Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5165–5169, July 1986
Genetics

Analysis of the human α-globin gene cluster reveals a highly informative genetic locus

(haplotype/genetic linkage/recombination/polymorphism)

D. R. HIGGS*, J. S. WAINSCOAT*, J. FLINT*, A. V. S. HILL*, S. L. THEIN*, R. D. NICHOLLS*, H. TEAL*, H. AYYUB*, T. E. A. PETO*, A. G. FALUSI†, A. P. JARMAN*, J. B. CLEGG*, AND D. J. WEATHERALL*

*Medical Research Council Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, England; and †Department of Haematology, University College Hospital, Ibadan, Nigeria

Communicated by Sir James Gowans, March 13, 1986

ABSTRACT **Extensive molecular studies have characterized 15 dimorphic and 2 multiallelic genetic markers within the human α-globin gene cluster. Analysis of these markers in 9 populations has shown that the α-globin locus is remarkably polymorphic and is therefore an ideal marker on chromosome 16 for the construction of a human genetic linkage map. The combined analysis of 9 polymorphic markers has established α-globin haplotypes that provide the means to study the molecular genetics and common mutants of this cluster. The novel association of a conventional restriction fragment length polymorphism haplotype and linked, hypervariable regions of DNA should allow a comparison of the rate of change of such markers.**

The human α-globin gene cluster is the only cloned DNA marker on the short arm (p12-pter) of chromosome 16. It includes two adult (*α2* and *α1*) and one embryonic (*ζ2*) gene separated by two pseudogenes (*ψα* and *ψζ1*) within a 30-kilobase (kb) segment of DNA (5′ *ζ2-ψζ1-ψα-α2-α1* 3′) (reviewed in ref. 1). Previously we have identified several polymorphic markers at this locus (2–5), including two hypervariable regions (HVRs), one between the *ζ2* and *ψζ1* genes (interzeta HVR, IZHVR) (3–5) and another at the 3′ end of the complex (3′ HVR) (5). To establish the maximum detectable variability in the α cluster we have systematically searched for polymorphisms of all types throughout the entire region. This has enabled us to construct multimarker haplotypes for this locus and hence to provide a basis for understanding the molecular genetics and evolution of the normal α-globin complex and its common mutants. The data indicate that although the α-globin locus is highly polymorphic, there are some common haplotypes in the nine populations that we have studied. The heterozygosity at this locus is at least 0.93 and it will therefore provide a valuable marker on the short arm of chromosome 16 for the construction and analysis of a human genetic linkage map.

MATERIALS AND METHODS

Individuals Studied and DNA Analysis. Several restriction fragment length polymorphisms (RFLPs) were identified in DNA from ≈1500 nonthalassemic, apparently healthy, unrelated individuals. DNA from 50 unrelated Jamaicans was systematically screened by using the restriction enzymes listed in the legend to Table 1 (usually 20 samples per enzyme site studied). Having identified the common RFLPs, all such markers were analyzed in individuals from 9 different populations (Table 1). Two selected groups of Mediterranean individuals homozygous for a common RFLP were also studied (10 subjects *Rsa* I ++ and 10 subjects *Rsa* I −−) (see below). Blot hybridization studies were performed as described (6). Nonthalassemic individuals (*αα/αα*) were identified by using a standard protocol (6).

Preparation of Radioactive Probes. The following probes were used in the detection of RFLPs: (*1*) a 0.6-kb *Bam*HI/*Eco*RI fragment ≈9 kb 5′ to the *ζ2* gene, isolated from the recombinant plasmid pJW5, which is a subclone of the cosmid cSG1 (7); (*2*) a 1.8-kb *Sac* I fragment from the recombinant pBR*ζ* (8, 9) or a 3.1-kb *Bam*HI/*Eco*RI fragment containing the *ζ* gene (9); (*3*) a 1.1-kb *Alu* I fragment from the plasmid pSG21, which contains the IZHVR (3, 4); (*4*) a 1.5-kb *Pst* I fragment from pDH6 containing the *ψα1* gene from a Chinese individual; (*5*) a 1.5-kb *Pst* I fragment from the recombinant pRBα1 (8) containing the *α1*-globin gene; (*6*) a 0.8-kb *Bam*HI fragment from the plasmid pDH12, which includes single-copy sequences ≈3 kb 3′ to the *α1*-globin gene (unpublished observation); and (*7*) a 4-kb *Hin*fI fragment from the recombinant pSEA1, which includes the 3′ HVR (10). The locations of these probes are indicated in Fig. 1. All probes were nick-translated and hybridized to nitrocellulose filters as described (2).

