Am. J. Hum. Genet. 60:1439–1447, 1997

Haplotype Analysis of Hemochromatosis: Evaluation of Different Linkage-Disequilibrium Approaches and Evolution of Disease Chromosomes

Richard S. Ajioka,[1] Lynn B. Jorde,[2] Jeffrey R. Gruen,[3] Ping Yu,[1] Diana Dimitrova,[1] Jalene Barrow,[1] Evette Radisky,[1] Corwin Q. Edwards,[1] Linda M. Griffen,[1] and James P. Kushner[1]

[1]Division of Hematology/Oncology and [2]Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah Health Sciences Center, Salt Lake City; and [3]Department of Pediatrics, Yale University School of Medicine, New Haven

Summary

We applied several types of linkage-disequilibrium calculations to analyze the hereditary hemochromatosis (hh) locus. Twenty-four polymorphic markers in the major histocompatibility complex (MHC) class I region were used to haplotype hh and normal chromosomes. A total of 169 hh and 161 normal chromosomes were analyzed. Disequilibrium values were found to be high over an unusually large region beginning 150 kb centromeric of HLA-A and extending nearly 5 Mb telomeric of it. Recombination in this region was ∼28% of the expected value. This low level of recombination contributes to the unusually broad region of linkage disequilibrium found with hh. The strongest disequilibrium was found at locus HLA-H (δ = .84) and at locus D6S2239 (δ = .85), a marker ∼10 kb telomeric to HLA-H. All disequilibrium methods employed in this study found peak disequilibrium at HLA-H or D6S2239. The cys282tyr mutation in HLA-H, a candidate gene for hh, was found in 85% of disease chromosomes. A haplotype phylogeny for hh chromosomes was constructed and suggests that the mutation associated with the most common haplotype occurred relatively recently. The age of the hh mutation was estimated to be ∼60–70 generations. Disequilibrium was maintained over a greater distance for hh-carrying chromosomes, consistent with a recent mutation for hh. Our data provide a reasonable explanation for previous difficulties in localizing the hh locus and provide an evolutionary history for disease chromosomes.

Received December 3, 1996; accepted for publication March 18, 1997.

Address for correspondence and reprints: Dr. Richard S. Ajioka, Division of Hematology/Oncology, University of Utah Health Sciences Center, 50 North Medical Drive, Salt Lake City, UT 84132. E-mail: rich_ajioka@hlthsci.med.utah.edu

© 1997 by The American Society of Human Genetics. All rights reserved.

0002-9297/97/6006-0022$02.00

Introduction

Hereditary hemochromatosis (hh) is one of the most common inherited metabolic abnormalities affecting Caucasians, occurring with a frequency of ∼.005. It is inherited as an autosomal recessive trait, and the disease locus has been mapped to the short arm of chromosome 6, where it is closely linked to the HLA-A locus (Simon et al. 1977; Cartwright et al. 1978; Lipinski et al. 1978; Edwards et al. 1981). The defect responsible for hemochromatosis is malregulation of intestinal iron absorption. Hyperabsorption eventually leads to organ damage due to iron overload, and the liver, heart, endocrine organs, skin, and joints all may be involved. Surprisingly little is known, however, about either the genes responsible for intestinal iron absorption or their regulation. This has made a candidate approach for the identification of the hh gene difficult.

The strong allelic association between HLA serotype A3 and hh is well recognized (Simon et al. 1976; Dadone et al. 1982). Approximately 70% of patients with hemochromatosis possess at least one HLA-A3 alloantigen (Edwards et al. 1982). It has been suggested that the association between hemochromatosis and HLA-A3 is related to a founder (or ancestral) mutation that occurred on a chromosome bearing an A3 allele (Simon et al. 1987). What is now considered a disease mutation originally may have provided a nutritional advantage in an iron-poor environment. A recent study reported that heterozygous women were less likely to be anemic (Crawford et al. 1995). A3-bearing chromosomes could have increased in frequency along with the hemochromatosis gene, through a genetic "hitchhiking" effect (Smith and Haigh 1974). The finding of HLA-A alloantigens other than A3 in patients with hh has been explained by infrequent recombination events between HLA-A and the hemochromatosis locus over many generations (Simon et al. 1987).

The ancestral mutation hypothesis predicts that alleles of loci close to the hh locus should be overrepresented in hh chromosomes possessing the A3 allele. Indeed,

haplotype analysis has shown that specific alleles at loci within the class I region are in linkage disequilibrium with hh (Jazwinska et al. 1993, 1995; Worwood et al. 1994; Yaouanq et al. 1994; Raha-Chowdhury et al. 1995*a,* 1995*b*). These reports confirm strong linkage of hh with HLA-A3 but disagree with respect to a telomeric border for the region containing the hh locus.

Disequilibrium analysis is sensitive to the mutational history of genes in populations, and some disequilibrium statistics can be affected by allele-frequency variation at different marker loci (Devlin and Risch 1995). These factors might be responsible for the discrepancies in the linkage-disequilibrium peaks reported for hh. In an effort to account for these factors, we used several different analytic approaches to determine if we could define a consistent location for the hh locus. These analyses utilized several recently described polymorphisms, including HLA-H, a candidate for the hh gene (Feder et al. 1996). In addition, we used phylogenetic analysis to analyze the evolutionary history of marker haplotypes in the hh region.

