Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 6085–6089, October 1984
Genetics

Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster

(restriction fragment length polymorphism/haplotypes/nonrandom association/gene deletion)

ARAVINDA CHAKRAVARTI*, JOHN A. PHILLIPS III†, KENNETH H. MELLITS†, KENNETH H. BUETOW*, AND PETER H. SEEBURG‡

*Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261; †Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205; and ‡Department of Molecular Biology, Genentech, Inc., South San Francisco, CA 94080

Communicated by Victor A. McKusick, May 14, 1984

ABSTRACT **Six restriction fragment length polymorphisms (RFLPs) detected in the human growth hormone–human chorionic somatomammotropin (hGH–hCS) gene cluster were studied in Mediterraneans, Northern Europeans, and American Blacks; the polymorphisms showed that, on the average, one of 500 bases in this cluster is variant. Haplotypes constructed for four of these RFLPs display strong nonrandom associations. However, the strongest associations were between RFLPs that are in homologous DNAs rather than between the physically closest RFLPs. From this and other evidence we argue that duplication of an ancestral *hCS* gene occurred at least twice, the second event being relatively recent. In other words, duplication of the *hCS-L* gene to produce the *hCS-A* gene occurred twice, so that *hCS-A* genes in humans may have independent origins. Our results imply that chromosomes with absent *hCS* genes (leading to hCS deficiency) may represent the nonduplicated ancestral unit rather than gene deletions.**

The human growth hormone gene cluster located on chromosome 17 (1, 2) spans about 50 kilobase pairs (kb) of DNA and contains the genes encoding growth hormone (hGH) and chorionic somatomammotropin (hCS). Recently, the order and arrangement of the five loci [5′-*hGH-N*–*hCS-L*–*hCS-A*–*hGH-V*–*hCS-B*-3′ (Fig. 1)] composing this cluster has been elucidated by using cosmid clones (3). The highly homologous coding sequences of these genes are thought to have arisen by repeated duplications of an ancestral gene (3–6). Barsh *et al.* (3) have recently shown that the 5′ flanking DNA is homologous for all five loci, while the 3′ flanking sequences are homologous for *hGH-N* and *hGH-V* or the *hCS-L*, *hCS-A*, and *hCS-B* loci (3).

By using ^{32}P-labeled cDNA probes for *hGH* genes, six restriction fragment length polymorphisms (RFLPs) have been discovered in DNA samples from Blacks, Mediterraneans, and North Europeans (7). Of these six RFLPs, five (two *Bgl* II, two *Msp* I, and one *Hin*cII) were common to all three populations. The sixth, a *Bam*HI polymorphism, was detected only in U.S. Blacks. All of these RFLPs are tightly linked, in agreement with the expectation that only about 0.05% recombination should occur over 50 kb of DNA (8). This situation is thus similar to that observed in the human β-globin cluster (9).

Analysis of these RFLPs in the *hGH* cluster enabled us to study the nucleotide variability, nonrandom association, and evolution of the gene cluster. From the patterns of nonrandom associations between the RFLPs we have obtained evidence that duplication of an ancestral *hCS* gene creating the *hCS-A* gene occurred on two different chromosomes. The second duplication event leading to *hCS-A* is relatively recent. This implies that hCS deficiency due to absent *hCS* genes (10, 11) may not be due to *hCS* gene deletions *per se* but rather represents the nonduplicated ancestral unit.

METHODS

Subjects. The individuals studied were broadly grouped into three classes (U.S. Black, Mediterranean, and North European). The first two groups included couples who desired prenatal diagnosis for sickle cell anemia or β-thalassemia among their offspring. All such Greek and Italian families were grouped as Mediterraneans. The North Europeans included individuals of Northern European descent who had isolated growth hormone deficiency (IGHD) type 1 segregating among their offspring but were themselves unaffected. In these families, the IGHD type 1 phenotype was previously shown by linkage analysis not to involve the hGH gene cluster (7).