RESULTS

Identification of Polymorphic Markers. During the systematic search for RFLPs we screened DNA from various individuals using 26 restriction enzymes (see legend to Table 1) and 7 probes (Fig. 1). Although not all digests were analyzed with all probes, a total of 203 restriction sites [1132 base pairs (bp)] was examined and 14 of them were polymorphic (Table 1 and Fig. 1). The estimate of nucleotide diversity in this segment of the genome is therefore 1 in 80 bp, which is similar to previous estimates for other loci (14). The positions of these 14 polymorphic sites are indicated in Fig. 1, which also includes three previously characterized polymorphic regions of the α-globin locus. Two of these are HVRs of DNA consisting of tandemly reiterated 36-bp (IZHVR in refs. 3 and 4) and 17-bp sequences (3′ HVR; A.P.J., R.D.N., and D.R.H., unpublished). The third is a common variation in the downstream ζ-like gene, the structure of which may resemble either a *ψζ1* (PZ) or *ζ1* (Z) gene (2).

All 17 polymorphic markers were analyzed in nine populations (Table 1 and Fig. 1 legend, excluding the 3′ HVR). They fall into three broad groups. Common polymorphisms (Table 1) are defined here as those in which the frequency of the less common allele is >0.05 in most populations. The major deviations from this pattern are seen in populations that may have been founded from relatively small numbers of

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: HVR, hypervariable region; IZHVR, interzeta HVR; RFLP, restriction fragment length polymorphism; kb, kilobase(s); bp, base pair(s).

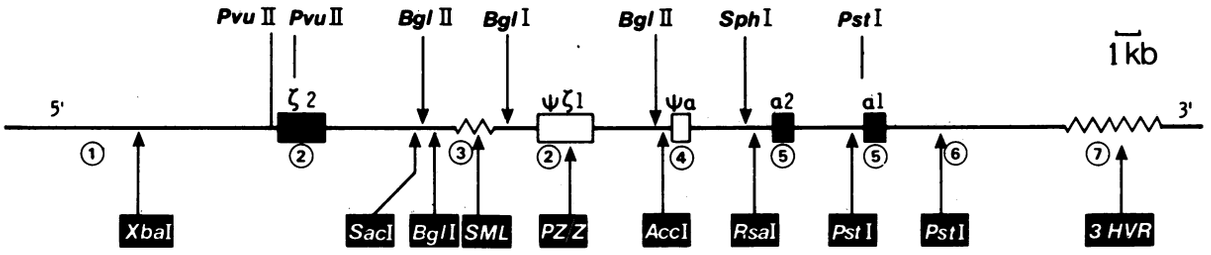

FIG. 1. α-Globin haplotype and genetic map. The IZHVR between the ζ2 and ψζ1 genes and the 3′ HVR are denoted by ~~. The probes described in the text are numbered (in circles). *Below*: common polymorphisms (white letters on black base) are shown. L, M, and S refer to large, medium, and small alleles (4) of the IZHVR. PZ and Z refer to the presence of either a pseudozeta or zeta-like sequence in this position (2). To improve the quality of the Southern blots used to detect the *Xba* I polymorphism, a double-digest (*Bam*HI/*Xba* I) was used to produce smaller fragments. For simplicity, the uncommon 2.1-kb ψζ1 and common 1.8-kb ψζ1 allele (2) are both referred to as PZ. The *Xba* I (11), *Sac* I (4), and *Rsa* I (12) site polymorphisms have been described in detail elsewhere. *Above*: rare polymorphisms (black letters). The precise position of the rare *Pst* I polymorphism detected with probe *5* has not yet been mapped. The *Pvu* II and 3′ *Bgl* II (refs. 2 and 13) polymorphisms have been described elsewhere.

individuals (e.g., Papua New Guinea and Island Melanesia). Uncommon polymorphisms (summarized in the legend to Table 1) are often limited in their population distribution and generally occur at frequencies of <0.05. Some markers in this group (e.g., *Pvu* II, *Bgl* II, *Sph* I, and *Pst* I) are population-specific. Included within this group are ζ-globin rearrangements (15) that give rise to chromosomes with single (ζ) or triplicated (ζζζ) configurations. Although the (ζ) chromosome is virtually limited to Jamaican blacks in this survey (gene frequency 0.04), the reciprocal (ζζζ) arrangement is found at a relatively high frequency among the Southeast Asian population (0.09). The third group of polymorphic markers (3′ HVR) is discussed below.

