Homology Requirements for Unequal Crossing Over in Humans

Aïda B. Metzenberg,*[,1] Gina Wurzer,*[,2] T. H. J. Huisman[†] and Oliver Smithies*[,3]

**Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, and [†]Department of Cell and Molecular Biology, Medical College of Georgia, Augusta, Georgia 30912*

Manuscript received October 16, 1990
Accepted January 14, 1991

ABSTRACT

To gain insight into mechanisms of unequal homologous recombination *in vivo*, genes generated by homologous unequal crossovers in the human β-globin gene cluster were examined by nucleotide sequencing and hybridization experiments. The naturally occurring genes studied included one δ-β Lepore-Baltimore fusion gene, one δ-β Lepore-Hollandia fusion gene, 12 δ-β Lepore-Boston genes, one $^{A}\gamma$-β fusion Kenya gene, one $^{A}\gamma$-$^{G}\gamma$ fusion (the central gene of a triplication) and one $^{G}\gamma$-$^{A}\gamma$ fusion. A comparison of the nucleotide sequences of three Lepore-Boston genes indicates that they were derived from at least two independent homologous but unequal crossover events, although the crossovers occurred within the same 58-bp region. Nine additional Lepore-Boston genes from individuals of various ethnic origins were shown, by hybridization to specific oligonucleotide probes, to have been generated by a crossover in the same region as the sequenced genes. Evidence for gene conversion accompanying a homologous unequal crossover event was found in only one case (although some of the single nucleotide differences observed in other genes in this study may be related to the crossover events in ways that we do not presently understand). Thus, as judged by this limited sample, concurrent gene conversions are not commonly associated with homologous but unequal exchange in humans *in vivo*. Classification of the recombinant chromosomes by their polymorphic restriction sites in the β-globin gene cluster indicated that the Lepore-Boston genes are found in at least six different haplotype backgrounds. Therefore the total number of independent examples in this study is at least 6, and at most 12. We have shown that in at least six cases of genes that have arisen by homologous but unequal crossing over *in vivo*, each event occurred in a relatively extensive region of uninterrupted identity between the parental genes. This preference cannot be explained by a mechanism whereby crossovers occur at random within misaligned related but not identical genes. In general, crossovers occur in regions that are among the largest available stretches of identity for a particular pair of mismatched genes. Our data are in agreement with those of other types of studies of homologous recombination, and support the idea that sequence *identity*, rather than general homology, is a critical factor in homologous recombination.

HOMOLOGOUS but unequal crossing over: Unequal crossing over is a genetic phenomenon that changes the number of similar genes or sequences on a chromosome. It was first described at the protein level in the haptoglobin system by SMITHIES *et al.* (1962), and in the globin system by BAGLIONI (1962, 1965). Studies of unequal crossing over at the level achievable by restriction fragment analyses have previously been conducted in a variety of multigene families, including the globin and haptoglobin genes, color vision genes, immunoglobulin genes, apolipoprotein B receptor genes, and major histocompatibility genes. Similar studies of genes with repetitive sequence compositions such as the zein genes (PEDERSON *et al.* 1982), the proline-rich salivary protein genes (AZEN *et al.* 1984; ZIEMER *et al.* 1984) and fibroin genes (MANNING and GAGE 1980) have also been conducted. However, until very recently, few structural studies at the nucleotide sequence level of the products of unequal crossing over events had been presented. The purpose of this study was to search for independent examples of genes resulting from homologous but unequal crossing over in the human β-globin gene cluster, and to characterize these rearrangements at the nucleotide sequence level, in the hope of deriving rules that would describe the molecular details of this class of recombinational events.

Hemoglobin Lepore (δ-β fusions): Hemoglobin (Hb) Lepore, first described by GERALD and DIAMOND in 1958, contains a β-type chain that results from homologous but unequal crossing over between the δ and β genes (BAGLIONI 1962, 1965), as shown in Figure 1A. Nine potential forms of Hb Lepore formed by single crossovers between the δ and β genes can in principle be distinguished by protein sequence determination, but only four would result in hybrid protein

[1] Current address: Howard Hughes Medical Institute, University of California San Francisco, San Francisco California 94143.
[2] Current address: The Upjohn Company, Kalamazoo Michigan 49001.
[3] Current address: Department of Pathology, University of North Carolina-Chapel Hill, Chapel Hill North Carolina 27599.

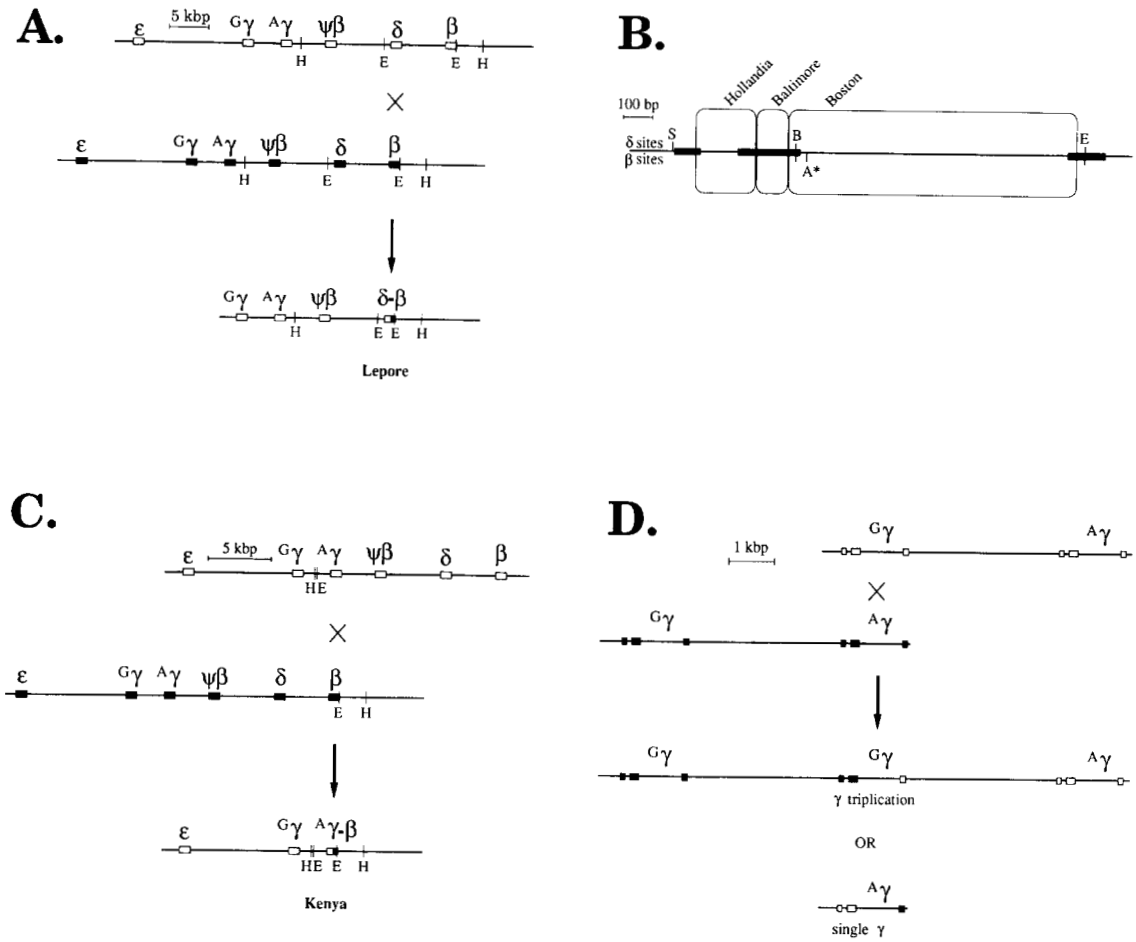


FIGURE 1.—A, Homologous unequal crossing over between δ- and β-globin genes of the type that gave rise to Lepore globin genes. The globin genes of one chromosome participating in the crossover are marked with open boxes, and the genes of the other are distinguished by filled boxes. The fusion Lepore gene is indicated by a box that is partly filled. Also shown are the positions of the recognition sites for *Hin*dIII and *Eco*RI (H and E, respectively), used for mapping and cloning the Lepore genes. B, Zones of crossover within misaligned δ and β globin genes that can generate and restriction sites for Lepore-Hollandia, Lepore-Baltimore, and Lepore-Boston genes. Black boxes represent exons. Restriction sites above the line represent sites in the normal δ sequence, and sites below the line are in β. Potential crossover zones that can generate the three Lepore types are enclosed in oval boxes. S, B, E and A represent sites for the enzymes *Sal*I, *Bam*HI, *Eco*RI and *Ava*II, with a polymorphic *Ava*II site marked with an asterisk. C, Homologous unequal crossing over between $^{A}\gamma$- and β-globin genes of the type that gave rise to a Kenya globin gene. Markings as for part A. D, Homologous unequal crossing over between $^{G}\gamma$- and $^{A}\gamma$-globin genes of the type that gave rise to γ and γγγ rearrangements. Markings as for part A.

molecules readily identifiable by gel electrophoresis (SMITHIES 1964). Three of these four have been described. The first and most common is Hb Lepore-Boston (BAGLIONI 1962, 1965) with a crossover, shown by peptide mapping, between codon 87 of the δ gene, δ87, and codon 116 of the β gene, β116. Hb Lepore-Baltimore (OSTERTAG and SMITH 1969), first described in an Afro-American family, is characterized by a crossover between δ50 and β86. Hb Lepore-Hollandia (BARNABAS and MULLER 1962), first described in a family in Papua, New Guinea, has a crossover between δ22 and β50. Although any δ-β fusion gene should be readily identifiable by modern genomic mapping techniques, no new forms of the Lepore globin have been reported to date. The potential crossover regions of the three Lepore types described as a consequence of amino acid sequence studies are shown in Figure 1B.

Hemoglobin Kenya genes ($^{A}\gamma$-β fusions): Hb Kenya, first described by HUISMAN *et al.* in 1972, contains a fusion chain with a γ chain amino terminus and a β chain carboxy terminus (HUISMAN *et al.* 1972). The gene for the fusion chain was proposed to have arisen via an unequal crossover between an $^{A}\gamma$ gene and a β gene in the β-globin cluster, as shown in Figure 1C. Protein studies of the Kenya globin chain established that the crossover was between codons 80 and 87 (HUISMAN *et al.* 1972; KENDALL *et al.* 1973; SMITH *et al.* 1973; NUTE *et al.* 1976), in Exon 2.

Single and triple γ-containing chromosomes ($^{G}\gamma$-$^{A}\gamma$ and $^{A}\gamma$-$^{G}\gamma$ fusions): The two fetal globin genes, $^{G}\gamma$ and $^{A}\gamma$, are the result of a 4.9-kbp tandem duplication (SHEN, SLIGHTOM and SMITHIES *et al.* 1981). The $^{G}\gamma$ and $^{A}\gamma$ gene products are identical at 145 of their 146 amino acid residues. The only difference is at position 136 where there is a glycine residue in $^{G}\gamma$ and an

TABLE 1

Plasmid probes of the β-globin cluster

Probe name	Length (kbp)	Restriction sites	Contents
ε 1.2BR	1.2	*Bam*HI-*Eco*RI	Part of ε and its 3′ flanking region
γ IVS2	0.46	*Bam*HI-*Pvu*II	5′ half of IVS2 of $^{A}\gamma$
γδ1.6BX	1.6	*Bam*HI-*Xba*I	ψβ
5′δ	0.35	*Alu*I-*Rsa*I	Between bipolar Alu repeats 5′ to δ
β IVS 2	1.0	*Bam*HI-*Eco*RI	β IVS 2
3′β	1.2	*Sal*I-*Eco*RI	Sequence 16 kbp 3′ to β

alanine residue in $^{A}\gamma$ (SCHROEDER *et al.* 1968), the result of the single nucleotide difference in the coding regions (SLIGHTOM, BLECHL and SMITHIES 1980). The original duplication event that produced the tandem arrangement of γ genes has been followed, in the chromosome for which the nucleotide sequence has been published, by a relatively recent (one million years past) and extensive gene conversion event. The result is that, on chromosomes carrying the gene conversion, the two tandem γ genes are nearly identical for a stretch of over 1 kbp, including the 5′ two-thirds of the genes (POWERS and SMITHIES 1986; STOECKERT, COLLINS and WEISSMAN 1983).

The first report of a triplication in place of the usual arrangement of duplicated γ-globin genes in humans came from TRENT *et al.* (1981) who observed an additional $^{G}\gamma$-like gene in a healthy Vanuatan newborn. Their data indicate that the triplication is the consequence of unequal but homologous crossing over between $^{G}\gamma$ and $^{A}\gamma$ genes (TRENT *et al.* 1981; SHIMIZU *et al.* 1986), as shown in Figure 1D.

A chromosome containing a single γ-globin gene, the predicted reciprocal product of unequal but homologous crossing over between $^{G}\gamma$ and $^{A}\gamma$ genes, was first reported by SUKUMARAN *et al.* (1983), in a healthy newborn from India, who was noted to have a low $^{G}\gamma/^{A}\gamma$ ratio in routine neonatal screening tests. Their data indicate that the deletion was the result of unequal crossover between the tandem normal genes, as shown in Figure 1D.

MATERIALS AND METHODS

Probes and primers: Six plasmid-derived probes spanning the β-globin gene cluster were used, as shown in Table 1.

Twelve oligonucleotides with sequences from the β-globin gene cluster were used for polymerase chain reactions, differential hybridization experiments, and/or as primers for double-stranded plasmid dideoxy sequencing. The sequences of these oligomers are shown in Table 2, along with a tabulation of their locations. In addition, the so-called M13 Universal and Reverse oligonucleotide primers were used for single- and double-stranded dideoxy sequencing (MESSING 1983). All oligonucleotides were synthesized by the Protein and DNA Synthesis Facility at the Biotechnology Center (University of Wisconsin-Madison).

