Am J Hum Genet 34:263-277, 1982

Evidence for Recessive and against Dominant Inheritance at the HLA-"Linked" Locus in Coeliac Disease

DAVID A. GREENBERG,[1] SUSAN E. HODGE,[2] AND JEROME I. ROTTER[3]

SUMMARY

It has been proposed that gluten sensitive enteropathy (GSE) results from the interaction of two loci: one locus linked to HLA and associated with dominant inheritance, and the other, a non-HLA-linked GSE-associated B-cell alloantigen, exhibiting recessive inheritance. We have shown in previous analyses that a two-locus, dominant-recessive model is less compatible with the existing population prevalence and observed familial segregation data than is a recessive-recessive two-locus model. Here we present additional analyses of reported population and familial HLA data that support the recessive mode of inheritance for the HLA-linked disease locus. Reported data from HLA typing of affected sib pairs, the association of GSE with *DR3* and *DR7* in different populations, and the proportions of different HLA phenotypes and genotypes were compared with expected data derived by three different methods. The HLA data analyses consistently reject a dominant mode of inheritance for the presumed HLA-linked disease allele but do not reject a recessive model. The affected sib-pair data also support a recessive model. These analyses are consistent with our previous prediction that the HLA-"linked" disease allele in GSE is recessively inherited.

Received March 2, 1981; revised July 31, 1981.

This work was supported in part by grants AM-25834, MH-33617 (to S. E. H.), and AM-17328 from the National Institutes of Health; Clinical Investigator Award AM-00523 from NIAMDD; a March of Dimes Basil O'Connor Award (to J. I. R.); a research grant from the Kroc Foundation; and by contract no. N01-NS-0-2332 from the Department of Health and Human Services.

[1] Wadsworth VA Medical Center, Room 3405, Los Angeles, CA 90073.

[2] Mental Retardation Research Center, UCLA School of Medicine, Los Angeles, CA 90024.

[3] Division of Medical Genetics, Departments of Pediatrics and Medicine, UCLA School of Medicine, Harbor/UCLA Medical Center, 1000 West Carson Street, Torrance, CA 90509.

 0002-9297/82/3402-0010$02.00

INTRODUCTION

Coeliac disease, also known as gluten sensitive enteropathy (GSE), or nontropical sprue, is a disorder characterized by intestinal malabsorption due to the toxic effects of gluten, a wheat protein [1–4]. The earliest evidence that genetic factors are of significance in coeliac disease consisted of isolated reports of multiple cases occurring within families. Subsequently, family studies using clinical criteria to determine which family members were affected clearly demonstrated that GSE occurred more frequently in the relatives of patients than in the general population (reviewed in [4, 5]). In addition, most reported monozygotic twin pairs were concordant for the disorder, emphasizing the importance of genetic factors [2, 5]. Despite these observations, the mode of inheritance remained unclear. At one time or another, both multifactorial inheritance and a single major gene with reduced penetrance have been proposed [2, 6]. The discovery that the disease was associated with the histocompatibility antigens B8 [7, 8] and DR3 [4, 9, 10] represented a major advance. The D (DR) antigen associations are felt to be stronger than the *B* locus ones. Several studies support the hypothesis that two unlinked loci control the inheritance of GSE [2, 11, 12]. Specifically, Peña et al. [13] proposed that the genetic basis of coeliac disease is a function of two genes: *DR3* (or a closely linked gene) and a non-HLA-linked GSE-associated B-cell alloantigen. They proposed at that time that *DR3* was necessary only in a single dose (dominantly inherited) and the GSE-associated B-cell alloantigen was necessary in a double dose (recessively inherited). Peña et al. [13] noted, however, that this dominant-recessive two-locus model was inconsistent with the known population prevalence data on GSE.

In [6] we showed that a two-locus model of inheritance for GSE is consistent with the known population prevalence and population segregation ratio data if recessive inheritance is assumed at both loci. Since that analysis was performed, additional data have become available that allow us to examine in more detail the inheritance at one of the presumed loci, the one in the HLA complex.

As mentioned above, coeliac disease has been shown to be associated with several HLA antigens. The frequency of *HLA-B8* and *DR3* (which are in linkage disequilibrium with each other) is increased in GSE patients in all populations studied [9]. *DR7* is increased in most populations studied [14–19]. The *DR3/7* heterozygote is increased (although not necessarily significantly), even in populations that do not show an increased *DR7* association with the disease [14, 15]. These association data can be used to examine the inheritance of GSE at the HLA-"linked" locus. By assuming either a dominant or recessive mode of inheritance at that locus, the expected frequencies of the different HLA haplotypes can be calculated and compared with the observed. The predictions of the particular model can then be statistically tested. We will show here that the dominant model can be consistently rejected no matter how the data are examined. The recessive model, on the other hand, cannot be rejected with the current data available. The results of this analysis are consistent with our previous prediction [6] that, assuming two loci are responsible for GSE, both the HLA-"linked" and the non-HLA-linked locus should exhibit recessive inheritance.

METHODS

Data from seven studies from different countries (France, Italy, Spain, Germany, Israel, the Netherlands, and Ireland) were analyzed [14–19]. As the reported data were in different forms and varied in the amount of detail given, several different analytic approaches were used to ensure the robustness of our conclusions. Computational details are in the APPENDIX. A summary of the data used in the present analyses is given in table 1.

