Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 3486–3490, June 1984
Genetics

Linkage disequilibrium and evolutionary relationships of DNA variants (restriction enzyme fragment length polymorphisms) at the serum albumin locus

(chromosome 4/nonhuman primates/molecular evolution/recombination rates/crossovers)

JEFFREY C. MURRAY*†, KATHLEEN A. MILLS‡, CHRISTINA M. DEMOPULOS*, SIGRID HORNUNG*†, AND ARNO G. MOTULSKY*†§

Departments of *Medicine (Medical Genetics), ‡Anthropology, and §Genetics, and †Center for Inherited Diseases, University of Washington, Seattle, WA 98195

Contributed by Arno G. Motulsky, February 28, 1984

ABSTRACT **Four additional DNA variants (restriction enzyme fragment length polymorphisms) making a total of eight polymorphic sites at the human albumin locus have been identified. These eight sites were found after screening 689 of 20,000 nucleotides by using cDNA probes for albumin with 27 different restriction enzymes. One in 85 nucleotides was therefore potentially polymorphic. The average nucleotide diversity between any two randomly chosen chromosomes was calculated to be 1/500. We observed marked linkage disequilibrium between the eight variants. Only 7 haplotypes among 256 possible combinations were observed in 160 chromosomes from Caucasoids, Blacks, and Asians. Two haplotypes were found in all three human races, indicating that their origin predated human racial divergence. The three rarest haplotypes appear to represent recombinational events between the more common haplotypes. All crossovers occurred in the same general region. Studies of several nonhuman primates indicated that the origin of one haplotype predated the human–African ape divergence. Although it is not possible to rule out maintenance of this tight linkage by selection or fixation, it is suggested that the limited number of haplotypes at the chromosomal site of the albumin gene near the centromere of chromosome 4 may be the result of decreased recombination.**

The study of human genetic variation has been enhanced by the discovery of restriction enzyme fragment length polymorphisms (RFLPs). With the increasing availability of probes these DNA variants are being discovered at an accelerating pace. Having detected four DNA polymorphisms at the albumin locus (1), we describe an additional four such variants here. The physical and genetic distances between these variants are much smaller than those conventionally studied in human genetics. The probability of recombination between such polymorphic DNA sites is therefore quite low and some linkage disequilibrium, as demonstrated by the existence of a smaller than expected number of haplotypic arrangements of the eight DNA variants, would not be too surprising. However, even with closely linked markers, a very large number of generations should lead to free reassortment of polymorphic alleles unless certain haplotypes were more common because of selection, genetic drift, or a reduced recombination rate. This study disclosed only 7 of the 256 (or 2^8) possible haplotypes of DNA variants at the albumin locus. To evaluate the significance of this strong linkage disequilibrium we studied the haplotypic frequencies of DNA variants at this locus in the three major human races as well as in a variety of nonhuman primates. Since the origin of some of these haplotypes predated human racial divergence, a significant deviation from free reassortment of markers in this region appears likely. Various considerations suggest reduced recombination as the most reasonable explanation.

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.

MATERIALS AND METHODS

DNA was prepared (2) from 30 ml of peripheral blook obtained following signed informed consent. Five- to 10-μg aliquots of DNA were digested with appropriate restriction enzymes (Bethesda Research Laboratories or New England Biolabs) and subjected to electrophoresis on 0.8% agarose gels. Southern transfers onto nitrocellulose, hybridization with ^{32}P-labeled probes, washing, and autoradiography were carried out as described (1). The cDNA probes for albumin (3) consisted of a 900-base-pair (bp) fragment from the 5′ end of the albumin coding region, a 1000-bp fragment from the 3′ end of the coding region, and a 250-bp probe to the middle of the coding region. The individuals studied were normal, unrelated, and included 105 Caucasoid individuals (i.e., 210 chromosomes), 22 American Blacks, and 18 Asians (10 Japanese and 8 Chinese). Family studies were done to trace a given RFLP in family members.

