Am. J. Hum. Genet. 40:001–014, 1987

Assessing the Role of HLA-linked and Unlinked Determinants of Disease

NEIL RISCH

Departments of Epidemiology and Public Health and of Human Genetics,
Yale University School of Medicine, New Haven, CT 06510

SUMMARY

The relationship between increased risk in relatives over population prevalence ($\lambda_R = K_R/K$) and probability of sharing zero marker alleles identical by descent (ibd) at a linked locus (such as HLA) by an affected relative pair is examined. For a model assuming a single disease-susceptibility locus or group of loci tightly linked to a marker locus, the relationship is remarkably simple and general. Namely, if ϕ_R is the prior probability for the relative pair to share zero marker alleles identical by descent, then P(sharing 0 markers|both relatives are affected) is just ϕ_R/λ_R. Alternatively, λ_{AR}, the increased risk over population prevalence to a relative R due to a disease locus tightly linked to marker locus A, equals the prior probability that the relative pair share zero A alleles ibd divided by the posterior probability that they share zero alleles ibd, given that they are both affected. For example, for affected sib pairs, P(sharing 0 markers|both sibs are affected) = $.25/\lambda_S$. This formula holds true for any number of alleles at the disease locus and for their frequencies, penetrances, and population prevalence. Similar formulas are derived for sharing one and two markers. Application of these formulas to several well-studied HLA-associated diseases yields the following results: For multiple sclerosis, insulin-dependent diabetes mellitus, and coeliac disease, a single-locus model of disease susceptibility is rejected, implying the existence of additional unlinked familial determinants. For all three diseases, the effect of the HLA-linked locus on familiality is minor: for multiple sclerosis, it accounts for only a 2.5-fold increased risk to sibs over the population prevalence, compared to an observed value

Received March 25, 1986; revision received August 1, 1986.

Address for correspondence and reprints: Dr. Neil Risch, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, P.O. Box 3333, New Haven, CT 06510.

© 1987 by the American Society of Human Genetics. All rights reserved. 0002-9297/87/4001-0001$02.00

of 20; for coeliac disease, it accounts for approximately a 5.25-fold increased risk to sibs, while the observed value is on the order of 60; for insulin-dependent diabetes mellitus, it accounts for a 3.42-fold increased risk in sibs, while the observed value is 15. In all cases, the secondary determinants must be outside the HLA region. For tuberculoid leprosy, an unlinked familial determinant is also implicated (increased risk to sibs due to HLA = 1.49; observed value = 2.38). For hemochromatosis and Hodgkin's disease, there is little evidence for HLA-unlinked familial determinants. With this formula, it is also possible to examine the hypothesis of pleiotropy versus linkage disequilibrium by comparing λ_{AS} with the increased risk to sibs due to the associated allele(s). The formula is also generalized to a two-locus multiplicative model, which allows for estimation of the potential impact of unlinked loci for this type of model.

INTRODUCTION

A wide range of diseases have been shown to be associated with HLA, the major histocompatibility complex of humans (Thomson 1981). This relationship is usually discovered through association with disease of one or more alleles at one or more of the HLA loci. Confirmatory evidence comes from family studies showing that affected relatives share HLA haplotypes more frequently than chance would predict. One of the most common of such paradigms uses affected sib pairs and measures the degree to which haplotypes are shared by affected sibs. Such information can also be used to draw inferences about mode of inheritance of an HLA-linked locus (assuming that recombination is negligible).

For a number of HLA-associated diseases, the assumption of a single HLA-linked disease locus underlying inheritance leads to serious discrepancies among observations. For example, for a number of diseases, family-recurrence data have been used to obtain a model of inheritance for the disease, which was then applied to linkage analysis with HLA, leading to very large estimates of the recombination fraction (Clerget-Darpoux et al. 1980; Suarez et al. 1982). Such large values of θ are inconsistent with the known associations of disease with certain HLA alleles, associations that require short recombination distances. To explain the discrepancy, several investigators (Thomson 1980; Ho et al. 1982) have suggested the possibility that non-HLA-linked alleles are also important in disease etiology. Such loci may act synergistically (e.g., in a multiplicative model) or additively (e.g., in genetic heterogeneity) with the HLA-linked locus.

In the present paper I directly examine the relationship between the increased risk in relatives of affected individuals vis-à-vis the population prevalence ($\lambda_R = K_R/K$) and the proportion of affected relative pairs that share zero

haplotypes identical by descent, i.e., P(sharing 0|both affected). It turns out that under any single HLA-linked-disease locus model, there is a very simple relationship between these two values, namely P(0 sharing|both affected) = ϕ_R/λ_R, where ϕ_R is the prior probability that the relatives share zero alleles identical by descent (ibd). This formula is valid irrespective of the number of distinct HLA-linked-disease alleles.

Analogous formulas are derived for the probability of affected sib pairs sharing two haplotypes and one haplotype ibd. These formulas can be used to determine the likelihood of there being HLA-unlinked familial determinants underlying disease susceptibility. By defining a two-locus, multiplicative model, it is also possible to assess the relative contribution to familiality of the HLA-linked locus and an unlinked locus (or loci) for this type of model.