The main object of the analyses was to compute the expected proportion of some observable quantity (e.g., the proportion of *DR3/7* heterozygotes among coeliac patients) assuming dominant or recessive inheritance at the HLA-"linked" locus and to compare this expected proportion with the observed under both models. The most discriminating data available were the *DR3/7* heterozygote rates among coeliacs since these data were reported in each of the studies. Also, *DR3/7* represents a genotype, whereas the simple associations of *DR3* and *DR7* with the disease do not.

The three parameters underlying the calculations are the *DR3* and *DR7* chromosomal associations (or, more explicitly, the population frequencies of the haplotypes, x and s [see APPENDIX]) with the presumably linked disease locus, and the disease gene frequency (q). The overall frequencies of both *DR3* and *DR7* are similar in most control populations. However, the population frequencies of *DR3* and *DR7*, as reported in the control groups of the different studies, were used in the calculations where appropriate (see APPENDIX).

The assumption of a two-locus model of inheritance for GSE is not important to these calculations. Based on the gene frequencies of the disease gene, the *DR3* allele and the *DR7* allele, we calculate the value of interest (e.g., the frequency of genotypes that can be affected and also type for *DR3/7*). This number is divided by the sum of the frequencies of all genotypes that can be affected. Including a second locus would entail multiplication of both numerator and denominator by the same number. Thus, in a two-locus model, the reduced penetrance of the disease in those who carry an HLA-linked susceptibility allele would be due to a non-HLA-linked second locus (as described in [6]). This calculation does assume, however, that people who carry the HLA-"linked" disease allele without the associated *HLA* allele are at the same risk as those who carry both the *HLA* and disease alleles. For example, someone who is *DSD3* (i.e., disease allele-*DR3* allele haplotype) is presumed to have the same risk as someone who is *DSD2* (disease allele-*DR2* haplotype). (See APPENDIX for a further explanation of the notation.)

The term "reduced penetrance" is usually used when a segregation ratio calculation fails to give .5 or .25 (i.e., simple dominant or recessive inheritance). Under this definition, reduced penetrance can be caused by environmental factors, by multiple loci, or by both.

TABLE 1

SUMMARY OF DATA

Study	*DR3** positive	*DR7** positive	*DR3/7*	*DR3/x* ($x \neq 7$)	*DR7/x* ($x \neq 3$)	*DRx/x* ($x \neq 3$ or 7)	N†
Netherlands [15]	23 (.96)	5 (.21)	5 (.21)	18 (.75)	0 (0)	1 (.04)	24
Spain [18]	20 (.77)	18 (.69)	12 (.46)	8 (.31)	6 (.23)	0 (0)	26
Italy [17]	27 (.60)	30 (.67)	14 (.31)	13 (.29)	16 (.36)	2 (.04)	45
Ireland [14]	51 (.88)	16 (.28)	13 (.22)	...	...	...	58
Israel [19]	17 (.40)	19 (.44)	8 (.19)	...	...	...	43
France [16]	13 (.68)	11 (.58)	6 (.32)	7 (.37)	5 (.26)	1 (.05)	19
Germany [19]	58 (.64)	40 (.44)	16 (.18)	...	...	...	91

NOTE: Nos. in parentheses are proportions.

* No. patients typing positive for this allele.

† No. patients typed in a given study.

The method described here uses only information from the population affected with the disease. It is, therefore, penetrance-independent, granted our assumption that the HLA marker plays no role in the disease etiology. Whatever is outside the HLA region or in the environment that may be conferring susceptibility, every affected person has it. Thus, even if multiple loci, exposure to environmental factors, or genetic heterogeneity were involved in the disease, the method would still be valid, although if many different HLA-linked loci were involved, interpretation would be complicated.

Dominant Model

The dominant model is more complicated than the recessive one since the gene frequency of the disease allele among coeliacs is a function of the disease allele frequency in the general population, whereas under a recessive model, the disease gene frequency among coeliacs is always unity. To explore the data, we used three approaches (see (1) in the APPENDIX).

Maximum-likelihood method. Four of the seven reported studies included enough data to construct mutually exclusive and exhaustive categories of HLA types. These data were then used to estimate the three parameters (disease gene frequency q and the two chromosomal associations x and s) by maximum likelihood. The categories are: (1) the number of individuals typing for *DR3* but not *DR7*; (2) the number typing for *DR7* but not *DR3*; (3) the number of *DR3/7* heterozygotes; and (4) the number typing for neither *DR3* nor *DR7*. The actual equation (A-11) is given in the APPENDIX, (3).

The two chromosomal associations x and s (i.e., the haplotype frequencies of the disease allele with the *DR3* or *DR7* alleles) and the disease gene frequency q were allowed to vary simultaneously, thus ensuring the best possible fit of the parameters. Estimating three parameters from three pieces of data leaves no degrees of freedom to test the observed against the expected. However, even this fitting procedure, which ensures the best possible fit of the theoretical with the observed, leads to a rejection of the dominant model, as shown below. Fixing the gene frequency (for which there are independent data—see below) would have the effect of worsening the derived parameters and making the agreement between the observed and expected even worse for the dominant model.

