Disequilibrium Pattern Analysis. II. Application to Danish HLA *A* and *B* Locus Data

William Klitz and Glenys Thomson

Department of Genetics, University of California, Berkeley, California 94720

Manuscript received September 3, 1986
Revised copy accepted May 2, 1987

ABSTRACT

Disequilibrium pattern analysis, a general method for analyzing evolutionary events acting on pairs of tightly linked polymorphic loci, is applied to a large sample of Danish individuals typed for *A* and *B* loci of the HLA (human leukocyte antigen) system. Cases of selection on particular haplotypes are revealed from patterns of linkage disequilibrium among the HLA haplotypes. These patterns cannot be explained by either population admixture or random genetic drift. Six haplotypes out of the total array of 273 haplotypes have been identified which show in varying extents the patterns indicating selection.

TWO population genetic features of the HLA (human leukocyte antigen) system make it ideally suited for the study of evolutionary forces acting on a population: the high level of polymorphism exhibited by many of the loci, and the existence of significant linkage disequilibrium (non-random association) between certain pairs of alleles at different loci. HLA is the major histocompatibility complex (MHC) in humans. This complex of genes codes for cell surface proteins which play a central role in governing the immune system (HOOD, STEINMETZ and MALISSEN 1983; LARSEN 1981; KLEIN, FIGUEROA and NAGY 1983; LEW *et al.* 1986; AUFFRAY and STROMINGER 1986). At least 23 antigens have been defined for the HLA *A* locus, 47 for the *B* locus, 8 for *C*, 19 for *D* and 14 for *DR* (see *Histocompatibility Testing 1984*, pp. 4–8, ALBERT, BAUR and MAYR 1984). In no case is there a single common allele; rather there are quite a few alleles with relatively even frequencies (DAUSSET and COLOMBANI 1972; BAUR and DANILOVS 1980; BODMER and THOMSON 1977; THOMSON 1981; HEDRICK and THOMSON 1983; BAUR *et al.* 1984; KLITZ, THOMSON and BAUR 1986). Consequently, the level of heterozygosity at each locus is very high. In addition, significant linkage disequilibrium is detected for many pairs of antigens (see for example BAUR and DANILOVS 1980; and BAUR, NEUGEBAUER and ALBERT 1984). The most widely quoted example is the high degree of association between the antigens A1 and B8 in Caucasian populations.

When linkage disequilibrium values for two locus HLA haplotypes were first calculated (CEPPELLINI *et al.* 1967; BODMER and BODMER 1970; DAUSSET and COLOMBANI 1973), haplotypes with a significant excess of numbers were noted, and assumed to be candidates for selection, and, therefore, of special interest. Although the haplotypes comprising the great majority of the population are those without strong positive disequilibria, these have not, until now, been considered useful in the search for the forces governing interlocus associations.

In this paper we apply a method, disequilibrium pattern analysis, for examining the disequilibrium distribution of the entire array of the HLA *A* and *B* locus haplotypes in the Danish study by HANSEN *et al.* (1979). The theoretical foundation for the method is outlined in the companion paper (THOMSON and KLITZ 1987). This method identifies the constraints on the values that a disequilibrium coefficient can assume, based on the distribution of the entire population of haplotypes in the disequilibrium space. Specific quantifiable patterns are predicted as a result of selective events. In particular, the following criteria reveal selection in a chromosomal region and are used to identify those two-locus haplotypes showing the effects of selection.

1. The presence of just one or a few haplotypes in the positive disequilibrium space when we plot the linkage disequilibrium for all haplotypes containing a given allele.

2. Haplotypes sharing one allele with a favored haplotype (termed related haplotypes) have an expected value of linkage disequilibrium, D, which is negative and proportional to the frequency of the unshared allele, and have an expected single standardized disequilibrium value, D'. The common related haplotypes will better estimate this value than the rarer related haplotypes, which will be more widely distributed around this point due to sampling effects.

In addition, the number of haplotypes having positive and negative disequilibrium values is used to predict the minimum number of selection events in a

two locus multiallelic system, and the less common of the two alleles in a selected two-locus haplotype will have arisen more recently in evolutionary time.

RESULTS

A population study of HLA *A* and *B* locus haplotypes: The large size (5,202 individuals), and relative homogeneity of the Danish population sample are both desirable characteristics for the application of disequilibrium pattern analysis.

The disequilibrium space is defined by either the linkage disequilibrium parameter D, or the standardized (or normalized) linkage disequilibrium parameter D' on one axis, and on the other axis by the expected haplotype frequency, in the absence of linkage disequilibrium for that haplotype, that is $p_{A_i}q_{B_j}$ (where p_{A_i} and q_{B_j} are the frequencies of the ith and jth alleles at each of the two loci). The normalized linkage disequilibrium D' *vs.* $p_{A_i}q_{B_j}$ for the array of 273 haplotypes from the 13 HLA *A* and 21 HLA *B* alleles from the Danish population is plotted in Figure 1. The haplotypes exhibit disequilibrium values forming a broadly distributed group of points, occupying both the positive and negative spaces. Many low frequency haplotypes are found in the two regions where D' equals either -1.0 or 0.0. There are 173 haplotypes with negative and 100 with positive disequilibrium values.

Patterns from the data suggesting selection: This total array of haplotypes was subdivided by considering one allele at one locus in combination with each of the alleles at another locus, for example *A1* with each of the *B* locus alleles (Figure 2). This subdivision revealed cases where patterns indicative of a strong selection event were apparent, as well as cases which did not exhibit such patterns. From the entire set of 34 graphs of the 13 HLA *A* and 21 HLA *B* allelic combinations, 5 have been chosen as illustrative of these cases.

Haplotype A1B8, which has the highest positive disequilibrium (and normalized disequilibrium) value in this population ($D = 0.0766$, $D' = 0.728$) reveals a pattern indicative of selection. In the B8 graph (Figure 3) A1B8 is the only haplotype in the positive space, while all related haplotypes (B8notA1) fall in a linear array in the negative space with disequilibrium values approximately proportional to the frequency of the unshared *A* allele. The same pattern is apparent with the A1 graph (Figure 2), except that two additional haplotypes (A1B17 and A1B37) are in the positive space.

The graph of normalized linkage disequilibrium values, D', shown in Figure 4 for A1 haplotypes reveals an alignment of the negative values for the commoner haplotypes. Most of the values fall between -0.6 and -0.8. Rarer haplotypes, for example (A1B47, A1B38 and A1B13) depart furthest from this alignment apparently due to sampling effects. The A1BX haplotype is an undefined mixture of *B* locus alleles, which available antisera were unable to define. The greater departure of the D' value for the A1BX haplotype (Figure 4) from the other values may be a reflection of its heterogeneous composition. The graph of the B8 haplotypes in the negative space (not shown) also displays this alignment for the commoner haplotypes, with the normalized values again falling mainly between -0.6 and -0.8. Note that, the normalized disequilibrium value of 0.728 for A1B8, matches with these negative normalized values, as predicted by the theory for selection (THOMSON and KLITZ 1987).

HLA alleles *A1* and *B8* are both relatively common, having frequencies of 0.17 and 0.13, respectively. The presence or absence of patterning is not a function of high or low allele frequencies of the constituent alleles. Comparison should be made between the linkage disequilibrium, D, graphs of *A1* and *B8* (Figures 2 and 3) which show the pattern of selection, with that of another common allele *A2* (frequency 0.31), which shows no particular pattern (Figure 5). Nine A2 haplotypes are in the positive space and 11 in the negative space. The largest positive disequilibrium value is 0.21 for A2B15 ($D' = 0.34$). No linear alignment of the unrelated haplotypes is apparent.