Genomic DNA samples: Samples were obtained as blood or as lymphoblastoid cell lines, from various sources as shown in Table 3.

Polymerase chain reactions: Polymerase chain reactions (PCRs) were performed on genomic DNAs (100 ng/reaction) and cloned DNAs (25 pg/reaction) as described by KIM and SMITHIES (1988). The reaction temperature for all reactions was 60°, for 5–10 min depending on the length of DNA to be amplified. The melting conditions were 90° for 15 sec, with annealing for 15 sec as the temperature fell from 90 to 60°.

Differential hybridizations: Differential hybridizations were performed as described by WOOD *et al.* (1985) upon Southern blots of cloned or amplified DNAs following electrophoresis in duplicate (after GITSCHIER *et al.* 1985) in 2% agarose. Oligomers used as probes for differential hybridization were first purified by electrophoresis.

DNA sequencing: DNA sequencing was performed either by the method of MAXAM and GILBERT (1980) with modifications as described by SLIGHTOM, BLECHL and SMITHIES (1980) or by the dideoxy DNA sequencing method of SANGER, NICKLEN and COULSON (1977), or using Sequenase (U.S. Biochemical).

DNA sequence analysis: The DNA sequences were analyzed using software provided by the University of Wisconsin Genetics Computer Group (DEVEREUX, HAEBERLI and SMITHIES 1984).

RESULTS

Lepore gene structure: The normal δ and β genes are located approximately 7.7 kbp apart on chromosome *11*. Six of our examples of Lepore-Boston DNAs were tested by genomic mapping for the presence of the 7.7-kbp deletion in the β-globin gene cluster expected for the Lepore rearrangement. All had the expected deletion (data not shown).

Polymorphic restriction site haplotyping of Lepore-Boston genes: In order to collect evidence for or against the independent origin of 12 Lepore-Boston genes in our collection, polymorphic restriction site haplotypes (ORKIN *et al.* 1982) were determined by genomic mapping. The positions of the nine informative restriction site polymorphisms that we used are shown in Figure 2A. The 5′ cluster of polymorphic restriction sites (region I in Figure 2A) is 34.6 kbp in length, and begins upstream of the ε gene. The 3′ segment (region II in Figure 2A) is 19.4 kbp in length, and begins at the β globin gene. The two clusters, represented by sites 1 through 4 and 6

TABLE 2

Globin oligonucleotide probes and primers

Name	Strand	Location	Sequence
A	Antisense	β exon 2-IVS 2 border (11907-11887)	AAGGGTCCCATAGACTCACCC
B	Antisense	Same, with one δ-like base (11907-11887)	AAGGGTCCCATGGACTCACCC
C	Antisense	δ exon 2 (11846-11826)	CTCACTCAGTGTGGCAAAGGT
E	Antisense	β IVS 2 (12004-11985)	ATCATTCGTCTGTTTCCCAT
F	Sense	δ IVS 1 (4256-4275)	CCTACCCTCAGATTACTGGT
G	Antisense	3′ to β (12887-12868)	TGGACAGCAAGAAAGCGAGC
H	Sense	5′ to δ (3553-3568)	**GAATTC**GTCCATTCCAACCTCTC
I	Sense	5′ to δ (3101-3120)	TATTTGTCTGCCATTGTGGC
J	Antisense	β exon 3 (12780-12766)	**GAATTC**AAAGTGATGGGCCAG
Pop2	Antisense	β IVS 2 (12637-12617)	GGTAGCTGGATTGTAGCTGCT
γ240	Sense	5′ to Aγ[6349-6365]	CAGTCATGATTTATTTC
γ1020R	Antisense	5′ to Aγ[7130-7113]	CTGTGAAATGACCCATG

The location of each sequence is shown with the coordinates on the sequence of Poncz *et al.* (1983) in parentheses (for the δ and β sequences) and coordinates on the sequence of Shen *et al.* (1981) (for the γ sequences). (The positions of Oligos A, B and C are also shown in Figure 3.) Non-globin sequences are shown in boldface, and the single internal base at which Oligo B differs from Oligo A is underlined.

TABLE 3

Sources of DNA samples

Sample	Origin	Source
Alg.	Algeria	MORLE *et al.* (1984)
An.A.	Italy	T. H. J. HUISMAN (unpublished)
D.R.	Yugoslavia	G. D. EFREMOV (unpublished)
G.W.	Africa-United States	B. SHARON and E. SCHWARTZ (unpublished)
L.B.	Italy	T. H. J. HUISMAN (unpublished)
L.P.	Italy	MEARS *et al.* (1978)
M.K.	Greece	H. KAZAZIAN (unpublished)
M.S.	Thailand	BOONTRAKOONPOONTAWEE *et al.* (1987)
P.W.	England	A. FINDLAY (unpublished)
R15	Yugoslavia	EFREMOV *et al.* (1976)
Sp14	Thailand	S. FUCHAROEN (unpublished)
Span.	Spain	GIMFERRER (1976)
Go. (Baltimore)	Spain	T. H. J. HUISMAN (unpublished)
Holl. (Hollandia)	Vanuatu-Canada	T. H. J. HUISMAN (unpublished)
F.M. (γ)	India	SUKUMARAN *et al.* (1983)
J3 (γγγ)	Vanuatu	TRENT *et al.* (1981)

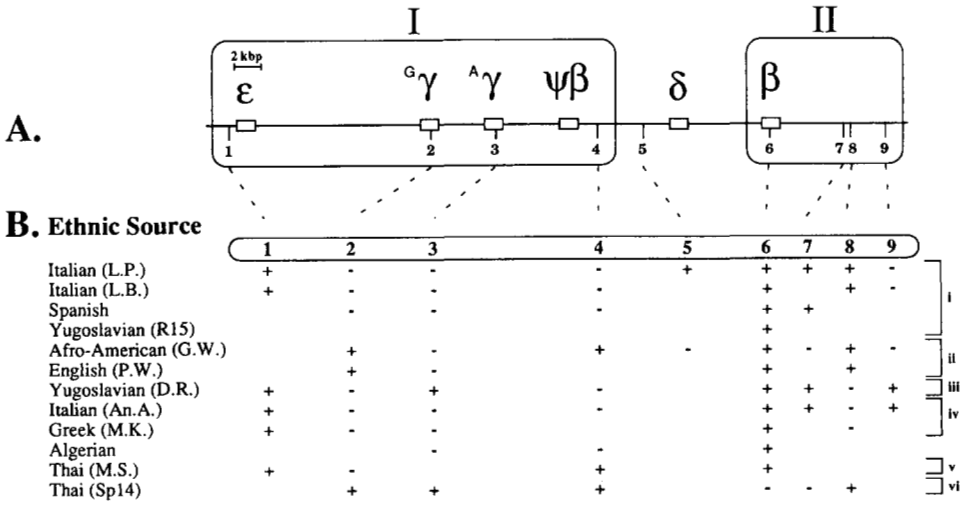


FIGURE 2.—A, Polymorphic restriction sites in the β-globin gene cluster. The arrangement of genes (open boxes) in the β-globin cluster is shown, with the locations of the polymorphic restriction sites studied numbered, as follows: 1, *Hinc*II; 2, *Hin*dIII; 3, *Hin*dIII; 4, *Hinc*II; 5, *Rsa*I; 6, *Ava*II; 7, *Hin*dIII; 8, *Bam*HI; 9, *Rsa*I. The clusters, I and II, of polymorphic restriction sites, are indicated. B, Polymorphic restriction haplotype data of chromosomes of different ethnic origin carrying Lepore-Boston genes. The presence of a site is marked with a +, an absence of a site is marked with a −, and noninformative sites are unmarked.

through 9, as shown in Figure 2A, are separated by 9.1 kbp of DNA, represented by site 5. This 9.1-kbp region is a hotspot for germline recombination, with a crossover frequency that is 3–30 times greater than expected for its physical length (CHAKRAVARTI *et al.* 1984). Since a Lepore gene is a fusion of a δ gene (located within the hotspot region) and a β gene (located in the 3′ cluster), haplotypes in the 3′ region are more likely to be in linkage disequilibrium with the fusion genes than those in the 5′ region, and are therefore of special interest in assessing possible independent origin of genes. In addition, since the crossover region in three Italian cases of Lepore homozygosity had been localized by genomic mapping of the β-specific polymorphic *Ava*II site in IVS 2 (BAIRD *et al.* 1981; CHEBLOUNE and VERDIER 1983; and MAVILIO *et al.* 1983), particular attention was paid to this polymorphism in all Lepore-Boston cases that were collected.

Most of the individuals in this study are Lepore heterozygotes, and DNAs from family members were not available. Consequently, assignment of a polymorphic site on a Lepore chromosome was only possible if the relevant individual was homozygous for the presence or absence of that particular site, or if nucleotide sequence analysis of the position was performed on cloned DNA from the Lepore chromosome.

The results of haplotyping experiments are shown in Figure 2B. The presence of a particular site is indicated with a +, and its absence with a − sign. Three categories of Lepore chromosome were fully informative for the four 3′-most sites (sites 6, 7, 8, and 9 in Figure 2) in cluster II. The first category (category i in Figure 2B) is represented by the Italian L.P. homozygote, who has a +++− haplotype. A second haplotype, +−+− (category ii in Figure 2B), is found in the Afro-American heterozygote, G.W. A third and a fourth category (categories iii and iv in Figure 2B), are both ++−+ for the four 3′-most sites, but differ in their 5′ haplotypes (+−+− *vs.* +−−−); they are represented by the the Yugoslavian D.R. sample and the Italian An.A. sample, respectively. One Thai Lepore chromosome, M.S., had a distinctive haplotype in Cluster I (+−?+). This sample is therefore in a fifth category (category v in Figure 2B). Finally, the Thai sample Sp14 is unique in both cluster I and cluster II, and is therefore in a sixth category (category vi in Figure 2B). Thus, these Lepore-Boston genes were present on six distinguishable haplotype backgrounds, which provided us with assurance that further study of representative samples from each category was warranted.

Lepores L.P., G.W., and M.K. from the first, second, and fourth categories, respectively, were chosen for further study via molecular cloning. The Lepore D.R. gene (third category) was previously examined by nucleotide sequencing because we had initially thought that it was a Lepore-Baltimore type gene. The sequence analysis showed that D.R. was, in fact, of the Boston type, and we have subsequently used it as the reference Lepore-Boston chromosome.

Sequencing of the reference Lepore-Boston gene, D.R.: The nucleotide sequence of the D.R. Lepore-Boston reference fusion gene was determined from the *Sal*I site just 5′ to the gene (position 1 in Figure 3A) to the *Eco*RI site in Exon 3. Comparison of this sequence to those of normal δ and β globin genes shows (Figure 3A) that the crossover zone for the D.R. Lepore-Boston gene is within a 58-bp region of identity at the border of Exon 2 and IVS 2.

Differential oligonucleotide hybridization: Hybridization experiments were carried out in order to compare other Lepore-Boston genes to the reference gene, D.R. The 58-bp region of identity between the δ and β genes, which is the crossover region in the reference Lepore-Boston gene, begins 4 bp 5′ to the polymorphic base pair that constitutes the *Ava*II polymorphism. We therefore asked whether any of the remaining Lepore-Boston genes were different from D.R. with respect to the presence of the *Ava*II polymorphism, and whether there were any other indications of a different crossover region from that of the reference gene. The differential hybridization method relies on stringent washing conditions, such that a one base mismatch between the probe and sequence of interest causes a significant reduction in the level of hybridization.

In order to determine whether the remaining Lepore-Boston genes were identical to that of D.R. (which is δ-like 5′ to the 58 bp region and β-like 3′), two 21-base oligomers were synthesized, differing at a single internal position. Probe A has the same sequence as the D.R. gene in the region spanning the 3′ border of the 58 bp region of δ-β identity at the Exon 2-IVS 2 junction, and probe B has a single base difference at the point of divergence between δ and β in IVS 2 (see Table 2). Probe A was used to look for identity with D.R., or lack thereof. Probe B was used as a control for the hybridization and washing conditions, to show that each set of tests was able to detect a single base mismatch. An additional 21-base oligomer, probe C, was synthesized, which has the sequence of a normal β gene at the 5′ border of the 58-bp region of identity. Lepore-Boston genes do not hybridize to this probe, and thus hybridization of a particular Lepore gene to probe C would indicate that it was not a Boston-type gene.