Distance minimization. As a second approach, we estimated the parameters by calculating the minimum distance [Σ(predicted $-$ observed)^2] between the predicted and observed population disease associations (i.e., v_{DR3} and v_{DR7} in the APPENDIX) and the *DR3/7* heterozygote rates. While this procedure does not lend itself to classical statistical testing, it has an immediate intuitive appeal and makes maximum use of the data by utilizing those data common to all of the published studies.

Exploratory method. The observed population associations reported (i.e., the proportion of disease patients typing for a specific marker, or v_{DR7} and v_{DR3}) were derived from a relatively small sample of coeliacs in each study. Because of the uncertainty due to these small sample sizes, we did not rely on the estimates of the disease associations in the studies. Instead, we considered a range for the observed associations between the disease and the *DR3* and *DR7* alleles. We examined all possible parameter values (x, s, and q) and selected those compatible with this range. From among these compatible sets of parameter values, we then chose that set leading to the minimum difference between the theoretical and observed *DR3/7* heterozygote frequency.

The studies in France, Germany, Italy, and the Netherlands reported disease associations above .5 (observed range: .6 to .95) for the *DR3* association, and above about .2 (observed range: .18 to .5) for the *DR7* association. As discussed above, the difference between the expected and observed *DR3/7* frequencies was minimized within the range allowed by the above constraints. Conceptually, this procedure corresponds to weighting the *DR3/7* frequency more than the simple *DR3* and *DR7* frequencies. The *DR3* allele population frequency in controls was fixed at .1 and the *DR7* control frequency at .15 [9]. A similar procedure was followed for the data in the remaining studies, with the following differences:

Spain, *DR7* allele frequency of .25; Ireland, *DR3* allele frequency of .25; and Israel, *DR3* allele frequency of .07 and a maximum *DR3*-disease association of .6. It must be emphasized that within the limits set (*DR3*-disease association greater than .5, except for the Israel data, and a *DR7* association greater than .2) the associations were allowed to assume any value. We hoped that by this procedure, the broadest possibilities for the dominant model would be explored. We chose the *DR3/7* heterozygote rate as our test parameter since it is the most informative datum common to all the studies. Since the predicted was calculated individually for each study, an exact binomial test was then used to compare the predicted and observed *DR3/7* heterozygote rates. The predicted *DR3/7* rate was assumed, and the probability of observing the number (or more) of *DR3/7* heterozygotes reported in the different studies was then calculated.

Recessive Model (See (2) in the APPENDIX)

(a) A maximum-likelihood calculation similar to the one for the dominant model (as above) was done. The disease gene frequency was fixed at .2 because this was consistent with previously reported immunologic data, our earlier calculations, and the affected sib-pair data (see below). In addition, by fixing one parameter (the disease allele frequency), we had the necessary degrees of freedom for a statistical test. (Varying the disease allele frequency between .1 and .3 did not change the result.)

(b) If we notice that the disease gene frequency among coeliacs is unity, assuming a recessive model, then we can calculate the proportions of the various HLA genotypes among coeliacs directly from the association data (see (2) (b) in the APPENDIX, for computational details). In this way, we can predict proportions of the various genotypes even in those studies in which there were not sufficient data to perform a maximum-likelihood calculation. In addition, other proportions were calculated for which there are as yet few reported data (e.g., *DR3* homozygote rate).

Since only four of the seven studies were used in the maximum-likelihood calculation, a chi-square test with 4 degrees of freedom was used in the cases in which the parameters were estimated by maximum likelihood. Seven degrees of freedom were used when parameters were estimated by minimizing the distance, since all seven reported studies could be used to test the models.

RESULTS

Table 2 shows the results of the maximum-likelihood calculations for the dominant model for each of the four studies in which there were sufficient data to estimate the parameters. (It should be noted that in two studies, one of the cells had a zero count. Those zero cells did not contribute to the likelihood calculation.) A chi-square with 4 degrees of freedom was calculated, testing the observed vs. the expected *DR3/7* heterozygote data (adding together the individual chi-square results), since these were the only observations that dealt directly with a genotype and would therefore give the greatest discrimination. As can be seen, we clearly reject the dominant model. The recessive model was not rejected (table 3).

Table 4 shows the results of minimizing the distance between the theoretical and observed frequencies as a function of the parameters. Again testing the the number of observed vs. expected *DR3/7* heterozygotes (for 7 degrees of freedom) leads to a clear rejection of the dominant model.

Table 5 shows the results of minimizing the predicted and observed *DR3/7* heterozygote rate difference alone with only the constraints on the associations discussed in METHODS (under *Dominant Model, Exploratory Method*). It can be seen

TABLE 2

COMPARISON OF OBSERVED AND MAXIMUM-LIKELIHOOD DERIVED GENOTYPE PROPORTIONS AMONG COELIACS ASSUMING DOMINANT INHERITANCE AT THE HLA-"LINKED" LOCUS