Our data support the hypothesis of an "ancestral" mutation and provide a molecular haplotype for a founder hemochromatosis chromosome. Haplotype analysis enabled us to construct a phylogenetic tree for hh chromosomes and provided a likely scenario for the evolution of mutation-bearing chromosomes. When HLA-H and flanking polymorphisms were included in the disequilibrium calculations, HLA-H and the nearby locus D6S2239 exhibited the highest disequilibrium values, regardless of the calculation method employed. This suggests that previous analyses were hindered primarily by lack of marker density in a region characterized by a low recombination rate.

Subjects and Methods

Subjects

Methods of collection and use of human samples were approved by the institutional review board at the University of Utah. Eighty-five unrelated hemochromatosis homozygous probands were studied. All displayed laboratory evidence of iron overload, and most had undergone liver biopsy establishing the presence of hepatic iron overload (Edwards et al. 1988). Eighty-seven normal controls were selected either from spouses marrying into our pedigrees or from pedigree members sharing no HLA haplotype in common with the proband. All controls had normal values for percent saturation of transferrin and for serum ferritin concentration. The population studied consisted of Caucasians primarily from Utah and neighboring states. We have determined in the past that the Utah population does not differ genetically from other northern European populations (McLellan et al. 1984).

Characterization of Markers in the hh Region

A YAC extending approximately from HLA-E to HLA-G was a gift from Dr. Daniel Cohen (CEPH, Paris). The YAC (225B1) was gel purified by use of contour-clamped homogeneous-electric-field (CHEF) electrophoresis and was used to probe a cDNA library from human placenta (Stratagene). Nine non–class I cDNA clones were isolated, and three, designated "Y104," "Y129," and "Y158," displayed informative polymorphisms by Southern analysis.

Probes specific to various class I loci were a gift from Dr. Harry Orr. The probes and their specificities are as follows: 114.5.32 (HLA-C), B1.1EH.11 (HLA-E), and 5.4SH (HLA-F) (Koller et al. 1989) and pHLA-6p1 (HLA-G) (Geraghty et al. 1987).

The following microsatellite-repeat polymorphisms were used to construct haplotypes: D6S265, D6S306, D6S464, and D6S105 (Sood et al. 1981); D6S1260 (Raha-Chowdhury et al. 1995*b*); D6S1558, D6S1621, GATA-p19326, D6S1545, and D6S1691 (Whitehead Institute Database: http://www-genome.wit.mit.edu/); and D6S2231, D6S2238, D6S2239, and D6S2241 (Feder et al. 1996). Typing was performed by use of standard PCR conditions. Alleles were identified by electrophoresis on sequencing gels.

HLA-H is a reported candidate gene for hh (Feder et al. 1996). The mutation associated with disease chromosomes involves a G→A transition that creates both a *Sna*BI and *Rsa*I restriction site. Patient DNA samples were typed at this locus by PCR amplification followed by restriction-enzyme digestion and electrophoresis on 2% agarose gels. PCR primers were those described by Feder et al. (1996).

Linkage-Disequilibrium Analysis

Haplotype data were used to estimate the standardized linkage-disequilibrium coefficient, r (Hill and Robertson 1968). A second measure of disequilibrium, D', also was estimated (Lewontin 1964). This measure has the advantage that it is less dependent on allele frequencies than is r (Hedrick 1987; Lewontin 1988). In the estimation of r for multiallelic systems, the most frequent allele in the population was designated as one allele, and all remaining alleles were added together to form a second allele. The r statistic also was estimated by use of the relationship $r = \sqrt{(\chi^2/n)}$, where n is the number of chromosomes and the χ^2 value is estimated from a contingency table in which the rows correspond to the alleles at one locus and in which the columns correspond to alleles at the second locus. This procedure allows all alleles of a multiallelic system to be used, although some lumping still may be necessary to avoid small expected cell sizes. In addition to these measures, the δ statistic, originally formulated by Bengtsson and Thomson (1981), was used. Unlike r, this statistic is

independent of allele frequencies and is not influenced by oversampling of disease allele–bearing chromosomes in case-control studies (Devlin and Risch 1995).

Two relatively new linkage-disequilibrium approaches also were applied to our data set. The method of Kaplan et al. (1995) assumes that a disease is caused by a single mutation that arose in a founder population that subsequently grew at an exponential rate. This method avoids some of the equilibrium assumptions inherent in traditional linkage-disequilibrium approaches and provides an estimate of the recombination fraction, θ_K, between each marker and a putative disease-causing mutation. It employs a replicate sampling procedure to generate an empirical distribution of θ_K values from which 95% confidence limits can be estimated. One thousand replicate samples were generated for each marker. This method requires an estimate of the age of the disease-causing mutation, in order to calculate θ_K. On the basis of the mutation date–estimation procedure described below, it was conservatively assumed that the ancestral mutation occurred 100 generations in the past. The mutation-age estimate is not a critical component of this analysis, since we primarily are interested in relative marker order, rather than in the actual θ_K values.

Terwilliger's (1995) method also was used, to estimate a map location for the hemochromatosis locus. This method accommodates multiallelic markers and provides an estimate (λ) of the degree to which a marker allele is elevated on disease chromosomes, relative to the marker-allele frequency in the general population. This measure is thus very similar to the δ statistic. Terwilliger's method can be applied to multiple marker loci simultaneously, to derive a maximum-likelihood estimate of λ. Likelihood ratios then can be used to derive a support interval for the maximum-likelihood estimate (a support interval corresponding to 1,000:1 odds was used here).

