fit ($r = 0.9$).

data on Caucasians and Pima Indians are also consistent with the above distance relationship. The Δ values for the comparisons *Rsa* I–*Taq* I (2 kb), *Taq* I–5′ FP (11.9 kb), and *Rsa* I–5′ FP (13.9 kb) are −0.281, 0.008, 0.200 in Caucasians and −0.286, −0.158, 0.060 in the Pima Indians.

Recombination in the Insulin Region. The estimation of ϕ depends on the assumption that all the RFLPs arose at approximately the same time, but Sved's (27) equation used to estimate ϕ is quite insensitive to violation of the condition that the ages of the RFLPs be the same. The *Rsa* I, *Taq* I, and 5′ FP RFLPs probably precede human racial divergence, since they are shared by the diverse populations we have studied. On the other hand, the *Hin*cII RFLP, despite the fact that it is polymorphic only in Blacks, is probably also ancient because it has a high frequency (50%). We presume that this polymorphism has been lost in non-Black populations. The *Sac* I RFLP, with a 3% frequency in U.S. Blacks, is perhaps the youngest polymorphism and, as Table 6 and Fig. 2 demonstrate, departs the most from the fitted curve. We have, however, retained the *Sac* I comparisons, since their effect would be to reduce the ϕ value.

At the insulin locus, we have calculated the recombination frequency to be 24 times greater than expected (95% confidence limits 16–33 times) if we let N_e = 8950. However, the 95% confidence limits on N_e are 130 (ϕ_e = 0.005) to 17,770 (ϕ_e = 0.711). If we consider the largest effective population size and the lowest estimated $\hat{\phi}$ value, then recombination occurs 8 times more frequently than expected; the smallest effective size is then equivalent to an increase of 2244 times expected for ϕ. On average, the total recombination frequency is 0.48% over the 20-kb segment, as compared to the uniform value of 0.02%. A 1% recombination usually corresponds to 1000 kb, but for this region it would correspond to 42 kb with 95% confidence limits of 30–63 kb. The putative recombinational hot spot around the insulin gene can be experimentally confirmed, as has recently been shown for the β-globin hot spot, using recombinational pathways in yeast (34).

Jeffreys *et al.* (35) have recently reported that the human genome contains many dispersed tandem-repetitive regions that share a 10- to 15-bp core sequence similar to the generalized recombination signal χ (5′ GCTGGTGG 3′) of *Escherichia coli*. These repetitive sequences show similarity to the tandem repeats of immunoglobulins, embryonic α-globin, and insulin. While low levels of nonrandom association at the insulin locus suggest that increased recombination is prevalent throughout the 20-kb region (Fig. 2), the disequilibrium is markedly low for the region *Rsa* I–*Taq* I–5′ FP, and consequently most of the increase in recombination may be restricted to this segment. This may be explained by the fact that the tandem repeats at 5′ FP may promote increased recombination (4, 35). Furthermore, χ sequences increase recombination 10- to 20-fold but 5′ to their location (36). It is intriguing that two recombinational hot spots on 11p, β-globin, and insulin are both related to χ.

The presence of recombinational hot spots on chromosome 11p has important implications for the discovery of closely linked disease susceptibility genes for diabetes or other disorders or for gene mapping. Even if disease susceptibility genes do exist near the insulin locus, or if specific mutations are present in the structural gene, increased recombination will decrease associations in population studies with insulin RFLPs. For example, in β-thalassemia, the associations between specific β^{thal} mutants and RFLP haplotypes are small because of a recombinational hot spot (ref. 37; Table 2). Thus, population studies will require large sample sizes. Associations, even if discovered, may be too weak to be of practical benefit. The discovery of disease susceptibility genes near the insulin gene or of structural mutations at the insulin locus by linkage analysis in particular families remains the best alternative. Fortunately, low levels of linkage disequilibrium increase haplotype heterozygosity and thus a majority of families will be informative for linkage.

We thank H. H. Kazazian and C. D. Boehm (Johns Hopkins Hospital, Baltimore, MD) for kindly supplying DNA from U.S. and African Black nuclear families and W. Knowler (NIAMDD, Phoenix, AZ) for supplying leukocyte nuclei from Pima Indians. We thank Lynn Corsetti for technical assistance, and Drs. D. C. Rao, J. V. Neel, and the reviewers for critical comments on the manuscript. This study was supported by National Institutes of Health Grants GM33771, AM13983, AM31866, and AM16746.