The sampling unit in our study was a nuclear family in which all parents and their offspring were typed for the *Bgl* II, *Hin*cII, and *Msp* I RFLPs. Studies of the inheritance patterns of the RFLPs usually enabled a complete description of parental haplotypes (7). Since none of the parents were related, we could obtain at most four independent haplotypes from each family studied.

Nuclear DNA Preparation. High molecular weight DNA was prepared from cultured amniotic fluid cells or from leukocytes in 10–15 ml of peripheral blood as described (12).

Preparation of Probes and Restriction Endonuclease Analyses. The recombinant plasmid chGH800/pBR322 contains nearly full-length complementary DNA (cDNA) to hGH mRNA (13). The 800-base-pair (bp) hGH cDNA insert was isolated and labeled with ^{32}P as described (7). Samples of nuclear DNA were digested to completion with various restriction endonucleases and the resulting DNA fragments were subjected to electrophoresis in 1% (wt/vol) agarose gels. The DNA fragments were then transferred to nitrocellulose filters and hybridized to the hGH probe. The filters were then washed and autoradiographed as previously described (7).

RFLPs. The six RFLPs examined in the three populations were termed *Bgl* II A and B (13/10.5 kb or 8.1/3.0 kb), *Hin*cII (6.7/4.5 kb), *Msp* I A and B (4.3/3.6 kb or 3.9/3.3 kb) (7), and *Bam*HI (6.7/3.8 kb); the latter was a private RFLP detected only in U.S. Blacks. For the description of each polymorphism, the longer fragment is denoted by a − and the shorter fragment by a +, which correspond to the absence or presence of a restriction site at a particular DNA location. The relative positions of the six different polymorphic sites with respect to the genes in the *hGH* cluster are shown in Fig. 1 (7, 10, 11). In the case of the *Bam*HI, *Bgl* II, and *Msp* I

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: hGH, human growth hormone; hCS, human chorionic somatomammotropin; RFLPs, restriction fragment length polymorphisms; kb, kilobase pair(s); bp, base pair(s).

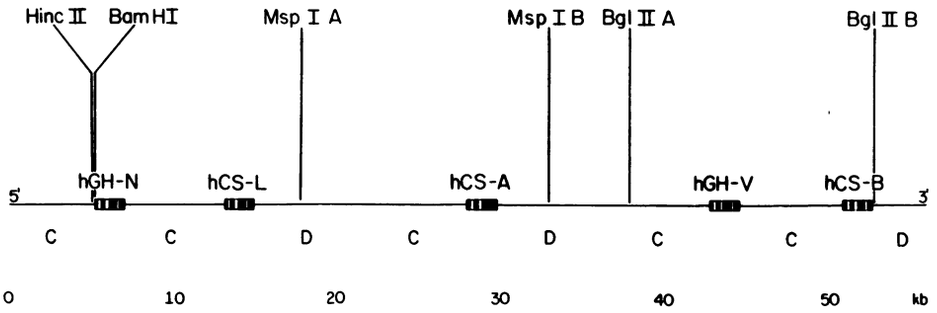

FIG. 1. Physical arrangement of the *hGH* gene cluster, showing the location of the six polymorphic restriction sites detected with hGH cDNA as a probe (3, 7, 10, 11). The relative locations of two flanking homologous sequence elements (*C* and *D*) are shown below the line (3). Note that the *Hinc*II, *Bgl* II A, and *Msp* I A and B sites lie within homologous *C* or *D* regions, respectively.

RFLPs, the structural locus adjacent to the polymorphic restriction site was determined by digesting large genomic *Hin*dIII fragments with each endonuclease (7). The distance of the polymorphic site from its adjacent locus was deduced by comparing the fragment lengths observed with the DNA sequence of each *hGH* or *hCS* locus and the physical map of the cluster (refs. 3, 7, 10, and 11 and unpublished data). The location of the *Hinc*II RFLP was known from prior experiments (7).