The age of the hh mutation was estimated by use of the following equation:

$$g = \log(\delta)/\log(1 - \theta) , \quad (1)$$

where δ is the disequilibrium coefficient estimated for HLA-H and locus HLA-A, θ is the observed recombination rate between HLA-H and HLA-A, and g is the age of the hh mutation in generations (Risch et al. 1995). By incorporating θ in equation (1), this method adjusts for the recombination rate in a given physical region. Confidence limits for g were estimated by substituting the upper and lower confidence limits for θ in equation (1) (Risch et al. 1995). The confidence limits for θ were obtained by use of the binomial formula.

A Mantel matrix-comparison test was used to determine whether there was a significant correlation between the matrix of pairwise disequilibrium (r) values and the matrix of physical distances between each pair of markers. This method has been used in previous comparisons of physical distance and disequilibrium and is described elsewhere (Jorde et al. 1994).

To assess the evolutionary history of the hemochromatosis haplotypes, a distance matrix was formed in which the elements of the matrix contained the percentage of alleles that differed between each possible pair of 24-locus haplotypes (thus, a 169 × 169 matrix was formed for the 169 haplotypes). A haplotype phylogeny was then estimated by use of the neighbor-joining method (an option of the PHYLIP package [Felsenstein 1993]).

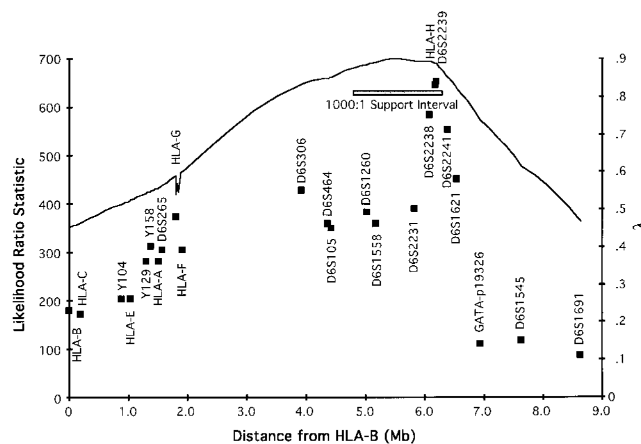

Figure 1 Plot of the likelihood-ratio statistic for multipoint linkage disequilibrium, by Terwilliger's (1995) method. The 1,000:1 support interval (*gray-shaded bar*) includes the HLA-H locus. Individual λ values also are shown (*black squares*).

Results

Characterization of Markers

Twenty-four polymorphic markers were analyzed on normal and hemochromatosis chromosomes. The markers and their relative physical locations are shown in figure 1. Physical distances between HLA class I markers are based on estimates from published data (Carroll et al. 1987; Lawrance et al. 1987; Geraghty et al. 1992; Gruen et al. 1992; Abderrahim et al. 1994). Physical placement of other markers is based on information derived from Southern analysis of contiguous YACs (data not shown) and on published data (Feder et al. 1996).

HLA-A and -B were characterized by serotype. Polymorphisms for the remaining class I loci were determined by use of specific DNA probes and Southern analysis. Alleles for microsatellite markers were determined by PCR amplification followed by electrophoresis on sequencing gels. Haplotypes were assigned by comparing genotypes from appropriate pedigree members.

Table 1

Comparison of Linkage-Disequilibrium Measures: Hemochromatosis versus 24 Polymorphisms

Marker[a]	*R* Lumped	*R* Multiallele	D'	δ	λ	θ_K
HLA-B (.35/.14)	.25	.32	.44	.25	.23	.016
HLA-C (.48/.65)	.19	.19	.26	.24	.22	.01
Y104 (.51/.71)	.2	.2	.26	.28	.26	.014
HLA-E (.76/.66)	.1	.1	.15	.27	.26	.016
Y129 (.83/.72)	.13	.13	.23	.38	.36	.004
Y158 (.47/.08)	.43	.43	.71	.42	.4	.012
HLA-A (.47/.12)	.37	.4	.61	.38	.36	.014
D6S265 (.46/.08)	.42	.43	.7	.41	.39	.008
HLA-G (.73/.46)	.28	.28	.32	.5	.48	.008
HLA-F (.51/.87)	.38	.38	.59	.4	.39	.01
D6S306 (.76/.46)	.32	.33	.39	.57	.55	.002
D6S464 (.73/.47)	.26	.28	.32	.49	.46	.002
D6S105 (.59/.21)	.38	.38	.48	.47	.45	.008
D6S1260 (.74/.46)	.28	.31	.34	.51	.49	.006
D6S1558 (.83/.68)	.19	.21	.31	.48	.46	.006
D6S2231 (.60/.13)	.47	.53	.69	.53	.5	.008
D6S2238 (.82/.23)	.59	.6	.6	.76	.75	.003
HLA-H (.85/.04)	.8	.8	.93	.84	.83	
D6S2239 (.91/.37)	.57	.57	.85	.85	.84	.002
D6S2241 (.91/.67)	.31	.3	.55	.74	.71	.004
D6S1621 (.65/.07)	.58	.59	.86	.6	.58	.006
GATA (.46/.36)	.1	.24	.13	.16	.16	.01
D6S1545 (.48/.38)	.11	.13	.13	.17	.14	.012
D6S1691 (.19/.28)	.11	.23	.18	.12	.1	.016

[a] Data in parentheses are frequencies of most common allele in hemochromatosis chromosomes/frequency of same allele in normal chromosomes.