Identification of Multimarker Haplotypes. All 9 common polymorphic markers (excluding the 3′ HVR) were characterized in 223 individuals. To identify the most frequent haplotypes we initially analyzed data from those individuals who were homozygous at all sites (Table 2); the combinations thus identified were arbitrarily grouped according to their 3′ haplotype (Z/PZ, *Acc* I, *Rsa* I, *Pst* I, *Pst* I). This analysis identified five predominant haplotypes (*Ia, IIa, IIIa, IVa*, and *VIIa*) that occur with predicted frequencies of ≈0.3–0.5 in the populations studied (see Table 2 and legend). Although some haplotypes are common in several, diverse population groups (e.g., *Ia* and *IIa*), others predominate in a single group (e.g., *IIIa, IVa, VIIa*).

To obtain further information, we analyzed those individuals who were homozygous for eight of the nine common markers. In such cases it is possible to assign unequivocally the linkage of all sites along each chromosome (Table 2). This confirmed the distribution and frequency of the common haplotypes but also revealed some minor haplotypes that are probably related to the major groups (see below).

It was apparent from this analysis that the haplotype *Ia*, which includes the *Rsa* I + polymorphism, is common in the British, Mediterranean, and Asian Indian populations. We analyzed 10 selected Mediterranean subjects who were homozygous for *Rsa* I (++) and 8 of them were homozygous for the *Ia* haplotype; 2 individuals were heterozygous at the *Xba* I site and therefore had the *Ia/Ib* genotype. All 4 British subjects, 1 Saudi Arabian, and all 6 Asian Indian subjects who were homozygous for *Rsa* I (++) were also homozygous for the common *Ia* haplotype. It therefore seemed reasonable to assume in these populations that the *Rsa* I + polymorphism marks the *Ia* haplotype and thus it was possible to derive the other haplotype in *Rsa* I +− compound heterozygotes. This analysis (Table 2) also confirmed our previous observations and revealed several more, minor haplotypes.

If a further assumption was made, that *Rsa* I − was most likely to occur as part of haplotype *IIa* in the Mediterranean population, it became possible to assign all but 4 of 62 haplotypes (94%) in this group to the list of 29 observed haplotypes in Table 2 (data not shown), suggesting that there is a limited (<30) number of haplotypes in this population.

Table 1. Common polymorphisms of the human α-globin complex

Population	*Xba* I	*Sac* I	*Bgl* I	IZHVR			Z/PZ, PZ	*Acc* I	*Rsa* I	*Pst* I	*Pst* I
				Small	Medium	Large					
United Kingdom	0.63 (30)	0.76 (38)	0.17 (42)	0.03 (40)	0.68 (40)	0.30 (40)	0.81 (42)	0.79 (42)	0.48 (42)	0.08 (38)	0.11 (38)
India	0.61 (74)	0.78 (74)	0.11 (74)	0.08 (74)	0.59 (74)	0.33 (74)	0.86 (74)	0.85 (74)	0.43 (75)	0.04 (72)	0.03 (60)
Island Melanesia	0.63 (40)	0.08 (40)	0.38 (34)	0.50 (40)	0.50 (40)	0.00 (40)	0.58 (40)	0.58 (40)	0.03 (40)	0.43 (40)	0.45 (20)
Jamaica	0.54 (50)	0.48 (96)	0.07 (72)	0.18 (93)	0.81 (93)	0.02 (93)	0.48 (66)	0.74 (98)	0.19 (108)	0.04 (74)	0.08 (38)
Mediterranean	0.53 (68)	0.80 (78)	0.13 (72)	0.08 (78)	0.62 (78)	0.31 (78)	0.85 (78)	0.84 (64)	0.45 (74)	0.04 (78)	0.04 (72)
Nigeria	0.50 (36)	0.42 (38)	0.03 (38)	0.21 (38)	0.79 (38)	0.00 (38)	0.55 (38)	0.90 (30)	0.29 (79)	0.08 (38)	0.13 (38)
Papua New Guinea	0.40 (60)	0.02 (58)	0.58 (62)	0.36 (58)	0.64 (58)	0.00 (58)	0.43 (58)	0.47 (62)	0.00 (60)	0.49 (77)	0.40 (58)
Saudi Arabia	0.47 (32)	0.47 (38)	0.16 (38)	0.11 (38)	0.76 (38)	0.13 (38)	0.71 (38)	0.67 (36)	0.32 (31)	0.03 (34)	0.06 (18)
Southeast Asia	0.41 (32)	0.69 (32)	0.19 (32)	0.16 (32)	0.41 (32)	0.44 (32)	0.79 (29)	0.72 (25)	0.31 (29)	0.03 (31)	0.03 (32)

