

# Relative power of linkage and transmission disequilibrium test strategies to detect non-HLA linked coeliac disease susceptibility genes

S Bevan, S Popat, R S Houlston

## Abstract

**Background**—Susceptibility to coeliac disease is genetically determined by possession of specific HLA DQ alleles, acting in concert with one or more non-HLA linked genes. The pattern of familial risk is most parsimonious with a multiplicative model for the interaction between these two classes of genes. Haplotype sharing probabilities across the HLA region in affected sibling pairs suggest that genes within the MHC complex contribute no more than 40% of the sibling familial risk of coeliac disease, making the non-HLA linked gene (or genes) the stronger determinant of coeliac disease susceptibility. Attempts to localise these non-HLA linked genes have been carried out using both linkage and association tests.

**Aims**—To review the evidence for the involvement of non-HLA linked genes in coeliac disease, and to compare the relative merits of linkage and transmission disequilibrium tests (TDT) to detect the non-HLA linked gene (or genes) contributing to the development of coeliac disease.

**Methods**—Under a range of genetic models the number of affected sibling pairs needed to detect linkage was compared with the number of families required to show a relation between marker and disease, adopting the TDT strategy.

**Results and conclusions**—Power calculations show that, if there is a single major non-HLA linked susceptibility locus, a non-parametric linkage approach may well prove effective. However, if there are a number of non-HLA susceptibility genes, each with small effect, the sample size necessary for linkage studies will be prohibitive and a systematic search for allelic association should be a more effective strategy.

(Gut 1999;45:668-671)

Keywords: coeliac disease; non-HLA linked genes; linkage; TDT test

Gluten sensitivity, or coeliac disease, is due to T cell sensitisation and results in a range of mucosal abnormalities that may lead to malabsorption.<sup>1</sup> Studies of small bowel biopsy specimens from first degree relatives of coeliac patients provide compelling evidence that genetic factors influence susceptibility to the

disease.<sup>2</sup> Reported recurrence risks vary between 5% and 20%, with most being around 10%.<sup>2</sup> This variation probably results not only from genetic and environmental heterogeneity among populations, but also from differing diagnostic criteria between studies. Further support for an inherited predisposition to develop coeliac disease comes from twin studies. The concordance rate of coeliac disease in monozygotic twins is around 70%.<sup>2</sup> Incomplete concordance between monozygotic twin pairs suggests that additional factors might be involved, although not all the twin pairs studied had proven monozygosity, and some of the twin pairs had insufficient long term follow up to preclude the future development of coeliac disease.

Coeliac disease shows a strong HLA association. The DQ2 molecule encoded by the alleles DQB1\*0201 and DQA1\*0501 is possessed by 95% of coeliac patients compared with 20-30% of controls.<sup>3</sup> This DQ(A1\*0501,B1\*0201) heterodimer can be encoded in *cis* or in *trans* configuration. The possibility that the DQ2 molecule conferring susceptibility to coeliac disease might be unique has been excluded by showing that the DQB1\*0201 and DQA1\*0501 alleles do not show any disease specific sequences in patients.<sup>3</sup> The difference in concordance rates between monozygotic twins and HLA identical siblings (70% versus 30%) clearly implicates the involvement of non-HLA genes in the genetic predisposition to coeliac disease.<sup>2,3</sup>

The precise mode of inheritance of coeliac disease is unknown. Using family data, Pena and colleagues<sup>4</sup> proposed that a prerequisite for developing coeliac disease was homozygosity at an HLA unlinked locus. This proposal has been supported by some, but not all studies.<sup>5,6</sup> It is clear, however, that the stronger determinant of disease susceptibility is the non-HLA linked component. Detecting the non-HLA linked gene or genes can be undertaken by a number of different strategies. The purpose of this article is to compare the relative power of linkage versus the transmission disequilibrium test (TDT) for identifying a non-HLA linked coeliac predisposition gene.

## Linkage using affected sibling pairs

Localisation of disease genes using the classical linkage approach is based on the demonstration of cosegregation of markers in families

**Abbreviations used in this paper:** ASP, affected sibling pair; IBD, identical by descent; TDT, transmission disequilibrium test.

Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK  
S Bevan  
S Popat  
R S Houlston

Correspondence to:  
Dr Houlston

Accepted for publication  
29 April 1999

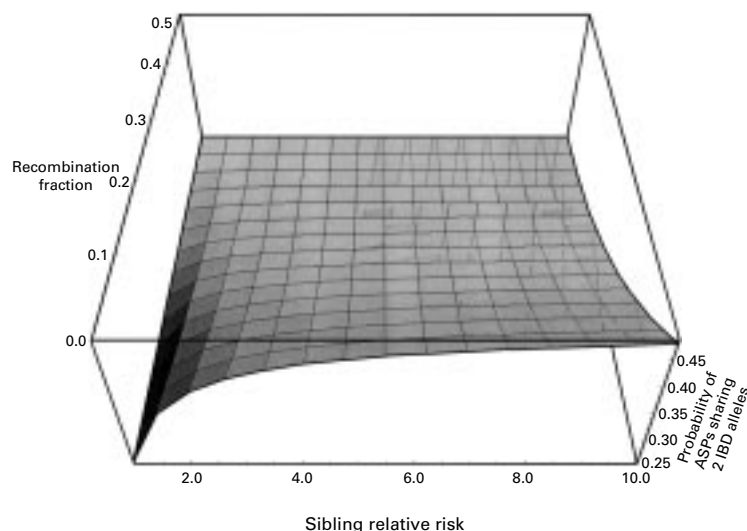


Figure 1 Effect of sibling relative risk and recombination on the probability of affected sibling pairs sharing two alleles identical by descent (IBD). Values based on the assumption that the sibling and parent-offspring relative risks are equal.

with the disease. The essential prerequisite for utilising this approach to detect disease genes is that the model of inheritance of the disease can be specified with some degree of certainty. Unfortunately many diseases like coeliac disease display a complex pattern of inheritance indicative of the interaction of a number of distinct susceptibility genes. To circumvent the requirement for a specified model of inheritance a number of non-parametric methods have been developed. These are based on determining which regions of the genome are identical by descent (IBD) in affected relatives. The most common paradigm of this approach utilises affected sibling pairs (ASPs) and is based on comparing the IBD allele sharing at a given marker with the expectation under the null hypothesis that no deleterious gene is present. A marker close to a susceptibility gene would be expected to display excess IBD sharing, with ASPs sharing both alleles IBD more frequently than sharing neither allele. The allele sharing probabilities of ASPs depend on the contribution any gene makes to the genetic variation of the trait.

This is generally measured in terms of the relative risk ( $\lambda$ ) which is the risk to relatives of affected probands compared with the population risk. The probability that ASPs share 0, 1, or 2 alleles IBD ( $Z_0$ ,  $Z_1$ , and  $Z_2$  respectively) is given by<sup>7</sup>:  $Z_0 = 1/4\lambda_s$ ,  $Z_1 = \lambda_o/2\lambda_s$ ,  $Z_2 = \lambda_{mz}/4\lambda_s$

where  $\lambda_s$ ,  $\lambda_o$ , and  $\lambda_{mz}$  are the sibling, parent-offspring, and monozygotic relative risks respectively. If a marker is unlinked to the disease, these allele sharing probabilities will be equal to those expected under the null hypothesis (0.25, 0.5, and 0.25). Figure 1 shows the effect of sibling relative risk on the probability that ASPs will share two IBD alleles. As the sibling relative risk increases, the probability that ASPs will share two alleles at the disease gene locus increases. These formulae hold true irrespective of the mode of inheritance at the disease locus, the number of alleles and their frequencies, penetrance, and population prevalence.<sup>7</sup> The only requirement is that

recombination between the marker and the disease gene is negligible. Recombination between the marker and disease gene leads to a reduction in allele sharing between ASPs. Therefore, for a given sibling relative risk an increase in recombination will lead to a reduced deviation of the marker sharing from its null expectation. This has the consequence that larger numbers of ASPs are required to detect linkage (fig 1).

**Transmission disequilibrium test**

An alternative strategy for identifying the location of a gene conferring susceptibility to coeliac disease is by allelic association based on showing over representation of a specific allele in affected individuals. The simplest means of undertaking this is to compare the frequency in affected individuals (cases) with the frequency in the general population (controls). Marker alleles that are positively associated with the disease are analogous to risk factors in epidemiology. A major problem inherent in this approach is that spurious associations can arise as a result of population stratification. One method of overcoming the problem of hidden population stratification is to use family based controls. The most common approach introduced by Spielman and Ewens<sup>8</sup> is the TDT, which is based on the McNemar test for matched pair data. It considers only parents whose transmitted and non-transmitted alleles are different (heterozygous parents) and assesses the evidence for preferential transmission of one allele over the other (table 1). One additional attractive feature of the TDT is that it is also a test of linkage and not merely of linkage disequilibrium, as only linkage disequilibrium can distort the distribution of marker genotypes among parents of affected individuals. In addition to the analysis of parent-child trios, the TDT approach can be extended to family data. Provided that the sampling scheme is confined to nuclear families such as parent affected sibpairs (TDT-ASP) the test will still be valid for both association and linkage.