Linkage-Disequilibrium Analysis

Estimates of linkage disequilibrium between hemochromatosis and each of the 24 markers are shown in table 1. The two estimates of *r* (multiallelic and diallelic after alleles are lumped together) are very similar to one another. The D' measure also yields a pattern very similar to that of *r*, as has been the case in previous analyses (Hegele et al. 1990; Jorde et al. 1993, 1994). In part, this reflects the fact that the frequency of the "rare" allele at most marker loci was $\geqslant$.30 (values this high and above exert little effect on statistics such as *r*). The λ and δ statistics are quite similar to one another, as expected, with minor differences arising from the fact that δ was estimated for lumped two-allele systems whereas λ is a multiallele measure. Like *r* and D', the λ and δ estimates reach a peak in the HLA-H region (at HLA-H itself for *r* and D' and at the closely linked marker D6S2239 for λ and δ). The θ_K estimate also reflects maximum disequilibrium at D6S2239 (θ_K = .002), but θ_K reaches the same value at markers D6S306 and D6S464. In general, all measures of disequilibrium reach their maximum at or near HLA-H, with minor variations seen among different estimates in other parts of the region.

Figure 1 shows the results of the multipoint disequilibrium mapping method of Terwilliger (1995). The single-locus λ values shown in table 1 are given along with the curve representing the multipoint likelihood-ratio statistic. The peak value of the likelihood-ratio statistic occurs at locus D6S2239, near HLA-H. The 1,000:1 support interval extends 4.8–6.2 Mb from HLA-B, spanning ~1.4 Mb.

Figure 2*A* illustrates the relationship between pairwise linkage disequilibrium (*r*) and pairwise physical distance (in kb) for all possible pairs of the 24 marker loci on normal chromosomes. The *r* statistic is used here for comparability with previous studies (Jorde et al. 1993, 1994; Watkins et al. 1994). The expected negative relationship between disequilibrium and physical distance is readily apparent, and significant ($P < .001$) disequilibrium extends for a distance of >1,000 kb. The Mantel matrix correlation for physical distance and disequilibrium was −.33 ($P < .0001$). Figure 2*B* displays the same relationship for the chromosomes carrying a hemochromatosis mutation. A negative relationship between distance and disequilibrium is seen again, and the correlation is somewhat higher (−.50; $P < .0001$). Compared with the normal chromosomes, significant disequilib-

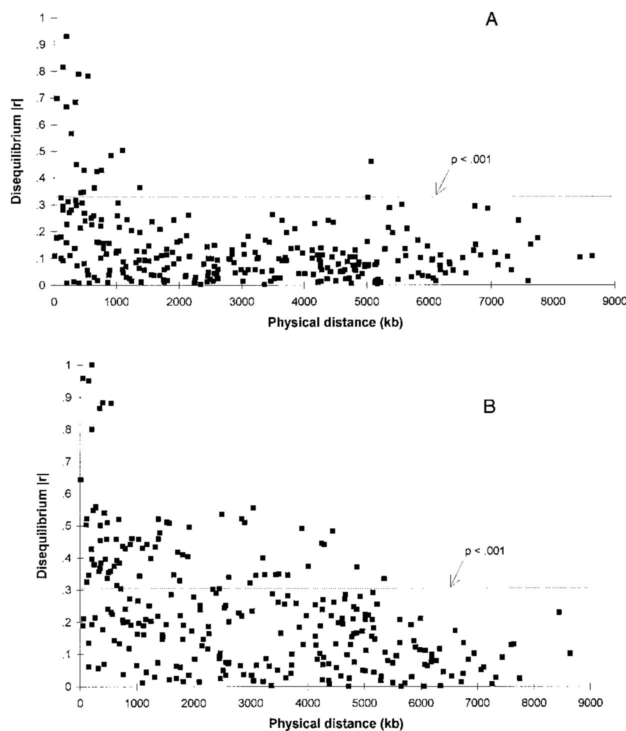

Figure 2 *A,* Plot of linkage disequilibrium ($|r|$) versus physical distance, for all pairs of 24 polymorphic markers on chromosomes that do not contain a hemochromatosis mutation. *B,* Plot of linkage disequilibrium versus physical distance, for chromosomes that contain a hemochromatosis mutation.

rium is maintained over a substantially larger region (~5,000 kb) for the hemochromatosis chromosomes.

To estimate the age of an hh mutation, the recombination rate between HLA-H and HLA-A was calculated first. Six recombinants were observed in 433 informative meioses in these families, yielding $\theta_K = .0139$ in this 5-Mb region. When this value of χ and a δ value of .38 (table 1) are substituted into equation (1), an estimate of 69 generations is obtained. The 90% confidence interval for this estimate is 35–161 generations, and the 95% confidence interval is 32–189 generations. Both estimates are relatively recent in an evolutionary sense. The G→A transition associated with hh in the HLA-H gene results in a conversion of a cysteine to tyrosine at amino acid 282 (cys282tyr). Because the cys282tyr mutation has been assayed in the present sample, it also is possible to restrict this analysis only to hh chromosomes containing this substitution. The δ value for HLA-A versus presence or absence of the substitution is .44, yielding an estimate of 59 generations (this degree of similarity with the previous estimate is expected, since 85% of hh chromosomes have the cys282tyr substitution). The 90% and 95% confidence intervals for this estimate are 30–136 generations and 27–161 generations, respectively. These estimates support the hypothesis that the major hemochromatosis-causing mutation is a recent event.