Gene frequencies of the + allele are shown for each common polymorphic enzyme. The number of haplotypes studied is shown in parentheses. These polymorphic sites were initially identified by systematically screening DNA from various individuals. The restriction enzymes (and number of sites examined) were *Acc* I, 7; *Dra* I, 3; *Ava* I, 5; *Bal* I, 3; *Bam*HI, 9; *Bcl* I, 6; *Bgl* I, 8; *Bgl* II, 7; *Bst*EII, 11; *Eco*RI, 6; *Hae* II, 4; *Hin*cII, 9; *Hin*dIII, 6; *Hin*fI, 5; *Nae* I, 2; *Nar* I, 3; *Nco* I, 13; *Pst* I, 12; *Pvu* II, 17; *Rsa* I, 16; *Sac* I, 13; *Sma* I, 9; *Sph* I, 8; *Stu* I, 9; *Taq* I, 7; *Xba* I, 5. Uncommon polymorphisms (see Fig. 1) were identified in a group of 3000 individuals from the 9 population groups studied and will be reported in detail elsewhere. The rare *Pvu* II polymorphism was only seen in Jamaicans and Southeast Asians. The interzeta *Bgl* II was only observed in Melanesians and Polynesians (A.V.S.H., unpublished). The uncommon *Bgl* I polymorphism was seen in several populations and reaches a frequency of 0.06 in Southeast Asia. The uncommon 3′ *Bgl* II polymorphism was also seen in several populations, being particularly frequent in the British (0.07) and Mediterranean (0.04) populations. The rare *Sph* I and *Pst* I polymorphisms were exclusively found in black individuals.