RESULTS

We were able to construct haplotypes for most couples in our study by examining the markers in their offspring or other relatives. In situations in which both parents and offspring were heterozygous at a site and no other relatives were available, the haplotype phase was ambiguous. In some cases, not all markers could be studied. The *Bgl* II B RFLP was not used because of our inability to distinguish between +− and ++ genotypes. The *Bam*HI RFLP was also not included because it occurred only in U.S. Blacks. Given these limitations, the distribution of observed haplotypes for four RFLPs in these populations is shown in Table 1, with the first eight rows giving the frequency data for complete haplotypes. These are the basic data on which our analyses are performed.

Variation in Restriction Sites. The data in Table 1 were used to compute the frequency of each site in each population. If n chromosomes are examined of which r are +, then the estimate of the frequency p of the + site is $\hat{p} = r/n$, with binomial variance $V(\hat{p}) = p(1 - p)/n$. $\hat{p}$ is a maximum-likelihood estimate and is identical to that obtained by gene counting among genotypes (14). In Table 2 we present values of n, $\hat{p}$, and its standard error $[V(\hat{p})]^{1/2}$. The frequency of the *Bam*HI RFLP was determined by examining 86 chromosomes of U.S. Blacks, 70 of which had the restriction site (i.e., +), and 30 chromosomes of Northern Europeans, of which all were of type +. This RFLP was not studied in the Mediterraneans. The data in Table 2 demonstrate that while all sites are highly polymorphic and that the frequencies of the *Bgl* II A and *Hinc*II sites are similar in all populations, the frequencies of the *Msp* I and *Bam*HI sites differ between populations.

The extent of polymorphism was evaluated by computing the gene diversity (heterozygosity h) for all three groups. For each group $\hat{h}_i = 2p_i(1 - p_i)$ is the estimate of heterozygosity for the ith site, with the average over m restriction sites being given by $\bar{h} = \Sigma\hat{h}_i/m$. In Table 2 we present the $\hat{h}_i$ values for each RFLP in each population. These values range from 0.20 to ≈0.50, with the average heterozygosities over polymorphic sites being 0.30, 0.45, and 0.45, in the U.S. Black, Mediterranean, and Northern European groups, respectively.

Nonrandom Association Between Sites. Since the *hGH* cluster contains duplicated genes and, moreover, since the polymorphic restriction sites are in regions (Fig. 1) previously shown to have homology (3), it was of interest to study the nonrandom associations (linkage disequilibrium) between them. For this purpose, we used the haplotype data presented in Table 1 and considered only the *Hinc*II, *Msp* I A and B, and *Bgl* II A sites. Since linkage disequilibrium can be studied for any pair of RFLPs there are six measurable associations in any population.

The association between a pair of polymorphic sites is measured by the Δ value, defined as,

$$\Delta = \frac{p_{11} - p_{10}p_{01}}{(p_{10}q_{10}p_{01}q_{01})^{1/2}}, \qquad [1]$$

Table 1. Distribution of haplotypes in the *hGH* cluster in three populations

Haplotype				No. of occurrences in		
*Hinc*II	*Msp* I A	*Msp* I B	*Bgl* II A	U.S. Black	Mediterranean	N. European
+	+	+	+	4		1
+	+	+	−	3	7	6
+	−	−	+	1		1
−	+	+	+	12	2	4
−	+	−	+			3
−	+	+	−	1		1
−	−	+	+	1		1
−	−	−	+	2	7	13
			Total	24	16	30
(+)	+	+	+			2
+	+	(+)	−			1
(−)	+	+	−			1
(−)	−	−	+			1
−	−	(−)	+			1
+	+	+				2
−	−	−				2
	+	+	+	1	2	
	+	+	−	1	1	
	−	−	+		1	
+			+	1		1
+			−	3		2
−			+	12	6	5
−			−	2		
	+	+				3
	−	−				1
−				2		
			+	2	2	
			Total	48	28	52

A parenthesis implies that the assignment of + or − with respect to the rest of the haplotype is ambiguous (see *Results*). However, it is still useful for computing the frequency of + at an *individual* RFLP.