Figure 3 displays a neighbor-joining tree that shows relationships among the 169 24-locus hh haplotypes. The tree reveals six major clusters of haplotypes, labeled as groups 1–6. Table 2 presents the frequencies of the most common hemochromatosis-associated alleles for the haplotypes included in each group (for brevity, some loci were omitted in the HLA-B–HLA-A region and in the region telomeric of HLA-H). The most common haplotype for the hh-containing chromosomes has alleles HLA-B7, HLA-A3, D6S306-6, D6S464-11, D6S105-8, CS5-8, ML3-7, D6S2231-3, D6S2238-5, HLA-H-2 (cys282tyr), D6S2239-3, D6S2241-4, and D6S1621-1 (B7-A3-6-11-8-8-7-3-5-2-3-4-1).

Group 1 consists of 45 hh-containing chromosomes, nearly all of which have HLA-A3 and HLA-B7, as well as the cys282tyr mutation at HLA-H. It is noteworthy that 28% (47/169) of all hh chromosomes have the A3-B7 haplotype, whereas only 3% (5/161) of the non-hh chromosomes have this haplotype. Group 2 (28 hh-containing chromosomes) has a low frequency of HLA-B7, but the frequencies of all other common markers,

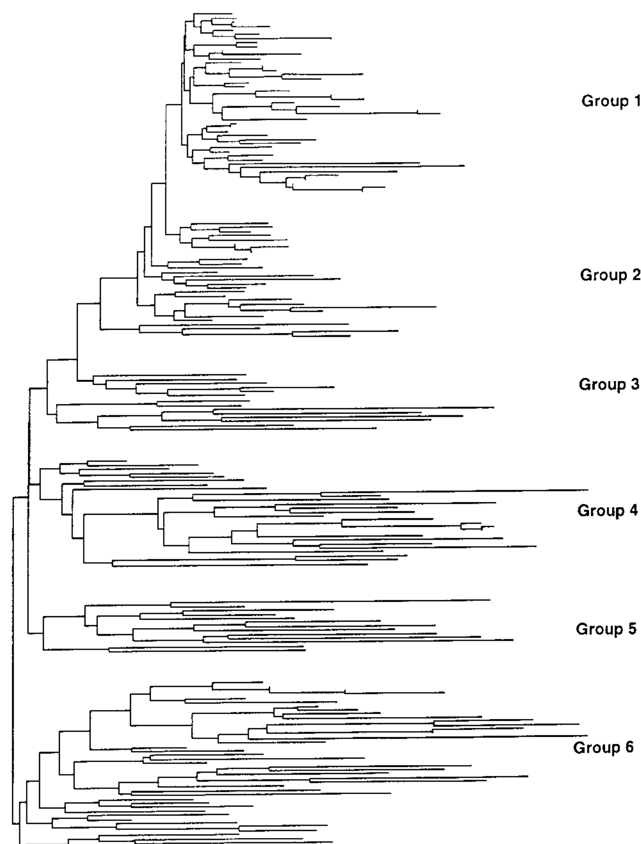

Figure 3 Neighbor-joining tree showing the genetic similarity of six major groups of chromosomes. For common allele frequencies for each of the six groups, see table 2.

Table 2

Allele Frequencies for Common Marker Polymorphisms on Hemochromatosis Chromosomes in Each of the Six Groups of Tree Shown in Figure 4

	Frequency[a]					
Marker Allele	Group 1 (*N* = 45)	Group 2 (*N* = 28)	Group 3 (*N* = 14)	Group 4 (*N* = 27)	Group 5 (*N* = 13)	Group 6 (*N* = 42)
HLA-B7	.98	.07	.29	.3	.08	0
HLA-A3	1	1	.21	0	0	0
D6S306-6	.92	.92	.75	.73	.64	.54
D6S464-11	.84	.88	.58	.77	.27	.67
D6S105-8	.83	.93	.75	.24	.2	.31
CS5-8	.95	.92	.85	.58	.5	.54
ML3-7	.95	.92	1	.64	.62	.73
D6S2231-3	.83	.92	.38	.38	.62	.29
D6S2238-5	.94	.88	.86	.58	.69	.83
HLA-H-2	.95	1	.71	.63	.91	.81
D6S2239-3	1	1	.86	.75	1	.83
D6S2241-4	.94	1	.93	.7	1	.94
D6S1621-1	.68	.84	.62	.42	.31	.67

[a] *N* = number of chromosomes in each group.

including HLA-A3, remain high (indeed, the cys282tyr mutation is seen with a frequency of 100% in this group). The major difference between groups 1 and 2, the presence versus the absence of HLA-B7, represents the historical recombination of HLA-B7 away from the hemochromatosis-containing haplotype. The remaining groups, 3–6, all have very low HLA-A3 and HLA-B7 frequencies, while maintaining relatively high frequencies of the HLA-H mutation and the common D6S2239 allele. The division between group 2 and groups 3–6 reflects a second historical recombination, in this case separating HLA-A3 from the hemochromatosis-containing haplotype. The cys282tyr mutation is substantially more frequent in groups 1 and 2 (95%–100%) than in groups 3–6 (63%–91%). In fact, nearly all of the alleles that are most common on hh-containing chromosomes have a higher frequency in groups 1 and 2 than in groups 3–6. These findings support the notion that the cys282tyr mutation took place relatively recently on an A3-B7 background, with haplotypes in groups 3–6 representing recombinations away from the A3-B7 background as well as other hemochromatosis-causing mutations.