DERIVATION: SINGLE-LOCUS MODEL

Suppose a single HLA-linked locus underlies disease susceptibility. Consider a relative of type R. Let ϕ_R equal the probability that two relatives of type R share zero marker alleles identical by descent (for example, $\phi_R = 0$ for parents or offspring; ¼ for sibs; ½ for half sibs, uncles/aunts, nieces/nephews, grandparents, and grandchildren; ¾ for first cousins; etc.). Let K equal the population prevalence of disease (more specifically, cumulative lifetime risk); let K_R equal the risk among relatives of type R of affected probands.

Under a single-locus model, if the two relatives share zero marker alleles identical by descent, then they are genetically independent, and the probability that both are affected is K^2. Without information on marker-allele sharing, the probability that both are affected is the probability that the proband is affected (K) times the probability that the relative is affected given that the proband is affected (K_R) (this requires certain assumptions about ascertainment, discussed below). Hence, letting P represent probability, we have P(2 relatives affected) $= K \cdot K_R$. Similarly, P(2 relatives affected and share 0 alleles ibd) = P(2 relatives share zero alleles ibd)$\cdot P$(2 relatives affected|share 0 alleles ibd) $= \phi_R \cdot K^2$. Therefore,

$$P(\text{2 relatives share 0 alleles ibd}|\text{both are affected}) = \phi_R \cdot K^2/(K \cdot K_R) = \phi_R/\lambda_R \ . \quad (1)$$

Note that this formula holds true irrespective of the mode of inheritance, number and frequency of alleles, penetrances at the disease-susceptibility locus, or population prevalence of disease. The only requirement is that recombination be negligible.

It is now clear why linkage analyses have yielded large values for recombination with HLA. If a model with a high recurrence risk vis-à-vis population prevalence is assumed, then a low probability of sibs sharing zero haplotypes is implied. In linkage analysis, only a large value of θ can reconcile the discrepancy between high values of λ_S and of sib haplotype discordance.

A similar formula can be derived relating the probability of affected sib pairs sharing two marker alleles to recurrence risks. Specifically, define K_S as the recurrence risk in siblings and K_{MZ} as the recurrence risk in monozygotic (MZ) twins. Again, P(2 sibs affected) = $K \cdot K_S$, while P(2 sibs affected|share 2 alleles ibd) = $K \cdot K_{MZ}$, provided that the locus of interest is the only inherited-disease determinant. In other words, if the sibs share two alleles at the disease-susceptibility locus, then they have the same genetic relationship as MZ twins. Then

$$P(\text{sibs share 2 alleles ibd}|\text{both affected}) = \tfrac{1}{4}K \cdot K_{MZ}/(K \cdot K_S) = \tfrac{1}{4}/(K_S/K_{MZ}) \; . \tag{2}$$

That is, the proportion of affected sib pairs sharing two alleles identical by descent should be .25 divided by the relative risk to sibs vis-à-vis MZ twins. The adequacy of a single-locus model can also be tested using this formula, provided reliable information exists about MZ-twin concordance. We expect this formula to be less useful in practice, since unbiased twin data are difficult to obtain.

The analogous formula for affected sib pairs sharing one marker allele ibd can also be derived. In this case, the probability of two sibs being affected, given that they share one marker allele ibd is just $K \cdot K_{P/O}$, where $K_{P/O}$ is the recurrence risk for parents and/or offspring. This is because sibs who share one disease allele ibd have the same genetic relationship as do parents/offspring. Therefore, the probability of a sib pair sharing one marker allele given that they are both affected is $.5/(K_S/K_{P/O})$. This formula ranges in value from .5 (no dominance) to 0 (complete dominance).

TWO-LOCUS MULTIPLICATIVE MODEL

We now consider a model of two unlinked disease-susceptibility loci in which the penetrance matrix can be characterized in a multiplicative fashion (see, e.g., Hodge 1981). Define two loci A and B (say A represents HLA). Suppose there are m possible genotypes at the A locus and n possible genotypes at the B locus. Define penetrance vectors $(f_1, \ldots, f_m)$ and $(g_1, \ldots, g_n)$ for the A and B loci, respectively. Enumerate genotypes as A_i, $i = 1, \ldots, m$; B_j, $j = 1, \ldots, n$. Then the penetrance of genotype A_iB_j is defined as f_ig_j. Let the population frequency of genotypes A_i and B_j be $P(A_i)$ and $P(B_j)$, respectively. This type of model is appropriate, for example, for a situation in which a genotype at one locus is required for susceptibility and penetrance is determined by the other locus. For example, suppose locus B has two alleles B and b and that only BB individuals are susceptible, with penetrance determined by genotypes at the A locus. This case corresponds to $g_1 = 1$, $g_2 = g_3 = 0$.