Table 2. Estimates of frequencies, standard errors, and heterozygosities of RFLPs in the *hGH* gene cluster

RFLP	Value of	U.S. Black	Mediterranean	N. European
*Hin*cII	*p*	0.273 ± 0.067	0.218 ± 0.099	0.333 ± 0.068
	n	44	22	48
	h	0.397	0.434	0.444
Msp I A	*p*	0.846 ± 0.071	0.600 ± 0.110	0.546 ± 0.075
	n	26	20	44
	h	0.260	0.480	0.496
Msp I B	*p*	0.885 ± 0.063	0.600 ± 0.110	0.500 ± 0.075
	n	26	20	44
	h	0.204	0.480	0.500
Bgl II A	*p*	0.783 ± 0.061	0.714 ± 0.085	0.750 ± 0.065
	n	46	28	44
	h	0.340	0.408	0.375
*Bam*HI	*p*	0.814 ± 0.042	—	1.000 ± 0.000
	n	86	—	30
	h	0.303	—	0.000
$\bar{h}$ (above sites)		0.301	0.451	0.454
$\bar{h}$ (all 98 sites)		0.016	0.020	0.025
% polymorphic sites		0.054	0.043	0.043

For each RFLP, line 1 gives the frequency ± standard error of the + site; line 2 is the number of chromosomes sampled; and line 3 is the heterozygosity.

where p_{11}, p_{12}, p_{21}, and p_{22} are the frequencies of ++, +−, −+, and −− haplotypes and $p_{10} = p_{11} + p_{12}$, $q_{10} = 1 - p_{10}$, $p_{01} = p_{11} + p_{21}$, and $q_{01} = 1 - p_{01}$. Since Δ is a correlation coefficient the hypothesis of linkage equilibrium ($\Delta = 0$) can be tested using the statistic $z = \tanh^{-1}\Delta$. Then, $(n - 3)^{1/2}z$ is a unit normal deviate under the null hypothesis $\Delta = 0$ (ref. 15).

Table 3 presents the calculated Δ values for pairwise comparisons of the four RFLPs in the three groups. All pairs are highly significant at the 1% level in the Mediterranean and Northern European groups but the disequilibrium is lower in the Blacks. In the Blacks, only the pairs *Hin*cII–*Bgl* II A and *Msp* I A–*Msp* I B are significantly associated.

Nucleotide Variability in the *hGH* Gene Cluster. Of the 16 enzymes used, only 4 detected RFLPs. Since no abnormal hybridizing fragments were detected, the RFLPs studied are probably all due to base substitutions rather than large insertions or deletions (9, 16). Assuming that all RFLPs are base substitutions and selectively neutral, it is possible to estimate the nucleotide diversity π (heterozygosity at a nucleotide site) in the *hGH* cluster by using the procedure of Ewens *et al.* (17). In this method, if the *i*th restriction endonuclease used had a recognition site of r_i bp, and if in a sample of n_i chromosomes, m_i cleavage sites were observed of which k_i were polymorphic, then the estimate of nucleotide diversity is

$$\hat{\pi} = \frac{\hat{\theta}}{1 + \hat{\theta}}, \text{ where } \hat{\theta} = \frac{\Sigma k_i}{2\Sigma r_i m_i \ln n_i}, \quad [2]$$

Table 3. Standardized pairwise linkage disequilibrium values (Δ) in three populations

Site 1	Site 2	U.S. Black	Mediterranean	N. European
*Hin*cII	*Msp* I A	+0.08 (24)	+0.78* (16)	+0.54* (36)
*Hin*cII	*Msp* I B	+0.00 (24)	+0.78* (16)	+0.60* (34)
*Hin*cII	*Bgl* II A	−0.44* (42)	−1.00* (22)	−0.75* (40)
Msp I A	*Msp* I B	+0.85* (26)	+1.00* (20)	+0.81* (42)
Msp I A	*Bgl* II A	−0.21 (26)	−0.67* (20)	−0.55* (36)
Msp I B	*Bgl* II A	−0.18 (26)	−0.67* (20)	−0.59* (34)

Figures in parenthesis are number of haplotypes.
*Significant at 1% level.