Figure 4 shows that, as expected, under the stringent conditions used, the cloned DNA from the reference Lepore-Boston gene (D.R.) hybridizes only to probe A. Three additional cloned Lepore-Boston

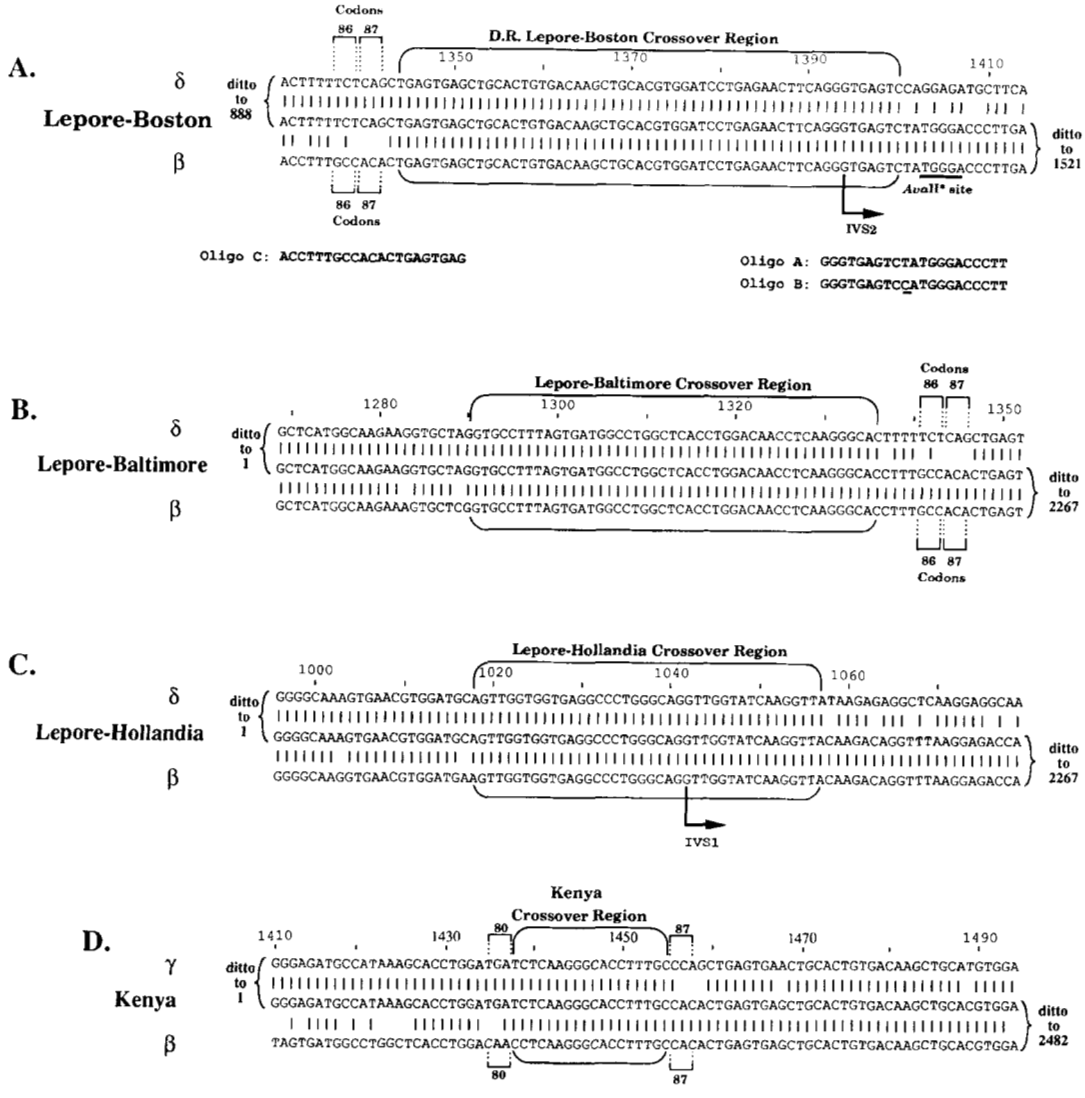


FIGURE 3.—Sequences of the crossover region of three Lepore-Boston genes and one Kenya gene. A, The sequence of the normal δ gene in the relevant region is shown on the upper sequence line aligned with the Lepore-Boston (D.R.) sequence, on the middle line, and the normal β sequence, on the lower line. Vertical lines indicate matched nucleotides between the sequences. The numbering system begins with the *Eco*RI site 5′ to the δ gene. The brackets labelled "ditto to 888" and "ditto to 1521" show that the D.R. sequence is identical to the δ and β sequences, respectively, over the indicated nucleotides. Codons 86 and 87, which encode divergent amino acids in the δ and β genes, are indicated. The crossover region is marked with a horizontal curved bracket. The beginning of IVS 2 is indicated with an arrow. The polymorphic *Ava*II site in IVS 2 of the β gene is underlined and marked with an asterisk. The sequences of 21-mer oligonucleotide probes used for differential hybridization experiments, Oligos A, B and C, are indicated in bold face under the sequences of the gene regions, in the appropriate positions. The single internal base mismatch between Oligos A and B is underlined. B, Lepore-Baltimore. As for part A. The brackets labeled "ditto to 1" and "ditto to 2267" are to show that the Lepore-Baltimore sequence is identical to the δ and β sequences, respectively, over the indicated nucleotides. The crossover region is marked with a horizontal curved bracket. C, Lepore-Hollandia. As for part B, except that the beginning of IVS 1 is indicated with an arrow. D, Kenya gene. The sequence of the crossover region for the Kenya gene, on the middle line, is aligned with the sequence of a normal $^{A}\gamma$ gene, on the upper line, and the normal β gene, on the lower line. The numbering system begins with the *Hin*fI site 5′ to the $^{A}\gamma$ gene. The positions of codons 81 and 86, which encode divergent amino acids in the normal $^{A}\gamma$ and β genes, are indicated by brackets. The beginning of IVS 1 is indicated with an arrow.

genes (M.K., G.W. and L.P.) also hybridized to probe A only. DNA from a Lepore-Hollandia gene (Holl.) hybridized to probe C (Figure 4). This same Lepore-Hollandia DNA did not hybridize to probe A, although it is β-like in this region, because a single noncentral base mismatch (at the *Ava*II site) prevents hybridization. Cloned DNA from two normal β genes (with intact *Ava*II sites) hybridized to both probes A and C, as expected.

Six additional genomic Lepore-Boston DNA samples, one Algerian (Alg.), two Italian (An.A. and L.B.), one English (P.W.) one Spanish (Span.), and one Thai, M.S.) were amplified by PCR using oligos F and G (Table 2). This set of primers will amplify a 1223 bp fragment that extends from within Exon 1 of the δ-derived portion of the Lepore-Boston genes to IVS 2 in the β-derived portion. The amplified DNAs were tested for hybridization to the three probes A, B and

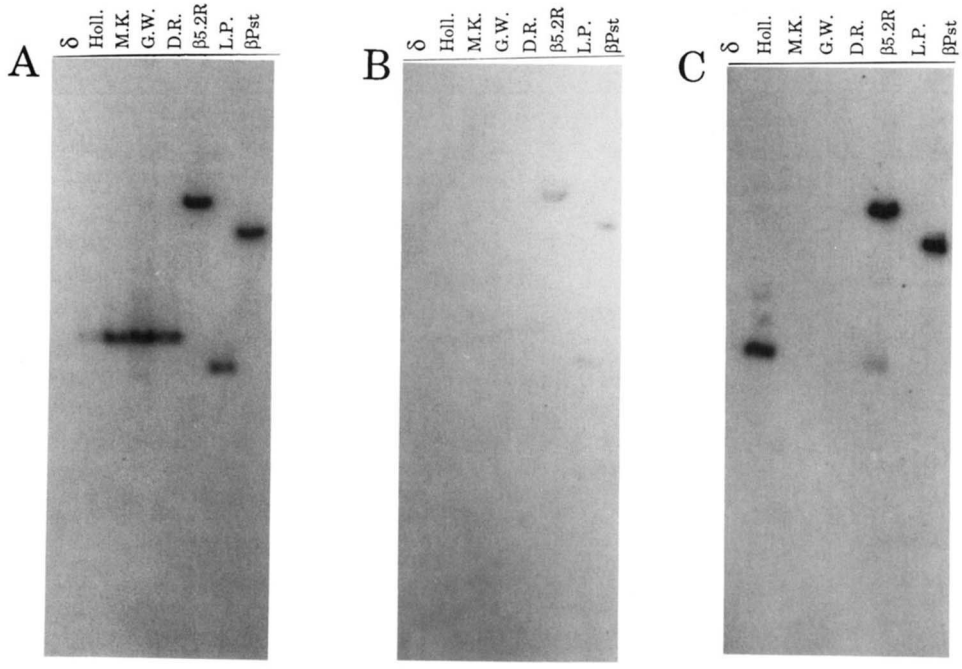


FIGURE 4.—Autoradiographs of cloned Lepore genes hybridized to Oligos A, B and C, as indicated. Samples from left to right are: cloned DNA from a normal δ gene, a Lepore-Hollandia gene, the Greek (M.K.), Afro-American (G.W.) and Yugoslavian (D.R.) Lepore-Boston samples, a normal β gene, the Italian (L.P.) Lepore-Boston gene, and another normal β gene.

C. These six samples hybridized to probe A only, as shown in Figure 5. One additional amplified genomic DNA sample, Sp14 (Thai), failed to hybridize to probe A, as shown in panel C.

These differential hybridization results indicate that nine of the samples tested, L.P., L.B., Span., G.W., P.W., An.A., M.K., Alg. and M.S. (representing categories i, ii, iii, iv and v in Figure 2) have crossover regions identical to that of the reference Lepore-Boston gene, D.R. No further study of these genes was undertaken. A tenth (Thai) sample, Sp14, failed to hybridize to probe A, indicating that it was different from the others in some way, and for that reason was studied further by nucleotide sequence analysis. In addition, an eleventh sample, R15, for which we had little DNA for other studies, was cloned and studied by nucleotide sequencing.

Nucleotide sequence analysis of Lepore-Boston genes Sp14 and R15: The 2.3-kbp *Eco*RI fragment containing about 1 kbp of 5′ flanking region, Exon 1, IVS 1, Exon 2, IVS 2 and part of Exon 3 of the Lepore genes Sp14 and R15 were cloned into the λ vector gt10 and sequenced.

The sequences of the Sp14 and R15 Lepore-Boston genes were compared to each other, to the reference Lepore-Boston gene, D.R., and to published sequences of normal δ- and β-globin genes (PONCZ *et al.* 1983). The crossover region in the Sp14 and R15 Lepore-Boston genes is the same as in D.R., having occurred at the 58-bp region of identity between the parental δ and β genes at the border of Exon 2 and IVS 2 (see Figure 3). Sp14, however, carries the same polymorphic base mismatch at the *Ava*II site in IVS 2 that prevented hybridization of the Hollandia gene to probe A. The crossovers were apparently clean breakage and reunion events, with no evidence for insertions, deletions, or nearby concurrent gene conversions. There were, however, several nucleotide substitutions in the sequences that require comment.

Table 4 compares the nucleotides found at 12 positions where differences were observed in the Lepore-Boston genes Sp14 and R15 compared to the published sequences of normal δ and β genes and with our data for Lepores Baltimore and Hollandia, and for the 3′ portion of the Kenya gene. Differences from the δ sequence are present in four locations in the regions 5′ to these genes, at positions 40, 76, 692 and 839. However, each of these differences is also found in at least one other δ-derived gene (Lepore-Baltimore or Lepore-Hollandia), and consequently they are likely to be normal and possibly polymorphic variations of the sequence. There are also four nucleotide positions 3′ to the crossover zones in Sp14 and R15 at which differences from the β gene are found. These are at positions 1409, 1467, 1682 and 2097. Three of these differences (at positions 1409, 1467 and 2097) have been previously reported in other individuals (ANTONARAKIS *et al.* 1982a), as part of various β gene "frameworks." However, the combination of these nucleotides in the Sp14 gene is unique. In addition, there is an extra nucleotide in the R15 gene not seen in other β and β-derived genes, at nucleotide position 1682 in Table 4.

Analysis of Lepore-Baltimore gene structure: The same 2.3-kbp *Eco*RI fragment that was studied in Sp14 and R15 was also isolated and sequenced from a Spanish Lepore-Baltimore heterozygote. Comparison of the Lepore-Baltimore gene sequence to the pub-

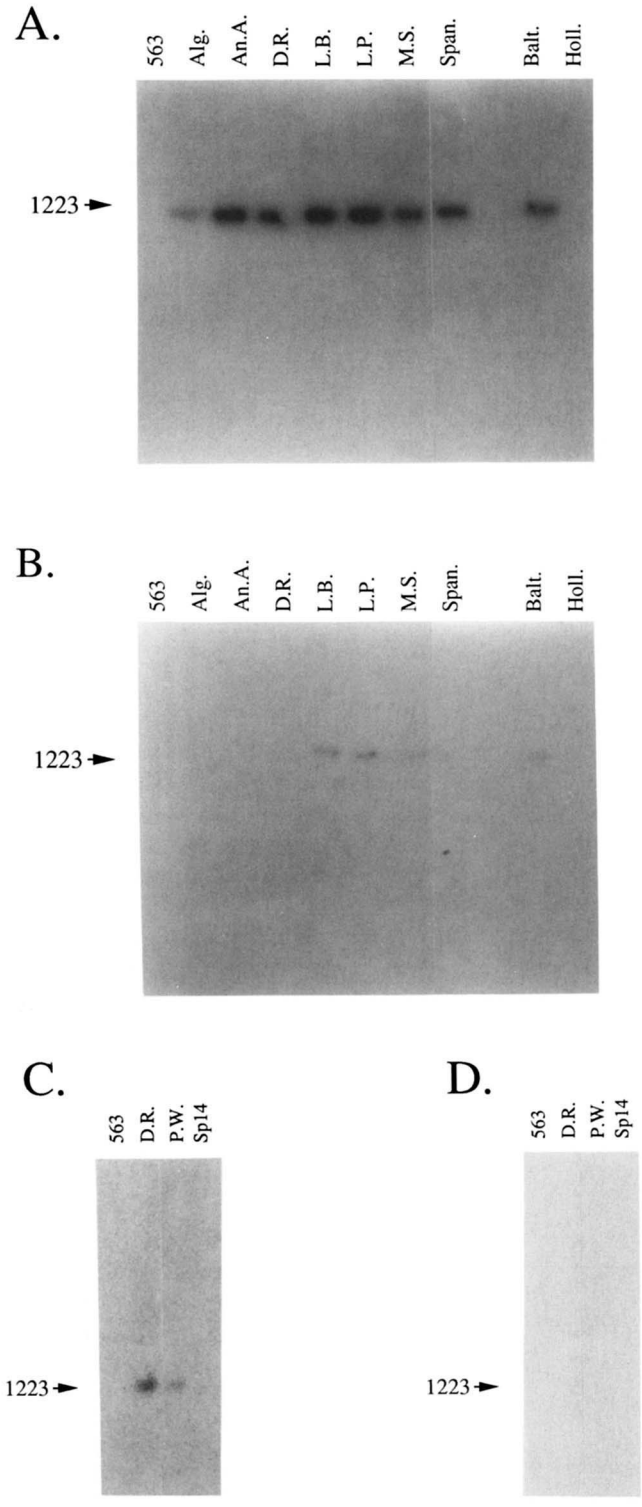


FIGURE 5.—Autoradiographs of Southern blots of gels containing electrophoresed 1223 bp amplified Lepore-Boston and Lepore-Baltimore fragments obtained using Oligos F and G as primers (see Table 2). Samples from left to right are: control DNA(563), the Algerian(Alg.), Italian (An.A.), Yugoslavian (D.R.), Italian (L.B. and L.P.), Thai (M.S.) and Spanish (Span.) Lepore-Boston samples, the Lepore-Baltimore (Balt.) and Lepore Hollandia (Holl.) samples, hybridized to Oligos A and B, respectively. The English (P.W.) and Thai (Sp14) Lepore-Boston samples were hybridized to oligos A and B in separate experiments (panels C and D, respectively).