As above, define K as population prevalence and K_R as the recurrence risk in relatives of type R of affected individuals. We make the following additional

definitions:

$$K_A = \sum_{i=1}^{m} P(A_i) f_i \ ;$$

$$K_B = \sum_{j=1}^{n} P(B_j) g_j \ ;$$

$$K_{AR} = \left(\sum_{i=1}^{m} P(A_i) f_i \sum_{k=1}^{m} P(\text{relative is } A_k | \text{proband is } A_i) \cdot f_k\right) / K_A \ ;$$

$$K_{BR} = \left(\sum_{j=1}^{n} P(B_j) g_j \sum_{l=1}^{n} P(\text{relative is } B_l | \text{proband is } B_j) \cdot g_l\right) / K_B \ ;$$

$$K_R = K_{AR} \cdot K_{BR} \ ;$$

and

$$\lambda_{AR} = K_{AR}/K_A, \ \lambda_{BR} = K_{BR}/K_B, \text{ and } \lambda_R = \lambda_{AR} \cdot \lambda_{BR} \ .$$

Note that K_{AR} and K_{BR} are defined analogously to the single-locus case using the "marginal" penetrances f and g. The parameters λ_A and λ_B can be interpreted as the multiplicative increase in risk to relatives of type R due to alleles at loci A and B, respectively. In the Appendix, we derive the following equation:

$$P(\text{proband and relative share 0 A-locus alleles} | \text{both are affected}) = \phi_R / \lambda_{AR} \ . \tag{3}$$

Knowing the proportion of relative pairs that share no A-locus alleles ibd, we can calculate λ_{AR} from equation (3). Then, since $\lambda_R = \lambda_{AR} \cdot \lambda_{BR}$, we have

$$\lambda_{BR} = \lambda_R / \lambda_{AR} \ . \tag{4}$$

Hence, we can calculate the remaining increase in risk to relatives of type R due to an unidentified locus (loci).

If a second disease-susceptibility locus B has been identified, λ_{BR} can also be calculated from the probability of affected relatives sharing no B alleles ibd, and the possibility of additional loci can be assessed. Note that in the derivation of equation (1), the critical equation is P(both relatives affected|share 0 alleles ibd) $= K^2$. This equation can be extended to any number of loci, where the conditioning is on sharing zero alleles ibd at any of the relevant loci. In this case, the relatives are genetically independent, and hence the probability of both being affected is K^2. Once a second (B) locus has been identified, the adequacy of the multiplicative model can be determined as follows: If associ-

ated markers are available, then an $m \times n$ contingency table (m genotypes at the A locus $\times$ n genotypes at the B locus) should show no deviation from independence if the multiplicative model is correct. Similarly, a 3×3 or 2×2 table of sharing of marker alleles by affected relatives (depending on the maximum possible number of alleles shared) for the A and B loci should reveal no deviation from independence if the model is correct (see, e.g., Hodge 1981).

An example of a model that deviates from multiplicativity is a genetic-heterogeneity model. In this case, we would expect sib pairs who share zero alleles ibd at one marker locus to be concordant for alleles ibd at the second locus, and vice versa.

LINKAGE DISEQUILIBRIUM VERSUS PLEIOTROPY

With the approach described above, it is also possible to assess in a simple way whether the effect of HLA can be attributed to a defined HLA allele(s) that is associated with disease (i.e., pleiotropy) or to an as yet undetected linked allele (linked disequilibrium). Basically, one can calculate the increased risk to sibs due to segregation of the defined "high-risk" allele(s) and compare this to the increased risk calculated from haplotype-discordance data. Relative penetrances for the various HLA genotypes can be obtained by comparing HLA genotype frequencies in cases with those in controls. With the formula of James (1971), these penetrances can then be used to estimate K_S. If this value is below that obtained from the haplotype-discordance data, then linkage disequilibrium is favored over pleiotropy.

APPLICATION

The most common type of relative examined both for recurrence risk and HLA haplotype sharing is the sib, although other relative types, such as uncle/aunt or cousin may also be examined. Even if the only information available for second- or third-degree relatives is for haplotype sharing, an analysis is still possible using recurrence-risk information on parents and/or offspring, for whom haplotype-sharing information is useless. When considering parents, one must always take selection effects into account. From James (1971) we have $K_R = K + ((s_2 + \frac{1}{2}s_1)V_A + s_2 V_D)/K$, where V_A is the additive genetic variance of the penetrances and V_D is the dominance variance; s_1 is the prior probability that the relatives share one allele ibd; s_2 is the prior probability that they share two alleles ibd. Therefore, for parents and offspring, $K_{P/O} = K + (\frac{1}{2}V_A)/K$; for uncles, $K_U = K + (\frac{1}{4}V_A)/K$; and for cousins, $K_C = K + (\frac{1}{8}V_A)/K$. Hence, $K_U = \frac{1}{2}(K_{P/O} + K)$ and $K_C = \frac{1}{4}(K_{P/O} + 3K)$; and $\lambda_U = \frac{1}{2}(\lambda_{P/O} + 1)$ and $\lambda_C = \frac{1}{4}(\lambda_{P/O} + 3)$. Thus, assuming a single-locus model, given values for λ and $\lambda_{P/O}$ we can derive a value for λ_U or λ_C that can be used in conjunction with the haplotype-sharing information.