The disequilibrium space of a less common allele *A11* (frequency 0.054) also shows no apparent pattern indicative of selection, with haplotypes scattered across the disequilibrium space (Figure 6). On the other hand, the allele *A29*, which is less common than *A11* (with a frequency of 0.025), does show a linear alignment of the disequilibrium with allele frequency for the haplotypes in the negative space (Figure 7). There are five haplotypes in the positive disequilibrium space, two of which have D' values which are quite high, namely *B44* ($D' = 0.537$) and *B45* ($D' = 0.387$). The allele *Aw32* which has the same frequency as *A29*, however, shows no pattern indicative of selection (not shown) and the largest D' value in this case is 0.10 for A32B40.

These examples illustrate the distinction between patterned and unpatterned distributions of haplotypes in the disequilibrium space.

The number of selected haplotypes: The number of favorable independent selective events in a two locus system, K, is given by the quadratic equation

$$K = \frac{b \pm \sqrt{b^2 - 4M}}{2}$$

where b is the total number of alleles at both loci minus one, and M is the number of haplotypes with negative D values (THOMSON and KLITZ 1987). The Danish sample has 13 HLA *A* alleles and 21 HLA *B*

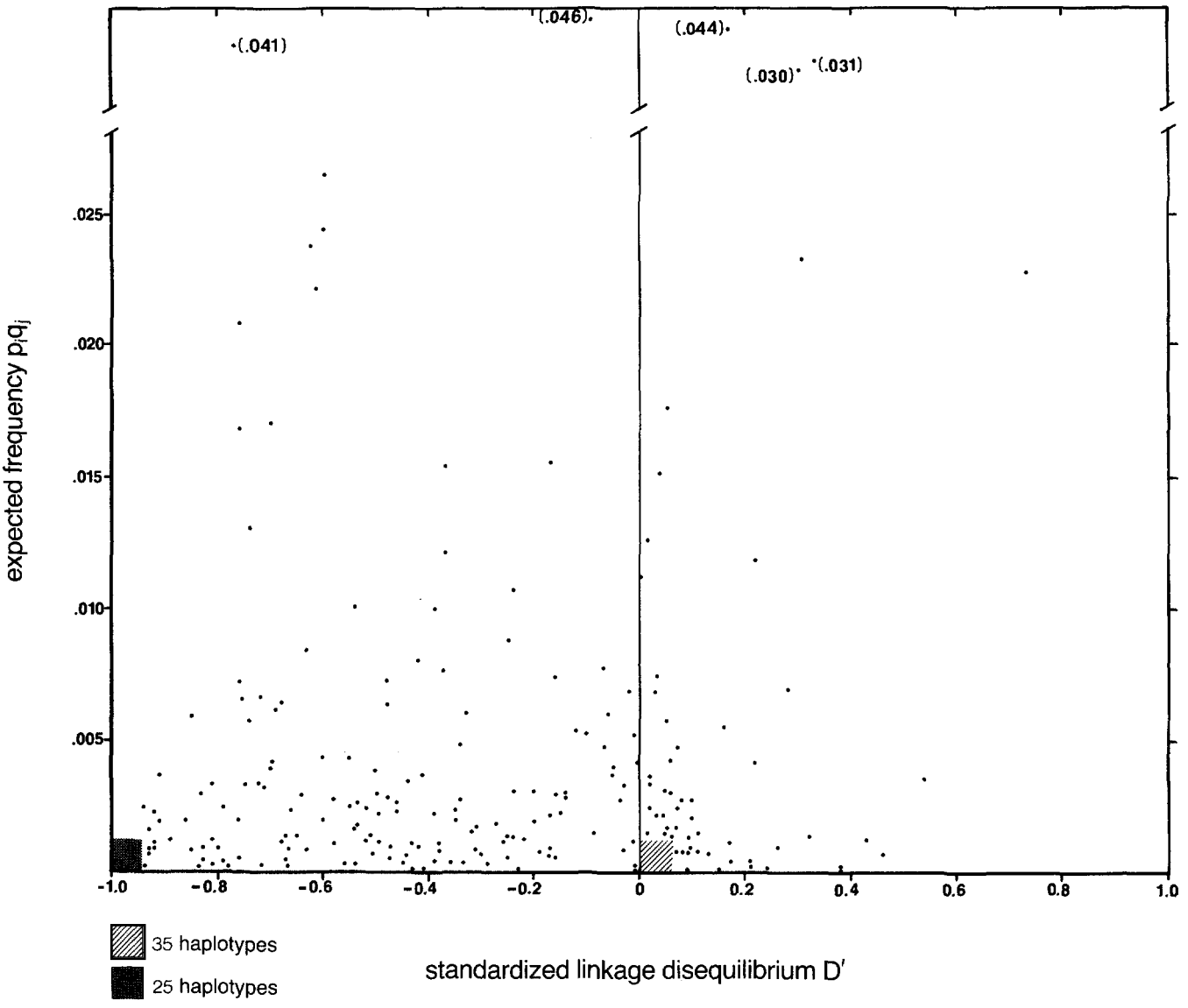

FIGURE 1.—Expected haplotype frequencies and D' values of the 273 HLA *A–B* locus haplotypes taken from a sample of 5202 Danes (HANSEN *et al.* 1979). These haplotypes, constructed from the 13 HLA *A* and 21 HLA *B* alleles, are plotted on the disequilibrium space, which is comprised of the frequency of a haplotype expected in the absence of disequilibrium, pq, and the standardized linkage disequilibrium, D'.

alleles with 173 haplotypes in the negative D space. This yields an estimate of 6.54 selective events. Note that this derivation assumes that the selective events are occurring concurrently and that a particular allele does not occur more than once in a selected haplotype. Thus, the estimate of six to seven selective events should be taken merely as a guide for understanding the evolutionary events acting on the region.

The influence of neutrality forces: The theoretical distribution of haplotypes in the disequilibrium space under the influence of neutrality and migration were considered in the companion paper (THOMSON and KLITZ 1987). This work showed that, in general, neutrality and migration produce different patterns of haplotype distribution than that found under selection. Here we also compare the sample of Danish HLA data with the overall distribution of standardized disequilibrium values generated under a Fisher-Wright neutral allele model (HUDSON 1983; HEDRICK and THOMSON 1986). The parameters of the two-locus neutrality simulation were chosen to imitate as nearly as possible those of the HLA loci in the Danish study. The recombination parameter, or actually $4Nc$, where N is the effective population size and c the recombination frequency between the two loci, was set to 160.0, making it comparable to the 0.8% recombination rate observed between the HLA *A* and *B* loci with an effective population size of 5000. HEDRICK and THOMSON (1986) showed that little change occurs in the distribution of disequilibrium under neutrality conditions when $4Nc$ is greater than 100, so that an exact estimate of population size is not important for our use here. The mutation parameter θ or $4N\mu$, where μ is the mutation rate, was set to 4.2 to generate