TABLE 4

Nucleotide differences in δ and β genes

	5′						IVS2				
			44	4	6	8	1	1	1	1	1
	4	7	22	6	9	3	4	4	4	6	0
	0	6	89	6	2	9	0	6	7	8	9
							9	7	4	2	7
δ	*	C	**	*	C	A	C	T	C	T	T
D.R.							C	G			
Sp14	C	*	**	*	T	G	G	G	C	*	C
R15	C	C	**	*	C	A	C	T	C	T	T
Balt.	*	C	**	*	T	G	C	T	C	*	T
Holl.	C	*	AT	A	C	A	G	T	T	*	C
β	*	C	**	*	C	A	C	T	C	T	T
Kenya							C	T	C	*	T

Numbers in vertical orientation give nucleotide coordinates beginning at the *Eco*RI site 5′ to δ. Gaps, represented by asterisks, were introduced to maximize alignments of the sequences (METZENBERG 1989). Positions left blank represent those at which the DNA sequence was not determined. Normal δ and β sequences are from PONCZ *et al.* (1983).

lished sequences of normal δ and β genes (PONCZ *et al.* 1983) confirmed that this Lepore gene is the result of the expected fusion between δ and β. The crossover region for this type of Lepore gene lies within a region of 47 bp of identity between the parental δ and β genes, within Exon 2, as shown in Figure 3B. This region is located between codons 69 and 84. This result confirms and refines the conclusions of earlier peptide mapping studies which placed the Lepore-Baltimore crossover between codons 50 and 86 (OSTERTAG and SMITH 1969).

Comparison of the nucleotide sequences 5′ and 3′ to the Lepore-Baltimore crossover region with the corresponding sequences from normal δ and β genes revealed two sequence variations 5′ to the crossover, at positions 692 (692 bp 3′ to the start of the *Eco*RI recognition site upstream of δ) and 839, as shown in Table 4. These variations were also seen in one other δ-derived sequence in the study, Lepore-Boston Sp14. The sequence structure of the Lepore-Baltimore gene 3′ to the crossover (including position 1467) has been described previously as a β gene Framework 1 (ANTONARAKIS *et al.*, 1982a). Thus the observed nucleotide variations are likely (but not demonstrated) to be polymorphisms that have been maintained in normal δ and β genes.

Analysis of Lepore Hollandia gene structure: The nucleotide sequence of the 2.3-kbp *Eco*RI fragment from a Lepore-Hollandia gene was determined with cloned DNA from a Lepore heterozygote of Vanuatuan extraction. Comparison of the Lepore-Hollandia gene sequence to the published sequences of normal δ and β genes (PONCZ *et al.* 1983) confirmed that the gene resulted from a fusion between a δ gene and a β gene. The crossover region was a 40-bp region of identity between the parental δ and β genes, at the

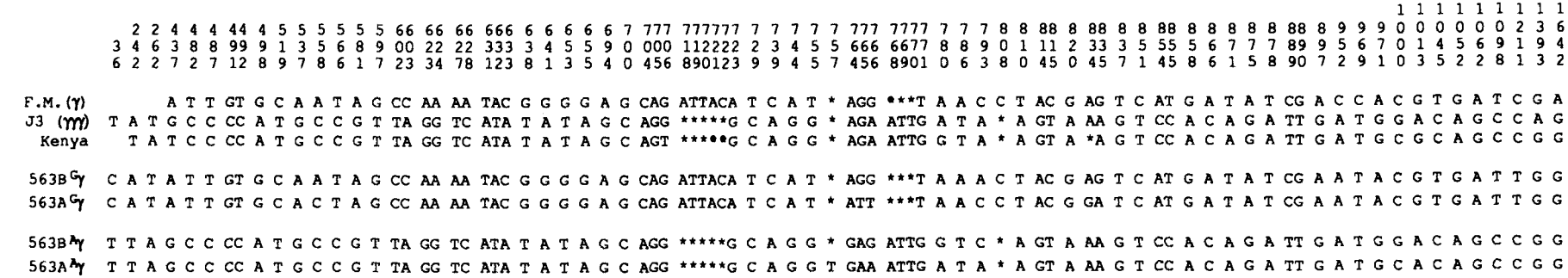


FIGURE 6.—Summary of variable sites 5′ to the single γ, central gene of the γγγ, and Kenya genes compared to other γ genes. Numbers in vertical orientation give nucleotide coordinates beginning at the *Ava*II site where sequencing commenced, approximately 1400 bp 5′ to the $^{A}\gamma$-$^{G}\gamma$ fusion gene (METZENBERG 1989).

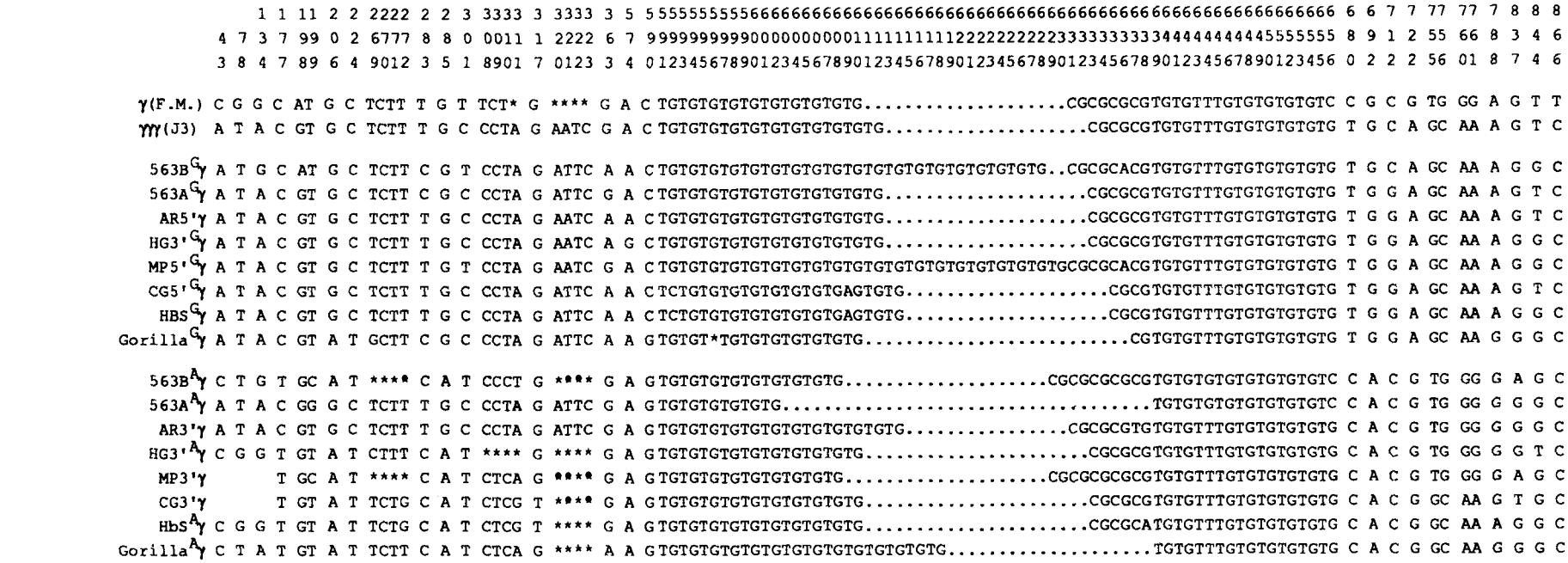


FIGURE 7.—Summary of variable sites in IVS 2 of the single γ gene and the central gene of the γγγ, compared to other γ genes. Numbers in vertical orientation give nucleotide coordinates corresponding to those of POWERS and SMITHIES (1986). Position 1 is in the *Bam*HI recognition sequence in Exon 2. Normal sequence data are from SLIGHTOM, BLECHL and SMITHIES (1980), STOECKERT *et al.* (1984), SCOTT *et al.* (1984) and POWERS and SMITHIES (1986).

border of Exon 1 and IVS 1, as shown in Figure 3C. This region begins in codon 25, and ends in the intron before Exon 2. This result confirms and refines the conclusions of previous peptide mapping studies which defined the Hollandia crossover region as the area between codons 22 and 50.

Comparison of the Lepore-Hollandia nucleotide sequences flanking the crossover region with the corresponding regions of normal δ and β genes revealed six notable features. As shown in Table 4, the Lepore-Hollandia sequence contains a C at position 40 that is absent in the normal δ gene, while at position 76, the Lepore-Hollandia sequence is missing a C. At positions 428 and 429, the Lepore-Hollandia sequence has an AT dinucleotide not found in the δ or β sequences, and at position 466, the Hollandia gene has an A that is absent in the δ sequence.

Most of the variations that we observed in the sequences 3′ to the crossover of the Lepore-Hollandia gene have been previously described, however there is a unique change at position 1474 in Table 4, where there is a T in the Hollandia sequence rather than the C found at this position in the normal δ and β genes.

The sequence differences found in this gene are not evidence for gene conversion accompanying the crossover, although it is possible that they represent errors of repair during the crossover.

Analysis of the Kenya gene structure: The Kenya chromosome from our propositus contains a deletion of approximately 22.5 kbp of DNA that includes the 3′ section of the $^{A}\gamma$ gene, the entire $\psi\beta$ gene, the δ gene, and the 5′ portion of the β gene (OJWANG *et al.* 1983). Our restriction mapping data (not shown) indicated that the *Xba*I site approximately 200 bp 3′ to the start of IVS 2 of $^{A}\gamma$ is absent, and placed the crossover 5′ to that region. The crossover itself occurred between codons 80 and 87, as mentioned in the Introduction.

The nucleotide sequence of a 2482-bp stretch of DNA was determined using cloned DNA from the Kenyan propositus, P.O. The sequence included 1050 bp of 5′ flanking sequence, Exon 1, IVS 1, Exon 2, IVS 2 and all but the final 25 codons of Exon 3. Comparison of this Hb Kenya gene sequence to the published sequences of normal $^{A}\gamma$- and β-globin genes (SLIGHTOM, BLECHL and SMITHIES 1980 and PONCZ *et al.* 1983) confirmed that this Kenya gene resulted from a fusion between an $^{A}\gamma$ gene and a β gene. The 5′-flanking region, as well as Exon 1, IVS 1 and part of Exon 2, were $^{A}\gamma$-like. The 3′ part of Exon 2, as well as IVS 2 and the portion of Exon 3 that was examined were all identical to the published sequence of β. The crossover region for this Kenya gene was an 18-bp region of identity between the parental $^{A}\gamma$ and β genes, as shown in Figure 3D. This is in agreement with the region predicted by the earlier peptide mapping studies (HUISMAN *et al.* 1972; KENDALL *et al.* 1973).

In Figure 6 a comparison is shown of nucleotides observed in the Kenya gene 5′ to the crossover with the nucleotides from four normal γ genes ($^{G}\gamma$ and $^{A}\gamma$) and from our $\gamma\gamma\gamma$ (the central gene of a triplication, J3) and γ (a single γ gene, F.M.) genes (see below). As shown in Figure 6, most of the differences between the Kenya sequence and the published $^{A}\gamma$ sequence 5′ to the crossover have been previously observed in other γ genes, and are probably common polymorphisms. However, an A is missing at position 834 in the Kenya sequence that is present in all previously reported γ genes. In the β-derived sequences 3′ to the crossover, the Kenya gene has the normal β gene Framework 2 (ANTONARAKIS *et al.* 1982a) configuration. It is not known whether the unique nucleotide described represents a conversion accompanying the unequal crossover event that produced the gene.

Analysis of the central gene of the Vanuatan γ triplication, J3: The three genes of the γ triplication were first isolated on a single 17.8-kbp *Bgl*II fragment using DNA from the λ vector Charon 35 with the bacterial host K802 recA^{-}. (The recA^{+} version of this strain allowed recombination in the clones and consistently yielded $\gamma\gamma$ rearrangements instead of the parental $\gamma\gamma\gamma$ molecules.) Restriction enzyme mapping of the cloned γ genes confirmed the $^{G}\gamma$ $^{G}\gamma$ $^{A}\gamma$ arrangement of the triplication described by TRENT *et al.* (1981).