Nonhuman primate DNAs were prepared in the same manner from 50 ml of blood sampled from one of each of the following species: pygmy chimpanzee (*Pan paniscus*), common chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and baboon (*Papio cynocephalus*); and from two orangutans (*Pongo pygmaeus*).

RESULTS

We have previously described four polymorphic sites at the albumin locus (1). One of these, the 5′ *Pst* I polymorphism, was initially considered to be a three-allele "system." After further study with a variety of double digests, including *Sac* I–*Pst* I, *Msp* I–*Pst* I, and *Hae* III–*Pst* I (data not shown), two different polymorphic sites were shown to account for the three alleles observed.¶ The three other polymorphisms discovered in this study included an *Hae* III site in the midportion of the locus giving fragments of 4.1 and 4.05 kb, a *Sac* I site in the 3′ portion of the locus giving fragments of 20 kb and 16 kb + 4 kb, and an *Eco*RV site in the 3′ flanking portion of the locus producing fragment sizes of 9.0 kb + 6.2 kb (Fig. 1).

Abbreviations: RFLP, restriction enzyme fragment length polymorphism; kb, kilobase(s); bp, base pair(s).

¶The two sites in the direction 5′ to 3′ generated fragments of 4 kilobases (kb) + 14 kb and 14 kb + 6 kb. The fact that only three alleles rather than the expected four were seen is explained by linkage disequilibrium between the two sites. Thus, in over 200 chromosomes the absence of the 14 kb + 6 kb site was always linked to the absence of the 4 kb + 14 kb site. The presence of the 14 kb + 6 kb site, however, was found both with and without the adjacent site.

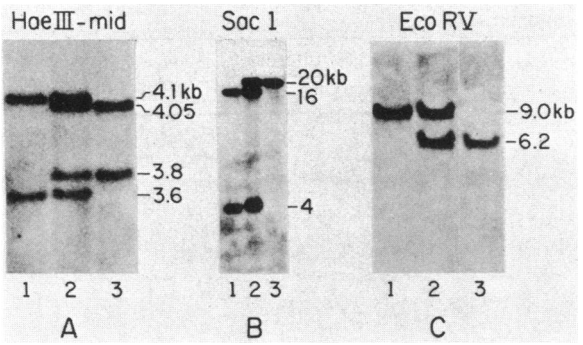

FIG. 1. (*A*) Lanes 1, 2, and 3 are father, child, and mother, respectively. The 3.8- and 3.6-kb bands have been reported (1). The 4.1- and 4.05-kb bands were disclosed by using the enzyme *Hae* III and any of the three probes described in the text. (*B*) Lanes 1, 2, and 3 are father, child, and mother, respectively. The 20- and 16-kb bands were disclosed by the enzyme *Sac* I and either the 5′ (F47) or 3′ (B44) probe. The 4-kb band was seen only when the 3′ (B44) probe was used. (*C*) Lanes 1, 2, and 3 are father, child, and mother, respectively. The 9.0- and 6.2-kb bands were seen only with *Eco*RV and the 3′ (B44) probe. The predicted 2.8-kb band has not been observed, suggesting the polymorphic site is distal to the end of the probe used.

The frequencies of each of the two alleles found at the eight different polymorphic sites have been calculated in Caucasoids, American Blacks, and Asians (Table 1). All variants showed autosomal codominant inheritance and the distribution of homozygotes and heterozygotes at each site was consistent with Hardy–Weinberg equilibria determined from the individual allele frequencies.

Using the 5′, middle, and 3′ probes, the arrangement of the eight polymorphic sites relative to one another was determined (Table 2). Although the specific locations of the polymorphic sites within the 5′ internal and 3′ internal regions may vary relative to one another, the positions of these sites with respect to the three other regions at this locus are firmly established.