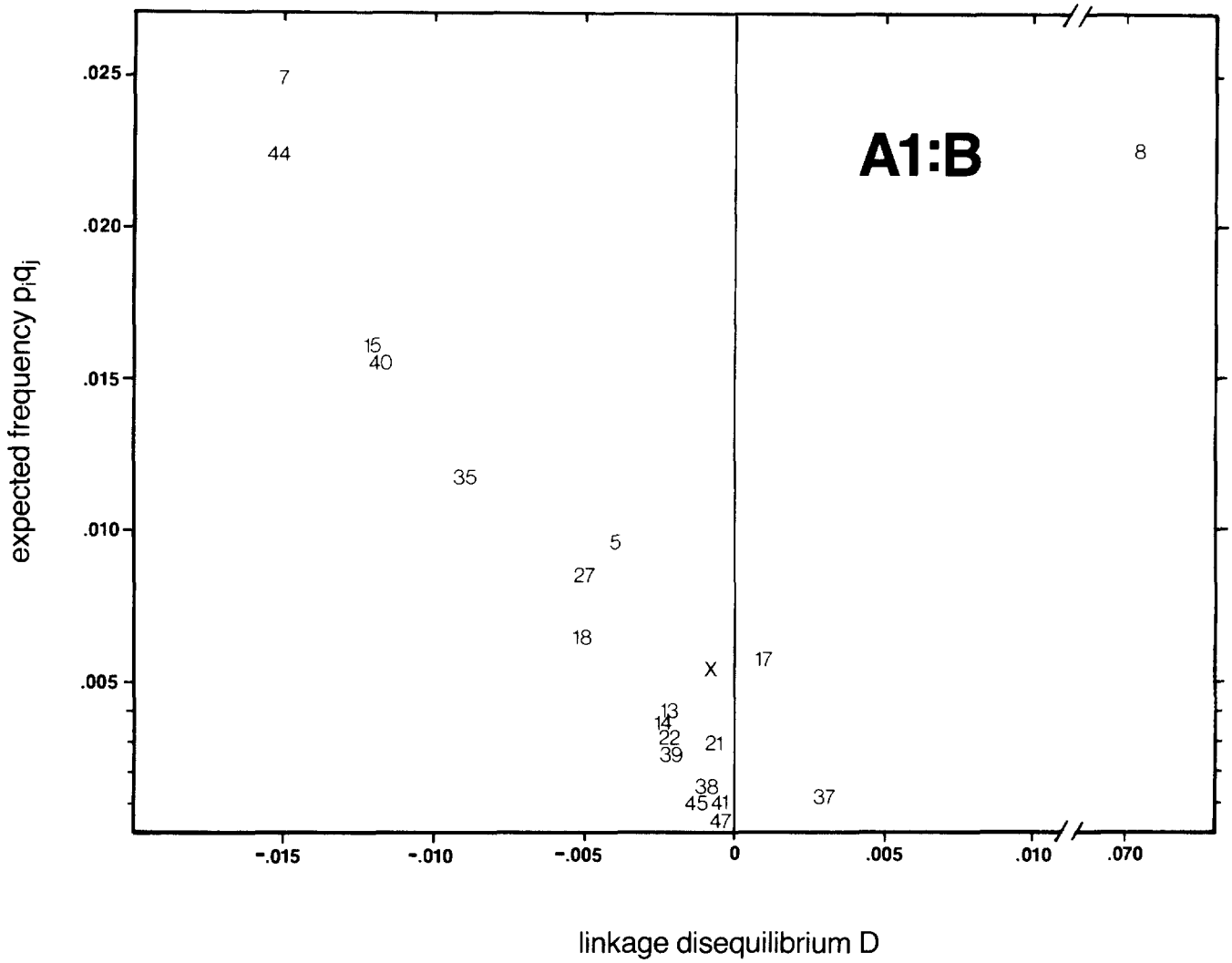

FIGURE 2.—All haplotypes containing the allele *A1* taken from the Danish population are plotted for their expected frequency in the absence of disequilibrium, pq, against the observed linkage disequilibrium parameter D. The *B* allele designation is used as the position of each haplotype.

the large numbers of alleles similar to those found at the HLA loci. This mutation rate, on the order of $1/N$, is quite high, although some experimental studies suggest that gene conversion, which may be the mechanism generating much of the variation at the MHC, does have very high rates of occurrence (LOH and BALTIMORE 1984).

We used the observed frequency of blank alleles at the HLA *A* and *B* loci (0.022 and 0.032, respectively) in the Danish study to create artificial blank classes in the simulations. The rarest allele at a locus was combined with consecutively less rare alleles until a frequency equal to or greater than the observed blank class at the HLA *A* and *B* loci was achieved. This procedure makes the single locus allele frequency distribution closer to that observed for the HLA data. It also reduces the number of alleles found in only one, or very few, copies on a single haplotype and thus reduces the number of missing haplotypes having D' of -1. The size of the $D' = -1$ haplotypic class is a hallmark of neutrality conditions: thus our procedure of reducing this class makes the test of neutrality *vs.* selection conservative.

Because it was found that many dozens of runs were necessary to generate one population with the $k = 13$ and $l = 21$ present in the Danish sample, 19 different populations with k from 10 to 14 and l from 17 to 23 were accepted. An examination of the D' distributions of these various simulated populations revealed no heterogeneity in the patterns generated, and we combined these populations with their limited range of k, l values.

In order to accurately characterize the variation in disequilibrium expected under neutrality, 1010 replicates were obtained under the conditions described above. Figure 8 shows the frequency distribution of D' for the total haplotype array for the neutrality case and the Danish data in intervals of 0.1 (plus the maximum and minimum classes, +1.0 and −1.0). For each D' interval in the neutrality case the median, the