Since affected sib pairs are the most commonly used paradigm in studying the effect of HLA on disease susceptibility, I have plotted in figure 1 the relationship between P(sharing 0 marker alleles|2 sibs affected) and λ_S. As can be seen in the figure, the probability of sharing zero alleles ibd initially drops very rapidly with λ_S; for example, for a fivefold increased risk in sibs over

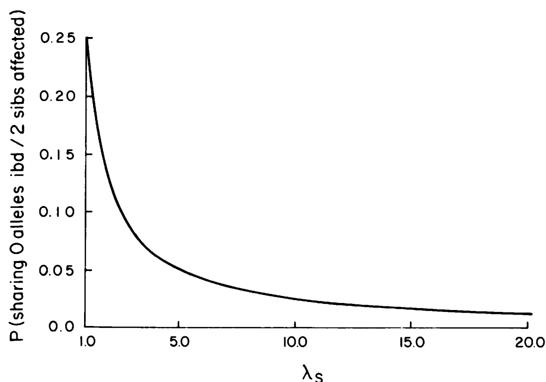

FIG. 1.—The relationship between the posterior probability of two sibs sharing zero marker alleles ibd given that they are both affected (vertical axis) and the relative risk of disease to sibs of an affected proband vs. the population prevalence ($\lambda_S = K_S/K$; horizontal axis). The formula is P(sharing 0 alleles ibd|2 sibs affected) = $.25/\lambda_S$.

population prevalence, the probability of sharing zero alleles ibd decreases to .05. For a 10-fold increased risk, λ_S decreases to .025.

Multiple Sclerosis (MS)

Using a model derived from recurrence-risk data for linkage analysis, a number of investigators have estimated large values of θ (Ho et al. 1982; Suarez et al. 1982). In fact, for MS, although the significance of the HLA effect in disease susceptibility is clear, the proportion of affected sib pairs sharing no HLA haplotypes ibd is large; for example, summarizing the literature, Ho et al. (1982) determined that 19 (10.44%) of 182 affected sib pairs shared zero haplotypes ibd. Therefore, according to equation (1) for a single-disease locus model, λ_S should be $.25/.104 = 2.4$. This value is far below estimates in the literature of relative risks for sibs; for example, Kuwert (1977) has summarized a number of studies estimating population prevalence and prevalence in sibs, giving a median ratio of approximately 20, with remarkable consistency among studies. For a λ_S value of 20, the expected probability of sharing zero haplotypes ibd is $.25/20 = .0125$. We assume a Poisson distribution for the number of sib pairs sharing zero haplotypes. Then the probability of observing that 19 or more of 182 sib pairs share zero haplotypes, assuming a prior probability of .0125, is 5.78×10^{-12}. Ho et al. (1982) also observed 76 (41.76%) of 182 affected sib pairs sharing two haplotypes ibd. If a sibling risk of 1% (Kuwert 1977) is assumed, equation (2) gives a value for K_{MZ} of $4(.4176)(.01) = 1.67\%$, which seems quite a bit lower than observed values (Spielman and Nathanson 1982). Once again, unlinked determinants are implicated.

It appears that the contribution of HLA alleles to the familial aggregation of susceptibility to MS is quite small; for example, assuming a two-locus multiplicative model and using equation (4), we obtain $\lambda_{BS} = 20/2.4 = 8.35$. Therefore, it seems worthwhile to search for other, unlinked susceptibility loci for

this disease, although residual familial aggregation may be due to nongenetic factors.

The strongest allelic association with MS is with the HLA allele DR2. Stewart et al. (1981) found that of 60 MS patients, 5% were DR2 homozygotes, 61.67% were DR2 heterozygotes, and 33.33% were non-DR2 homozygotes. In a control series, in which the Hardy-Weinberg equilibrium was assumed, the corresponding values were 1.74%, 22.92%, and 75.34%. Therefore, assuming a population prevalence of .001 (the actual value is irrelevant), the penetrance vectors for the genotypes DR2/DR2, DR2/DRx, and DRx/DRx are .002874, .002691, and .000442, respectively. Using the formula of James (1971) for recurrence risk to sibs and assuming a population frequency of .132 for DR2, we obtain a risk to sibs of .00146, or $\lambda_S = K_S/K = 1.46$. This value is below the value of 2.40 estimated from HLA haplotype discordance. Hence, with respect to allele DR2, these results are more consistent with linkage disequilibrium than with pleiotropy.

Insulin-Dependent Diabetes Mellitus (IDDM)

The evidence linking HLA to susceptibility to IDDM is prodigious. The existence of more than two distinct susceptibility alleles in the HLA region appears likely (Hodge et al. 1980; Svejgaard and Ryder 1981; Risch 1984). The number of such alleles is irrelevant to assessing the possibility of unlinked determinants, however, as demonstrated in the derivation of equation (1). Analysis of IDDM has also been characterized by large estimates of θ in linkage when models derived from recurrence risks have been used (Barbosa et al. 1979; Clerget-Darpoux 1980).

Payami et al. (1985) summarized affected sib-pair sharing from the literature and discovered that 55 (7.74%) of 711 pairs shared zero haplotypes ibd when truncate selection was assumed and that 44.7 (7.30%) of 612 pairs shared zero haplotypes when single selection was assumed. Spielman et al. (1980) have estimated a population lifetime incidence of .004 and a recurrence risk in sibs of patients of .060. These values are similar to those obtained by Wagener et al. (1982) from the Pittsburgh study of IDDM. Therefore, λ_S is approximately 15. According to equation (1), for a single-locus model, the expected proportion of affected sib pairs sharing zero haplotypes is then $.25/15 = .0167$, far below the observed incidence. If a Poisson distribution for the number of sib pairs sharing zero haplotypes and a prior probability of .0167 are assumed, the probability of observing at least 44 of 612 pairs sharing zero haplotypes is 4.61×10^{-15}. Hence, an unlinked familial determinant is implicated.