Table 4. Restriction enzymes used for detecting polymorphisms

			U.S. Black		N. European	
Enzyme	*r*	*f*	*k*	*n*	*k*	*n*
Dde I	4	3	—	—	0	16
Hae III	4	1	—	—	0	12
Msp I	4	7	2	26	2	54
*Sau*3A I	4	5	—	—	0	34
Ava II	5	2	—	—	0	16
*Hin*cII	5	4	1	59	1	58
*Bam*HI	6	6	1	86	0	30
Bgl I	6	5	—	—	0	36
Bgl II	6	6	2	61	2	60
*Eco*RI	6	6	0	6	0	16
*Hin*dIII	6	3	0	2	0	20
Hpa I	6	1	—	—	0	6
Pst I	6	3	0	2	0	16
Pvu II	6	2	0	6	0	36
Sst I	6	4	—	—	0	14
Xba I	6	3	—	—	0	16

where $\theta = 4N_e u$, N_e is the effective population size, and u is the average mutation rate per bp (16). Usually, m cannot be directly observed, but if f fragments were consistently scored then m has upper limit $2f$ (none of the cut sites are contiguous) and lower limit $f + 1$ (all cut sites are contiguous). We use m as the average $(3f + 1)/2$. The 16 enzymes studied in Northern Europeans and the 8 studied in U.S. Blacks together with the m, k, and n values are presented in Table 4. The data on the Mediterraneans were too sparse to be included in this analysis. Since several endonucleases used to screen the U.S. Blacks were chosen because they were discovered to yield RFLPs in Northern Europeans, the data on the Blacks are biased. In any case, Eq. 2 gives $\hat{\pi}$ = 0.0014 for Northern Europeans and $\hat{\pi}$ = 0.0032 for U.S. Blacks. Both of these values are of the same order as the $\hat{\pi}$ = 0.0017 calculated for the human β-globin cluster (9, 18). Finally, for comparative purposes we selected only those endonucleases studied in Northern Europeans that were also studied in the Blacks. This analysis gave $\hat{\pi}$ = 0.0022, a value closer to the U.S. Black sample.

DISCUSSION

Polymorphism in the *hGH* Cluster. The frequency data on individual RFLPs in the three groups or the heterozygosities at these RFLPs (Table 2) show that considerable variation exists within the *hGH* cluster. Table 2 also demonstrates that despite the occurrence of a private polymorphism in the U.S. Black sample, the Mediterraneans and Northern Europeans have higher average heterozygosities. One should note that in our study an estimated total of 98 restriction sites (Table 4) were screened, using 16 restriction endonucleases, and 5 sites were found to be polymorphic. Since the majority of these restriction sites were monomorphic, it is more realistic to compute the average heterozygosity ($\bar{h}$) over all 98 sites. Then $\bar{h}$ is approximately 1.6–2.5%, a figure far below the 10% average heterozygosity found for protein and blood group loci in humans (19). Furthermore, the percentage of polymorphic sites was only 4.3–5.4%, which is also lower than the corresponding figure of 31–40% obtained for protein and blood group loci (19). These comparisons are valid since both polymorphic and monomorphic RFLPs and proteins/blood groups were compared. Nevertheless, the RFLPs all occurred in noncoding DNA, whereas the classical markers represent polymorphisms in coding sequences. However, the values for the hGH cluster are similar to those obtained for the human β-globin cluster, as were the calculations of the nucleotide diversity ($\pi \approx 0.002$) (9, 18). This implies that approximately 1 in 500 nucleotides is variant, so that the

hGH cluster should contain about 100 such variants. Since the chance is low that any of these variants would yield an RFLP (16, 17) it follows that a large number of restriction enzymes have to be screened to detect an RFLP. In contrast, the increased efficiency of electrophoretic/immunological detection of heterozygosity of proteins/blood groups probably accounts for the greater variability at the latter loci.