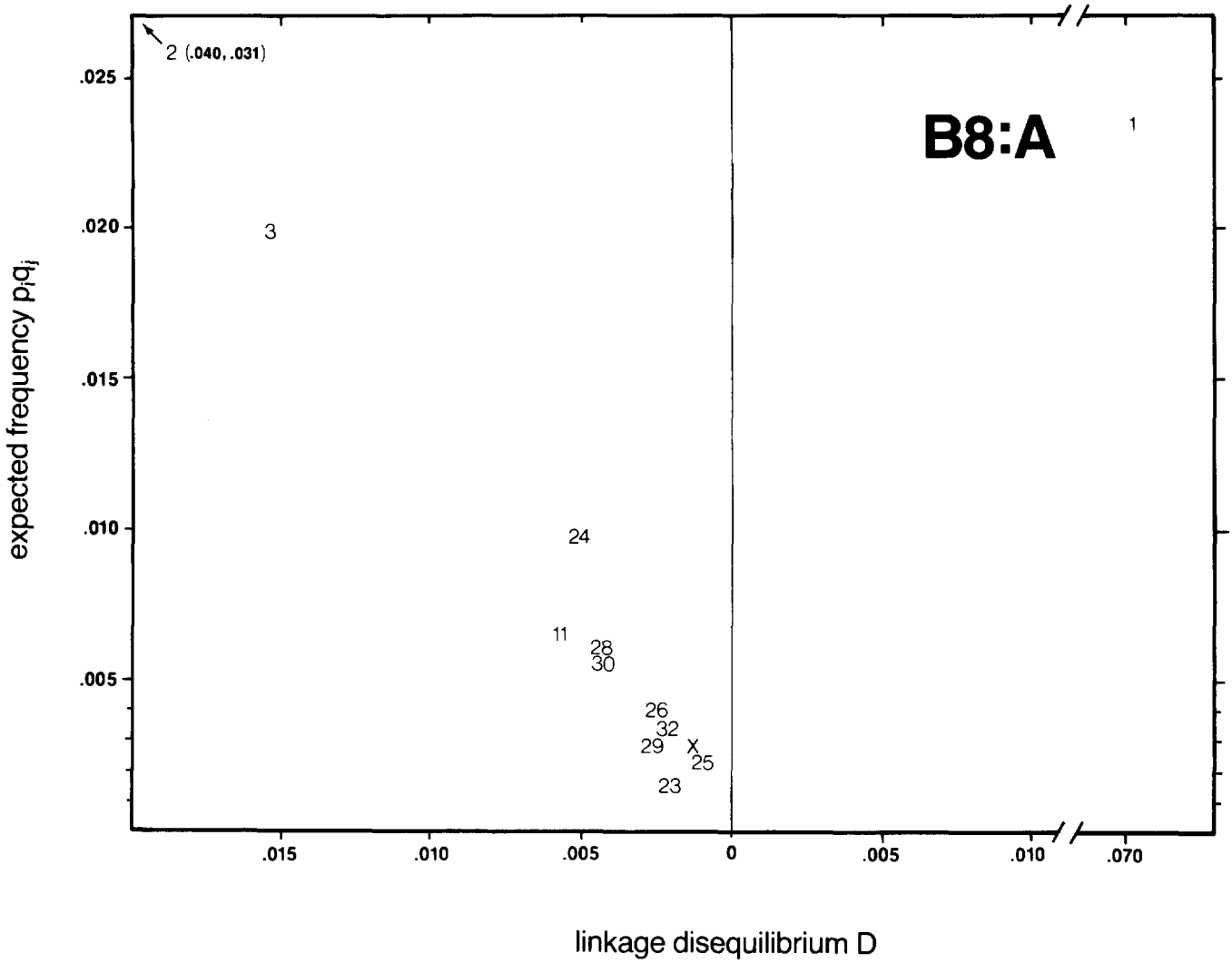

FIGURE 3.—All haplotypes containing the allele *B8* in the disequilibrium space. The *A* allele designation is the position of each haplotype.

95% interval, and the range are indicated. The median frequency of the number of haplotypes at −1.0 under neutrality is very high (40.8%). Under neutrality in the positive space, the frequency of the number of haplotypes declines rapidly from its maximum of 12.1% in the 0 to +0.1 interval. It does not drop quite to zero, however, but maintains a low and constant incidence of haplotypes with high positive D' values. These haplotypes with high positive D's are all quite rare, and are responsible for the large $D' = -1$ class, consisting of related haplotypes.

The observed distribution of D' for the sample of 10,404 Danish HLA *A–B* haplotypes are superimposed on the neutrality distribution in Figure 8. The Danish distribution is clearly different, falling close to or beyond the 95% interval in ten regions and entirely outside the range of observations from the simulations in three of these. The distributions differ in three areas. First, the incidence of haplotypes with $D' = -1.0$ at 7.3% for the Danish population is far below that observed under neutrality. Second, for D' ranging from −0.8 to 0.0 the Danish distribution exceeds that of neutrality. Third, at 25.3% the proportion of Danish haplotypes in the D' interval 0 to +0.1 significantly exceeds that expected under neutrality.

Disequilibrium pattern analysis predicts that selection favoring a small fraction of the 273 possible Danish haplotypes could produce this pattern. A small number of selected haplotypes would not be expected to stand out in the moderate to high positive D' space (Figure 8). The related haplotypes are present in excess numbers in the negative space, while the unrelated haplotypes cluster in the low positive D' space.

DISCUSSION

Disequilibrium pattern analysis is a general method useful for revealing the evolutionary dynamics of variation in tightly linked highly polymorphic loci in natural populations. It has been applied here to antigen data from the MHC in humans, but is equally applicable to other systems and data including restriction fragment length polymorphisms.

Identification of selected haplotypes: Examining two-locus haplotypes of HLA data at the *A* and *B* loci

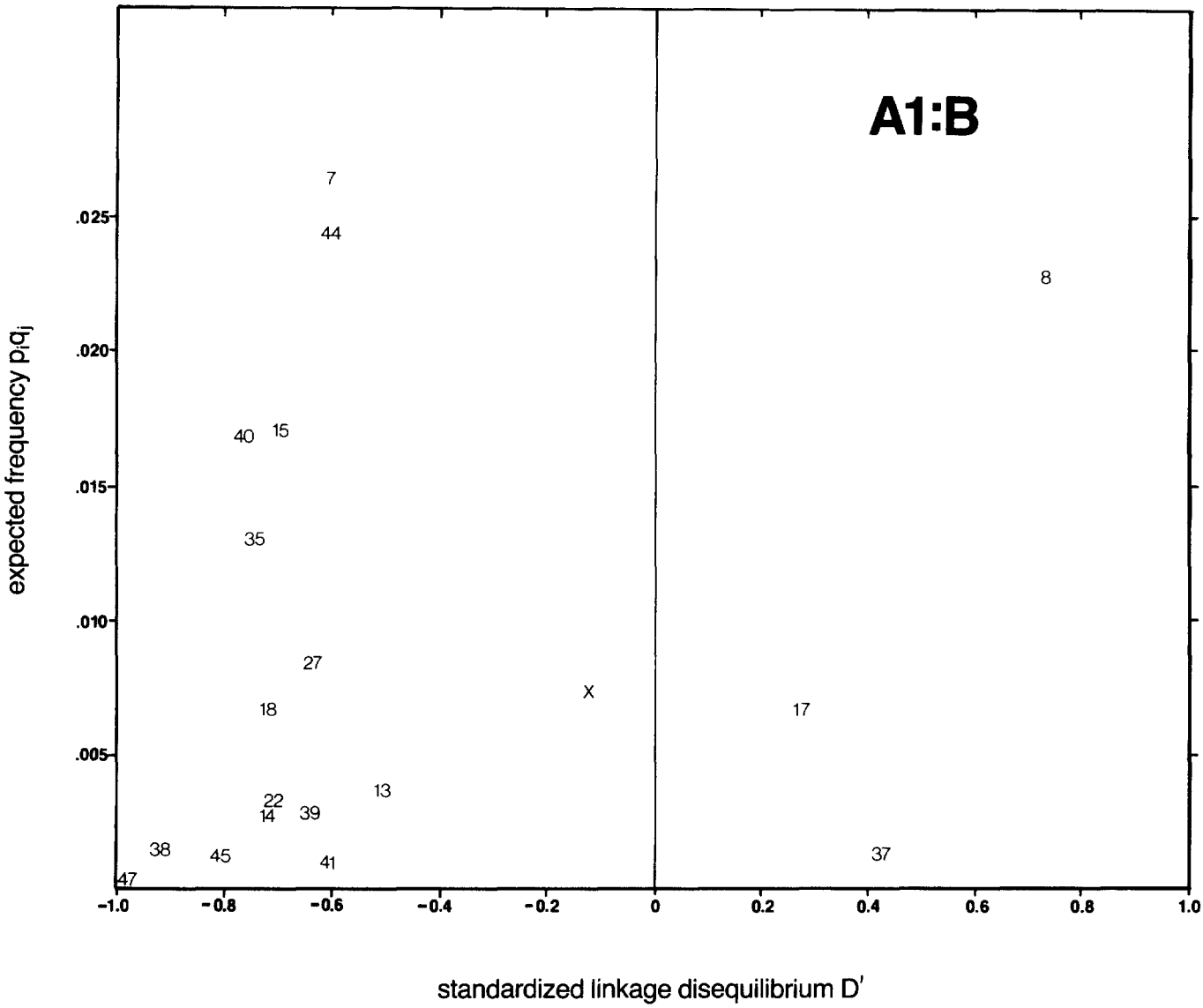

FIGURE 4.—All haplotypes containing the allele *A1* plotted against the standardized disequilibrium *D'*.