STUDY	% *DR3* POSITIVE PATIENTS		% *DR7* POSITIVE PATIENTS		*DR3/x* ($x \neq 7$)		*DR7/x* ($x \neq 3$)		*DR3/7**		*DRx/x* ($x \neq 3$ or 7)		q†
	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	
Italy	.66	.60	.56	.67	17.1 (.38)	13 (.29)	18.9 (.42)	16 (.36)	6.3 (.14)	14 (.31)	2.7 (.06)	2 (.04)	.27
Spain	.74	.77	.48	.69	13.5 (.52)	8 (.31)	6.8 (.26)	6 (.23)	5.7 (.22)	12 (.46)	1.8 (.07)	0 (0)	.13
France	.58	.68	.47	.58	8.9 (.47)	7 (.37)	6.8 (.36)	5 (.26)	2.1 (.11)	6 (.32)	0 (0)	1 (.05)	.23
Netherlands	.93	.96	.13	.21	19.4 (.81)	18 (.75)	.2 (.01)	0 (0)	2.9 (.12)	5 (.21)	1.2 (.05)	1 (.04)	.12

NOTE: Nos. in parentheses are proportions.
* χ^2 significant at $P < .0001$ for 4 df ($\chi^2 = 29.7$).
† The HLA-"linked" disease gene frequency, q, was allowed to assume any value. Value reported is the maximum-likelihood estimated disease allele frequency.

TABLE 3

COMPARISON OF OBSERVED AND MAXIMUM-LIKELIHOOD DERIVED GENOTYPE PROPORTIONS AMONG COELIACS ASSUMING RECESSIVE INHERITANCE AT THE HLA-"LINKED" LOCUS

STUDY	% *DR3* POSITIVE PATIENTS		% *DR7* POSITIVE PATIENTS		*DR3/x* ($x \neq 7$)		*DR7/x* ($x \neq 3$)		*DR3/7**		*DRx/x* ($x \neq 3$ or 7)	
	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed
Italy	.58	.60	.70	.67	11.8 (.26)	13 (.29)	17.2 (.38)	16 (.36)	14.2 (.32)	14 (.31)	1.8 (.04)	2 (.04)
Spain	.80	.77	.64	.69	7.8 (.30)	8 (.31)	6.4 (.25)	6 (.23)	11.7 (.45)	12 (.46)	.10 (0.0)	0 (0)
France	.70	.68	.58	.58	7.3 (.38)	7 (.37)	5.0 (.26)	5 (.26)	6.0 (.32)	6 (.32)	0.8 (.04)	1 (.05)
Netherlands	.94	.96	.19	.21	18.9 (.79)	18 (.75)	10 (.04)	0 (0)	3.6 (.15)	5 (.21)	0.6 (.02)	1 (.04)

NOTE: $q = .2$; nos. in parentheses are proportions.
* χ^2 not significant (4 df) ($\chi^2 = 0.66$).

TABLE 4

COMPARISON OF OBSERVED AND MINIMIZATION-DERIVED *DR3* ASSOCIATION, *DR7* ASSOCIATION, AND *DR3/7* HETEROZYGOTE FREQUENCIES ASSUMING A DOMINANT MODEL

STUDY	% *DR3* POSITIVE PATIENTS		% *DR7* POSITIVE PATIENTS		*DR3/7* HETEROZYGOTE FREQUENCY*		q†
	Predicted	Observed	Predicted	Observed	Predicted	Observed	
Netherlands	.92	.96	.21	.21	.13 (3.1)	.21 (5)	.10
Spain	.70	.77	.55	.69	.21 (5.5)	.46 (12)	.14
Italy	.64	.60	.51	.67	.15 (6.8)	.31 (14)	.20
Ireland	.89	.88	.28	.28	.21 (12.2)	.22 (13)	.30
Israel	.41	.40	.44	.44	.08 (3.4)	.19 (8)	.21
France	.61	.68	.51	.58	.12 (2.3)	.32 (6)	.12
Germany	.64	.64	.49	.44	.13 (11.8)	.18 (16)	.14

NOTE: In parentheses are nos. individuals.
* $P < .0001$ with 7 df ($\chi^2 = 31.25$).
† Disease allele frequency estimated from the minimization.

that even with those loose constraints, it is not possible, in three out of the seven cases, to get the theoretical *DR3/7* heterozygote frequency within nonrejectable limits. In the other four cases, while it was possible to get acceptable agreement between the theoretical and observed *DR3/7* heterozygote rates, the predicted HLA-disease associations were considerably different from the observed. The differences in predicted and observed *DR3* associations (among those studies in which the *DR3/7* heterozygote rate was not significantly different from the observed) ranged from .13 to .37. The range of differences between expected and observed for the *DR7* association was .12 to .43. Clearly it is virtually impossible to force both the predicted values of the associations and the *DR3/7* heterozygote frequencies to agree with the observed data if a dominant model is assumed.

Table 6 shows the results of computing the *DR3/7* heterozygote rate for all seven reported studies, assuming a recessive model, as described in METHODS [under *Recessive Model*, (b)]. Again, the recessive model cannot be rejected.

DISCUSSION

It is apparent from our theoretical analysis that not only does the dominant model for the HLA-"linked" disease susceptibility gene not fit the known coeliac HLA data, but even allowing for liberal limits on the known data, the dominant model cannot be forced to fit. The recessive model, despite having a greater number of constraints, fits well and correctly predicts the number of *DR3/7* heterozygotes.