Mapping of the cloned genes also showed that the 5′ γ gene carried the polymorphic *Xmn*I site (the "up" promoter) at position −158, 5′ to the cap site (GILMAN *et al.* 1985), but the central and 3′ genes did not. Therefore, this triplication was *Xmn*I +−−.

The three γ genes of the triplication were subcloned independently, and the sequence of a 2798-bp segment of DNA from the central γ gene of the triplication was determined. This includes 1431 bp of 5′ flanking sequence, Exon 1, IVS 1, Exon 2, IVS 2 and part of Exon 3.

Sequence differences between the central γ gene and published sequences for $^{G}\gamma$ and $^{A}\gamma$ genes are summarized in Figures 6 and 7. These results confirmed that the central γ gene was a fusion between $^{A}\gamma$-like (5′ end) and $^{G}\gamma$-like sequences (3′ end). The crossover that resulted in the $\gamma\gamma\gamma$ arrangement was within the 1.0-kbp region of identity between $^{G}\gamma$ and $^{A}\gamma$ between −281, 5′ to the promoter and IVS 2.

There were no nearby deletions or insertions observed; however there were several notable sequence features. The first, an apparent region of gene conversion between ancestral $^{G}\gamma$ and $^{A}\gamma$ genes, begins at position 242, and consisted of two $^{G}\gamma$-like base pairs, separated by 19 bp, four of which were $^{A}\gamma$-like in character, and surrounded by $^{A}\gamma$-like sequence. The

second region of apparent gene conversion was at position 757 in Figure 6 (−611 bp, 5′ to the cap site), and was identical to the one seen 5′ to the Hb Kenya gene. Finally, at position 1000, there was one additional unexpected base pair difference, where the central γ gene had a G in the coding strand instead of the C found in this position in the published $^{G}\gamma$ and $^{A}\gamma$ sequences. All of these differences have been observed in other γ genes, as shown in Figures 6 and 7, and are probably polymorphisms.

Analysis of a single γ gene: Previous genomic mapping studies of the single γ gene from individual F.M. from Karnatka State in India (SUKUMARAN *et al.* 1983) indicated that this gene resulted from a fusion between a $^{G}\gamma$ and an $^{A}\gamma$ gene, with the 5′ flanking restriction sites corresponding to published sequences for $^{G}\gamma$, and the 3′ sites corresponding to $^{A}\gamma$.

A 2370-bp segment of cloned DNA from F.M., including 938 bp of 5′ flanking sequence, Exon 1, IVS 1, Exon 2, IVS 2 and most of Exon 3 was subjected to nucleotide sequence analysis.

Tables 5 and 6 summarize sequence differences between the single γ gene and the published sequences of $^{G}\gamma$ and $^{A}\gamma$ genes (SLIGHTOM, BLECHL and SMITHIES 1980). The 5′-flanking sequence is almost entirely $^{G}\gamma$-like in character, and the third exon is $^{A}\gamma$-like. These results confirm that the gene resulted from a $^{G}\gamma$-$^{A}\gamma$ fusion. However, there are many differences from the published sequences in IVS 2 of this unusual gene, which will be discussed below.

Differences in the $^{G}\gamma$-derived sequences 5′ to the single γ gene are shown in Figure 6. Several nucleotide differences from the published sequences that have been found in other γ genes (SLIGHTOM, BLECHL and SMITHIES 1980; STOLLE *et al.* 1990; J. L. Slightom, personal communication) were also found in the $^{G}\gamma$-derived sequences 5′ to the single γ gene, at positions 558, 834 and 835 in Figure 6. There were also four sequence differences in the region 5′ to IVS 2 that have not been found in other γ genes, at positions 952, 969, 1211 and 1642. This latter difference produces a silent codon change (a CTA instead of the CTG seen in codon 32 in normal $^{G}\gamma$ and $^{A}\gamma$ genes).

Differences in the $^{A}\gamma$-derived sequences in IVS 2 of the single γ gene in comparison to the normal $^{A}\gamma$ gene are shown in Figure 7, and compared to published sequences from the $^{G}\gamma$ and $^{A}\gamma$ genes of humans and other primates. This comparison revealed a number of differences. Most of these differences, at positions 43, 78, 134, 188–189, 301, 311, 320–323, 611–612, 631–632, 680, 712, 722, 755–756 and 760–761, have been observed in other published or unpublished γ gene sequences, some in $^{G}\gamma$ genes, and others in $^{A}\gamma$ genes. Two of the sequence differences, at positions 308 and 866, have not been found previously in other γ genes.

The region 5′ to the crossover in this single γ gene is largely $^{G}\gamma$-like (except for some previously unreported changes). However, the sequence 3′ to the crossover, in IVS 2, has features characteristic of both $^{G}\gamma$ and γ genes, in an alternating fashion. A comparison of the sequence of IVS 2 of the F.M. single γ gene with sequences of human and other primate γ genes was revealing. The complex patterns in this gene differed from the "canonical" $^{G}\gamma$ and $^{A}\gamma$ sequences derived from human *and* gorilla γ gene sequences, particularly in the diagnostic simple sequence region in IVS 2 [the published chimpanzee and orangutan sequences are not useful for this derivation as they have each undergone extensive gene conversion in IVS 2 subsequent to the evolutionary divergence of the great apes (SLIGHTOM *et al.* 1985, 1987)].

Many of the unexpected sequence differences resemble polymorphisms found in other $^{A}\gamma$ genes [for example, nucleotides found in IVS 2 positions 43, 78, and 134 of F.M. (Figure 7) are like those found in an $^{A}\gamma$ gene reported by STOECKERT, COLLINS and WEISMAN (1984) and the C at IVS 2 position 177 is found in two other γ genes]. These will be designated type A features. Other differences have only been found in $^{A}\gamma$ genes (for example, the A at IVS 2 position 188 and the C at position 590), and these will be designated type G differences. Other features were found in both $^{G}\gamma$ and $^{A}\gamma$ genes (for example, IVS 2 positions 177, 189, 206, 224, and 269 to 272), and will be indicated with a dot. Others were unique to this gene (at IVS 2 positions 308 and 866), and will be called type *X* differences. The unusual sequence features of IVS 2 of the F.M. single γ gene can then be summarized as follows:

$^{G}\gamma$-(crossover)-AAA.G.....*X*....A..GAAG.AAA...*X*-$^{A}\gamma$

The overall pattern of the F.M. gene was $^{G}\gamma$-like in the region 5′ to the 1-kbp region of identity where the crossover occurred. In the region 3′ to the crossover, the pattern was characterized by first $^{A}\gamma$-like, then $^{G}\gamma$-like sequences for several cycles before ending up $^{A}\gamma$-like at the 3′ end of the gene and in the flanking region. This single γ gene has quite an unusual array of nucleotide differences as compared to other γ genes. However, despite these differences, the crossover region in this single γ gene can be identified as having occurred within the 1.0-kbp stretch of identity between the parental $^{G}\gamma$ and $^{A}\gamma$ genes beginning 281 bp 5′ to the cap site and ending in IVS 2.

DISCUSSION

There are relatively few situations in higher eukaryotes in which the exact crossover points of germline recombinational events can be identified and in which molecular details of the relevant crossovers can still

be deciphered at the nucleotide sequence level by examining the recombinants. In lower organisms such as *Escherichia coli coli*, *Ascobolus immersus* and *Saccharomyces cerevisiae* (POTEETE and VOLKERT 1988; ROSSIGNOL, PAQUETTE and NICKLAS 1978; SZOSTAK *et al.* 1983), genetic studies have been extremely successful in sorting out multiple recombinational pathways. In these organisms, homologous and nonhomologous events are probably the result of distinct mechanisms.

It is reasonable to suppose that multiple recombinational pathways also exist in mammals, where events involving very short stretches of homology might be the products of either the homologous or nonhomologous recombinational pathway(s), or perhaps of other pathways. In some examples of large deletions in humans, there is evidence for very limited stretches (up to 5 bp) of homology in the normal DNA at the sites of the deletion endpoints (VANIN *et al.* 1983; RULEY and FRIED 1983; HENTHORN *et al.* 1986; GILMAN and ABRAHAM 1987). It may be that such "illegitimate" recombinational events are promoted by homologies, or that these very short regions of homology reflect the way that randomly broken DNA sequences are repaired.

A few examples have been presented in the earlier literature in which duplications or deletions were shown to involve recombination between regions with more extensive homology (SHEN *et al.* 1981; MICHAELSON and ORKIN 1983). In those situations it is difficult to avoid the conclusion that the recombination was in some way facilitated by the region of homology. Events involving longer stretches of homology might be the products of a specific homology-dependent pathway.

Our understanding of recombination processes in higher eukaryotes would be helped, for example, if we knew whether homologous but unequal crossing over events are the result of clean breakage and reunion of DNA strands, or are accompanied by insertions, deletions or nearby gene conversion events. We would also like to know what sequence requirements, if any, must be met in order to allow homologous crossovers to occur. Our data provide some answers to these various questions.

Evidence for a possible instance of unequal crossover with gene conversion: There are two possible explanations for the observed unusual features in IVS 2 of the single γ gene, with alternating $^{G}\gamma$-like and $^{A}\gamma$-like sequences. The first possibility is that the crossover which produced this fusion gene occurred as a simple event between chromosomes that already had the unexpected features. The second possibility is that these deviations from the nucleotide sequence expected from a fusion of $^{G}\gamma$ and $^{A}\gamma$ sequences were produced by a series of gene conversions (and possibly point mutations as well) as part of the homologous but unequal crossover event that led to the fusion. The pattern is strongly suggestive of a "patchy" gene conversion that could well be the consequence of the crossover event between a $^{G}\gamma$ and $^{A}\gamma$ gene that created the single γ gene. However, in the absence of a survey of normal γ genes in the Indian population, we cannot formally rule out the possibility that this gene conversion occurred before or after the time of the unequal crossover. It is rather interesting to note that in yeast, gene conversions occur over an average of 300–600 bp from the site of initiation of crossover events (STRUHL 1987; ORR-WEAVER, NICOLAS and SZOSTAK 1988), and therefore crossover-associated conversions over the length observed in the F.M. single γ crossover are plausible. However, we emphasize that we found no evidence for such conversions associated with the other six types of crossover events studied, so they are not common associations.

Evidence for independent origin of examples of homologous unequal crossover: In order to address the question of the distribution of crossover sites, it is necessary to first ask how many independent recombinational events are represented in our collection. The minimum number is six, since we have examined recombinant genes with six different structures: these are the Lepore-Boston, Lepore-Baltimore, Lepore-Hollandia, Kenya, single γ, and γ triplication fusion products. However, in weighting the contribution of the structural analysis of each of these gene types to our understanding of unequal homologous crossing over, it is necessary to consider whether there are any additional examples of independent events among the many examples of Lepore-Boston genes that we examined.

At the nucleotide sequence level, comparison of sequence differences between two Lepore-Boston genes, Sp14 and R15, the normal δ and β genes, the Lepore-Baltimore and the Lepore-Hollandia genes (Figure 5A) is strongly suggestive of the possibility that the two Lepore-Boston genes may have arisen independently. For example, although the crossover positions of the two genes are identical, the Sp14 and R15 sequences differ at three sites in the region 5′ to the crossover, while the Baltimore and R15 Boston sequences, which must have arisen from independent fusion events due to their different crossover sites, differ at only two nucleotide positions in the same region. Therefore, the flanking sequences of the two Lepore-Boston genes are no more closely related than those of clearly unrelated Lepore genes. Furthermore, in the 3′ regions, the Sp14 gene has a haplotype that has not been described previously, such that at least one additional recombinational or mutational event must be invoked to derive the observed pattern from any previously described haplotype.

Other support for the plausibility of independent

origins for Sp14 and R15 comes from population surveys that indicate that the Thai population is relatively distant, genetically, from the European population (TRIANTIPHYLLIDIS, KOUVATSI and KAPLANOGLOU 1983; CAVALLI-SFORZA *et al.* 1988). Furthermore, it has been shown that the types of (non-Lepore) thalassemia mutations found among Thai individuals differ from those of Europe as well as from those of the geographically closer Chinese and Indian populations (LYNCH *et al.* 1988). In addition, it should be noted that the unique β framework of the Sp14 gene is most similar to the 3-Asian framework, which is not found in Mediterranean populations. There is no reason to suppose that the Lepore mutations in Thailand differ from the other thalassemia mutations with respect to independence of origin. The alternative explanation, that Sp14 and R15 are descendants of a single unequal crossover, requires that the Lepore gene found in Thailand is the result of migration of a Lepore-carrying individual from the Mediterranean to Thailand, followed by interbreeding. There is not any historical evidence for such migration. Taken together, these observations make it likely, but do not prove, that the Sp14 and R15 genes were generated by independent crossover events.

Consideration of the haplotypes of restriction site polymorphisms along the entire length of the β-globin cluster in chromosomes containing Lepore-Boston genes (Figure 2) shows that the Lepore-Boston genes are present on at least six types of chromosomes. This sort of result has previously been taken as evidence of independent origin of identical mutations (ANTONARAKIS *et al.* 1982b; KAZAZIAN *et al.* 1984; PAGNIER *et al.*, 1984; KULOZIK *et al.* 1986; NAKASHIMA *et al.* 1990; OKANO *et al.* 1990).

Thus, while none of our observations constitutes absolute proof that the Lepore-Boston crossover happened more than once in human evolution, they can be interpreted as indicating that this event may have occurred as many as six times in the genes ancestral to those in our collection.