Patterns of Polymorphism and Linkage Disequilibrium. An essential feature of the *hGH* RFLPs is that the *Hin*cII and *Bgl* II A sites are 5′ to *hGH-N* and *-V*, respectively, while the *Msp* I A and *Msp* I B sites are 3′ to *hCS-L* and *hCS-A* (Fig. 1). Barsh *et al.* (3) have recently shown that the 5′ DNA sequences adjacent to *hGH-N* and *-V* are highly homologous, as are the 3′ flanking regions adjacent to the hCS genes. Thus, the *Hin*cII and *Bgl* II A RFLPs are in homologous 5′ DNA regions, while the two *Msp* I RFLPs are in homologous 3′ regions. Interestingly, the frequency of *Hin*cII + is similar to the frequency of *Bgl* II A − in all three groups, as are the frequencies of + sites for *Msp* I A and B (Table 2). Thus, the frequency changes for the above pairs of RFLPs seem to be occurring in concert. Furthermore, the strongest associations are between those RFLPs that are in regions that share homology (*Hin*cII and *Bgl* II A, or *Msp* I A and B) rather than between RFLPs that are physically the closest. Finally, the direction (sign) and *relative magnitude* of linkage disequilibrium are similar in all three groups, suggesting that the RFLPs are sufficiently ancient to predate the divergence of the major races of man.

Linkage Disequilibrium and Duplication Events. Having a set of closely linked markers, it was not surprising to observe strong linkage disequilibria between all sites in the Caucasian populations (20). In contrast, a marked reduction of associations was found in the U.S. Blacks. This is possibly due to recent admixture (21), as also observed for RFLPs in the β-globin cluster (18). As explained above, the two strongest associations were between sites that are physically separated but evolutionarily related by virtue of gene duplication (*Hin*cII–*Bgl* II A and *Msp* I A–B) (3). The recent studies of Barsh *et al.* (3) suggest that an ancestral duplication of an *hGH-hCS* unit initially created the cluster *5′-hGH-N, hCS-L, hGH-V, hCS-B-3′*. Subsequently, the *hCS-A* gene appeared by unequal crossing over between repeated *Alu* elements in the hGH cluster. This led us to ask the question: If the *Msp* I sites are related and presumably homologous due to the duplication process, is there a high frequency of ++ or −− chromosomes (depending on whether the duplication occurred on an *Msp* I + or − chromosome)?

In spite of the limited sample of haplotypes that were available, our data provide interesting answers to this question. First, we computed the frequency of *Msp* I A–B ++ and −− chromosomes from Table 1 and found them to be 96%, 100%, and 90% of all chromosomes in the U.S. Black, Mediterranean, and Northern European groups. If this concordance resulted from a single rightward duplication of an ancestral *hCS* gene (*hCS-L*) then either the ++ or −− chromosomes would predominate. The data in Table 1 show that nearly 85% of U.S. Black chromosomes are ++, with only 12% being −−. On the other hand, in the Mediterraneans 60% are ++, 40% are −−, whereas the Northern Europeans show 48% ++ and 43% −−. These findings suggest that there were at least two independent duplications of the *hCS-L* gene leading to the *hCS-A* genes. The low frequency of −− chromosomes in the U.S. Black population suggests that the −− chromosome may have arisen later in Blacks than in Caucasians or arose in a Caucasian population after the divergence of the major races of man and was introduced into the Black population by recent gene admixture.

To test the concept of multiple independent origins of the *hCS-A* gene we next compared the distribution of the *Hin*cII and *Bgl* II A sites on *Msp* I A–B ++ and −− chromosomes. These data, derived from Table 1, are shown in Table 5. Interestingly, the *Bgl* II A site is polymorphic only on *Msp* I A–B ++ chromosomes; on other chromosomes *Bgl* II A is always +. Furthermore, an excess of repulsion (+− and −+) types for the variant *Hin*cII and *Bgl* II A sites is seen on *Msp* I A–B ++ chromosomes. In the Blacks, the nonrandom association between these *Hin*cII and *Bgl* II A sites is not significant ($\chi^2 = 3.52$, 1 df, $0.05 < P < 0.10$), but in the Caucasians (Northern Europeans and Mediterraneans) it is highly significant ($\chi^2 = 12.96$, 1 df, $P < 0.0005$). Thus, the distributions of the *Hin*cII and *Bgl* II A sites are very different on the *Msp* I A–B ++ and −− chromosomes. These findings provide additional evidence for independent origins of the *hCS-A* gene and thus the *hGH* cluster.