Table 2. Distribution of α-globin haplotypes

	Xba I	*Sac* I	*Bgl* I	IZHVR	Z/PZ	*Acc* I	*Rsa* I	*Pst* I	*Pst* I	British	Mediterranean	Asian Indian	Saudi Arabian	Papua New Guinean	Island Melanesian	Southeast Asian	Jamaican	Nigerian
**Ia*	+	+	−	M	PZ	+	+	−	−	16 (8)	28 (9)	26 (13)	5 (3)	0	0	0	3	0
Ib	−	+	−	M	PZ	+	+	−	−	0	2	0	0	0	0	0	1	0
Ic	+	−	−	M	PZ	+	+	−	−	0	0	0	0	0	0	0	1	0
**Id*	−	+	−	L	PZ	+	+	−	−	0	0	0	0	0	0	2	0	0
**IIa*	−	+	−	L	PZ	+	−	−	−	4 (2)	7 (5)	6 (5)	3 (1)	0	0	1	0	0
IIb	+	+	−	L	PZ	+	−	−	−	1 (1)	0	1	0	0	0	0	0	0
IIc	+	+	−	M	PZ	+	−	−	−	0	1	1	0	0	0	0	0	0
IId	−	−	−	S	PZ	+	−	−	−	0	1 (1)	1	0	0	0	0	0	0
IIe	+	−	−	S	PZ	+	−	−	−	0	0	2 (2)	0	0	0	0	0	0
IIf	+	−	+	S	PZ	+	−	−	−	0	0	1 (1)	0	0	0	0	0	0
IIg	+	−	−	M	PZ	+	−	−	−	0	0	0	0	0	0	0	1	0
**IIIa*	−	−	+	M	Z	−	−	−	−	0	1 (1)	1 (1)	1 (1)	14	1	1	0	0
**IIIb*	+	−	+	M	Z	−	−	−	−	3 (3)	0	1 (1)	0	4	3	0	1	0
IIIc	−	+	−	L	Z	−	−	−	−	0	+	0	0	0	0	0	0	0
IIId	+	−	−	M	Z	−	−	−	−	0	0	0	0	0	0	0	1	0
IIIe	−	+	−	M	Z	−	−	−	−	0	1 (1)	0	0	0	0	0	0	0
**IIIf*	−	−	−	M	Z	−	−	−	−	0	0	0	0	0	0	0	2	2
**IVa*	+	−	−	S	PZ	+	−	+	+	0	0	0	0	2	6	0	0	0
IVb	−	−	−	S	PZ	+	−	+	+	0	0	0	0	1	0	0	0	0
IVc	−	+	−	L	PZ	+	−	+	+	1 (1)	0	0	0	0	0	0	0	0
Va	−	+	−	L	PZ	+	−	−	+	0	0	1	0	0	0	0	0	0
Vb	−	−	+	M	PZ	+	−	−	+	0	0	1 (1)	0	0	0	0	0	0
Vc	+	−	−	S	PZ	+	−	−	+	0	0	0	0	2	0	0	0	0
Vd	−	−	−	S	PZ	+	−	−	+	0	0	0	0	1	0	0	0	0
VIa	−	+	−	L	PZ	−	−	−	−	1 (1)	0	0	0	0	0	1	0	0
VIb	+	+	−	L	PZ	−	−	−	−	0	0	1 (1)	0	0	0	0	0	0
**VIIa*	−	+	−	M	Z	+	−	−	−	0	0	0	0	0	0	0	2	0
VIIIa	+	−	−	S	PZ	+	−	+	−	0	1 (1)	0	1 (1)	0	0	0	0	0
VIIIb	−	+	−	L	PZ	+	−	+	−	0	0	1 (1)	0	0	0	0	0	0
Unclassified										6	20	18	4	36	8	14	26	26
Incomplete										12	16	12	24	4	22	13	0	10
Total										44	78	74	38	64	40	32	38	38

The five major haplotypes are boxed. The total number of haplotypes identified in each population is shown. The number of haplotypes derived by assuming that *Rsa* I + marked the *Ia* combination is shown in parentheses. When homozygotes were identified for particular haplotypes (marked by an asterisk) it was possible to estimate the frequency of that haplotype (q) from the frequency of homozygotes in that population (q^2). This analysis showed that the *Ia* haplotype occurred at the following frequencies: British, 0.50; Mediterraneans, 0.51; Asian Indians, 0.44; Saudi Arabians, 0.38; and Jamaicans, 0.23. The *Id* haplotype occurred in Southeast Asians (0.34). The *IIa* haplotype occurred in British (0.25), Mediterraneans (0.18), and Saudi Arabians (0.38). The *IIIa* haplotype occurred in Papua New Guineans (0.45), *IIIb* in Papua New Guineans (0.18) and Island Melanesians (0.32), *IIIf* in Jamaicans (0.23) and Nigerians (0.27), *IVa* in Island Melanesians (0.55), and *VIIa* in Jamaicans (0.23). A "+" indicates a haplotype identified by analysis of Mediterranean subjects assuming *Rsa I* − to indicate the *IIa* haplotype as described in the text; the numerical results of this particular analysis are not included in the table.

Further assumptions were not made about the haplotypes in other populations, and those that could not be clearly assigned were considered unclassified. Some haplotypes were incomplete, often because only one site was missing. The data from these partial haplotypes were largely consistent with the expected distribution of haplotypes for the population studied (data not shown).

Linkage Disequilibrium Within the α-Globin Cluster. The haplotypes described here include eight dimorphic markers and one region that can be broadly classified into three alleles (4); thus, there is a theoretical total of 768 different combinations. In fact, we have observed only 29, suggesting that linkage disequilibrium exists between these polymorphic markers. The large number of unclassified haplotypes could harbor several more combinations, although the analysis of Mediterranean haplotypes (see above) suggests that the number is limited. The observed frequency of the common haplotype within any population exceeds the predicted frequency calculated from the product of the frequencies of the component polymorphic markers in that population by a factor varying between 5- and 50-fold. This again supports the general argument for linkage disequilibrium in the α-globin cluster.