Longer regions of identity are preferred for crossing over: Our data indicate that the unequal crossovers that we have studied occurred in relatively large stretches of sequence identity. In order to investigate the significance of this observation, we tabulated the distribution of sizes of regions of identity between the δ and β genes (Figure 8A). As shown by inspection of the figure, this distribution is at least bimodal and possibly trimodal. We therefore defined "long stretches" as all stretches in the middle and last peaks of the frequency distribution, namely, those containing more than 10 target crossover sites (10 or more matching bases). "Short stretches" contain fewer than 10 target sites. The number of sites in the long category is 366 internucleotide bonds out of a total of 1299. The probability that all of the three certainly independent Lepore crossovers occurred by chance in long stretches of identity is therefore $(366/1299)^3 = 0.022$. We therefore propose, as a working hypothesis, that homologous unequal crossing over occurs preferentially in long stretches of identity.

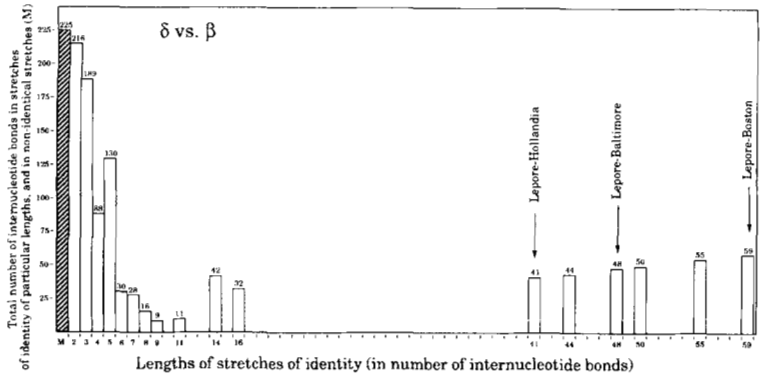
Two additional instances of unequal crossover between the δ and β genes have been studied previously by sequence analysis. The crossover event leading to the β-δ fusion P-Nilotic occurred in a 54-bp region of identity in Exon 2 (LIU *et al.* 1987). The crossover that produced the β-δ fusion Miyada occurred in a 14-bp region of identity in IVS 1 (KIMURA *et al.* 1984). These results are in agreement with our hypothesis that longer stretches of identity are preferred.

We performed a similar analysis for the $^{A}\gamma$-β Kenya crossover. As shown in Figure 8B, the distribution of numbers of targets in stretches of identity of particular lengths is again bimodal or possibly trimodal. The number of target internucleotide bonds in stretches of identity greater than 10 in size (the middle and last peaks) is 155 out of a total of 1390 internucleotide bonds. The probability that a crossover between $^{A}\gamma$ and β would occur in a long region of identity on the random hypothesis is therefore $(155/1390) = 0.11$. Combining the two probabilities for the Lepore and Kenya crossovers gives an overall probability for random crossover of 0.0025. If, as we have argued, the Lepore-Boston Thai crossover class is also an independent event, the overall probability that the observed results would have been obtained assuming random crossover sites is 0.00069. We therefore conclude that longer uninterrupted stretches of identity are preferred as crossover targets in homologous unequal crossing over.

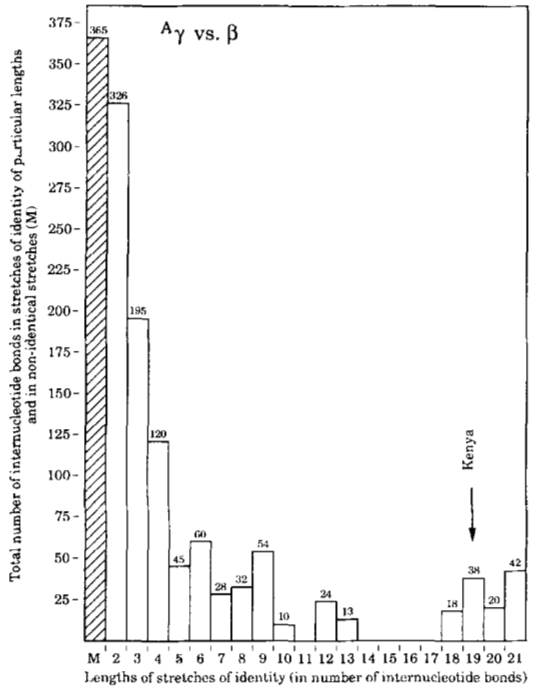
Analysis of the $^{A}\gamma$-$^{G}\gamma$ (single γ) and $^{G}\gamma$-$^{A}\gamma$ (γ triplication) crossovers does not contribute much to this aspect of our work. As shown in Figure 8C, over 99% of the target internucleotide bonds for recombination are in larger stretches of identity. However, note that in both cases the crossover occurred in the largest available region of identity. Also in one additional example of crossover between the γ genes, the γ triplication studied by LIU *et al.* (1988), the crossover occurred in a 73-bp region of identity 400 bp 5′ to the cap sites of $^{G}\gamma$ and $^{A}\gamma$. Since the homology between the two γ genes is considerably reduced 5′ to this 73-bp region, our conclusions are fully consistent with this example as well.

Recombination frequencies: A possible extension, not directly addressed in this investigation, to our conclusion that longer stretches of identity are preferred is that frequency of homologous unequal crossover is likely to be positively correlated with the lengths of the shared sequences. Support for this possibility comes from the increasing numbers of re-

A.



B.



C.

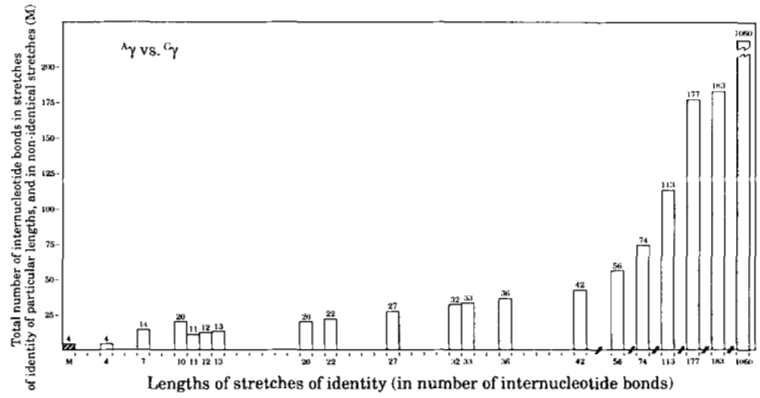


FIGURE 8.—Distribution of regions of identity between nonhomologous globin genes. A, δ *vs.* β. The numbers of internucleotide bonds within and next to regions of *identity* of particular lengths between the δ and β genes were tabulated (see text) and are illustrated with open bars. The number of internucleotide bonds in stretches of *nonidentity* of any length is shown with a hatched bar. Sequences are from PONCZ *et al.* (1983). Crossover regions of the applicable lengths for the Lepore-Hollandia, Lepore-Baltimore, Lepore-Boston, P-Nilotic and Miyada crossovers are indicated with arrows. B, Distribution of regions of identity between $^{A}\gamma$ and β. Sequences are from SLIGHTOM, BLECHL and SMITHIES (1980) and PONCZ *et al.* (1983). The crossover region for the Kenya crossover is indicated with an arrow. C, Distribution of regions of identity between $^{G}\gamma$ and $^{A}\gamma$. The crossover region for the single γ and γ-triplication crossovers is indicated with an arrow. Sequences are from SLIGHTOM, BLECHL and SMITHIES (1980).

ports of chromosomes resulting from unequal crossovers between tandem genes with extensive identity in various human population. Many of these events result in different structures, and are therefore of independent origin. Some examples are the single γ, triple γ, quadruple γ and quintuple γ configurations (TRENT *et al.*, 1981, 1986; NAKATSUJI *et al.* 1986; HILL *et al.* 1986; SHIMASAKI AND IUCHI 1986; SHIMIZU *et al.*, 1986; DAENEN *et al.* 1987; FEI *et al.* 1988), the single α, triple α and quadruple α globin gene arrangements (GOOSENS *et al.* 1980; PHILLIPS *et al.* 1980; HIGGS *et al.* 1980, 1984; LIE-INJO, HERRERA and KAN 1981; DOVER *et al.* 1987; EL-HAZMI 1987; GU, LANDMAN and HUISMAN 1987; HENNI *et al.* 1987; AKERMANN *et al.* 1990; NAKASHIMA *et al.* 1990), the single ζ-, triple ζ- and quadruple ζ-globin gene arrangements (WINICHAGOON *et al.* 1982; HILL *et al.* 1986; TITUS, HSIA and HUNT *et al.* 1988), and the multiple *Hpr*

genes (MAEDA *et al.* 1986). Such examples do not prove that nearly identical genes are more subject to unequal crossing over, because the frequencies have not been directly measured, but are an indication that it may be the case.

Homologous recombination studies in other systems: Considerable attention has been given to characterizing homologous recombination events both *in vitro* and *in vivo* in recent years with one area of special interest being the amount of homology required in the recombining molecules. Although the *E. coli* recombinase *Rec A* has been observed to promote branch migration through regions of considerable heterology *in vitro* (BIANCHI and RADDING 1983), studies on recombination in *E. coli* (WATT *et al.* 1985; SHEN and HUANG 1986) and bacteriophage T4 (SINGER *et al.* 1982) show that a region of identity of about 30 bp between the recombining molecules is required for efficient homologous recombination, and that even a 1-bp mismatch in this region substantially reduces the recombination efficiency (WATT *et al.* 1985; SHEN and HUANG 1986; SINGER *et al.* 1982; KITTS and NASH 1987). SHEN and HUANG (1986) referred to this required minimum amount of homology as the minimum efficient processing segment (MEPS) for each system.

Other studies have concluded that recombination systems will tolerate considerable heterologies (PEARSON and FOX 1988); however, the efficiencies are considerably reduced. A requirement for MEPS has been observed in mammalian cells in plasmid-plasmid recombination experiments (RUBNITZ and SUBRAMANI 1984; AYARES *et al.* 1986), as well as for recombination between chromosomally inserted plasmids and native chromosomal genes (LISKAY, LETSOU and STACHELEK 1987; WALDMAN and LISKAY 1988). The consensus of these studies indicates that approximately 200 bp of uninterrupted identity are required for efficient recombination in mammalian cells. As in the prokaryotic studies, mismatches (*i.e.* shorter stretches of identity) appear to reduce efficiencies.

In an interplasmid recombination experiment, using monkey cos cells, GOMEZ-PEDROZO, HU and SHEN (1988) found that a majority of crossovers between two α-globin gene fragments occurred in the largest available regions of identity. These regions were 339 and 435 bp in length. One crossover observed in this study occurred in a 5-bp region of identity, bounded by more extensive regions of identity. Events of this sort are not surprising, as mismatches may greatly reduce the efficiency of branch migration, but not interrupt it in every case. No gene conversions were observed in seven recombinants studied by these authors, who suggest that regions of identity are preferred in unequal crossing over in cultured primate cells.

A number of studies of sequences resulting from homologous recombination in mammals *in vivo* have been reported. Although the fluctuation tests necessary for measurements of MEPS values are not practical in these systems, it is of interest to compare the lengths of the identity regions in which crossovers have been observed. Analysis of mouse strains with variant major histocompatibility loci by KOBORI *et al.* (1986) revealed that crossovers occurred either in a 410 bp identity region (one example) or in a 1-kbp identity region (three examples).

Several examples of unequal crossing over in humans between homologous genes other than the globin genes have been presented in recent years. LYONS, STEIN and SMITHIES (1988) presented evidence from studies of human proline-rich protein genes for intragenic unequal crossing over that is highly sensitive to reduction in lengths of sequence identity. These investigators observed 20 unequal crossover variants in PRB loci which contain highly homologous 36-bp repeat units, but only one unequal crossover in a PRH locus, which has shorter identity regions than the PRB loci. NEITZ, NEITZ and JACOBS (1989) examined a fusion gene that produces red-green color-blindess, and showed that it resulted from crossing over in a 113-bp region of identity between two visual pigment genes. ZIMRAN *et al.* (1990) showed that the mutation in a case of Gaucher disease was a fusion of the glucocerebrosidase gene and its pseudogene, in a 339-bp region of identity between the two genes. Crossing over in 26-bp regions of identity between Alu repeats have also been reported (LEHRMAN *et al.* 1987; MARKERT *et al.*, 1988). Our findings are consistent with all of these observations of crossing over in relatively large regions of identity.

An increasing body of experimental data [see, for example, ABASTADO *et al.* (1987) and WALDMAN and LISKAY (1987, 1988)] support the inference that homologous recombination *in vivo* initiates preferentially in regions of sequence identity, and that branch migration proceeds, until a region of divergence in the interacting molecules is reached, at which point resolution of the crossover occurs. The recombination event will usually resolve at this point within the identity region. Resolution may occur within the neighboring region of decreased homology, with the probability of the latter occurrence in some inverse proportion to the degree of divergence. Our sequence studies of the products of homologous unequal crossovers in humans are readily understandable in these terms.

This work was supported by a grant from the National Institutes of Health research grants GM 20069 and HL 37001 and by United States Public Health Service Research Grant HLB-05168. This is paper number 3185 from the Laboratory of Genetics, University of Wisconsin-Madison.

We thank M. BAIGET, Barcelona, Spain; A. BANK, New York; A.

B. FINDLAY, Brisbane, Australia; G. EFREMOV, Skopje, Yugoslavia; S. FUCHAROEN, Bangkok, Thailand; H. KAZAZIAN, Baltimore; L. MORLE, Lyon, France; E. SCHWARTZ, Philadelphia; and R. J. A. TRENT, Sydney, Australia, for kindly providing DNA and blood samples for this study. We thank BILL SUGDEN for making the Kenya lymphoblastoid cell line. We thank RUSSELL KAUFMAN for the gift of the plasmid pRK28. We are also grateful to JOHN BERMINGHAM, BRUCE ELDER, JANE GITSCHIER, BARBARA LEVINSON, J. LYNCH, SUE KENWRICK, ROBERT METZENBERG, STAN METZENBERG and JERRY SLIGHTOM for support and helpful discussions during the preparation of this manuscript.