Other data that would discriminate between the two models would be *DR3/3* and *DR7/7* homozygote frequencies. Table 7 shows predicted values of the *DR3/3*

TABLE 5

RESULTS OF MINIMIZING ONLY THE OBSERVED AND CALCULATED *DR3/7* FREQUENCY DIFFERENCES ASSUMING A DOMINANT MODEL WITH ALLOWED RANGES FOR THE ASSOCIATIONS

STUDY	% *DR3* POSITIVE PATIENTS		% *DR7* POSITIVE PATIENTS		*DR3/7* FREQUENCY		PREDICTED DISEASE GENE FREQUENCY	BINOMIAL PROBABILITY (SEE TEXT)
	Predicted	Observed	Predicted	Observed	Predicted	Observed		
Netherlands	0.58	0.95	0.56	0.21	0.14	0.21	0.23	0.22
Spain	0.10	0.77	0.26	0.69	0.26	0.46	0.09	0.02
Italy	0.87	0.66	0.28	0.48	0.15	0.31	0.13	0.01
Ireland	0.75	0.88	0.43	0.27	0.22	0.22	0.29	0.50
Israel	0.59	0.39	0.52	0.44	0.11	0.19	0.13	0.09
France	0.61	0.64	0.51	0.55	0.12	0.32	0.21	0.02
Germany	0.87	0.64	0.27	0.48	0.14	0.18	0.13	0.38

TABLE 6

COMPARISON OF OBSERVED AND ASSOCIATION-DATA-DERIVED *DR3/7* HETEROZYGOTE RATES ASSUMING A RECESSIVE MODEL

	Predicted	Observed
Netherlands	4.1 (.17)	5 (.21)
Spain	12.0 (.46)	12 (.46)
Italy	9.0 (.20)	14 (.31)
Ireland	11.0 (.19)	13 (.22)
Israel	5.2 (.12)	8 (.19)
France	4.8 (.25)	6 (.32)
Germany	20.0 (.22)	16 (.18)

NOTE: χ^2 not significant with 7 df (χ^2 = 7.3).

homozygote frequency and the *DR3/3* plus *DR3/blank* frequency (i.e., those individuals typing for *DR3* alone. *DR3/3* is indistinguishable from *DR3/blank* without family studies) [see equation (A-4) in the APPENDIX]. As can be seen from the table, the population data for those typing only for *DR3* but which consist of *DR3/3* and *DR3/blank* together cannot readily distinguish between the models. One can predict, however, that a much greater proportion of those individuals typing for *DR3* alone will be *DR3* homozygous under the recessive model than under the dominant model. As those data become available from family studies, they can be used as a further means of testing the inheritance at the HLA-"linked" locus.

For the recessive model, deriving the *DR3/7* rate directly from the association data provides a way to check the consistency of the association and *DR3/7* data. The nature of the recessive model is such that we know that every affected person must be homozygous for the disease allele. It is this fact that allows us to calculate the *3/7* rate from the association data, which cannot be done if a dominant model is assumed. The fact that the observed *DR3/7* rate can be correctly derived from

TABLE 7

COMPARISON OF PREDICTED FREQUENCIES OF GENOTYPES ONLY FOR *DR3* ASSUMING EITHER A DOMINANT OR RECESSIVE MODEL

	PREDICTED *DR3* HOMOZYGOTES		PREDICTED *DR3/3* AND *DR3/blank*	
STUDY	Dominant	Recessive	Dominant	Recessive
Netherlands	.05	.60	.24	.66
Spain	.06	.27	.22	.31
Italy	.05	.13	.17	.16
Ireland	.20	.19	.40	.24
Israel	.02	.11	.02	.13
France	.04	.15	.17	.18
Germany	.05	.17	.18	.20

NOTE: *Blank* allele frequency = .2 (after Bodmer and Bodmer [20]).

the *DR3* and *DR7* population associations with the disease is compelling evidence in favor of the recessive model.

Additional support for a recessive mode of inheritance at the HLA-"linked" locus is provided by studies on affected sib pairs. Table 8 shows a summary of existing data on the haploidentity of affected sib pairs as collected in [19]. The dominant model would predict that a maximum of 50% of the affected sib pairs would share both HLA haplotypes, assuming tight linkage between HLA and the disease locus. Using an exact binomial test, there is a probability of only .04 of observing 27 or more HLA identical affected sibs out of 42 sib pairs given that the expected frequency is .5, as the dominant model would require. Also, an HLA identical affected sib rate of .5 assumes a low disease gene frequency in the population ($q < .05$) [21]. A low disease gene frequency in the population is not only contrary to the analysis of the population data [6], and the original immunologic findings of Peña et al. [13], but is also lower than any predicted disease gene frequency in the above analysis. The lowest predicted disease gene frequency in any of the analyses, assuming a dominant model, was .09 (table 5). The affected sib-pair data do, however, fit a recessive model with a disease gene frequency of about .2, which agrees with the analysis of the population data [6], the original study of Peña et al., and the current analysis. The probability of observing 27 or more HLA identical sib pairs out of 42 is about .61, assuming an HLA identical affected sib-pair frequency of .65, as the recessive model would predict for a disease gene frequency of .2 [21].