By assigning a "+" for the presence of a particular restriction enzyme site and a "−" for its absence, haplotype arrangements for all eight polymorphic sites taken as a unit have been determined in 160 chromosomes from members of three races. Haplotype status was assigned by family studies (16 chromosomes) or by homozygote status for a given haplotype (82 chromosomes). The haplotype assignment of 62 chromosomes from heterozygotes was consistent with the 7 haplotypes established in families or homozygotes, although definitive assignment was not possible. An additional 130 chromosomes have been partially scored for only some enzymes. All findings were consistent with the 7 observed haplotypes, designated *A*, *B*, C_1, C_2, *D*, *J*, and *L* (Table 2).

Since each of the eight polymorphic sites had two alleles (+ and −) there were 2^8 or 256 possible haplotypic arrangements. To date we have observed only 7 of these 256 possibilities in the 160 chromosomes studied. This is a highly significant discrepancy ($\chi^2_1 = 74.8$, $P < 0.001$). Table 2 shows the 7 observed haplotypes, and Table 3 indicates their frequencies in each of the populations studied. Haplotype *A* was not observed in the sample of 22 chromosomes from Asians and was found in only 4% of chromosomes from Caucasoids and 12% of chromosomes from Blacks. Haplotype C_2 was only found in Caucasoids; it differed from C_1 only by the presence of an *Eco*RV site. Haplotypes *D* and *L* are rare variants, each found on only 1 of 110 chromosomes from Caucasoids. Haplotype *J*, seen only in Blacks, occurs with a frequency of 8%. Haplotypes *B* and C_1 were found with somewhat similar frequencies in all three races.

The eight polymorphic sites were found in 689 nucleotide sites screened by 27 restriction enzymes; thus, the average frequency with which any nucleotide site is polymorphic is 1 in 85 at the albumin locus. Because some alleles are relatively infrequent and because of linkage disequilibrium the actual number of nucleotides that differ between any two randomly chosen chromosomes is much less. Calculating this measure of nucleotide diversity in Caucasoids (4, 5) gives values of 0.0022[‖] and 0.0030,[**] respectively. These values indicate an average of 1/500 and 1/300 nucleotide differences between random chromosomes.

To discern evolutionary origins of these haplotypes we

[‖]Nucleotide diversity (4), $\pi = \Sigma_{ij}/n_c$, in which n_c = the number of sites differing between any two haplotypes taken from Table 2 by inspection (example, haplotypes *A* and *B* differ at four sites) divided by the total number of nucleotide sites studied (689), = (0.04)(0.49)(4/689) + (0.04)(0.32)(5/689) + (0.04)(0.14)(6/689) + (0.04)(0.01)(5/689) + (0.04)(0.01)(4/689) + (0.49)(0.32)(5/689) + (0.49)(0.14)(6/689) + (0.49)(0.01)(1/689) + (0.49)(0.01)(4/689) + (0.32)(0.14)(1/689) + (0.32)(0.01)(4/689) + (0.32)(0.01)(1/689) + (0.14)(0.01)(5/689) + (0.14)(0.01)(2/689) + (0.01)(0.01)(5/689) = 0.0022.

[**]Heterozygosity per position (5), $H = (nc - \Sigma C_i^2)/(jc)(n - 1)$, in which n = number of homologous segments studied = 20; c = total cuts of all cleavage sites = 2543; j = average length recognition sequence = 5.22; C_i = number of members of sample cut at each site by one of the enzymes; c_i = 20 for 124 sites, 10 for 5 sites, 3 for 1 site, and 1 for 2 sites. $H = [(20)(2543) - 124(20^2) - 5(10^2) - 1(3)^2 - 2(1)^2]/(5.22)(2543)(19) = 0.0030$.