LITERATURE CITED

ABASTADO, J.-P., S. DARCHE, F. GODEAU, B. CAMI and P. KOURILSKY, 1987 Intramolecular recombination between partially homologous sequences in *Escherichia coli* and *Xenopus laevis* oocytes. Proc. Natl. Acad. Sci. USA **84:** 6496–6500.

AKERMAN, B. R., T. M. FUJIWARA, G. A. LANCASTER, K. MORGAN and C. R. SCRIVER, 1990 Identification of deletion and triple α-globin gene haplotypes in the Montreal β-thalassemia screening program: implications for genetic medicine. Am. J. Med. Genet. **36:** 76–84.

ANTONARAKIS, S. E., C. D. BOEHM, P. J. V. GIARDINA and H. H. KAZAZIAN, JR., 1982a Nonrandom association of polymorphic restriction sites in the β-globin cluster. Proc. Natl. Acad. Sci. USA **79:** 137–141.

ANTONARAKIS, S. E., S. H. ORKIN, H. H. KAZAZIAN JR., S. C. GOFF, C. D. BOEHM, P. G. WABER, J. P. SEXTON, H. OSTRER, V. F. FAIRBANKS and A. CHAKRAVARTI, 1982b Evidence for multiple origins of the β^E-globin gene in Southeast Asia. Proc. Natl. Acad. Sci. USA **79:** 6608–6611.

ANTONARAKIS, S. E., C. D. BOEHM, G. R. SERJEANT, C. E. THEISEN, G. J. DOVER and H. H. KAZAZIAN, JR., 1984 Origin of the β^S globin gene in blacks: the contribution of recurrent mutation and/or gene conversion. Proc. Natl. Acad. Sci. USA **81:** 853–856.

AYARES, D., L. CHERKURI, K-Y. SONG and R. KUCHERLAPATI , 1986 Sequence homology requirements for intermolecular recombination in mammalian cells. Proc. Natl. Acad. Sci. USA. **83:**599–5203.

AZEN, E. A., K. M. LYONS, T. MCGONIGAL, N. L. BARRETT, L. S. CLEMENTS, N. MAEDA, E. F. VANIN, D. M. CARLSON and O. SMITHIES, 1984 Clones from the human gene complex coding for salivary proline-rich proteins. Proc. Natl. Acad. Sci. USA. **81:** 5561–5565.

BAGLIONI, C., 1962 The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. Proc. Natl. Acad. Sci. USA. **48:** 1880–1886.

BAGLIONI, C., 1965 Abnormal human hemoglobins X. A study of hemoglobin Lepore Boston. Biochim. Biophys. Acta **97:** 37–46.

BAIRD, M., H. SCHREINER, C. DRISCOLL and A. BANK, 1981 Localization of the site of recombination in formation of the Lepore Boston globin gene. J. Clin. Invest. **68:** 560–564.

BARNABAS, J., and C. J. MULLER, 1962 Haemoglobin Lepore Hollandia. Nature **194:** 931–932.

BIANCHI, M. E., and C. M. RADDING, 1983 Insertions, deletions and mismatches in heteroduplex DNA made by RecA protein. Cell **35:** 511–520.

BOONTRAKOONPOONTAWEE, P., J. SVASTI, S. FUCHAROEN and P. WINICHAGOON, 1987 Identification of hemoglobin Lepore-Washington-Boston in association with hemoglobin E [Beta-26(B8 Glu → Lys] in a Thai female. Hemoglobin **11:** 309–316.

CAVALLI-SFORZA, L. L., A. PIAZZA, P. MENOZZI and J. MOUNTAIN, 1988 Reconstruction of human evolution: bringing together genetic, archeological, and linguistic data. Proc. Natl. Acad. Sci. USA **85:** 6002–6006.

CHAKRAVARTI, A., K. H. BUETOW, S. E. ANTONARAKIS, P. G. WABER, E. D. BOEHM and H. H. KAZAZIAN, 1984 Nonuniform recombination within the human β-globin gene cluster. Am. J. Hum. Genet. **36:** 1239–1258.

CHEBLOUNE, Y., D. PONCET and G. VERDIER, 1984 S1-nuclease mapping of the genomic Lepore-Boston DNA demonstrates that the entire large intervening sequence of the fusion gene is of β-type. Biochem. Biophys. Res. Commun. **120:** 116–123.

CHEBLOUNE, Y., and G. VERDIER, 1983 The delta-beta-crossing-over site in the fusion gene of the Lepore-Boston disease might be localized in a preferential recombination region. Acta Haematol. **69:** 294–302.

CHEHAB, F. F., G. R. HONIG and Y. W. KAN, 1986 Spontaneous mutation in β-thalassemia producing the same nucleotide substitution as that in a common hereditary form. Lancet **1:** 3–5.

DAENEN, S., A. L. REESE, F. KUTLAR and T. H. J. HUISMAN, 1987 An Indonesian family with the Southeast Asian type of α-thal-1 and a γ globin gene triplication. Acta Haematol. **78:** 23–27.

DEVEREUX, J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

DOBKIN, C., J. CLYNE, A. METZENBERG and A. BANK, 1986 Expression of a cloned Lepore globin gene. Blood **67:** 168–172.

DOVER, G. J., V. T. CHANG, S. H. BOYER, G. R. SERJEANT, S. ANTONARAKIS and D. R. HIGGS, 1987 The cellular basis for different fetal hemoglobin levels among sickle cell individuals with two, three, and four alpha-globin genes. Blood **69:** 341–344.

EFREMOV, G. D., 1978 Hemoglobin Lepore and anti-Lepore. Hemoglobin **2:** 197–233.

EFREMOV, G. D., R. RUDIVIC, G. A. NIAZI, E. HUNTER, JR., T. H. J. HUISMAN and W. A. SCHROEDER, 1976 An individual with hemoglobin-Lepore-Baltimore delta-beta thalassemia in a Yugoslavian family. Scand. J. Haematol. **16:** 81–89.

EL-HAZMI, M. A. F., 1987 Haplotypes of alpha-globin gene in the Saudi population-the triplicated gene. Hum. Hered. **37:** 314–318.

FEI, Y. J., K. D. LANCLOS, F. KUTLAR, E. L. WALKER III and T. H. HUISMAN, 1988 A chromosome with five gamma-globin genes. Blood **72:** 827–829.

GERALD P. S., and L. K. DIAMOND, 1958 A new hereditary hemoglobinopathy (the Lepore trait) and its interaction with thalassemia trait. Blood **13:** 835–844.

GILMAN, J. G., T. HARANO, T. NAKATSUJI, I. BAKIOGLU, A. L. REESE, M. B. GARDINER and T. H. J. HUISMAN, 1985 The ratio of the $^G\gamma$ and $^A\gamma$ chains: variations due to anomalies at the molecular level. Ann. NY Acad. Sci. **445:** 235–247.

GILMAN, J. G., and J. ABRAHAM, 1987 DNA sequence analysis of the Dutch beta zero-thalassemia deletion. Biomed. Biochim. Acta **46:** S131-S135.

GIMFERRER, E., 1976 Haemoglobin Lepore in a Spanish family. Acta Haematol. **56:** 234–240.

GITSCHIER, J., W. I. WOOD, M. A. SHUMAN and R. M. LAWN, 1986 Identification of a missense mutation in the factor VIII gene of a mild hemophiliac. Science **232:** 1415–1416.

GOMEZ-PEDROZO, M., W. S. HU and C. K. SHEN, 1988 Recombinational resolution in primate cells of two homologous human DNA segments with a gradient of sequence divergence. Nucleic Acids Res. **16:** 11237–11247.

GOOSENS, M., A. M. DOZY, S. H. EMBURY, Z. ZACHARIADES, M. G. HADJIMINAS, G. STAMATOYANNOPOULOS and Y. W. KAN, 1980 Triplicated α globin loci in humans. Proc. Natl. Acad. Sci. USA **77:** 518–521.

GU, Y. C., H. LANDMAN and T. H. HUISMAN, 1987 Two different quadruplicated alpha globin gene arrangements. Br. J. Haematol. **66:** 245–250.

HENNI, T., F. MORLE, B. LOPEZ, P. COLONNA and J. GODET, 1987 Alpha-thalassemia haplotypes in the Algerian population. Hum. Genet. **75:** 272–276.

HENTHORN P. S., D. L. MAGER, T. H. J. HUISMAN and O. SMITHIES, 1986 A gene deletion ending within a complex array of repeated sequences 3′ to the human beta-globin gene cluster. Proc. Natl. Acad. Sci. USA **83:** 5194–5198.

HIGGS, D. R., J. M. OLD, L. PRESSLEY, J. B. CLEGG and D. J. WEATHERALL, 1980 A novel α globin gene rearrangement in man. Nature **284:** 632–635.

HIGGS, D. R., A. V. S. HILL, D. K. BOWDEN, D. J. WEATHERALL and J. B. CLEGG, 1984 Independent recombination event between the duplicated human α globin genes, implications for their concerted evolution. Nucleic Acids Res. **12:** 6965–6977.

HILL, A. V. S., D. K. BOWDEN, D. J. WEATHERALL and J. B. CLEGG, 1986 Chromosomes with one, two, three and four fetal globin genes, molecular and hematologic analysis. Blood **67:** 1611–1618.

HUISMAN, T. H. J., R. N. WRIGHTSTONE, J. B. WILSON, W. A. SCHROEDER and A. G. KENDALL, 1972 Hemoglobin Kenya, the product of fusion of γ and β polypeptide chains. Arch. Biochem. Biophys. **153:** 850–853.

KAZAZIAN, H. H., JR., S. H. ORKIN, S. E. ANTONARAKIS, J. P. SEXTON, C. D. BOEHM, S. C. GOFF and P. G. WABER, 1984 Molecular characterization of seven β-thalassemia mutations in Asian Indians. EMBO J. **3:** 593–596.

KENDALL, A. G., P. J. OJWANG, W. A. SCHROEDER and T. H. J. HUISMAN, 1973 Hemoglobin Kenya, the product of a γ-β fusion gene: studies of the family. Am. J. Hum. Genet. **25:** 548–563.

KIM, H-S., and O. SMITHIES, 1988 Recombinant fragment assay for gene targetting based on the polymerase chain reaction. Nucleic Acids Res. **16:** 8887–8903.

KIMURA, A., Y. OHTA, Y. FUKUMAKI and Y. TAKAGI, 1984 A fusion gene in man: DNA sequence analyses of the abnormal globin gene of hemoglobin Miyada. Biochem. Biophys. Res. Commun. **119:** 968–974.

KITTS, P. A., and H. A. NASH, 1987 Homology-dependent interactions in phage lambda site-specific recombination. Nature **329:** 346–348.

KOBORI, J. A., E. STRAUSS, K. MINARD and L. HOOD, 1986 Molecular analysis of the hotspot of recombination in the murine major histocompatability complex. Science **234:** 173–179.

KULOZIK, A. E., J. S. WAINSCOAT, G. R. SERJEANT, B. C. KAR, B. AL-AWAMY, G. J. ESSAN, A. G. FALUSI, S. K. HAQUE, A. M. HILALI and S. KATE, 1986 Geographical survey of beta S-globin gene haplotypes: evidence for an independent Asian origin of the sickle-cell mutation. Am. J. Hum. Genet. **39:** 239–244.

LANCLOS, K. D., J. PATTERSON, G. D. EFREMOV, S. C. WONG, A. VILLEGAS, P. J. OJWANG P. J., J. B. WILSON, F. KUTLAR and T. H. J. HUISMAN, 1987 Characterization of chromosomes with hybrid genes for hemoglobin Lepore-Washington, hemoglobin Lepore-Baltimore, hemoglobin P-Nilotic and hemoglobin Kenya. Hum. Genet. **77:** 40–45.

LEHRMAN, M. A., J. L. GOLDSTEIN, D. W. RUSSEL and M. S. BROWN, 1987 Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell **48:** 827–835.

LIE-INJO, L. E., A. R. HERRERA and Y. W. KAN, 1981 Two types of triplicated alpha globin loci in humans. Nucleic Acids Res. **9:** 3707–3717.

LISKAY, R. M., A. LETSOU and J. L. STACHELEK, 1987 Homology requirement for efficient gene conversion between duplicated chromosomal segments in mammalian cells. Genetics **115:** 161–167.

LIU, J. Z., T. HARANO, K. D. LANCLOS and T. H. HUISMAN, 1987 The beta-delta crossover leading to the beta delta hybrid gene of hemoglobin P-Nilotic is located within 54 base-pairs of the 5′ end of exon 2 or between codons 31 and 50. Biochim. Biophys. Acta **909:** 208–212.

LIU, J. Z., J. G. GILMAN, Q. CAO, I. BAKIOGLU and T. H. J. HUISMAN, 1988 Four categories of γ globin gene triplications: DNA sequence comparison of low G-γ and high G-γ triplications. Blood **72:** 480–484.

LYNCH, J., V. E. TATE, D. J. WEATHERALL, S. FUCHAROEN, V. S. TANPHAICHITR, P. ISARANGKURA, P. SEKSARN, V. LAOSOMBAT, P. KULAPONGA and P. WASI P., 1988 Molecular basis of β thalassemia in Thailand. B.D.O.A.S. **23:**71–79.

LYONS, K. M., J. H. STEIN and O. SMITHIES, 1988 Length polymorphisms in human proline-rich protein genes generated by intragenic unequal crossing-over. Genetics **120:** 267–278.