The value of .0730 for sharing zero haplotypes corresponds to a value of $.25/.730 = 3.42$. Therefore, applying equation (4) for a second locus B, we obtain $\lambda_{BS} = 15/3.42 = 4.39$. Payami et al. (1985) also found that, assuming single selection, 327.8 (53.56%) of the 612 affected sib pairs shared two haplotypes. On the basis of equation (2) and the assumption that $K_S = .06$, the predicted MZ recurrence risk would be 13%, somewhat below observed values (Thomson 1980).

Hence, for IDDM, it appears that HLA may not be the primary familial determinant and that an additional disease locus may be involved. This conclu-

sion holds true irrespective of the number of HLA-linked disease alleles. It is also consistent with recent evidence on nonobese diabetic mice, in which at least two distinct loci are involved in disease susceptibility (Hattori et al. 1986).

The two HLA alleles most strongly associated with susceptibility to IDDM are DR3 and DR4. Barbosa et al. (1982) have given the genotype distribution of HLA DR genotypes for 134 IDDM patients and 263 controls. Using these frequencies and assuming both a pleiotropic model and a disease prevalence of .004, we can calculate the penetrance of each of the six relevant genotypes: DR3/DR3 (.0113), DR3/DR4 (.0229), DR4/DR4 (.0117), DR3/DRx (.0030), DR4/DRx (.0070), and DRx/DRx (.0004). Using these penetrance values along with gene frequencies of .136 for DR3, .168 for DR4, and .696 for DRx (from Barbosa et al. 1982), we obtain $V_A = .0000214$ and $V_D = .0000060$. Using the formula of James (1971), we find that $K_S = 1.77$, a value well below the 3.42 estimated from the haplotype discordance data. Therefore, linkage disequilibrium is favored over pleiotropy of the DR3 and DR4 alleles.

Coeliac Disease (CD)

Although the data for coeliac disease are not as extensive, the pattern of association with HLA appears similar to that for IDDM, with at least two distinct HLA-linked disease-susceptibility alleles in evidence (Tiwari et al. 1984). Several investigators (see, e.g., Greenberg et al. 1982) have already suggested the likelihood of a second, HLA-unlinked disease-susceptibility locus for this disease.

Tiwari et al. (1984) assumed a population prevalence of .0005 for CD and a recurrence risk in sibs of .03. This gives a value of 60. Summarizing the literature, Greenberg et al. (1982) found 2 (4.76%) of 42 sib pairs sharing zero haplotypes ibd. A value of 60 for λ_S corresponds to an expected value of $.25/60 = .0042$ for the probability of sharing zero haplotypes ibd. If a Poisson distribution is assumed, the probability of observing at least 2 pairs of 42 sharing zero haplotypes is .014. Hence, unlinked familial disease determinants are implicated. The observed value of .0476 for sharing zero haplotypes corresponds to a λ_{AS} value of $.25/.0476 = 5.25$. Therefore, for a second locus in a multiplicative model, $\lambda_{BS} = 60/5.25 = 11.40$. Hence, a second locus may be a stronger determinant of familial aggregation of disease susceptibility than is HLA.

Idiopathic Hemochromatosis (IH)

There has been less controversy about both the inheritance of hemochromatosis and its relationship to HLA. Two studies (Kravitz et al. 1979; Lalouel et al. 1985) found both recessive inheritance at an HLA-linked locus with partial expression in the heterozygote and little evidence for recombination. Hors and Dausset (1983), summarizing affected sib-pair data, found no pairs of 127 sharing zero haplotypes. From the segregation-analysis results of Lalouel et al. (1985), I derived λ_S estimates of 65 for females and 41 for males. Using the value 41, we would expect $(.25/41) \times 127 = 0.77$ sib pairs to share zero haplotypes ibd. Therefore, it appears that the HLA-linked locus can entirely account for the familial aggregation of this disorder and that there is little evidence for the existence of HLA-unlinked familial determinants.

Tuberculoid Leprosy (TL)

Haile et al. (1985) examined 72 multiplex pedigrees with TL from southern India. Of 101 affected sib pairs, 17 (16.83%) were found to share no HLA haplotype. According to equation (1), HLA is responsible for an increased risk of .25/.1683 = 1.485-fold in sibs vis-à-vis the population prevalence. These researchers also performed segregation analysis on the family material. Using the estimates from their best-fitting model, I derived an increased risk to sibs vis-à-vis population prevalence (λ_S) of 2.38. On the basis of this value, the probability of observing at least 17 of 101 sib pairs sharing zero haplotypes is .043. In this case, the residual familial aggregation may well be due to clustering of the environmental agent (bacterium) responsible for this disease.