Table 1. Frequency of DNA variants (RFLPs) at human serum albumin locus

Polymorphism			Frequency (no. of chromosomes studied)			
Enzyme	Position	Fragment size, kb	Caucasoid	Black	Asian	Total
Hae III	5′	3.8	0.46 (82)	0.54 (25)	0.44 (14)	0.47 (123)
		3.6	0.54 (97)	0.46 (21)	0.56 (18)	0.53 (138)
Hae III	Midportion	4.1	0.51 (69)	0.33 (10)	0.50 (14)	0.47 (93)
		4.05	0.49 (65)	0.67 (20)	0.50 (14)	0.53 (103)
Hae III	3′	0.9	0.48 (92)	0.50 (16)	0.41 (14)	0.48 (124)
		0.77	0.52 (98)	0.50 (16)	0.59 (20)	0.52 (134)
Msp I	5′	18	0.97 (200)	0.91 (40)	1.00 (32)	0.96 (282)
		13	0.03 (7)	0.09 (4)	0 (0)	0.04 (11)
Pst I	5′ flank	24, 18	0.57 (104)	0.36 (15)	0.50 (18)	0.52 (141)
		14	0.43 (78)	0.64 (27)	0.50 (18)	0.48 (129)
Pst I	5′	24	0.03 (6)	0.07 (3)	0 (0)	0.03 (9)
		18, 14	0.97 (176)	0.93 (39)	1.00 (36)	0.97 (261)
Sac I	3′	20	0.54 (72)	0.33 (10)	0.60 (18)	0.52 (102)
		16	0.46 (62)	0.67 (20)	0.40 (12)	0.48 (96)
*Eco*RV	3′ flank	9.0	0.86 (144)	1.00 (32)	1.00 (36)	0.90 (222)
		6.2	0.14 (23)	0 (0)	0 (0)	0.10 (23)

Table 2. Haplotypic arrangements of RFLPs at the human serum albumin locus

Haplotype or species	5′ flanking	5′ internal			Midportion	3′ internal		3′ flanking
	Pst I	*Msp* I	*Pst* I	*Hae* III	*Hae* III	*Sac* I	*Hae* III	*Eco*RV
A	−	+	−	+	+	+	+	−
B	−	−	+	+	−	−	+	−
C_1	+	−	+	−	+	+	−	−
C_2	+	−	+	−	+	+	−	+
D	+	−	+	+	−	−	+	−
J	+	−	+	+	+	+	+	−
L	−	−	+	−	+	+	−	−
P. paniscus	N_A	N_C	−	+	+	+	+	−
P. troglodytes	N_A	N_C	−	+	+	+	+	−
G. gorilla	N_B	N	+	N	+	+	+	−
P. pygmaeus	N_B	N	+	+	N_D	−	+	N
P. cynocephalus	N	N	+	N	N_D	N	N	N

+ indicates the presence of a given restriction enzyme site; − indicates absence of the restriction enzyme site. For nonhuman primates + or − indicates that bands identical in size to those found in humans were found with a particular enzyme and human probe combination. N indicates that a band of different size was found with that enzyme–probe combination. Subscripts A–D indicate bands of different size from human, but identical to each other, in more than one nonhuman primate.

studied the restriction enzyme patterns of a few nonhuman primates. The results shown in Table 2 are consistent with well-established phylogenetic affinities. The two species of chimpanzees had identical haplotypes and closely resembled both the human haplotype *A* and that of the gorilla. The orangutan haplotype was less similar to those of humans than were the African apes, and the baboon, an Old World monkey, showed little haplotypic resemblance to humans or apes. To determine possible origins for the human types a parsimony analysis (6) was performed. This method involves determining the minimal number of nucleotide substitutions that have accumulated in different lineages since their current representatives last shared an assumed common ancestor. The high degree of similarity between the human haplotype *A* and the two chimpanzee patterns (with only two of eight differences) suggests a common ancestral origin. Other patterns were consistent with these data.