Neither the affected sib-pair data nor either minimization analysis (tables 4 and 5) will readily distinguish between the dominant and recessive model if the Irish data alone are examined. Population prevalence data indicate that the prevalence of GSE in Ireland is 1:300 [22]. The population frequency of *DR3* is also higher than reported in studies of other populations [14]. The data presented in [6] show that one could expect such a high prevalence if a two-locus model was assumed in which one of the loci exhibited dominant and one recessive inheritance, as originally proposed by Peña et al. [13]. The population segregation ratio data, how-

TABLE 8

DISTRIBUTION OF HLA IDENTITY IN AFFECTED SIB PAIRS AND PREDICTED DISEASE GENE FREQUENCY

	Identical	Share one haplotype	Nonidentical	Predicted disease gene frequency (assuming recessive model)*
Austria/Germany	5 (.71)	2 (.29)	0	.2
France	6 (.86)	1 (.16)	0	.08
Ireland	10 (.56)	6 (.34)	2 (.1)	.3
Switzerland	1	1	...	...
Holland	2	3	...	...
England	3	0	...	...
Total	27 (.64)	13 (.31)	2 (.05)	.22
Total excluding Irish data	17 (.71)	7 (.29)	0	.2

NOTE: Nos. in parentheses are percentages; data from [19].

* Gene frequencies calculated according to [21].

ever, would still force rejection of a dominant-recessive model. If one examines the population prevalence graphs in [23] for a double recessive model, one sees that, assuming a non-HLA-linked disease gene frequency of .1 (as was proposed by Peña et al. [13]), in order for the prevalence to be about .003, the disease allele frequency at the HLA-"linked" locus must be about .3, which also agrees with the affected sib-pair data from Ireland (from table 8). Both affected sibs would be expected to be HLA identical 59% of the time with a disease gene frequency of .3 [21] assuming a recessive model. If we exclude the Irish data (table 8), then the expected frequency of HLA identical affected sib pairs is .71, which would make a dominant model even less likely, and a recessive model with a disease gene frequency of .2 even more likely.

Coeliac disease is also associated with the HLA antigen B8. Ideally, this association should be more discriminating than the DR3 association, since only about 2% of the *B* alleles exhibit no typing response, as opposed to 20% for the *DR* locus. Therefore, someone who types for *B8* alone would most likely be a *B8* homozygote, and family studies would not be as essential as with the *DR* locus. However, since the *B8* association with GSE seems to be less strong, the *B8* homozygote data will usually not discriminate between the models (see table 9).

We conclude with an important caveat. The state of the art with regard to HLA-disease associations is such that caution should be exercised before drawing any conclusions about the inheritance of HLA-associated diseases. As in this study, the usual assumption is that the HLA-disease association is due to a "disease gene" in linkage disequilibrium with an HLA marker. For the great majority of HLA-associated diseases, however, linkage has not been demonstrated, and, in fact, cannot be demonstrated by the usual techniques (e.g., lod scores) in the presence of the association [26]. The very fact of so many disease-HLA associations should cause some hesitation before making some of the assumptions usually made in HLA-associated disease studies (i.e., the assumption that a "disease gene" is in linkage disequilibrium with a marker; the assumption that the marker is not involved in the etiology of the disease; and the assumption that only one locus is involved). If the *DR3/3* or *7/7* homozygote rate in GSE were found to be signifi-

TABLE 9

PREDICTED AND OBSERVED *B8/8* HOMOZYGOTE FREQUENCY

STUDY	PREDICTED		OBSERVED
	Dominant	Recessive	
Netherlands [15]	.10	.33	.33 (8/24)
Spain [18]	.02	.05	.08 (2/26)
Italy [17]	.04	.05	...
France [16]	.09	.22	...
Ireland [14]	.37	.32	...
Israel [24]	.02	.02	...
Germany [19]	.08	.09	...
England [25]	.16	.20	.22 (10/45)

NOTE: Nos. in parentheses are nos. observed *B8/8* homozygotes over the sample size.

cantly higher or lower than the recessive model would predict, the assumption of simple Mendelian inheritance at an HLA-"linked" disease locus might have to be discarded. Such a finding, however, would not invalidate the two-locus, double recessive analysis for inheritance of GSE proposed by Greenberg and Rotter [6]. In that work, no assumptions were made about the locations of the disease loci, except that they were in linkage equilibrium with each other, and that model explained the population data reasonably well.

The above analysis indicates that, no matter how one examines the HLA data for coeliac disease, there is little support for a dominant mode of inheritance at the HLA-"linked" locus. Furthermore, the recessive model is not excludable with the current data and predicts with reasonable accuracy the *DR3/7* heterozygote frequency, the affected sib-pair data, and, assuming the reduced penetrance is due to a second non-HLA-linked recessive locus, the previously reported population segregation ratio and population prevalence data [6].

ACKNOWLEDGMENT

We thank Dr. Ken Lange for helpful discussions.

APPENDIX

Designation of Alleles

DS = disease allele; *D3* = *DR3* allele; *D7* = *DR7* allele; *BL* = *blank* at *DR* locus (i.e., no typing response); -- = any other allele at either the disease or *DR* locus (exact meaning changes with context); q = the disease gene frequency.

(1) Dominant Inheritance

(a) *DR3* calculations. The following genotypes type for *DR3* and are "affected" at the HLA-"linked" locus under a dominant model. (Here, the designation "other alleles" at the *DR* locus [e.g., *DS*--] includes *DR7* and *DRBL* but not *DR3*.)