Examination of the 23 hh chromosomes that do not have the HLA-H cys282tyr mutation shows that the marker alleles that generally are common on hh chromosomes have low frequencies on the non–cysteine-tyrosine hh chromosomes (similar to their frequencies in normal chromosomes; see table 1). In addition, there is a large degree of marker haplotype diversity in these 23 chromosomes: only two chromosomes have the same haplotype, and no subset of markers shows significant conservation on these chromosomes. This is in marked contrast to the haplotype backgrounds for the hh chromosomes carrying the cys282tyr mutation, in which there is a substantial reduction in haplotype diversity.

Throughout the tree, haplotypes are found in which one or two microsatellite loci differ from the allele found in the predominant haplotype (i.e., B7-A3-6-11-8-8-7-3-5-2-3-4-1). This difference usually consists of the loss or gain of a single repeat unit (e.g., B7-A3-6-11-8-8-7-3-5-2-3-4-2) and reflects the relatively high mutation rate of the microsatellite systems.

Discussion

Data collected over the past 20 years have confirmed Simon et al.'s (1976, 1987) original hypothesis for a founding hh mutation on a chromosome with the HLA haplotype A3,B7. Linkage of hh to HLA is clear, although further localization has been confusing because alleles for markers known to be physically distant from the class I region also were found to be strongly associated with hh (Jazwinska et al. 1993; Yaouanq et al. 1994; Raha-Chowdhury et al. 1995*b*). A recent report characterized nearly 30 new polymorphisms, as well as a candidate for the hh gene (Feder et al. 1996). The candidate gene, HLA-H, was found to have a G→A transition that results in a cys282tyr substitution, and it was present in 85% of disease chromosomes. Our results show that the cysteine-tyrosine substitution also is found in 85% of hh-containing chromosomes. This is consistent with other recently published reports (Feder et al. 1996; Jouanolle et al. 1996), although another study

(Jazwinska et al. 1996) has reported that 100% of hh-containing chromosomes in Australia have the cysteine-tyrosine substitution. Four percent of chromosomes among our normal population contained the cys282tyr mutation. These were healthy people with no evidence of iron overload who either married into pedigrees or were siblings of affected individuals and shared no HLA haplotype. It is possible that these individuals represent undetected heterozygotes. If these are truly disease-bearing chromosomes, then the value for δ would be slightly higher but would not otherwise alter any of our conclusions. The possibility also exists that the cys282tyr mutation is a very closely linked polymorphism, and it would be unfair to bias normal versus affected chromosomes solely on the basis of this mutation. A second mutation in the HLA-H gene was found to result in a his63asp conversion (Feder et al. 1996). This mutation was reported as being enriched in heterozygotes. Subsequently, others have reported no correlation between this mutation and hh (Jazwinska et al. 1996; Jouanolle et al. 1996).

Linkage-disequilibrium mapping played a key role in identifying HLA-H as the likely hh-causing locus (although it still is possible that another nearby locus, in strong association with HLA-H, is the actual hh-causing gene). This method now has been used extensively to localize disease genes to confined chromosomal regions. Linkage disequilibrium performs optimally when there is a common mutation responsible for the disease and additional mutations in the disease gene are relatively rare. Because linkage disequilibrium decays through time as a function of recombination frequency, linkage-disequilibrium analysis incorporates the effects of many past generations of recombination. In effect, this increases the number of meioses available for analysis. Applied judiciously, linkage disequilibrium can be useful in narrowing the location of disease genes (Jorde et al. 1993; Jorde 1995). Thus, under appropriate conditions, linkage disequilibrium has the potential to overcome the limits of standard recombinational analysis.

We applied several disequilibrium methods to estimate the most likely location of the hh gene. It is encouraging that there was broad agreement among all measures: all of them indicated peak disequilibrium at or very near the HLA-H locus. Some minor discrepancies were observed (see table 1), however, reflecting differences in the various disequilibrium measures. For example, r and D' show a secondary peak of disequilibrium at loci Y158, HLA-A, and D6S265, whereas δ and λ do not. r and D' are relative-risk measures and are quite sensitive to the low frequency (.08–.12 for these three markers) of the common hh-associated allele in normal chromosomes. In contrast, δ and λ, which are population-attributable risk measures (Devlin and Risch 1995), are relatively insensitive to the frequency of the common allele in normal chromosomes and instead give more weight to the frequency of the common allele in disease chromosomes. This difference is especially evident when these measures are compared at loci HLA-H and D6S2239. r and D' are much higher for HLA-H (where the disease-associated allele is seen in only 4% of normal chromosomes) than for D6S2239 (where the disease-associated allele is seen in 37% of normal chromosomes). δ and λ, on the other hand, are almost identical for these two markers and in fact are slightly higher for D6S2239. The latter difference is due to the fact that the common allele at D6S2239 has a frequency of 91% whereas the common allele at HLA-H has a frequency of 85% on hh chromosomes. In general, δ and λ are less sensitive to allele-frequency variation and oversampling effects in case-control studies and thus are more likely to provide more accurate estimates of a disease gene's location (Devlin and Risch 1995).