The origins of the polymorphic variants may be assessed by ascertaining the time of origin of two sites relative to one another. The number of generations that have elapsed during which two variant sites have been present in the same breeding population was calculated. Assuming Hardy–Weinberg conditions, i.e., no selective advantage, no drift, a large population size, and a crossover frequency of 1 per 10^8 bp per generation (7),

$$n = \frac{\log (\text{probability of no crossovers})}{\log (1 - D/10^8)},$$

in which n = number of generations since origin; probability of no crossovers = 1 − % recombination observed between any two polymorphic sites; D = distance in base pairs between the two polymorphic sites. We have computed the number of generations since origin for a pair of polymorphic sites (the 5′ *Hae* III site and the 3′ *Hae* III site) in linkage disequilibrium at the albumin locus. The distance between these sites as assessed from the albumin map was no more than 4050 bp, as established by the size (Fig. 1*A*) and the position (Table 2) of the relevant *Hae* III-midportion fragment. Since no recombination was observed in any of the 196 chromosomes studied in three races, 95% and 99% confidence limits were used to estimate the maximal amount of recombination that could have occurred (8). Using minimal distance and maximal recombination between any two sites will provide values for the maximal number of generations since origin (n) under the ideal conditions stipulated. These values were estimated to be 505 generations (or 12,600 years) with 95% confidence limits and 736 generations (18,400 years) with 99% confidence limits.

Table 3. Frequency of RFLP haplotypes at human serum albumin locus

Haplotype	Caucasoid (110)	Black (26)	Asian (22)	Other (2)	Total
A	0.04	0.12	—	—	0.04
B	0.49	0.35	0.68	—	0.49
C_1	0.32	0.46	0.32	1.00	0.35
C_2	0.14	—	—	—	0.09
D	0.01	—	—	—	0.006
J	—	0.08	—	—	0.01
L	0.01	—	—	—	0.006

Values in parentheses represent numbers of chromosomes studied.

DISCUSSION

Nucleotide Diversity. Previous studies at the β-globin locus have shown that about 1 nucleotide site in 100 is potentially polymorphic and that 1 nucleotide in 500 will actually differ between any two randomly chosen chromosomes (4). This discrepancy between potential and actual data is due to the rarity of some alleles and to the linkage disequilibrium between different polymorphic sites. Combining published data with our data presented here, we have demonstrated similar values for the albumin locus (1/85 maximal difference and 1/500 average difference), which suggest that this degree of nucleotide diversity may apply to other loci.

If there are 10^6 bp of DNA on the average in 1 centimorgan (a 1% recombination distance) (7), it would mean that within 2 centimorgans of any given site of interest (1 centimorgan on either side) there would be 20,000 potentially polymorphic markers (1/100 of 2×10^6); 4000 of these (1/500 of 2×10^6) would be informative for distinguishing between any two randomly chosen chromosomes. This suggests enormous numbers of markers useful for study of human disorders. The actual number may, in fact, be much larger if the number of polymorphisms in noncoding regions exceeds that of coding loci (9). Since there is disequilibrium between markers within very short distances (100 kb or 0.1 centimorgan or less) in β-globin (10), immunoglobulins (11, 12), and albumin, such linkage disequilibrium may be a general phenomenon. However, free recombination for the RFLPs associated with Huntington disease on the short arm of chromosome 4 has been reported (13). Extensive searches for multi-

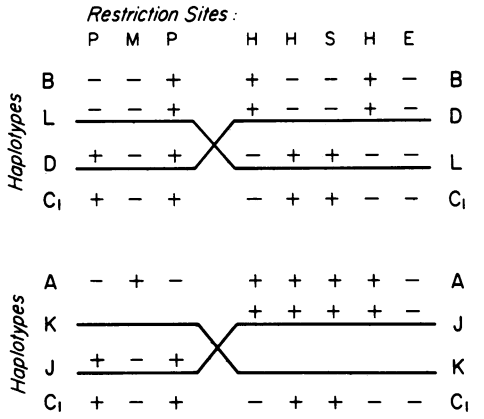

FIG. 2. (*Upper*) Origin of haplotypes *D* and *L* by a single crossover involving a *B* and a C_1 chromosome between the internal *Pst* I site and the 5′ *Hae* III site. (*Lower*) Origin of the *J* haplotype by a crossover in the same region as *Upper* but between an *A* and a C_1 chromosome. Haplotype *K* is the hypothesized reciprocal crossover chromosome but it has not been observed. See Table 2 for a full description of restriction sites (P = *Pst* I, M = *Msp* I, H = *Hae* III, S = *Sac* I, and E = *Eco*RV).

ple DNA polymorphisms with a single unique-sequence probe may not lead to the detection of more markers useful for linkage studies after the first two or three polymorphic sites have been found. It would be more productive to use probes located 100 or 200 kb distant from the initial site; these will be more likely to show polymorphisms not in linkage disequilibrium with the initial site. Yet at this distance, the recombination fraction between the polymorphic sites and the site of interest would not increase significantly.