**Statistical analysis**

To compare the sample sizes required to detect a predisposition gene using linkage and association methods requires consideration of the level of acceptable type I (false positive) and type II (false negative) error rates (denoted by  $\alpha$  and  $\beta$  respectively). The probability (power) that a test will correctly detect a deleterious locus is given by  $1-\beta$ . Most linkage searches are based on approximately 250–400 markers at a density of 10–20 cM throughout the genome. Lander and Kruglak have proposed that a Lod score of 3.4 be used to define the appropriate level of significance (equating to  $\alpha$  of  $2.2 \times 10^{-5}$ ) for genome wide searches using ASPs.<sup>9</sup> Studies using the TDT method can be based on a candidate gene approach or can be on a genome wide basis. Risch and Merikangas<sup>10</sup> proposed that  $5 \times 10^{-8}$  be adopted as a critical value for  $\alpha$  in these genome wide tests. This is based on a Bonferroni correction to account for testing of five biallelic markers in 100 000 genes.<sup>10 11</sup>

Table 1 Contingency table of transmitted and non-transmitted parental alleles

		Non-transmitted allele	
		M	N
Transmitted allele	M	a	b
	N	c	d

TDT = (b-c)<sup>2</sup>/(b+c).

Table 2 Sample sizes necessary to gain 80% power in affected sibling pair (ASP) linkage analysis and transmission disequilibrium tests (TDT)

Genotypic risk	Gene frequency	Sibling relative risk	ASP	TDT	TDT-ASP
16.0	0.01	3.4	59	34	<10
	0.1	6.9	31	14	<10
	0.2	4.5	43	18	<10
	0.5	1.9	161	49	45
8.0	0.01	1.5	417	119	<10
	0.1	3.1	66	29	<10
	0.2	2.8	76	30	12
	0.5	1.7	235	62	51
4.0	0.01	1.1	7 912	523	59
	0.1	1.5	342	86	19
	0.2	1.6	264	72	25
	0.5	1.4	549	106	72
2.0	0.01	1.01	550 000	4 154	1039
	0.1	1.1	9 995	533	164
	0.2	1.1	4630	360	131
	0.5	1.1	4630	348	192
1.5	0.01	1.002	86 000 000	16 369	5484
	0.1	1.02	130 000	1 952	720
	0.2	1.03	48532	1 197	486
	0.5	1.04	33 374	971	503

Data derived using the formulae given in Camp.<sup>11</sup> Values are based on a multiplicatively acting non-HLA linked disease gene.

The sample size required to detect any disease gene depends on its frequency and associated genotypic risk. Both these parameters are unknown for the non-HLA linked component of coeliac disease. However, the sibling relative risk can be derived from estimated overall and HLA linked sibling relative risks. Estimates of coeliac disease prevalence based on disease presenting symptomatically vary considerably in different areas of Europe and the USA.<sup>12-16</sup> However, population studies based on antibody screening have all shown that the prevalence of coeliac disease is approximately 1 in 200.<sup>17-20</sup> Based on this population prevalence rate and a 10% recurrence risk, the sibling relative risk of coeliac disease will be 20. The sibling relative risk of coeliac disease contributed by the HLA linked genes is approximately 3.<sup>21-22</sup> Hence the non-HLA linked gene or genes will be the stronger determinant of coeliac disease compared with those linked to HLA. The actual relative risk will depend on the mode of interaction between the HLA linked and unlinked genes. These genes could interact either additively (the penetrance of the disease is represented by the sum of the penetrances contributed by two or more loci) or multiplicatively (the penetrance of the disease is the product of the penetrances contributed by two or more loci). The familial risks seen in siblings and monozygotic twins are most parsimonious with a multiplicative model. Based on this model the relative risk associated with the non-HLA linked genes will be approximately 6.0, accounting for approximately 60% of the familial risk of coeliac disease.

## Results

Table 2 shows the sample sizes necessary to detect a multiplicatively acting predisposition gene with 80% power using ASP linkage analysis and TDT approaches. A comparison of the relative power was made using the formulae derived by Camp.<sup>11</sup> Power calculations for ASPs assume a highly polymorphic marker tightly linked to the disease locus. Estimates are based on gene frequencies of 0.01, 0.1, 0.2, and

0.5, and risk ratios of 1.5 to 16. These provide sibling relative risks compatible with the range of risk associated with an HLA unlinked susceptibility gene predisposing to coeliac disease.