The linkage of groups of polymorphic markers, as described here, may enable the identification of regions of the cluster that are favored for recombination. None of the five major haplotypes (*Ia, IIa, IIIa, IVa, VIIa*) is related to each other by single mutation events. However, 15 of the 24 minor haplotypes could arise by a single recombination between 2 of the major haplotypes. Analysis of the putative crossover points shows that although there is a tendency for more of them to be at the 5′ end of the α complex, they are distributed throughout its entire length.

By inspection, no subsegment of the haplotype appeared to be randomly associated with respect to another, as occurs in some previously characterized regions of the genome (reviewed in ref. 16). To assess this more fully it will be necessary to identify further high-frequency polymorphic markers between the *Rsa* I site and the 3′ HVR. However, at present there is no evidence for grossly different rates of recombination within the segment analyzed, although these data do not allow us to assess the rate of recombination within the α cluster compared with the overall rate of recombination in the genome.

Analysis of the 3′ HVR Locus. Because of the complex, multiallelic nature of the 3′ HVR, its distribution and linkage properties have been analyzed separately from the main haplotype. This region was assessed in three ways by using (*i*) *Sac* I and probe *5* (13.5- to 19.0-kb fragments), (*ii*) *Eco*RI and probe *6* (6.0- to 12.0-kb fragments), and (*iii*) *Pvu* II and probe *7* (1.5- to 7.0-kb fragments). The 3′ HVR locus was analyzed in all nine populations by using one or more of these strategies and in the Jamaican population all three approaches were used. Some similar-sized alleles that were not clearly resolved with *Eco*RI or *Sac* I were detected with *Pvu* II. However, in general, there was a good correlation between allele size estimated by these three strategies (results not shown), enabling comparison of the data from all populations (Fig. 2).

To determine the relationship between the 3′ HVR and the remainder of the haplotype we examined the allele sizes in individuals homozygous for haplotypes *Ia, IIa*, and *IIIa*. Individuals with the genotype *Ia/Ia* showed a remarkable degree of uniformity in the size of the 3′ HVR compared to other genotypes (Fig. 2). Differences in the 3′ HVR of individuals with this genotype often only became apparent when a *Pvu* II digest was used to assess allele sizes. The allele size distribution in *IIa* and *IIIa* homozygotes was wider than that observed for *Ia* homozygotes.

Observed Heterozygosity at the α-Globin Locus. The average heterozygosity observed at this locus using the entire haplotype is 0.93, with a range of 0.85–1.00 in the 9 populations studied. The average heterozygosity must be even higher than this since we have shown previously that alleles of the IZHVR locus described here as small, medium, and large can be further subdivided at a higher level of resolution (4), as can the 3′ HVR. Furthermore, this analysis does not include the readily detectable variation within the introns of the $\zeta 2$ and $\psi\zeta 1$ genes (2, 4).

DISCUSSION

This study has established that the human α-globin gene complex is highly polymorphic and, therefore, an extremely useful genetic marker on chromosome 16 (17). Recently, it has been accurately localized to band p13.1. (R.D.N. and D.R.H., unpublished observation). Since meiotic recombination seems to occur predominantly at the distal regions of chromosome 16 (18), it could be that a large proportion of loci on 16 p lies within 20 centimorgans of the polymorphic α cluster. This suggests that many important genetic loci could be localized to chromosome 16 with this single genetic marker. Using such an approach the loci for adult polycystic kidney disease (19) and the enzyme phosphoglycollate phosphatase (20) have already been mapped to within 11 centimorgans and 9 centimorgans (99% confidence limits) of the α-globin cluster, respectively. It could be that to construct a human genetic map, relatively few highly polymorphic markers such as that described here, within regions of low recombination, may suffice to map a large proportion of markers to the human genome.