in a population of Danes, we have been able to identify six haplotypes which clearly show the patterning indicative of a selected event. The 6 haplotypes are A1B8, A3B7, A25B18, A29Bw44, A2B15 and A1B17 (Table 1). Only in two cases, A1B8 and A3B7, is the patterning apparent in both the A allele *vs.* all B alleles graph, and vice versa. For A25B18 and A29Bw44, the pattern is observed only in the *A* allele *vs.* all *B* alleles graph. For the cases, A2B15 and A1B17, the pattern is observed only in the *B* allele *vs.* all *A* alleles graph. The conditional graphs of A1, A25, B7 and B8 show the strongest patterns indicative of selection, with the negative *D'* values closely clustered around a single *D'* value (see Figure 4 for A1).

For the four cases in which patterning is only exhibited in one graph combination, patterning is always for the rarer allele *vs.* all the alleles at the other locus (see Table 1). This is in agreement with deterministic predictions of a new allele appearing at one locus creating a new haplotype with the existing allele at a second locus. The new haplotype then increases in frequency due to direct selection or hitchhiking with patterning most clearly expressed in the new allele.

The two cases which show patterning for both graph combinations, A1B8 and A3B7, have high, and relatively equal frequencies, for both the *A* and *B* locus alleles. The more frequent of the two alleles in these two cases, *A1* and *A3*, both have other significant ($P < 0.001$) positive disequilibrium values. These results are all compatible with deterministic predictions of selective events.

In Table 1 the six haplotypes displaying the selective pattern are given, as well as haplotypes showing weaker positive associations with the alleles of these haplotypes, the allele frequencies, and normalized linkage disequilibria values. As indicated above, note that when the disequilibrium of a secondary haplotype differs significantly from zero it tends to be associated

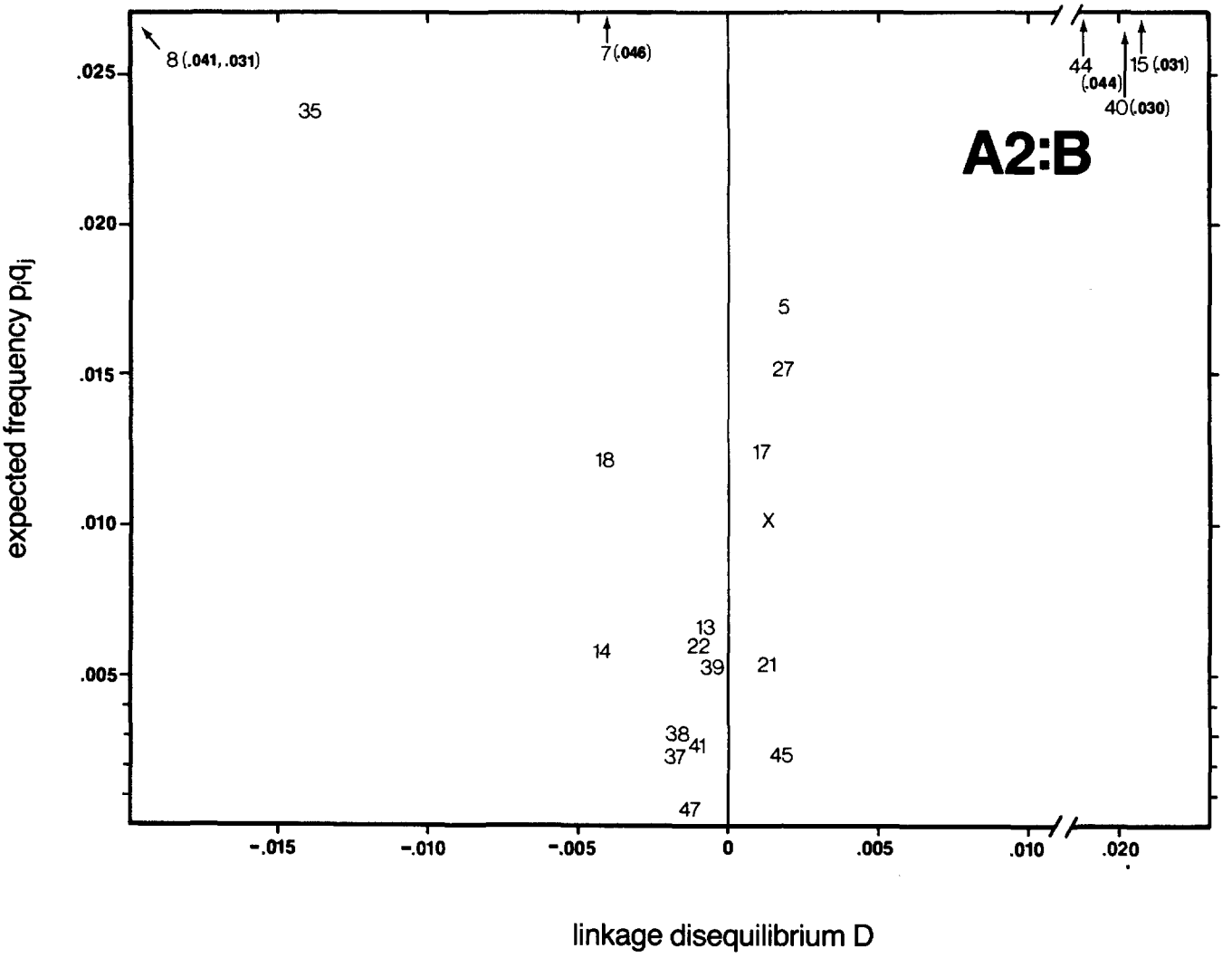

FIGURE 5.—All haplotypes containing allele *A2* in the disequilibrium space.

with the more common allele of the favored haplotype (Table 1). The only exception is in the case of A29Bw44 where the rarer allele of the pair, *A29*, does have a significant secondary positive (A29Bw45). However, Bw44, the commoner allele has four significant secondary positives, affirming the basic tendency. We have not attempted to classify as patterned or unpatterned the graphs of alleles with frequencies less than 1.5%, due to problems with large sampling errors for rare alleles and normalized negative disequilibrium values close to −1.

Population admixture and Danish HLA *A–B* disequilibrium: Population movement and mixing is a major theme of European history. In fact, allele frequency data from HLA and other loci sampled from modern European populations apparently retain evidence of major population movements in the region (MENOZZI, PIAZZA and CAVALLI-SFORZA 1978). The clearest migratory pattern implied from this work is the movement of near-eastern neolithic farmers northward over the period from 9 to 5 thousand years ago, bringing both lifestyle and genetic changes to the mesolithic Europeans in the process. Because of this history of admixture, a preliminary examination of the possible role of migration in influencing HLA haplotype and allele frequencies is necessary.