It is remarkable here that linkage disequilibrium on normal chromosomes was maintained over a physical distance of significantly >1 Mb, since disequilibrium typically dissipates more quickly than this in continental European populations (Jorde et al. 1994; Watkins et al. 1994). A recent study of a large number of meioses in CEPH kindreds, however, indicated that the recombination rate in the MHC class I region is ~1/5 of the rate expected under the usual rule that 1 Mb = 1 cM (Martin et al. 1995). Among other things, this could reflect natural selection for specific combinations of class I alleles, resulting in a lack of tolerance of recombination in this region. We calculated a recombination frequency of ~1.4% across the nearly 5-Mb region telomeric of HLA-A, indicating that this lack of recombination extends beyond the class I region. The fact that disequilibrium was maintained over an even *larger* distance for hh chromosomes than for normal chromosomes is consistent with a relatively recent mutational event. This conclusion is supported further by the mutation-age analysis (eq. [1]), which is independent of the amount of recombination in the chromosome region under study. Thus, the extreme disequilibrium seen on hh chromosomes is the result of both reduced recombination in this chromosome region (which equally affects normal chromosomes) and the recent age of the disease-causing mutation. A reduction in haplotype diversity among disease-bearing chromosomes has been observed for a number of genetic diseases, including cystic fibrosis (Cutting et al. 1990), myotonic dystrophy (Harley et al. 1991), phenylketonuria (Eisensmith et al. 1992), and Wilson disease (Petrukhin et al. 1993).

The phylogenetic analysis of hh-containing haplotypes supports Simon et al.'s (1987) hypothesis that a common hh mutation occurred on a haplotype containing the HLA-A3 and HLA-B7 alleles. Inclusion of the HLA-H locus in this analysis suggests strongly that

this mutation is the cys282tyr substitution. Alternatively, a very closely linked mutation could have occurred on a haplotype bearing the cys282tyr substitution. An estimate of the date of this mutation indicates that it took place quite recently, likely <100 generations ago. Hemochromatosis haplotypes that contain mutations other than the cys282tyr substitution at HLA-H are much more diverse than those that do not. This implies two possible evolutionary scenarios: either there is one other hh-causing mutation ancient enough so that many recombinations and marker mutations have subsequently occurred, or multiple additional hh-causing mutations have occurred on a variety of haplotype backgrounds. The latter case would be similar to what occurs in several other recessive diseases, such as cystic fibrosis and phenylketonuria, in which there are one or a few common mutations and many other rare mutations. Further characterization of hh-causing mutations will resolve this issue.

Acknowledgments

The authors gratefully acknowledge Drs. Jerry Kaplan, Bernie Devlin, and Koji Lum for their helpful comments and discussion. Some genotyping was performed in the Genomics Core Facility of the Huntsman Cancer Institute. J.P.K. and R.S.A. were supported by NIH grant DK-20630; L.B.J. was supported by NSF grant DBS-9310105; and J.R.G. was supported in part by NIH grant 1R29DK45819-03, March of Dimes grant 6-FY95-0050, and the Lucille P. Markey Charitable Trust.

References

Abderrahim H, Sambucy J, Iris F, Ougen P, Billault A, Chumakov I, Dausset J, et al (1994) Cloning the major histocompatibility complex in YACs. Genomics 23:520–527

Bengtsson BO, Thomson G (1981) Measuring the strength of associations between HLA antigens and diseases. Tissue Antigens 18:356–363

Carroll MC, Katzman P, Alicot EM, Koller BH, Geraghty DE, Orr HT, Strominger JL, et al (1987) Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. Proc Natl Acad Sci USA 84: 8535–8539

Cartwright GE, Skolnick M, Amos DB, Edwards CQ, Kravitz K, Johnson A (1978) Inheritance of hemochromatosis: linkage to HLA. Trans Assoc Am Physicians 91:273–281

Crawford DHG, Powell LW, Leggett BA, Francis JS, Fletcher LM, Webb SI, Halliday JW, et al (1995) Evidence that the ancestral haplotype in Australian hemochromatosis patients may be associated with a common mutation in the gene. Am J Hum Genet 57:362–367

Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui L, Antonarakis SE, Kazazian HH (1990) A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. Nature 346: 366–369

Dadone MM, Kushner JP, Edwards CQ, Bishop DT, Skolnick MH (1982) Hereditary hemochromatosis: analysis of laboratory expression of the disease by genotype in 18 pedigrees. Am J Clin Pathol 78:196–207

Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29:311–322

Edwards CQ, Dadone MM, Skolnick MH, Kushner JP (1982) Hereditary hemochromatosis. Clin Haematol 11:411–435

Edwards CQ, Griffen LM, Goldgar D, Drummond C, Skolnick MH, Kushner JP (1988) Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. N Engl J Med 318:1355–1362

Edwards CQ, Skolnick MH, Kushner JP (1981) Hereditary hemochromatosis: contribution of genetic analysis. Prog Hematol 12:43–71

Eisensmith RC, Okano Y, Dasovich M, Wang T, Güttler F, Lou H, Guldberg P, et al (1992) Multiple origins for phenylketonuria in Europe. Am J Hum Genet 51:1355–1365

Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, et al (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package), version 3.5c. Department of Genetics, University of Washington

Geraghty DE, Koller BH, Orr HT (1987) A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. Proc Natl Acad Sci USA 84:9145–9149

Geraghty DE, Pei J, Lipsky B, Hansen JA, Taillon-Miller P, Bronson SK, Chaplin DD (1992) Cloning and physical mapping of the HLA class I region spanning the HLA-E-to-HLA-F interval by using yeast artificial chromosomes. Proc Natl Acad Sci USA 89:2669–2673

Gruen JR, Goei VL, Summers KM, Capossela A, Powell L, Halliday J, Zoghbi H, et al (1992) Physical and genetic mapping of the telomeric major histocompatibility complex region in man and relevance to the primary hemochromatosis gene (HFE). Genomics 14:232–240