1. Shows, T. B., McAlpine, P. J. & Miller, R. L. (1984) *Cytogenet. Cell Genet.* **37**, 340–393.
2. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5759–5763.
3. Bell, G. I., Selby, M. J. & Rutter, W. J. (1982) *Nature (London)* **295**, 31–35.
4. Lebo, R. V., Chakravarti, A., Buetow, K. H., Cheung, M.-C., Cann, H., Cordell, B. & Goodman, H. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4808–4812.
5. Walker, M. D., Edlund, T., Boulet, A. M. & Rutter, W. J. (1983) *Nature (London)* **306**, 557–661.
6. Bell, G. I., Horita, S. & Karam, J. H. (1984) *Diabetes* **33**, 176–183.
7. Rotwein, P., Chyn, R., Chirgwin, J., Cordell, B., Goodman, H. M. & Permutt, M. A. (1981) *Science* **213**, 1117–1120.
8. Owerbach, D. & Nerup, J. (1982) *Diabetes* **31**, 275–277.
9. Rotwein, P., Chirgwin, J., Province, M., Knowler, W. C., Pettit, D. J., Cordell, B., Goodman, H. M. & Permutt, M. A. (1983) *N. Engl. J. Med.* **308**, 65–71.
10. Elbein, S., Rotwein, P., Permutt, M. A., Bell, G. I., Sanz, N. & Karam, J. H. (1985) *Diabetes* **34**, 433–439.
11. Knowler, W. C., Pettit, D. J., Vasquez, B., Rotwein, P. S., Andreone, T. L. & Permutt, M. A. (1984) *J. Clin. Invest.* **74**, 2129–2135.
12. Permutt, M. A., Andreone, T., Chirgwin, J., Elbein, S. & Rotwein, P. (1985) in *Proceedings of the Seventh International Congress on Endocrinology*, eds. Labrie, F. & Proulx, L. (Excerpta Medica,New York), p. 245.
13. Permutt, M. A., Rotwein, P., Andreone, T., Ward, W. K. & Porte, D. (1985) *Diabetes* **34**, 311–314.
14. Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. (1982) *Cell* **29**, 319–328.
15. Owerbach, D., Billesbotle, P., Schroll, M., Johansen, K., Paulsen, S. & Nerup, J. (1982) *Lancet* **ii**, 1291–1293.
16. Mandrup-Poulsen, T., Owerbach, D., Mortensen, S. A., Johansen, K., Meinertz, H., Sorensen, H. & Nerup, J. (1984) *Lancet* **i**, 250–252.
17. Jowett, N. I., Williams, L. G., Hitman, G. A. & Galton, P. J. (1984) *Br. Med. J.* **288**, 96–99.
18. Elbein, S. C., Corsetti, L. & Permutt, M. A. (1985) *Diabetes* **34**, 1139–1144.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Chakravarti, A., Phillips, J. A., Mellits, K. H., Buetow, K. H. & Seeburg, P. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6085–6089.
21. Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Waber, P. G., Boehm, C. D. & Kazazian, H. H. (1984) *Am. J. Hum. Genet.* **36**, 1239–1258.
22. Ewens, W. J., Spielman, R. S. & Harris, H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3748–3750.
23. Kazazian, H. H., Chakravarti, A., Orkin, S. & Antonarakis, S. E. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 137–146.
24. Matteson, K. J., Ostrer, H., Chakravarti, A., Buetow, K. H., O'Brien, W. E., Beaudet, A. L. & Phillips, J. A. (1985) *Hum. Genet.* **69**, 263–267.
25. Hedrick, P.W. & Thompson, G. (1986) *Genetics* **112**, 135–156.
26. Hill, W. G. & Robertson, A. (1968) *Theor. Appl. Genet.* **38**, 226–231.
27. Sved, J. A. (1971) *Theor. Popul. Biol.* **2**, 125–141.
28. Ohta, T. & Kimura, M. (1969) *Genet. Res.* **13**, 47–55.
29. Terasaki, P. I., ed. (1980) *Histocompatibility Testing* (UCLA Tissue Typing Laboratory, Los Angeles).
30. Weitkamp, L. R. & Lamm, L. (1982) *Cytogenet. Cell Genet.* **32**, 130–142.
31. Chakravarti, A. & Buetow, K. H. (1985) *Am. J. Hum. Genet.* **37**, 984–997.
32. Murray, J. C., Mills, K. A., Demopulos, C. M., Hornung, S. & Motulsky, A. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3486–3490.
33. Barker, D., Holm, T. & White, R. L. (1984) *Am. J. Hum. Genet.* **36**, 1159–1171.
34. Treco, D., Thomas, B. & Arnheim, N. (1985) *Mol. Cell. Biol.* **5**, 2029–2038.
35. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **314**, 67–73.
36. Kobayashi, I., Stahl, M. M., Leach, D. & Stahl, F. W. (1983) *Genetics* **104**, 549–570.
37. Kazazian, H. H., Orkin, S. H., Markham, A. F., Chapman, C. R., Youssoufian, H. & Waber, P. G. (1984) *Nature (London)* **310**, 152–154.