Hodgkin's Disease (HD)

Hors and Dausset (1983) have summarized HLA family studies of Hodgkin's disease. They found that of 23 affected sib pairs identified, there were either 1 or 2 who shared zero HLA haplotypes (discrimination was ambiguous in one pair). Hence, the value for λ_{AS} is either .25/.0435 = 5.75 or .25/.087 = 2.87. They were also able to identify 11 pairs of affected first cousins, of whom six were HLA concordant and five were discordant. In this case, λ_{AC} = .75/.4545 = 1.65.

Epidemiologic studies report sevenfold increased risk to sibs vis-à-vis population prevalence (λ_S), whereas that to parents and offspring ($\lambda_{P/O}$) is threefold (Hors and Dausset 1983). Although the λ_S value of 7 is greater than the 5.75 or 2.87 obtained from the HLA sib-discordance frequency, it is not significantly so. Assuming that $\lambda_S = 7.0$, the probability of observing at least one discordant sib pair is .56; the probability of observing at least two discordant sib pairs is .20. The expected increased risk to first cousins vis-à-vis population prevalence is $\lambda_C = \frac{1}{4}(\lambda_{P/O} + 3) = \frac{1}{4}(3.0 + 3) = 1.50$. This value is quite similar to the value of 1.65 obtained from the HLA haplotype-discordance data for first cousins.

From this limited data set, it appears that HLA alone can account for the familial aggregation of HD. If this finding continues to hold true in larger samples, it would suggest that an environmental agent, such as a virus, may contribute little to the familial aggregation of this disease (e.g., it may be very widespread).

DISCUSSION

In the derivation of equation (1), we have made certain assumptions about ascertainment of affected relative pairs. Basically, we assumed that an affected individual is selected at random from the population (designated as the proband) and that then one of the proband's affected relatives of type R is selected at random. For sibs, this coincides with the scheme described by Thomson and Bodmer (1977). This approach may provide a reasonable approximation of the true ascertainment schemes for unilineal relatives, e.g., uncles/aunts and cousins. For sibs, however, it may misrepresent the true method by which affected pairs are ascertained. However, Risch (1983) has shown that the

Thomson and Bodmer (1977) ascertainment scheme leads to a maximal amount of haplotype discordance among affected sib pairs (i.e., the highest probability of sharing zero haplotypes), although differences due to alternate ascertainment schemes were not large. Therefore, the observed amount of discordance is at best an *underestimate* of the amount of discordance assumed for our analysis. Hence, the results derived above are conservative; that is, the HLA effect would be overestimated and the effect of other loci would be underestimated. Again, discrepancies are likely to be small.

In discussing a two-locus model in the present paper, I do not mean to imply that the discrepancy between the sibling risk ratio and HLA discordance proves the existence of an additional locus (loci). What it does prove is the existence of a non-HLA familial factor(s) and a logical place to look is for additional segregating loci. This would be especially true for MS, CD, and IDDM, for which HLA seems to have a relatively minor role in familiality. If secondary loci do exist, they cannot be either within or tightly linked to the HLA region. If such a locus were in this region, then discordance at HLA would also imply discordance at the second locus. Hence, a second locus would have to exhibit at least a moderate amount of recombination with HLA.

The recent results on nonobese diabetes in mice (Hattori et al. 1986), which imply the existence of at least two distinct loci, further enhance the possibility of observing such a model in the study of human IDDM. Evidence for a second locus involving the insulin gene has also been reported for humans (Bell et al. 1984; Hitman et al. 1985; Raffel et al. 1985). One of the alleles at a locus flanking the insulin gene on chromosome 11 appears to be more frequent in IDDM ($q = .87$) than in controls ($q = .70$). Because the disease-associated allele is so common in the general population, it can only account for an increased risk to sibs vis-à-vis population prevalence of 1.16-fold. There are several possible explanations for this: (1) the disease-associated allele may be in linkage disequilibrium with the actual disease allele; (2) the disease-associated allele may actually subsume several distinct alleles, of which only one is truly related to disease predisposition; and (3) this locus may contribute only in a very minor way to the familial aggregation of IDDM. Resolution of these alternatives will be obtained through further studies of affected family members and identification of additional alleles at either this locus or nearby loci.

These results indicate that accurate inferences about mode of inheritance, penetrances, and gene frequencies of HLA-associated disease alleles (at least for MS, CD, and IDDM) require the use of a two-locus model—or at least the use of some model that allows for additional familiality. The approach of combined segregation and linkage analysis recently described by Risch (1984*a*, 1986, in press) and MacLean et al. (1984) provides one such method. In segregation-linkage analysis, haplotype-sharing (and -association) information among various types of relatives can be analyzed simultaneously.

Although I have shown how to use the derived formula to examine the question of pleiotropy versus linkage disequilibrium, there are other, more direct approaches to this question; for example, one can estimate coupling frequencies in combined segregation-linkage analysis—if coupling frequencies

of unity for associated alleles are rejected, then linkage disequilibrium is favored (MacLean et al. 1984; Risch 1984). Alternatively, one can also look at haplotype associations; if the disease is preferentially associated only with certain haplotypes bearing the high-risk allele, linkage disequilibrium is favored (Risch 1984; Thomson et al., in press).