Harley HG, Brook JD, Floyd J, Rundle SA, Crow S, Walsh KV, Thibault M-C, et al (1991) Detection of linkage disequilibrium between the myotonic dystrophy locus and a new polymorphic DNA marker. Am J Hum Genet 49:68–75

Hedrick PW (1987) Gametic disequilibrium measure: proceed with caution. Genetics 117:331–341

Hegele RA, Plaetke R, Lalouel J-M (1990) Linkage disequilibrium between DNA markers at the low-density lipoprotein receptor gene. Genet Epidemiol 7:69–81

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Jazwinska EC, Cullen LM, Busfield F, Pyper WR, Webb SI, Powell LW, Morris PC, et al (1996) Haemochromatosis and HLA-A. Nat Genet 14:249–251

Jazwinska EC, Lee SC, Webb SI, Halliday JW, Powell LW (1993) Localization of the hemochromatosis gene close to D6S105. Am J Hum Genet 53:347–352

Jazwinska EC, Pyper WR, Burt MJ, Francis JL, Goldwurm S, Webb SI, Lee SC, et al (1995) Haplotype analysis in Australian hemochromatosis patients: evidence for a predominant

ancestral haplotype exclusively associated with hemochromatosis. Am J Hum Genet 56:428–433

Jorde LB (1995) Linkage disequilibrium as a gene mapping tool. Am J Hum Genet 56:11–14

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 54:884–898

Jorde LB, Watkins WS, Viskochil D, O'Connell P, Ward K (1993) Linkage disequilibrium in the neurofibromatosis 1 (NF1) region: implications for gene mapping. Am J Hum Genet 53:1038–1050

Jouanolle AM, Gandon G, Jezeqeul P, Blayau M, Campion ML, Yaouanq J, Mosser J, et al (1996) Haemochromatosis and HLA-H. Nat Genet 14:251–252

Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. Am J Hum Genet 56:18–32

Koller BH, Geraghty DE, DeMars R, Duvick L, Rich SS, Orr HT (1989) Chromosomal organization of the human major histocompatibility complex class I gene family. J Exp Med 169:469–480

Lawrance SK, Smith CL, Srivastava R, Cantor CR, Weissman SM (1987) Megabase-scale mapping of the HLA gene complex by pulsed field gel electrophoresis. Science 235:1387–1390

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49:49–67

——— (1988) On measures of gametic disequilibrium. Genetics 120:849–852

Lipinski M, Hors J, Saleun JP, Saddi R, Passa P, Lafaurie S, Feingold N, et al (1978) Idiopathic hemochromatosis: linkage with HLA. Tissue Antigens 11:471–474

Martin M, Mandd D, Carrington M (1995) Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. Hum Mol Genet 4:423–428

McLellan T, Jorde LB, Skolnick MH (1984) Genetic distances between the Utah Mormons and related populations. Am J Hum Genet 36:836–857

Petrukhin K, Fischer SG, Piratsu M, Tanzi RE, Chernov I, Devoto M, Brzustowicz LM, et al (1993) Mapping, cloning and genetic characterization of the region containing the Wilson disease gene. Nat Genet 5:338–343

Raha-Chowdhury R, Bowen DJ, Burnett AK, Worwood M (1995*a*) Allelic associations and homozygosity at loci from HLA-B to D6S299 in genetic haemochromatosis. J Med Genet 32:446–452

Raha-Chowdhury R, Bowen DJ, Stone C, Pointon JJ, Terwilliger JD, Shearman D, Robson KJH, et al (1995*b*) New polymorphic microsatellite markers place the haemochromatosis gene telomeric of D6S105. Hum Mol Genet 4: 1869–1874

Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, Singer B, Fahn S, et al (1995) Genetic analysis of idiopathic torsion dystronia in Ashkenazi Jews and their recent descent from a small founder population. Nat Genet 9:152–159

Simon M, Bourel M, Fauchet R, Genetet B (1976) Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis. Gut 17:332–334

Simon M, Bourel M, Genetet B, Fauchet R (1977) Idiopathic hemochromatosis: demonstration of recessive inheritance and early detection by family HLA typing. N Engl J Med 297:1017–1021

Simon M, LeMignon L, Fauchet R, Yaouanq J, David V, Edan G, Bourel M (1987) A study of 609 HLA haplotypes marking for the hemochromatosis gene: (1) mapping of the gene near the HLA-A locus and characters required to define a heterozygous population and (2) hypothesis concerning the underlying cause of hemochromatosis-HLA association. Am J Hum Genet 41:89–105

Smith M, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35

Sood AK, Pereira D, Weissman SM (1981) Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility HLA-B by use of an oligodeoxynucleotide primer. Proc Natl Acad Sci USA 78:616–620

Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56: 777–787

Watkins WS, Zenger R, O'Brien E, Nyman D, Eriksson AW, Renlund M, Jorde LB (1994) Linkage disequilibrium patterns vary with chromosomal location: a case study from the von Willebrand factor region. Am J Hum Genet 55: 348–355

Whitehead Institute Database (1997) http://www-genome.wit.mit.edu

Worwood M, Raha-Chowdhury R, Dorak MT, Darke C, Bowen DJ, Burnett AK (1994) Alleles at D6S265 and D6S105 define a haemochromatosis-specific genotype. Br J Haematol 86:863–866

Yaouanq J, Perichon M, Chorney M, Pontarotti P, Le Truet A, El Kahloun A, Mauvieux V, et al (1994) Anonymous marker loci within 400 kb of *HLA-A* generate haplotypes in linkage disequilibrium with the hemochromatosis gene (*HFE*). Am J Hum Genet 54:252–263