FELDMAN and CHRISTIANSEN (1975) showed that with sufficient differences in allele frequencies between two populations at each of two linked loci, a stepping stone model of population admixture can generate high levels of disequilibrium, and that a cline in both allele frequencies and disequilibrium levels will result. In general, the HLA *A–B* haplotypes in Europe seem to satisfy these conditions with genetic and historical evidence of population movement and admixture from south and north and corresponding clines in both allele and haplotype frequencies for A1–B8 and A3–B7 (DAUSSET and COLOMBANI 1973). The disequilibrium levels for these haplotypes is highest in the north, which is just as expected where the contrast in allele frequencies of the mixing populations would have been the greatest.

However, a closer examination of HLA data has suggested that migration is not sufficient to account

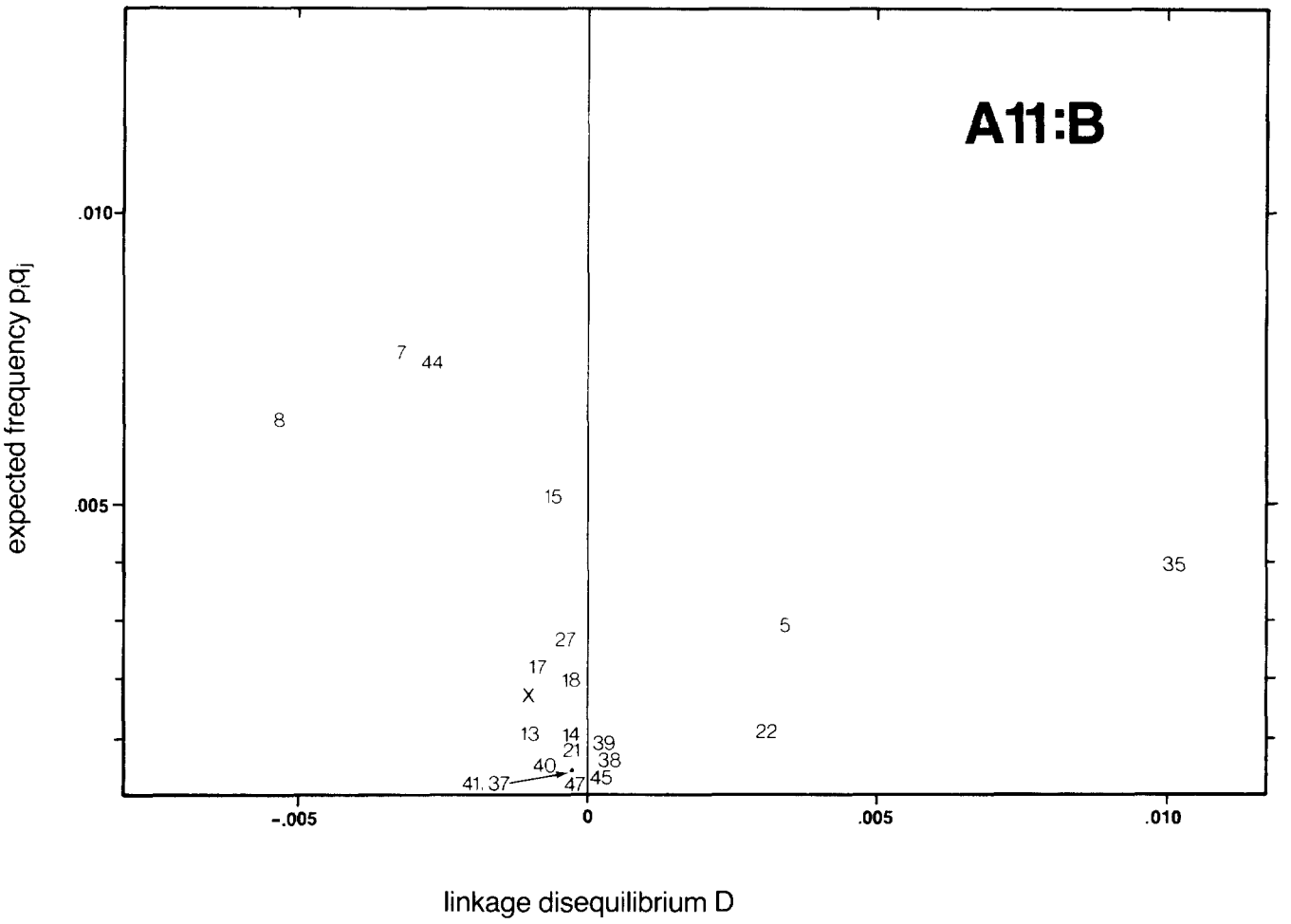

FIGURE 6.—All haplotypes containing allele *A11* in the disequilibrium space.

for observed HLA allele frequency and disequilibrium data in Europe. HEDRICK and THOMSON (1983) have shown that the distribution of *single* locus allele frequencies at both HLA *A* and HLA *B* are not easily explained by migration. In discussing HLA *A–B* disequilibrium in Europe, THOMSON, BODMER and BODMER (1976) pointed out that even if the admixture of mesolithic and neolithic peoples might have generated high disequilibrium levels and clines, the recombination between HLA *A* and *B* over the subsequent 5000 years since farming was established in Denmark (AMMERMAN and CAVALLI-SFORZA 1984) would have diminished the disequilibrium level (generated in this case by admixture) by 80%. This implies that the high positive disequilibrium levels observed in several Danish haplotypes must be due to more recent events. In addition, this reminds us that disequilibrium generated in some discrete period disappears over time, while clines in allele frequencies can persist much longer.

The predictions of disequilibrium pattern analysis enrich the discussion of the role of migration in generating HLA *A–B* disequilibrium in the Danish by specifying the required differences in parental populations necessary to generate the observed selection patterns in the disequilibrium space. The proportionality pattern of related haplotypes in the disequilibrium space, characteristic of selection, can be mimicked by population admixture only when three conditions are met (THOMSON and KLITZ 1987). (For comparison to the selection model we refer to a single haplotype whose alleles have extreme frequency differences between the parental populations as "selected," etc.) First, both of the alleles of the selected haplotype must be more common in say, population 1, than any of the other alleles in population 2. Second, all remaining alleles in population 1 must be rarer in population 1 than in population 2. Finally, the frequencies of all of the nonselected alleles at each of the loci in population 1 must be proportional to those in population 2. The only feasible case meeting these conditions occurs when a haplotype is fixed in one of the parental populations. These strict requirements for generating the selection pattern by means of admixture argues against the role of admixture in creating the major patterns of disequilibrium seen among HLA loci of the modern Danish population.