In practice, it may be difficult to distinguish among different multilocus models without identifying the specific loci involved or closely linked markers. Data on twins may be helpful in this regard; for example, Ebers et al. (1984) have recently discovered an 11-fold increased risk to MZ twins vis-à-vis dizygotic twins of MS patients. Such a result is more suggestive of a nonadditive interaction with HLA than of admixture of a genetically distinct HLA-unlinked form. Under simple genetic heterogeneity, the maximum increased risk to MZ vis-à-vis dizygotic twins is fourfold.

The results in the present paper can be expected to apply beyond the realm of HLA-associated diseases. With the powerful new molecular-genetic methodology and restriction fragment–length polymorphisms, we can anticipate finding linkages with a number of important complex familial disorders. Once such a linkage is found, provided that recombination is minimal, the present paper indicates a way of determining whether such a linked locus alone can account for the entire familial aggregation of the disease or whether secondary loci should be sought.

ACKNOWLEDGMENTS

This work was supported by NIH Research Career Development Award HD00648.

APPENDIX

DERIVATION OF EQUATION (3)

Following the definitions in the text, I derive the following equations:

$$K = \sum_{i=1}^{m} \sum_{j=1}^{n} P(A_i)P(B_j)f_i g_j = \sum_{i=1}^{m} P(A_i)f_i \cdot \sum_{j=1}^{n} P(B_j)g_j = K_A \cdot K_B \;;$$

and

$$\begin{aligned} &P(\text{proband and relative of type R affected}) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} P(A_i)P(B_j)f_i g_j \cdot \sum_{k=1}^{m} \sum_{l=1}^{n} P(\text{relative is } A_k | \text{proband is } A_i) \\ &\quad \cdot P(\text{relative is } B_l | \text{proband is } B_j) \cdot f_k g_l \\ &= \sum_{i=1}^{m} P(A_i)f_i \cdot \left[\sum_{k=1}^{m} P(\text{relative is } A_k | \text{proband is } A_i) \cdot f_k \right] \\ &\quad \cdot \sum_{j=1}^{n} P(B_j)g_j \cdot \left[\sum_{l=1}^{n} P(\text{relative is } B_l | \text{proband is } B_j) g_l \right] \\ &= (K_A \cdot K_{AR}) \cdot (K_B \cdot K_{BR}). \end{aligned}$$

Similarly,

$$P(\text{proband and relative affected}|\text{they share 0 A-locus alleles ibd}) = K_A^2 \cdot K_B \cdot K_{BR} \ .$$

Thus,

$$P(\text{proband and relative share 0 A-locus alleles ibd and both are affected}) = \phi_R \cdot K_A^2 \cdot K_B \cdot K_{BR} \ .$$

Therefore,

$$P(\text{proband and relative share 0 A-locus alleles ibd}|\text{both are affected}) = (\phi_R \cdot K_A^2 \cdot K_B \cdot K_{BR})/(K_A \cdot K_{AR} \cdot K_B \cdot K_{BR}) = \phi_R/(K_{AR}/K_A) = \phi_R/\lambda_{AR} \ .$$

REFERENCES

Barbosa, J., F. H. Bach, and S. S. Rich. 1982. Genetic heterogeneity of diabetes and HLA. Clin. Genet. **21**:25–32.

Barbosa, J., M. M. Chern, E. V. Anderson, H. Noreen, S. Johnson, N. Reinsmoen, R. McCarty, R. King, and L. Greenberg. 1979. Linkage analysis between the major histocompatibility system and juvenile insulin-dependent diabetes in families with patients in two consecutive generations. J. Clin. Invest. **65**:592–601.

Bell, G. I., S. Horita, and J. H. Karam. 1984. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes **33**:176–183.

Clerget-Darpoux, F., C. Bonaiti-Pellie, J. Hors, I. Deschamp, and N. Feingold. 1980. Application of the lod score method to detection of linkage between HLA and juvenile insulin-dependent diabetes. Clin. Genet. **18**:51–57.

Ebers, G. C., D. E. Bulman, A. D. Sadovnick, D. W. Paty, S. Warren, W. Hader, T. J. Murray, T. P. Seland, P. Duquette, T. Gray, and R. Nelson. 1984. A population based twin study in multiple sclerosis. Am. J. Hum. Genet. **36**:49S.

Greenberg, D. A., S. E. Hodge, and J. I. Rotter. 1982. Evidence for recessive and against dominant inheritance at the HLA "linked" locus in coeliac disease. Am. J. Hum. Genet. **34**:263–277.

Haile, R. W. C., L. Iselius, P. E. M. Fine, and N. E. Morton. 1985. Segregation and linkage analysis of 72 leprosy pedigrees. Hum. Hered. **35**:43–52.

Hattori, M., J. B. Buse, R. A. Jackson, L. Glimcher, M. E. Dorf, M. Minami, S. Makino, K. Moriwaki, H. Kuzuya, H. Imura, W. M. Strauss, J. G. Seidman, and G. S. Eisenbarth. 1986. The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. Science **231**:733–735.

Hitman, G. A., A. C. Tarn, R. M. Winter, V. Drummond, L. G. Williams, N. I. Jowett, and G. F. Bottazzo. 1985. Type 1 (insulin-dependent) diabetes and a highly variable locus close to the insulin gene on chromosome 11. Diabetologia **28**:218–222.