DSD3	*DSD3*	*DS--*	*DSD3*	*DSD3*
----	*--D3*	*--D3*	*DS--*	*DSD3*

The haplotype frequencies are defined as follows: *DSD3* = x; *DS--* = y_1; *--D3* = z; ---- = $1 - x - y_1 - z = w_1$; $x + z$ = the frequency of *DR3* in the general population; $x + y_1 = q$ = disease allele frequency. The fraction of coeliacs typing for *DR3*:

$$v_{DR3} = \frac{x^2 + 2xy_1 + 2zy_1 + 2xz + 2xw_1}{x^2 + 2xy_1 + 2xw_1 + 2y_1w_1 + 2zy_1 + 2xz + y_1^2} . \qquad \text{(A-1)}$$

Note that the denominator can be more simply expressed as: $a = q^2 + 2q(1 - q)$, or simply as the Hardy-Weinberg formula for the frequency of a dominant disease.

(b) To calculate the proportion of coeliacs typing for *DR7*, we define the following frequencies (note that now the designation "other alleles" includes the *DR3* and *DRBL* alleles, but not *DR7*): *DSD7* = s; $q - s = y_2$ = *DS--*; *--D7* = t; $s + t$ = the frequency of *DR7* in the general population; ---- = $w_2 = 1 - s - t - y_2$. The fraction of coeliacs typing for *DR7*:

$$v_{DR7} = \frac{s^2 + 2sy_2 + 2st + 2ty_2 + 2sw_2}{a} . \qquad \text{(A-2)}$$

(c) The fraction of coeliacs typing for *DR3* and *DR7*:

$$v_{DR3/7} = \frac{2sx + 2sz + 2xt}{a} . \tag{A-3}$$

(d) The fraction typing for *DR3/3* plus *DR3/blank*: v = *--BL*; u = *DSBL*; *--BL* + *DSBL* = .2 (after [10]).

$$v_{DR3/3} + v_{DR3/BL} = \frac{x^2 + 2xu + 2uz + 2xv + 2xz}{a} . \tag{A-4}$$

(2) Recessive Model

(a) The possible affected genotypes that type for *DR3* are:

DSD3	*DSD3*
DSD3	*DS--*

The fraction of all coeliacs who type for *DR3* is:

$$v_{DR3} = \frac{x^2 + 2xy_1}{x^2 + 2xy_1 + y_1^2} ,$$

or equivalently

$$v_{DR3} = \frac{x^2 + 2xy_1}{q^2} , \tag{A-5}$$

where q^2 = Hardy-Weinberg frequency of a recessive trait. Similarly

$$v_{DR7} = \frac{s^2 + 2sy^2}{q^2} , \tag{A-6}$$

$$v_{DR3/7} = \frac{2sx}{q^2} . \tag{A-7}$$

(b) Calculation of the *DR3/7* rate from the association data. As before, the population frequencies of the haplotypes are: *DSD3* = x; *DSD7* = s; *DS--* = y. The possible affected genotypes are:

DSD3	*DSD3*	*DSD3*	*DSD7*	*DS--*	*DSD7*
DSD3	*DSD7*	*DS--*	*DS--*	*DS--*	*DSD7*

and $b = x^2 + 2xs + 2xy + 2sy + y^2 + s^2$. (Note that here, *DS--* includes all *DR* alleles except *DR3* and *DR7*).

$$v_{DR3} = \frac{x^2 + 2xs + 2xy}{b} , \tag{A-8}$$

$$v_{DR7} = \frac{s^2 + 2xs + 2sy}{b} . \tag{A-9}$$

We also know:

$$q = x + y + s \ . \tag{A-10}$$

We now make the restriction that we look only at the population of coeliacs. Assuming a recessive model $q_{(\text{coeliacs})} = 1$, since all coeliacs have a double dose of the disease gene by definition. Thus, $b = 1$. We therefore have three equations and three unknowns and can solve explicitly for x, y, and s. Note that we cannot estimate the disease gene frequency in the general population from these data, only the expected proportions of a given genotype assuming a recessive model. Note also that by making the restriction that $q = 1$, the meaning of s, y, and x change so that they now refer only to the haplotype frequency in the population of coeliacs and no longer refer to the frequencies in the general population.

(3) Maximization of the Likelihood

The proportions used in the likelihood equation can all be derived from the two associations and the *DR3/7* rate. Let x denote all *DR* alleles that are neither 3 nor 7, then define

$$\begin{aligned} A &= \text{The frequency of } DR3/3 + DR3/x = v_{DR3} - v_{DR3/7} \\ B &= \text{The frequency of } DR7/7 + DR7/x = v_{DR7} - v_{DR3/7} \\ C &= \text{The frequency of } DR3/7 = v_{DR3/7} \\ D &= \text{The frequency of } DRx/x = 1 - A - B - C \ . \end{aligned}$$

Let uppercase letters denote probabilities and lowercase letters the observed number in each category. The likelihood equation is

$$L = A^a \times B^b \times C^c \times D^d \ . \tag{A-11}$$

For all maximizations or minimizations the parameters (i.e., q, x, and s in the dominant model and x and s in the recessive) were systematically varied in steps of .01, and the entire allowed surface was searched. The highest (or lowest) value found as a result of the search was then printed. Searching the surface in steps of .001 resulted in little change in the value. The maximum value of the frequency of the disease-marker haplotype (e.g., *DSD3*) was taken as the frequency of the marker in the control population of each study. Thus, if the frequency of *DR3* in the control population for a given study was, say, .1, it was assumed that the frequency of the *DSD3* haplotype in that general population could be no greater than .1.