Selection: One interpretation of the disequilibrium among HLA haplotypes concordant with the results of disequilibrium pattern analysis is viability selection

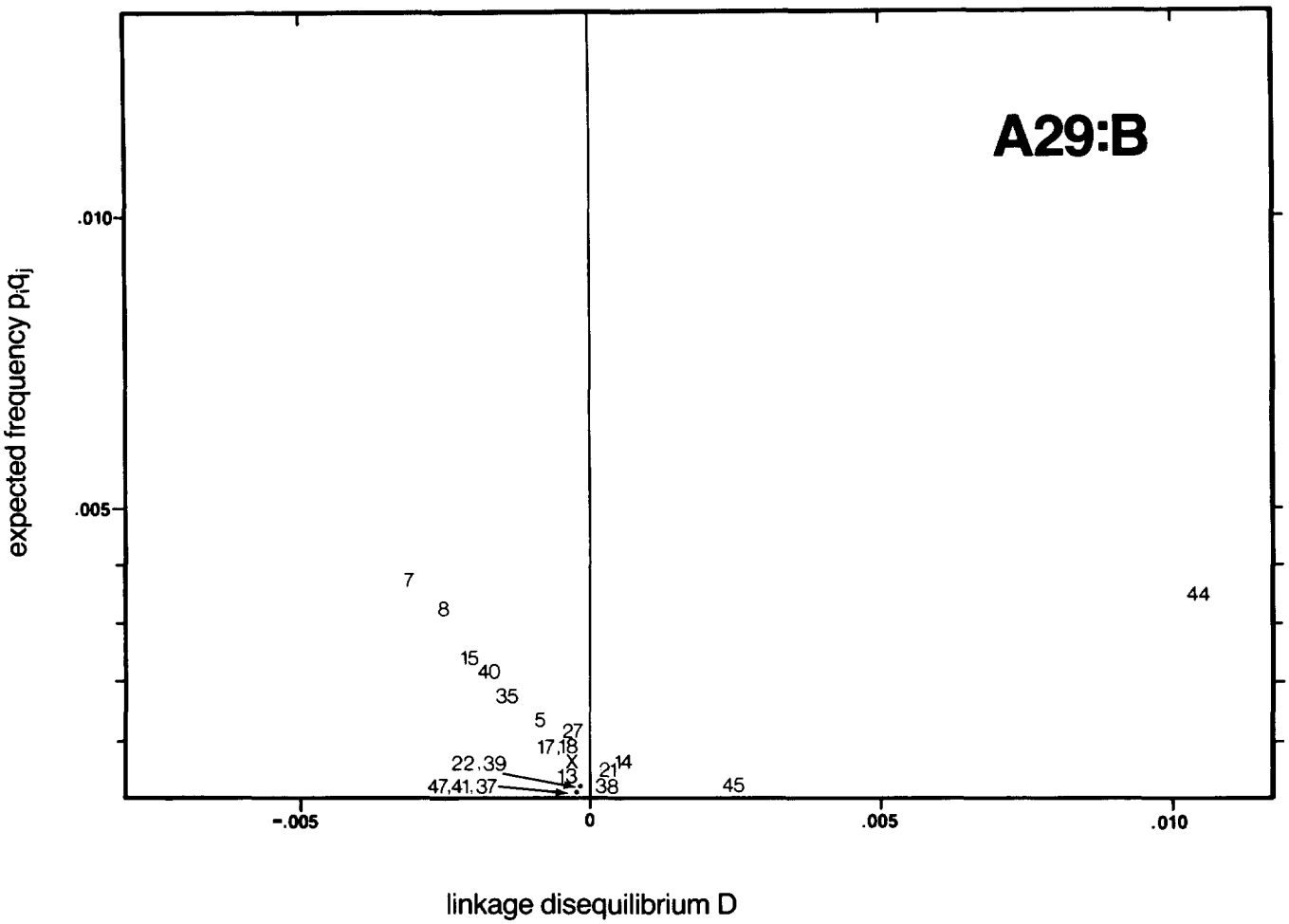

FIGURE 7.—All haplotypes containing allele *A29* in the disequilibrium space.

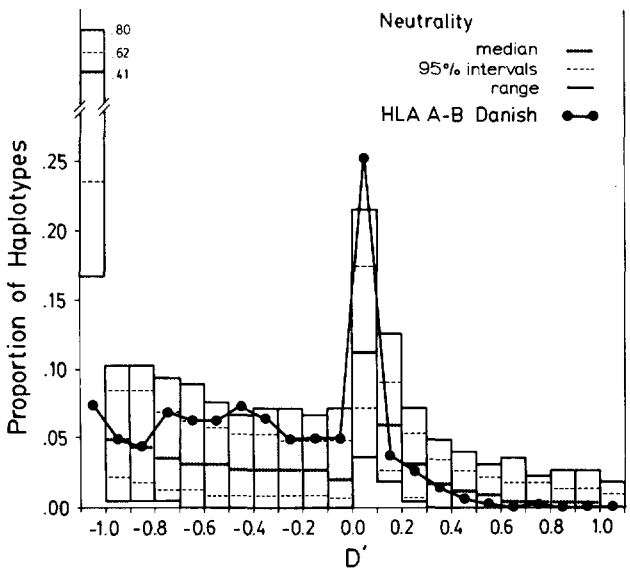

FIGURE 8.—The distribution of standardized linkage disequilibrium *D'* for haplotypes at the Danish HLA *A*, *B* loci, and for 1010 replicate runs of a two locus neutrality model.

due to disease pathogens. This is supported by two independent lines of evidence. First, different HLA antigens have different patterns of immune response to infectious agents. This work, outlined by VAN ENDEN, DE VRIES and VAN ROOD (1982), shows a rich array of individual variation in response to major infectious agents based on HLA antigen type. Second, the change in human demography brought about by the introduction of agriculture beginning 5000 years ago created conditions suitable for the establishment and spread of the infectious agents responsible for the epidemics of recorded history (BLACK 1975). Disequilibrium pattern analysis may be a useful aid in revealing the particular alleles and haplotypes which allowed an evolutionary response to these new mortality agents.

In addition to viability selection, other agents have been proposed to influence the evolution of the MHC, including fetal maternal effects (WARBURTON 1968), segregation distortion (ALPER *et al.* 1985), and nonrandom mating (YAMAZAKI *et al.* 1976). The relative impact and combined influence of these various forces on allelic variation and on the patterns and extent of linkage disequilibrium awaits further attention.

A distinctive feature of the HLA data at the single locus level is the high degree of polymorphism of the loci in combination with a relatively even distribution

TABLE 1

The six HLA *A-B* haplotypes which show, in varying extent, the pattern of a selected event

Favored haplotype	D'		Pattern	Allele frequency	Additional positive disequilibria
A1B8	0.728	*A1*	Yes	0.172	B17*, B37*
		B8	Yes	0.127	None
A3B7	0.310	*A3*	Yes	0.157	Bw35*, B14
		B7	Yes	0.143	Aw24, AX
A25B18	0.470	*A25*	Yes	0.019	B17, Bw37
		B18	No	0.037	A(w30, w31)*, Aw24, A26, Aw32, AX
A29Bw44	0.537	*A29*	Yes	0.025	Bw45*, B14, B21, Bw38
		Bw44	No	0.136	A2*, Aw23*, A28*, A29*, Aw32, AX
A2B15	0.336	*A2*	No	0.313	B40*, Bw44*, B5, B17, B21, B27, BX
		B15	Yes	0.095	Aw24
A1B17	0.271	*A1*	Yes (secondary)	0.172	B8*, Bw37*
		B17	Yes	0.039	A25, A(w30, w31)

For each of these haplotypes the allele frequencies and normalized linkage disequilibrium are given. In each case it is noted whether the pattern of selection is observed for both allele combination graphs or not. Also given are the alleles which show positive associations with an allele of the favored pair.

* Differs from $D = 0$ at $P < 0.001$, all other positive disequilibria are non-significant ($P > 0.05$).

of allele frequencies. Previous studies have shown that the class I and II HLA *A*, *B*, *C*, *DQ* and *DR* loci, as well as the *Glo-I* (glyoxylase) locus, have single locus homozygosity levels significantly less than neutrality expectations (HEDRICK and THOMSON 1983, KLITZ, THOMSON and BAUR 1984, 1986; HEDRICK, THOMSON and KLITZ 1986). After consideration of various evolutionary factors, HEDRICK and THOMSON (1983) suggested that some form of balancing selection is the explanation most consistent with the level of homozygosity at the HLA *A* and *B* loci in the populations they studied. Our results are not contradictory to the notion that the high levels of variation of the HLA loci is maintained by a selective mechanism, and that possibly all the HLA alleles have been subject to some degree of selection.