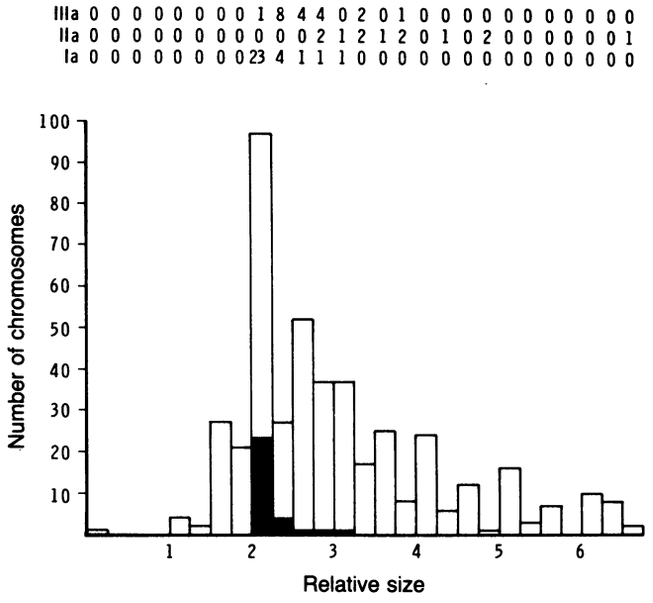

FIG. 2. Relative sizes of fragments spanning the 3′ HVR derived from one or another of the three strategies outlined in the text. White bars indicate the numbers of haplotypes of a particular size, including data from all nine populations studied. A similar distribution was seen in each separate population studied. Black bars indicate the size of the 3′ HVR estimated from *Pvu* II digests and probe *7*, in *Ia/Ia* homozygotes. (*Top*) Relative sizes of haplotypes from *Ia, IIa*, and *IIIa* homozygotes; figures indicate the number of chromosomes with a particular-sized 3′ HVR. *IIa* "homozygote" data include 3′ HVRs from four incomplete *IIa* haplotypes, and *IIIa* homozygote data include 3′ HVRs from six incomplete *IIIa* haplotypes.

Despite the usefulness of this and similar loci in studying human genetics, the mechanisms that underlie the observed polymorphism between and within populations are poorly understood. The potential contributions of migration, effective population sizes, rates of recombination or mutation, natural selection, and genetic drift, for which appropriate data are generally unavailable, have been discussed (16, 21–23). Nevertheless, some observations are common to the α-globin cluster and to many of the loci that have been studied previously. The common polymorphisms and some common haplotypes must have preceded human racial divergence (estimated to be 50,000–100,000 yr ago) and have been inherited as stable linkage groups since that time. The rare polymorphisms and haplotypes often provide useful population-specific markers and they may or may not have preceded the racial divergence. Nevertheless, it seems likely that many of the rare haplotypes are derived from the common haplotypes by recombination or mutation at a single marker within the locus.

The question of recombination rates within specific, relatively small (<100 kb) regions of the genome is of general interest and is clearly important to understanding the variation observed here. By using a similar approach in the analysis of the human β-globin cluster it has been noted that there is a region between the $\psi\beta$ and β-globin genes in which recombination occurs at a rate estimated to be 3–30 times that of the genome in general (reviewed in refs. 16 and 23), which

is calculated to be one recombination per 10^5 meioses per kb. Population data show that the two haplotype subsegments either side of this region in the β-globin cluster are randomly associated with each other. It has been suggested that minisatellite regions structurally similar to the IZHVR may also recombine at high rates [5–15 recombinations per 10^5 meioses per kb (24)]. However, we have observed no evidence for grossly different rates of recombination throughout the entire α-globin complex and, in particular, markers either side of the IZHVR exist in linkage disequilibrium.

Since no other highly polymorphic markers are available at the 3′ end of the α complex or beyond the 3′ HVR it is not possible to analyze recombination in this region. However, there is considerably more variation in the 3′ HVR than the IZHVR and, at the highest level of resolution, many individuals homozygous for *Ia*, *IIa*, or *IIIa* haplotypes are nevertheless heterozygous at the 3′ HVR locus. This supports the concept that the rate of change of this region is different from the adjacent haplotype and may be measurable (for example, see ref. 24). If this were so, the combined analysis of the haplotype and the adjacent 3′ HVR could provide an interesting way of analyzing recent evolutionary events, such as the human racial divergence or the evolution of particular variants within the linked haplotype. In this respect the restricted 3′ HVR distribution in *Ia/Ia* homozygotes is particularly interesting. If we could predict the rate of change within a 3′ HVR of this particular length it might be possible to estimate the time of origin of the common *Ia* haplotype. Similarly, we have noted variants in the 3′ HVR attached to a common α-globin mutant found in Southeast Asia (25). Again, an estimate of the time of origin of this mutation could be made if we could predict the rate of change within its linked 3′ HVR.