Note from Table 5 that nearly all individuals (22/24) who are *Msp* I A–B −− are *Hin*cII–*Bgl* II A −+, while these sites vary on *Msp* I A–B ++ chromosomes. This suggests that the duplication events yielded *Hin*cII, *Msp* I A, *Msp* I B, *Bgl* II A −−−+ and ±++± chromosomes.

The ages of these two types of ancestral duplications can be estimated from the frequency of the *Hin*cII + site on *Msp* I A–B ++ (0.70) and −− (0.23) chromosomes. By the method of Kimura and Ohta (22), the average age of the *Hin*cII polymorphism is 2.8 N_e for the ±++± duplication and 0.9 N_e for the −−−+ duplication, where N_e is the effective population size. This suggests that the younger duplication that yielded −−−+ is only ⅓ as old as the original duplication that produced the ±++± haplotype. Since the frequency of the *Hin*cII RFLP is significantly different on these two types of chromosomes, it is unlikely that gene conversion was responsible. In this case, the frequency of the *Hin*cII site would be expected to be the same.

From the above, several testable predictions can be made. First, the *Msp* I A–B +− or −+ chromosomes arose through recombination. Recombinations between ancestral chromosomes yielding *Msp* I A–B +− chromosomes could be either ++−+ or −+−+, of which only −+−+ has been observed. However, those yielding *Msp* I A–B −+ chromosomes would be of the type −−+±, of which both −−++ and −−+− have been observed. Second, since the predominant ancestral chromosome was *Msp* I A–B ++, we should observe this chromosome in all human populations. The six

Table 5. Distribution of *Hin*cII and *Bgl* II A sites on *Msp* I A and B ++, +−, −+, and −− chromosomes

*Hin*cII, *Bgl* II A	*Msp* I A and B sites											
	U.S. Black				Mediterranean				N. European			
	++	−−	+−	−+	++	−−	+−	−+	++	−−	+−	−+
++	4	1	0	0	0	0	0	0	1	1	0	0
+−	3	0	0	0	7	0	0	0	6	0	0	0
−+	12	2	0	1	2	7	0	0	4	13	3	1
−−	1	0	0	0	0	0	0	0	1	0	0	0
Totals	20	3	0	1	9	7	0	0	12	14	3	1

chromosomes we studied in the Japanese were four of the type +++− and one each of −++− and −−−+, while two Asian Indian chromosomes we studied were −+++ and ++++. Third, and perhaps the most crucial prediction, we posit that since one of the duplications occurred relatively recently, unduplicated chromosomes may still exist. These chromosomes should lack the *hCS-A* gene and probably be on a *Hin*cII–*Msp* I A ±+ chromosome (Table 5). Wurzel *et al.* (10) described a family with deletion of *hCS* genes in the 3′ end of the cluster having a ++ *Hin*cII–*Msp* I A haplotype. This suggests that such cases of "hCS gene deletions" may not represent deletions *per se* but may represent remnants of chromosomes that have yet to obtain the duplicated *hCS-A* gene segment *de novo* or by recombination. The presence of over 20 *Alu* sequences in the *hGH* cluster may be a factor promoting such repeated duplications (3).

Deletions of the *hGH-N* gene are, of course, another matter. We presume that most of them occurred *de novo* as evidenced by the fact that *hGH-N* gene deletions may be associated with either the *Msp* I A–B ++ or −− chromosomes in Europe; in Japan, however, it is associated with a ++ chromosome (23). Such spontaneous deletions may also be promoted by the numerous *Alu* sequences present in the *hGH* gene cluster (3).