We thank P. HEDRICK for his helpful commentary on the manuscript, and R. HUDSON for the use of his elegant simulation program. This research was supported by National Institutes of Health grants HD12731 and GM35326 and a UC Berkeley IBM Academic Information Systems grant.

LITERATURE CITED

ALBERT, E. D., M. P. BAUR and W. R. MAYR (Editors), 1984 *Histocompatibility Testing 1984.* Springer Verlag, Berlin.

ALPER, C. A., A. L. AWDEH, D. D. RAUM and E. J. YUNIS, 1985 Possible human analogs of the murine T/t complex. Expl. Clin. Immunogenet. **2:** 125–136.

AMMERMAN, A. J. and L. L. CAVALLI-SFORZA, 1984 *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton University Press, Princeton, N.J.

AUFFRAY, C. and J. L. STROMINGER, 1986 Molecular genetics of the human major histocompatibility complex. pp. 197–247. In: *Advances in Human Genetics*, Vol. 15, Edited by H. HARRIS and K. HIRSCHORN. Plenum Press, New York.

BAUR, M. P. and J. A. DANILOVS, 1980 Reference tables of two and three locus haplotype frequencies for HLA-A, B, DR, BF and GLO. In: *Histocompatibility Testing 1980*, Edited by P. I. TERASAKI. UCLA Tissue Typing Laboratory, Los Angeles.

BAUR, M. P., M. NEUGEBAUER and E. D. ALBERT, 1984 Reference tables of two-locus haplotype frequencies for all MHC marker loci. pp. 677–755. In: *Histocompatibility Testing 1984*, Edited by E. D. ALBERT, M. P. BAUR and W. R. MAYR. Springer-Verlag, Berlin.

BAUR, M. P., M. NEUGEBAUER, H. DEPPE, M. SIGMUND, T. LUTON, W. R. MAYR and E. D. ALBERT, 1984 Population analysis on the basis of deduced haplotypes, pp. 333–341. In: *Histocompatibility Testing 1984*, Edited by E. D. ALBERT, M. P. BAUR and W. R. MAYR. Springer-Verlag, Berlin.

BLACK, F. L., 1975 Infectious diseases in primitive societies. Science **187:** 515–518.

BODMER, J. G. and W. F. BODMER, 1970 Studies on Africa pygmies. IV. A comparative study of the HL-A polymorphism in the Babinga pygmies and other African and Caucasian populations. Am. J. Hum. Genet. **22:** 396–411.

BODMER, W. and G. THOMSON, 1977 Population genetics and evolution of the HLA system. pp. 280–295. In: *HLA and Disease.* Edited by J. DAUSSET and A. SVEJGAARD. Munksgaard, Copenhagen.

CEPPELLINI, R., E. S. CURTONI, P. C. MATTIUZ, V. MIGGIANO, G. SCUDDELER and A. SERRA, 1967 Genetics of leukocyte antigens. A family study of segregation and linkage. In: *Histocompatibility Testing 1967*, Edited by E. S. CURTONI, P. L. MATTIUZ and R. M. TOSI. Munksgaard, Copenhagen.

DAUSSET, J. and P. J. COLOMBANI (Editors), 1973 *Histocompatibility Testing 1972.* Munksgaard, Copenhagen.

FELDMAN, M. W. and F. B. CHRISTIANSEN, 1975 The effect of population subdivision on two loci without selection. Genet. Res. **24:** 151–162.

HANSEN, H. E., S. O. LARSEN, L. P. RYDER and L. S. NIELSON, 1979 HLA-A, B haplotype frequencies in 5,202 unrelated Danes by a maximum likelihood method of gene counting. Tissue Antigens **13:** 143–153.

HEDRICK, P. W. and G. THOMSON, 1983 Evidence for balancing selection at HLA. Genetics **104:** 449–456.

HEDRICK, P. W. and G. THOMSON, 1986 A two-locus neutrality test: applications to humans, *E. coli* and lodgepole pine. Genetics **112:** 135–156.

HEDRICK, P. W., G. THOMSON and W. KLITZ, 1986 Evolutionary genetics; HLA as an examplary system. pp. 583–606. In: *Evo-*

lutionary Processes and Theory, Edited by S. KARLIN and E. NEVO. Academic Press, New York.

HOOD, L., M. STEINMETZ and B. MALISSEN, 1983 Genes of the major histocompatibility complex of the mouse. Annu. Rev. Immunol. **1:** 529–568.

HUDSON, R. R., 1983 Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

KLEIN, J., F. FIGUEROA and Z. A. NAGY, 1983 Genetics of the major histocompatibility complex: the final act. Annu. Rev. Immunol. **1:** 119–142.

KLITZ, W., G. THOMSON and M. P. BAUR, 1984 The nature of selection in the HLA region based on population data from the ninth workshop. pp. 330–332. In: *Histocompatibility Testing 1984*, Edited by E. D. ALBERT, M. P. BAUR and W. R. MAYR. Springer-Verlag, Berlin.

KLITZ, W., G. THOMSON and M. P. BAUR, 1986 Contrasting evolutionary histories among tightly linked HLA loci. Am. J. Hum. Genet. **39:** 340–349.

LARSEN, B., 1981 Basic features of the HLA systems. In: *HLA in Endocrine and Metabolic Disorders*, Edited by N. R. FARID. Academic Press, New York.

LEW, A. M., E. P. LILLEHOJ, E. P. COWAN, W. L. MALOY, M. R. JAN SCHRAVENDIJK and J. E. COLIGAN, 1986 Class I genes and molecules: an update. Immunol. Today **57:** 3–18.

LOH, D. Y. and D. BALTIMORE, 1984 Sexual preference of apparent gene conversion events in MHC genes of mice. *Nature* **309:** 639–640.

MENOZZI, P., A. PIAZZA and L. L. CAVALLI-SFORZA, 1979 Synthetic maps of human gene frequencies in Europeans. Science **201:** 786–792.

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

THOMSON, G., W. F. BODMER and J. BODMER, 1976 The HL-A system as a mode for studying the interaction between selection, migration and linkage. In: *Population Genetics and Ecology*, Edited by S. KARLIN and E. NEVO. Academic Press, New York.

THOMSON, G. and W. KLITZ, 1987 Disequilibrium pattern analysis. I. Theory. Genetics **116:** 623–632.

VAN ENDEN, W., R. R. P. DE VRIES and J. J. VANROOD, 1982 HLA and infectious diseases. pp. 37–54. In: *Human Genetics, Part B: Medical Aspects*, Edited by B. BONNE-TAMIR. Alan R. Liss, New York.

WARBURTON, F. E., 1968 Maintenance of histocompatibility polymorphisms. Heredity **23:** 151–152.

YAMAZAKI, K., E. A. BOYSE, V. MIKE, H. T. THALER, B. J. MATHIESON, J. ABBOTT, J. BOYSE, Z. A. ZAYAS and L. THOMAS, 1976 Control of mating preferences in mice by genes of the major histocompatibility complex. J. Exp. Med. **144:** 1324–1335.

Communicating editor: W. J. EWENS