Crossovers. The rarest haplotypes—*D*, *J*, and *L*—are each explainable as having arisen by crossover events between the more common haplotypes. These crossovers are postulated to have occurred 5′ of the 5′ *Hae* III site and possibly between the internal *Pst* I site and the 5′ *Hae* III site. Fig. 2 shows the likely origin for haplotypes *D* and *L* by a crossover between *B* and one of the *C* haplotypes and the most probable origin for haplotype *J* by a crossover between *A* and one of the *C* haplotypes. Although haplotypes *D* and *L* appear to be reciprocal products of a single event, we do not know if they arose simultaneously or if they were the results of two separate events.

Nonhuman Primates. The pattern of resemblances between humans and other primates examined was in agreement with the data based on other molecular findings, anatomical comparisons, and the fossil record (14, 15). Since only a small number of chromosomes from nonhuman primates were examined (two to four for each species), it is likely that as yet undetected polymorphic sites will be found. It is therefore noteworthy that the human haplotype *A* over its 3′ half strongly resembles the patterns of the gorilla and chimpanzee (Table 2). Since this resemblance is one of similarity at a number of restriction sites over about 20 kb, it is unlikely to be due to convergence. Human haplotype *A* may thus represent the state of this chromosomal region prior to the divergence of hominids and the African apes.

Linkage Disequilibrium. Linkage disequilibrium over relatively short genetic distances such as reported here for albumin has previously been found at the *HLA* locus (0.8% recombination = 800 kb) by using antigenic markers and at the immunoglobulin (200 kb) and β-globin (60 kb) loci by using RFLPs. Several explanations have been offered. For *HLA*, a selective advantage for certain haplotypes has been postulated (16), but drift associated with effects of migration may also have played a role in the preservation of certain haplotypes at frequencies higher than expected (17). For immunoglobulin (11) and β-globin (18), both a recent origin for variant sites and gene conversion have been proposed to explain the observed disequilibria. For β-globin, founder effects, recurrent mutation (18), and selection (19) have also been proposed.

Calculating the time of origin of these polymorphisms yielded values ranging from 12,600 to 18,400 years. This period is more recent than the time hypothesized for human racial divergence, which was under way at least 20,000 years ago (20). The inconsistency between our maximum of 18,400 years and the minimum of 20,000 years would suggest that one or more of the assumptions required by the computation do not apply. These are discussed below.

First, the very similar degrees of linkage disequilibrium observed in haplotypes *B* and C_1 in Caucasoids, Blacks, and Asians suggest that migration or founder phenomena cannot be cited as invalidating the calculations. These factors are highly unlikely to have produced identical effects in three races over a period of at least 20,000 years. Another assumption of the calculation—no selection—is difficult to consider when the exact location of the polymorphic sites within introns, exons, or flanking sequences is not yet known. We have no direct evidence either for or against selection; some selective advantage that has now disappeared cannot be ruled out.

It is possible that recombination at the albumin locus is lower than average, as has been hypothesized for the immunoglobulin region (12). Centromeres are regions of low recombination frequency and the albumin locus is near the centromere of chromosome 4 (4q11 to 4q13). Furthermore, it has been shown that recombination frequencies can vary by as much as 60-fold between sites on the same chromosome (21). In fact, a relatively decreased frequency of meiotic recombination has been observed in the centromeric region of chromosome 4 (22). If the recombination frequencies were 1/10 that of the average (or now 1 crossover per 10^9 bp per generation), the calculated time of origin for the polymorphisms would increase from at least 12,600 years ago to at least 126,000 and to as much as 180,000 years ago. This interval could accommodate most hypotheses regarding human racial divergence.