REFERENCES

1. Strober W, Falchuk ZM, Rogentine GN, Nelson DL, Klaeveman HL: The pathogenesis of gluten-sensitive enteropathy. *Ann Intern Med* 83:242–256, 1975
2. Strober W: Gluten-sensitive enteropathy. *Clin Gastroenterol* 5:429–450, 1976
3. McConnell RB, ed.: *The Genetics of Coeliac Disease*. Lancaster, England, M.T.P. Press, 1981, pp 251–262
4. Strober W: Genetic factors in gluten-sensitive enteropathy, in *Genetics and Heterogeneity of Common Gastrointestinal Disorders*, edited by Rotter JI, Samloff IM, Rimoin DL, New York, Academic Press, 1980, pp 243–259
5. Ellis A: Coeliac disease: family and twin studies and disease associations, in ibid.
6. Greenberg DA, Rotter JI: Two-locus models for gluten sensitive enteropathy: population genetic considerations. *Am J Med Genet* 8:205–214, 1981
7. Falchuk ZM, Rogentine GN, Strober W: Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *J Clin Invest* 51:1602–1605, 1972

8. STOKES PL, ASQUITH P, HOLMES GKT, MACKINTOSH P, COOKE WT: Histocompatibility antigens associated with adult coeliac disease. *Lancet* ii:162–164, 1972
9. STROBER W: Abnormalities of the HLA system and gastrointestinal disease, in *HLA and Disease*, edited by DAUSSET J, SVEJGAARD A, Copenhagen, Munksgaard, 1977, pp 168–185
10. MACKINTOSH P, ASQUITH P: HLA and coeliac disease. *Br Med Bull* 34:291–294, 1978
11. FALCHUK ZM, KATZ AJ, SHWACHMAN H, ROGENTINE GN, STROBER W: Gluten-sensitive enteropathy: genetic analysis and organ culture study in 35 families. *Scand J Gastroenterol* 13:839–843, 1978
12. MANN DL, KATZ SI, NELSON DL, ABELSON LD, STROBER W: Specific B-cell antigens associated with gluten-sensitive enteropathy and dermatitis herpetiformis. *Lancet* ii: 110–111, 1976
13. PEÑA AS, MANN DL, HAGUE WE, ET AL.: Genetic basis of gluten sensitive enteropathy. *Gastroenterology* 75:230–235, 1978
14. MCKENNA R, STEVENS FM, BOURKE M, MCNICHOLL B, ALBERT ED, MCCARTHY CF: B-cell alloantigens associated with coeliac disease in the west of Ireland, in *The Genetics of Coeliac Disease*, edited by MCCONNELL RB, Lancaster, England, M.T.P. Press, 1981, pp 153–160
15. PEÑA AS, BIEMOND I, ROSEKRANS PCM, VAN LEEUWEN A, SCHREUDER I, VAN ROOD JJ: DR locus-controlled B-cell alloantigens in coeliac disease in the Netherlands, in ibid., pp 161–170
16. BETUEL H, GEBUHRER L, DESCOS L, PERCEBOIS H, MINARE Y, BERTRAND J: Adult coeliac disease associated with HLA-DRw3 and DRw7. *Tissue Antigens* 15:231–238, 1980
17. DEMARCHI M, BORELLI I, OLIVETTI E, ET AL.: Two HLA-D and DR alleles are associated with coeliac disease. *Tissue Antigens* 14:309–316, 1979
18. POLANCO I, BIEMOND I, VAN LEEUWEN A, ET AL.: Gluten sensitive enteropathy in Spain: genetic and environmental factors, in *The Genetics of Coeliac Disease*, edited by MCCONNELL RB, Lancaster, England, M.T.P. Press, 1981, pp 211–231
19. SCHOLZ S, ROSSIPAL E, BRAUTBAR C, ET AL.: HLA-DR antigens in coeliac disease, in ibid., pp 143–149
20. BODMER WS, BODMER JG: Evolution and function of the HLA system. *Br Med Bull* 34:309–316, 1978
21. THOMSON G, BODMER W: The genetic analysis of HLA and disease associations, in *HLA and Disease*, edited by DAUSSET J, SVEJGAARD A, Copenhagen, Munksgaard, 1977, pp 84–93
22. MYLOTTE M, EGAN-MITCHELL B, MCCARTHY CF, MCNICHOLL B: Incidence of coeliac disease in the west of Ireland. *Br Med J* i:703–705, 1973
23. GREENBERG DA: A simple method for testing two-locus models of inheritance. *Am J Hum Genet* 33:519–530, 1981
24. GAZIT E, AVIGAD S, ZFAT Z, EFTER T, MIZRACHI Y, ROTEM Y: The association of HLA-B8 and childhood coeliac disease in an Israeli population. *Isr J Med Sci* 13:400–407, 1977
25. ROBINSON BN, ROBERTS DF, MATHER BA, NELSON R, ROWLATT AS: Coeliac disease and HLA: a family study. *J Immunogenet* 7:381–391, 1980
26. HODGE SE, SPENCE MA: Some epistatic two-locus models of disease. II. The confounding of linkage and association. *Am J Hum Genet* 33:396–406, 1981