Haplotypes similar to the α-globin haplotype described here have been used to study the evolution, origins, and molecular mechanisms of diseases known to be caused by mutations within the functional genes of such haplotypes (reviewed in ref. 16). This information has provided the basis for genotype prediction and developing programs to enable appropriate antenatal diagnosis. Defects of the human α-globin genes (α-thalassemia) are extremely common throughout the world and in some countries cause a considerable degree of morbidity and mortality (reviewed in ref. 1). Based on previous strategies used to analyze common mutants of the β-globin cluster (β-thalassemia and sickle-cell disease), the α haplotype described here will be of value in elucidating the molecular basis and geographical distribution of the common forms of α-thalassemia.

1. Higgs, D. R. & Weatherall, D. J. (1983) in *Current Topics in Hematology*, eds. Piomelli, S. & Yachnin, S. (Liss, New York), Vol. 4, pp. 37–97.
2. Hill, A. V. S., Nicholls, R. D., Thein, S. L. & Higgs, D. R. (1985) *Cell* **42**, 809–819.
3. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5022–5026.
4. Goodbourn, S. E. Y., Higgs, D. R., Clegg, J. B. & Weatherall, D. J. (1984) *Mol. Biol. Med.* **2**, 223–238.
5. Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. & Weatherall, D. J. (1981) *Nucleic Acids Res.* **9**, 4213–4224.
6. Hill, A. V. S., Bowden, D. K., Trent, R. J., Higgs, D. R., Oppenheimer, S. J., Thein, S. L., Mickleson, K. N. P., Weatherall, D. J. & Clegg, J. B. (1985) *Am. J. Hum. Genet.* **37**, 571–580.
7. Goodbourn, S. E. Y. (1983) Dissertation (University of Oxford).
8. Lauer, J., Shen, C.-K. J. & Maniatis, T. M. (1980) *Cell* **20**, 119–130.
9. Proudfoot, N. J., Gil, A. & Maniatis, T. M. (1982) *Cell* **31**, 553–563.
10. Nicholls, R. D., Hill, A. V. S., Clegg, J. B. & Higgs, D. R. (1985) *Nucleic Acids Res.* **21**, 7569–7578.
11. Felice, A. E., Cleek, M. P., McKie, K. & Huisman, T. H. J. (1984) *Blood* **63**, 1253–1257.
12. Higgs, D. R., Hill, A. V. S., Bowden, D. K., Weatherall, D. J. & Clegg, J. B. (1984) *Nucleic Acids Res.* **12**, 6965–6977.
13. Wainscoat, J. S., Higgs, D. R., Kanavakis, E., Cao, A., Georgiou, D., Clegg, J. B. & Weatherall, D. J. (1983) *Am. J. Hum. Genet.* **35**, 1086–1089.
14. Jeffreys, A. J. (1979) *Cell* **18**, 1–10.
15. Winichagoon, P., Higgs, D. R., Goodbourn, S. E. Y., Lamb, J., Clegg, J. B. & Weatherall, D. J. (1982) *Nucleic Acids Res.* **10**, 5853–5868.
16. Antonarakis, S. E., Kazazian, H. H., Jr., & Orkin, S. H. (1985) *Hum. Genet.* **69**, 1–14.
17. Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977) *Cell* **12**, 205–218.
18. Saadallah, N. & Hulten, M. (1983) *J. Med. Genet.* **20**, 290–299.
19. Reeders, S. T., Breuning, M. M. H., Davies, K. E., Nicholls, R. D., Jarman, A. P., Higgs, D. R., Pearson, P. L. & Weatherall, D. J. (1985) *Nature (London)* **317**, 542–544.
20. Reeders, S. T., Breuning, M. H., Corney, G., Jeremiah, S. J., Meera Khan, P., Davies, K. E., Hopkinson, D. A., Pearson, P. L. & Weatherall, D. J. (1986) *Br. Med. J.* **292**, 851–853.
21. Murray, J. C., Mills, K. A., Demopulos, C. M., Hornung, S. & Motulsky, A. G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3486–3490.
22. Chakravarti, A., Phillips, J. A., III, Mellits, K. H., Buetow, K. H. & Seeburg, P. H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6085–6089.
23. 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.
24. Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985) *Nature (London)* **314**, 67–73.
25. Winichagoon, P., Higgs, D. R., Goodbourn, S. E. Y., Clegg, J. B., Weatherall, D. J. & Wasi, P. (1984) *EMBO J.* **3**, 1813–1818.