Ho, H.-Z., J. L. Tiwari, R. W. Haile, P. I. Terasaki, and N. E. Morton. 1982. HLA-linked and unlinked determinants of multiple sclerosis. Immunogenetics **15**:509–517.

Hodge, S. E. 1981. Some epistatic two-locus models of disease. I. Relative risks and identity by descent distributions in affected sib pairs. Am. J. Hum. Genet. **33**:381–395.

Hodge, S. E., J. I. Rotter, and K. Lange. 1980. A three-allele model for heterogeneity of juvenile onset insulin-dependent diabetes. Ann. Hum. Genet. **43**:399–409.

Hors, J., and J. Dausset. 1983. HLA and susceptibility to Hodgkin's disease. Immunol. Rev. **70**:167–191.

James, J. W. 1971. Frequency in relatives for an all-or-none trait. Ann. Hum. Genet. **35**:47–49.

Kravitz, K., M. Skolnick, C. Cannings, D. Carmelli, B. Baty, B. Amos, A. Johnson, N. Mendell, C. Edwards, and G. Cartwright. 1979. Genetic linkage analysis between hereditary hemochromatosis and HLA. Am. J. Hum. Genet. **31**:601–619.

Kuwert, E. K. 1977. Genetical aspects of multiple sclerosis with special regard to histocompatibility determinants. Acta Neurol. Scand. Suppl. **55**:23–42.

Lalouel, J.-M., L. LeMignon, M. Simon, R. Fauchet, M. Bourel, D. C. Rao, and N. E. Morton. 1985. Genetic analysis of idiopathic hemochromatosis using both qualitative (disease status) and quantitative (serum iron) information. Am. J. Hum. Genet. **37**:700–718.

MacLean, C. J., N. E. Morton, and S. Yee. 1984. Combined analysis of genetic segregation and linkage under an oligogenic model. Comput. Biomed. Res. **17**:471–480.

Payami, H., G. Thomson, U. Motro, E. J. Louis, and E. Hudes. 1985. The affected sib method. IV. Sib trios. Ann. Hum. Genet. **49**:303–314.

Raffel, L. J., G. A. Hitman, G. I. Bell, J. Karam, P. H. Yen, D. J. Galton, G. F. Bottazzo, and J. I. Rotter. 1985. The insulin gene polymorphism in type 1 diabetes (IDDM) families. Ped. Res. **19**:252A.

Risch, N. 1983. The effects of reduced fertility, method of ascertainment, and a second unlinked locus on affected sib pair marker allele sharing. Am. J. Med. Genet. **16**:243–259.

———. 1984*a*. Genetic analysis workshop II: multiple-locus segregation analysis incorporating linkage markers. Genet. Epidemiol. **1**:207–211.

———. 1984*b*. Segregation analysis incorporating linkage markers. I. Single locus models with an application to type 1 diabetes. Am. J. Hum. Genet. **36**:363–386.

———. Genetic analysis workshop IV: conditional linkage analysis of multiple sclerosis families. Genet. Epidemiol. (in press).

Spielman, R. S., L. Baker, and C. M. Zmijewski. 1980. Gene dosage and susceptibility to insulin dependent diabetes. Ann. Hum. Genet. **44**:135–150.

Spielman, R. S., and N. Nathanson. 1982. The genetics of susceptibility to multiple sclerosis. Epidemiol. Rev. **4**:45–65.

Stewart, G. J., J. G. McLeod, A. Basten, and H. V. Bashir. 1981. HLA family studies and multiple sclerosis: a common gene, dominantly expressed. Hum. Immunol. **3**:13–29.

Suarez, B., D. O'Rourke, and P. Van Eerdewegh. 1982. Power of the affected sib pair method to detect disease susceptibility loci of small effect: an application to multiple sclerosis. Am. J. Med. Genet. **12**:309–326.

Svejgaard, A., and L. P. Ryder. 1981. HLA genotype distribution and genetic models of insulin-dependent diabetes mellitus. Ann. Hum. Genet. **45**:293–298.

Thomson, G. 1980. A two locus model for juvenile diabetes. Ann. Hum. Genet. **43**:383–398.

———. 1981. A review of theoretical aspects of HLA and disease associations. Theor. Popul. Biol. **20**:168–208.

Thomson, G., and W. F. Bodmer. 1977. The genetic analysis of HLA and disease associations. Pp. 84–93 *in* J. Dausset and A. Svejgaard, eds. HLA and disease. Munksgaard, Copenhagen.

Thomson, G., W. Klitz, E. J. Louis, S. K. Lo, J. Bertrans, M. Baur, and M. Neugeberger. HLA and IDDM predisposition: new aspects. Genet. Epidemiol. (in press).

Tiwari, J. L., H. Betuel, L., Gebruhrer, and N. E. Morton. 1984. Genetic epidemiology of coeliac disease. Genet. Epidemiol. **1**:37–42.

Wagener, D. K., J. M. Sacks, R. E. LaPorte, and J. M. MacGregor. 1982. The Pittsburgh study of insulin-dependent diabetes mellitus: risk for diabetes among relatives of IDDM. Diabetes **31**:136–144.