If such recombinational "cold spots" exist in the genome there are important implications for relating physical distances on chromosomes to genetic distances. The probability of success in "walking" experiments, in which one attempts to move from a marker site to another site of potential interest, is determined by their relative distance from one another. It will be necessary to know whether the distance between sites involves 100,000 bp or 1,000,000 bp. Such physical distances will be apparent only if the specific relationship to genetic distance at a given chromosomal location is known.

Richard Lawn of Genentech (South San Francisco, CA) was most generous in allowing us use of the albumin cDNA probes. Nonhuman primate blood samples were provided by Dr. J. W. Foster (Woodland Park Zoological Gardens, Seattle), Robert Veith (Division of Hematology, University of Washington), and Yerkes Regional Primate Research Center (Emory University, Atlanta). We are grateful for blood samples provided by numerous human donors, for the skillful assistance of Tak Kurachi, Ryan Murray, and Sylvia Waggoner, and to George Stamatoyannopoulos and Peter Nute, who provided significant support. This research was supported by U.S. Public Health Service Center Grant GM 15253, U.S. Public Health Service Training Grant GM 07454, and a grant (to C.M.D.) from the Ahepa Cooley's Anemia Foundation.

1. Murray, J. C., Demopulos, C. M., Lawn, R. M. & Motulsky, A. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5951–5955.
2. Poncz, M., Solowiejczyk, D., Harpel, B., Mory, Y., Schwartz, E. & Surrey, S. (1983) *Hemoglobin* **6**, 27–36.
3. Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E.,

Houch, C. M., Najarian, R. C., Seeburg, P. H. & Wion, K. L. (1981) *Nucleic Acids Res.* **22,** 6103–6114.
4. Nei, M. & Koehn, R. K., eds. (1983) *Evolution of Genes and Proteins* (Sinauer, Sunderland, MA).
5. Engels, W. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6329–6333.
6. Felsenstein, J. (1983) in *Statistical Analysis of DNA Sequence Data*, ed. Weir, B. S. (Dekker, New York), pp. 133–150.
7. Kurnit, D. M. (1979) *Lancet* **i,** 104.
8. Rohlf, R. J. & Sokal, R. R. (1969) *Statistical Tables* (Freeman, San Francisco).
9. Cooper, D. N. & Schmidtke, J. (1984) *Hum. Genet.* **66,** 1–16.
10. Orkin, S. H., Kazazian, H. H., Jr., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G. & Giardina, P. J. V. (1982) *Nature (London)* **296,** 627–631.
11. Migone, N., Feder, J., Cann, H., van West, B., Hwang, J., Takahashi, N., Honjo, T., Piazza, A. & Cavalli-Sforza, L. L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 467–471.
12. Bech-Hansen, N. T., Linsley, P. S. & Cox, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6952–6956.
13. Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonilla, E. & Martin, J. B. (1983) *Nature (London)* **306,** 234–238.
14. Andrews, P. (1982) *Nature (London)* **295,** 185–186.
15. Andrews, P. & Cronin, J. E. (1982) *Nature (London)* **297,** 541–546.
16. Bodmer, W. F. & Bodmer, J. G. (1978) *Br. Med. Bull.* **34,** 309–316.
17. Ohta, T. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1940–1944.
18. Kazazian, H. H., Orkin, S. H., Antonarakis, S. E., Youssoufian, H., Markham, A. F., Boehm, C. D. & Waber, P. G. (1983) *Am. J. Hum. Genet.* **35,** 175A (abstr.).
19. Naeda, N., Bliska, J. B. & Smithies, O. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5012–5016.
20. Weiss, K. M. & Maruyama, T. (1976) *Am. J. Phys. Anthropol.* **44,** 31–50.
21. Saadallah, N. & Hulten, M. (1983) *J. Med. Genet.* **20,** 290–299.
22. Hulten, M. (1974) *Hereditas* **76,** 55–78.