MAEDA, N., S. M. MCEVOY, H. F. HARRIS and T. H. J. HUISMAN, 1986 Polymorphisms in the human haptoglobin cluster: chromosomes with multiple haptoglobin-related (Hpr) genes. Proc. Natl. Acad. Sci. USA **83:** 7395–7399.

MANNING, R. F., and L. P. GAGE, 1980 Internal structure of the silk fibroin gene of *Bombyx mori*. II. Remarkable polymorphism of the organization of crystalline and amorphous coding sequences. J. Biol. Chem. **255:** 9451–9457.

MARKERT, M. L., J. T. HUTTON, D. A. WIGINTON, J. C. STATES and R. G. KAUFMAN, 1988 Adenosine deaminase (ADA deficiency due to a deletion of the ADA gene promoter and first exon by homologous recombination between two Alu elements. J. Clin. Invest. **81:** 1323–1327.

MAVILIO, F., GIAMPAOLO A., A. CARE, N. M. SPOSI and M. MARINUCCI , 1983 The β crossover region in Lepore Boston hemoglobinopathy is restricted to a 59 base pairs region around the 5′ splice junction of the large globin gene intervening sequence. Blood **62:** 230–233.

MAXAM, A. M., and W. GILBERT, 1980 Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. **65:** 499–559.

MEARS, J. G., F. RAMIREZ, D. LEIBOWITZ, F. NAKAMURA, A. BLOOM, F. KONOTEY-AHULER and A. BANK, 1978 Changes in restricted human cellular DNA fragments containing globin gene sequences in thalassemias and related disorders. Proc. Natl. Acad. Sci. USA **75:** 1222–1226.

MESSING, J., 1983 New M13 vectors for cloning. Methods Enzymol. **101:** 20–78.

METZENBERG, A., 1989 Homologous unequal crossing-over *in vivo* in the human β-globin gene cluster. Ph.D. thesis, University of Wisconsin-Madison.

MICHAELSON, A. M., and S. H. ORKIN, 1983 Boundaries of gene conversion within the duplicated human alpha globin genes. J. Biol. Chem. **258:** 15245–15254.

MORLE, L., F. MORLE, E. DORLEAC, F. BAKLOUTI, C. BAUDONNET, J. GODET and J. DELAUNEY 1984 The association of Hemoglobin Knossos and Hemoglobin Lepore in an Algerian patient. Hemoglobin **8:** 229–238.

NAKASHIMA, H., A. FUJIYAMA, S. KAGIYAMA and T. IMAMURA, 1990 Genetic polymorphisms of gene conversion within the duplicated human α-globin loci. Hum. Genet. **84:** 568–570.

NAKATSUJI, T., A. KUTLAR A., F. KUTLAR and T. H. J. HUISMAN, 1986 Haplotypes among Vietnamese hemoglobin G homozygotes including one with a γ-globin gene triplication. Am. J. Hum. Genet. **38:** 1981–1983.

NEITZ, J., M. NEITZ and G. H. JACOBS, 1989 Analysis of fusion gene and encoded photopigment of colour-blind humans. Nature **342:** 679–682.

NUTE, P. E., W. G. WOOD, G. STAMATOYANNOPOULOS, C. OLWENY and P. J. FAILKOW, 1976 The Kenya form of hereditary

persistence of fetal haemoglobin: structural studies and evidence for homogeneous distribution of haemoglobin F using fluorescent anti-haemoglobin F antibodies. Br. J. Haematol. **32:** 55–63.

OJWANG, P. J., T. NAKATSUJI, M. B. GARDENER, A. L. REESE, J. G. GILMAN and T. H. J. HUISMAN, 1983 Gene deletion as the molecular basis for the Kenya-$^{G}\gamma$-HPFH condition. Hemoglobin **7:** 115–123.

OKANO, Y., T. WANG, R. C. EISENSMITH, F. GÜTTLER and S. L. C. WOO, 1990 Recurrent mutation in the human phenylalanine hydroxylase gene. Am. J. Hum. Genet. **46:** 919–924.

ORKIN, S. H., H. H. KAZAZIAN, JR., S. E. ANTONARAKIS, S. C. GOFF, C. D. BOEHM, J. P. SEXTON, P. G. WABER, and P. V .J. GIARDINA, 1982 Linkage of β-thalassemia mutations and β-globin gene polymorphisms with DNA polymorphisms in the human β-globin gene cluster. Nature **296:** 627–631.

ORR-WEAVER, T. L., A. NICOLAS and J. W. SZOSTAK, 1988 Gene conversion adjacent to regions of double-strand break repair. Mol. Cell. Biol. **8:** 5292–52 198.

OSTERTAG, W., and E. W. SMITH, 1969 Hemoglobin-Lepore-Baltimore, a third type of delta, beta cross-over. Eur. J. Biochem. **10:** 371–376.

PAGNIER, J., J. G. MEARS, O. DUNDA-BELKHODJA, K. E. SCHAEFER-REGO, C. BELDJORD, R. L. NAGEL and D. LABIE, 1984 Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. Proc. Natl. Acad. Sci. USA **81:** 1771–1773.

PEARSON, R. K., and M. S. FOX, 1988 Effects of DNA heterologies on bacteriophage lambda recombination. Genetics **118:** 13–19.

PEDERSON, K., J. DEVEREUX, D. R. WILSON, E. SHELDON and B. A. LARKINS, 1982 Unequal crossing-over in zein genes by sequence analysis. Cell **29:** 1015–1026.

PHILLIPS, J. A., III, T. A. VIK, A. F. SCOTT, K. E. YOUNG, H. H. KAZAZIAN, JR., K. D. SMITH, V. F. FAIRBANKS and H. M. KOENIG, 1980 Unequal crossing-over, a common basis of single alpha globin genes in Asians and American blacks with hemoglobin H disease. Blood **55:** 1066–1069.

PONCZ, M., E. SCHWARTZ, M. BALLANTINE and S. SURREY, 1983 Nucleotide sequence analysis of the $\delta\beta$-globin gene region in humans. J. Biol. Chem. **258:** 11599–11609.

POTEETE, A. R., and M. R. VOLKERT, 1988 Activation of Rec F-dependent recombination in *Escherichia coli* by bacteriophage lambda- and P22- encoded functions. J. Bacteriol. **170:** 4379–4381.

POWERS, P. A., and O. SMITHIES, 1986 Short gene conversions in the human fetal globin gene region: a by-product of chromosome pairing during meiosis? Genetics **112:** 343–358.

ROSSIGNOL, J. L., N. PAQUETTE and A. NICOLAS, 1978 Aberrant 4:4 asci, disparity in the direction of conversion, and frequencies of conversion in *Ascobolus immersus*. Cold Spring Harbor Symp. Quant. Biol. **43:** 1343–1351.

RUBNITZ, J., and S. SUBRAMANI, 1984 The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. **4:** 2253–2258.

RULEY, H. E., and M. FRIED, 1983 Clustered illegitimate recombination events in mammalian cells involving very short sequence homologies. Nature **304:** 181–184.

SANGER, T., S. NICKLEN and A. R. COULSON , 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCOTT, A. F., P. HEATH, S. TRUSKO, S. H. BOYER, W. PRASS, M. GOODMAN, J. CZELVSNIAK, L.-Y. E. CHANG and J. SLIGHTOM, 1984 The sequence of the gorilla fetal globin genes: Evidence for multiple gene conversions in human evolution. Mol. Biol. Evol. **1:** 371–389.

SCHROEDER, W. A., T. H. J. HUISMAN, J. R. SHELTON, J. B. SHELTON, E. F. KLEIHAUER, A. M. DOZY and B. ROBBERTSON, 1968 Evidence for multiple structural genes for the γ chain of human fetal hemoglobin. Proc. Natl. Acad. Sci. USA **60:** 537–544.

SHEN, P., and H. V. HUANG, 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics **112:** 441–457.

SHEN, S., J. L. SLIGHTOM and O. SMITHIES, 1981 A history of the human fetal globin gene duplication. Cell **26:** 191–203.

SHIMASAKI, S., and I. IUCHI, 1986 Diversity of human γ-globin gene loci including a quadruplicated arrangement. Blood **67:** 784–788.

SHIMIZU, K., T. HARANO, K. HARANO, S. MIWA, Y. AMENOMORI, Y. OHBA, F. KUTLAR and T. H. J. HUISMAN, 1986 Abnormal arrangements in the a and γ globin gene clusters in a relatively large group of Japanese newborns. Am. J. Hum. Genet. **38:** 45–58.

SINGER, B. S., L. GOLD, P. GAUSS, and D. H. DOUGHERTY, 1982 Determination of the amount of homology required for recombination in bacteriophage T4. Cell **31:** 25–33.

SLIGHTOM, J. L., A. E. BLECHL and O. SMITHIES, 1980 Human fetal G-γ and A-γ globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21:** 627–638.

SLIGHTOM, J. L., L.-Y. E. CHANG, B. F. KOOP and M. GOODMAN, 1985 Chimpanzee fetal G-γ and A-γ globin gene nucleotide sequences provide further evidence of gene conversion in homonine evolution. Mol. Biol. Evol. **2:** 370–389.

SLIGHTOM, J. L., T. W. THEISEN, B. F. KOOP and M. GOODMAN, 1987 Orangutan fetal globin genes: nucleotide sequences reveal multiple gene conversions during hominid phylogeny. J. Biol. Chem. **262:** 7472–7483.

SMITH, D. H., J. B. CLEGG, D. J. WEATHERALL and H. M. GILLES , 1973 Hereditary persistence of foetal haemoglobin associated with a gamma-beta fusion variant, Haemoglobin Kenya. Nature New Biol. **246:** 184–186.

SMITHIES, O., 1964 Chromosomal rearrangements and protein structure. Cold Spring Harbor Symp. Quant. Biol. **29:** 309–319.

SMITHIES, O., G. E. CONNELL and G. H. DIXON, 1962 Chromosomal rearrangements and the evolution of haptoglobin genes. Nature **196:** 232–236.

STOECKERT, C. J., JR., F. S. COLLINS and S. M. WEISSMAN, 1984 Human fetal globin DNA sequences suggest novel conversion event. Nucleic Acids Res. **12:** 4469–4479.

STRUHL, K., 1987 Effect of deletion and insertion on double-strand-break repair in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **7:** 1300–1303.

STOLLE, C. A., L. A. PENNY, S. IVORY, B. G. FORGET and E. J. BENZ, JR., 1990 Sequence analysis of the gamma-globin gene locus from a patient with the deletion form of hereditary persistence of fetal hemoglobin. Blood **75:** 499–504.

SUKUMARAN, P. K., T. NAKATSUJI, M. B. GARDINER, A. L. REESE, J. G. GILMAN and T. H. J. HUISMAN, 1983 γ-Thalassemia resulting from the deletion of a γ globin gene. Nucleic Acids Res. **11:**4635–4643.

SZOSTAK, J. W., T. L. ORR-WEAVER, R. J. ROTHSTEIN and F. W. STAHL, 1983 The double-strand-break repair model for recombination. Cell **33:** 25–35.

TITUS, E. A., Y. E. HSIA and J. A. HUNT, 1988 Alpha-thalassemia screening reveals quadruple zeta-globin genes in a Laotian family. Hemoglobin **12:** 539–50.

TRENT, R. J., D. K. BOWDEN, J. M. OLD, J. S. WAINSCOAT, J. B. CLEGG and D. J. WEATHERALL, 1981 A novel rearrangement of the human β-like globin gene cluster. Nucleic Acids Res. **9:** 6723–6733.

TRENT, R. J., K. N. P. MICHLESON, T. WILKINSON, J. YAKAS, M. W. DIXON, P. J. HILL and H. KRONENBERG, 1986 Globin genes in Polynesians have many rearrangements including a recently described γ-γ-γ-γ. Am. J. Hum. Genet. **39:** 350–360.

TRIANTAPHYLLIDES, C. D., A. KOUVATSI and L. KAPLANOGLOU, 1983 The genetic distances between the inhabitants of nine Mediterranean countries and the three major human races. Hum. Hered. **33:** 137–139.

VANIN, E. F., P. S. HENTHORN, D. KIOUSSIS, F. GROSVELD and O. SMITHIES, 1983 Unexpected relationships between four large deletions in the human β-globin gene cluster. Cell **35:** 701–709.

WALDMAN, A. S., and R. M. LISKAY, 1987 Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc. Natl. Acad. Sci. USA **84:** 5340–5344.

WALDMAN, A. S., and R. M. LISKAY, 1988 Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol. Cell. Biol. **8:** 5350–5357.

WATT, V. M., C. J. INGLES, M. S. URDEA and W. J. RUTTER, 1985 Homology requirements for recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82:** 4768–4772.

WINICHAGOON P., D. R. HIGGS, S. E. Y. GOODBOURN, J. LAMB, J. B. CLEGG and D. J. WEATHERALL, 1982 Multiple arrangements of the human embryonic zeta globin genes. Nucleic Acids Res. **10:** 5853–5868.

WOOD, W. I., J. GITSCHIER, L. A. LASKY and R. M. LAWN, 1985 Base composition-independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA **82:** 1585–1588.

ZIEMER, M. A., W. F. SWAIN, W. J. RUTTER, S. CLEMENTS, D. K. ANN and D. M. CARLSON, 1984 Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homologies of cat and human proline-rich proteins. J. Biol. Chem. **25:** 10475–10480.

ZIMRAN, A., J. SORGE, E. GROSS, M. KUBITZ, C. WEST and E. BEUTLER, 1990 A glucocerebrosidase fusion gene in Gaucher disease. J. Clin. Invest. **85:** 219–222.

Communicating editor: P. J. PUKKILA