A final conclusion of our study is that the heterogeneity observed is caused by multiple independent origins of the *hGH* cluster. This is the probable cause of the greater nonrandom associations observed in Caucasians as compared to U.S. Blacks. Furthermore, our finding that the *Msp* I A–B −− haplotype is generally restricted to Caucasians suggests that it arose after the divergence of the major races of man, that is, within the last 115,000 years (19). These questions can be more thoroughly answered by DNA sequencing to give precise information regarding the extent of the duplication units and the origin of the *hGH* and *hCS* genes and to more completely examine the heterogeneity within the duplicated segments.

This research was supported by Grants AM13983 and AM28246, and Research Career Development Award AM00958 from the National Institutes of Health.

1. Owerbach, D., Rutter, W. J., Martial, J. A., Baxter, J. D. & Shows, T. B. (1980) *Science* **209**, 289–292.
2. George, D. L., Phillips, J. A., III, Francke, U. & Seeburg, P. H. (1981) *Hum. Genet.* **57**, 138–141.
3. Barsh, G. S., Seeburg, P. H. & Gelinas, R. E. (1983) *Nucleic Acids Res.* **11**, 3939–3958.
4. Fiddes, J. C., Seeburg, P. H., Denoto, F. M., Hallewell, R. A., Baxter, J. D. & Goodman, H. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4294–4298.
5. Goodman, H. M., Denoto, F. M., Fiddes, J. C., Hallewell, R. A., Page, G. S., Smith, S. & Tischer, E. (1980) in *Miami Winter Symposium: Mobilization and Reassembly of Genetic Information*, eds. Scott, W. A., Werner, R., Joseph, D. R. & Schultz, J. (Academic, New York), Vol. 17, pp. 155–179.
6. Seeburg, P. H. (1980) in *Polypeptide Hormones*, eds. Beers, R. F. & Bassett, E. G. (Raven, New York), pp. 19–31.
7. Phillips, J. A., III, Parks, J. S., Hjelle, B. L., Herd, J. E., Plotnick, L. P., Migeon, C. J. & Seeburg, P. H. (1982) *J. Clin. Invest.* **70**, 489–495.
8. Kurnit, D. M. & Hoehn, H. (1979) *Annu. Rev. Genet.* **13**, 235–258.
9. Kazazian, H. H., Chakravarti, A., Orkin, S. H. & Antonarakis, S. E. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 137–146.
10. Wurzel, J. M., Parks, J. S., Herd, J. E. & Nielsen, P. V. (1982) *DNA* **1**, 251–257.
11. Parks, J. S. & Phillips, J. A., III (1984) in *Human Molecular Genetics*, ed. Ramirez, F. (Dekker, New York), in press.
12. Kunkel, L. M., Smith, K. D., Boyer, S. H., Borgaonkar, D. S., Wachtel, S. S., Miller, O. J., Breg, W. R., Jones, H. W. & Rary, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1245–1249.
13. Martial, J. A., Hallewell, R. A., Baxter, J. D. & Goodman, H. M. (1979) *Science* **205**, 602–607.
14. Ceppellini, R., Siniscalco, M. & Smith, C. A. B. (1955) *Ann. Hum. Genet. (London)* **20**, 97–115.
15. Rao, C. R. (1965) *Linear Statistical Inference and its Applications* (Wiley, New York).
16. Nei, M. & Tajima, F. (1981) *Genetics* **97**, 145–163.
17. Ewens, W. J., Spielman, R. S. & Harris, H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3748–3750.
18. Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Boehm, C. D. & Kazazian, H. H. (1981) *Am. J. Hum. Genet.* **33**, 134A (abstr.).
19. Nei, M. & RoyChowdhury, A. K. (1974) *Am. J. Hum. Genet.* **26**, 421–443.
20. Li, C. C. (1978) *A First Course in Population Genetics* (Boxwood, Pacific Grove, CA).
21. Morton, N. E. (1982) *Outline of Genetic Epidemiology* (Karger, New York).
22. Kimura, M. & Ohta, T. (1973) *Genetics* **75**, 199–212.
23. Phillips, J. A., III (1983) in *Banbury Report 14: Recombinant DNA Applications to Human Disease*, eds. Caskey, C. T. & White, R. L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 305–315.