The number of TDT trios is smaller in all cases than that necessary for linkage analysis using ASPs. However table 2 shows that relatively small numbers of ASPs are required to detect a predisposition gene when it is less common and the sibling relative risk is high. Under this scenario, linkage analysis based on ASPs will offer the most efficient strategy for detecting a disease gene, particularly given the limited amount of genotyping required compared with adopting a TDT approach. As the sibling relative risk becomes smaller the deviation of allelic sharing in ASPs from the null expectation becomes smaller (fig 1). As a consequence of this the numbers of siblings required to detect linkage increases dramatically. Once the disease allele is relatively common and its risk ratio is small the number of ASPs required will be prohibitively large. Table 2 also shows the sample sizes required to detect a susceptibility gene using ASPs for TDT analysis (TDT-ASP). In all cases the numbers of TDT-ASPs required to detect a disease gene is smaller than the numbers of TDT singletons, in many cases by a factor of four.

## Discussion

The sibling relative risk associated with the non-HLA linked genes in coeliac disease is comparatively large; however, the relative power of linkage and association detection methods will depend on whether this risk can be ascribed to a single locus or a number of different genes acting in concert. If the sibling risk is conferred by a single locus a linkage search will clearly be the best strategy. However, if susceptibility to coeliac disease is controlled by a number of non-HLA linked genes, each having a small effect, then a TDT based strategy should prove more effective.

Two genome wide linkage searches of coeliac disease families have been reported.<sup>23-24</sup> The study reported by Zhong and colleagues<sup>23</sup> was based on 15 Irish nuclear families containing 45 affected sibships. Linkage of coeliac disease to five new chromosome regions outside HLA were detected: 6p23 (telomeric to HLA), 7q31.3, 11p11, 15q26, and 22cen. These regions were evaluated in 28 northern European families reported by Houlston *et al.*<sup>25</sup> No significant evidence of linkage was found except for chromosome 15. The other genome wide study which has been undertaken, reported by Greco *et al.*<sup>24</sup> was based on an analysis of 108 Italian ASPs. In addition to HLA there was some evidence for linkage to 5qter and 11qter. The presence of a predisposition gene conferring a sibling relative risk greater than 1.8 in the regions of 15q26 and 11p11, and a risk of 1.6 in the regions 7q31 and 22cen were excluded. It is possible that the discrepancies in findings between studies are due to differences in the contribution of genetic factors to coeliac disease in the different populations analysed. Alternatively, the significant

findings reported by Zhong and colleagues<sup>23</sup> may reflect in part the structure of the families studied. Specifically, 31 of the affected sibships belonged to three families. The inclusion of large sibships in which no typing information is available has been shown to lead to artificially inflated support to specific regions.

It is clearly far too early to dismiss linkage as a strategy for detecting coeliac disease susceptibility genes on the basis of the published data. However, if further linkage searches do prove to be ineffective, then an association study may offer a more attractive strategy for locating novel predisposition genes. At present these studies will be restricted to the evaluation of candidate genes for which there is a priori evidence to support their role as susceptibility genes. Three classes of genes involved in the pathogenesis of coeliac disease include those involved in antigen presentation, antigen recognition, and antigen modification. Adopting a candidate gene approach will obviously require smaller sample sizes, and have more power to show involvement of individual loci than to detect loci on a genome wide basis. The limitation of genome wide association and TDT studies is not, as Risch and Merikangas pointed out, a statistical one,<sup>10</sup> but relies on the availability of markers at sufficient density. Marker reduction can only be achieved in the presence of strong disequilibrium, as less than maximum linkage disequilibrium between a marker and disease gene leads to a greatly increased sample size requirement.<sup>26</sup> The outlook for genome wide studies may not be so negative however, since as Risch and Merikangas indicated, both the apo-E region conferring susceptibility to Alzheimer's disease and the insulin VNTR region on chromosome 11p associated with diabetes, display strong linkage disequilibrium.<sup>27</sup>

The information derived from the human genome project coupled with the availability of microarray technology will, in the relatively near future, allow genome wide association studies to be undertaken. If no genes conferring susceptibility to coeliac disease are identified prior to the introduction of this technology, then their detection will ultimately only be limited by the availability of sufficient numbers of affected individuals and controls necessary for the identification of genes conferring small risks.

We are grateful to the two anonymous reviewers whose comments improved this paper. S Bevan is in receipt of a Fellowship from the Coeliac Society.

- 1 Marsh MN. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ("celiac sprue"). *Gastroenterology* 1992;102:330-54.
- 2 Houlston RS, Ford D. Genetics of coeliac disease. *Q J Med* 1996;89:737-43.
- 3 Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993;105:910-22.
- 4 Pena AS, Mann DL, Hague NE, et al. Genetic basis of gluten-sensitive enteropathy. *Gastroenterology* 1978;75:230-5.
- 5 Greenberg DA, Hodge SE, Rotter JL. Evidence for recessive and against dominant inheritance at the HLA-linked locus in coeliac disease. *Am J Hum Genet* 1982;34:263-77.
- 6 Hernandez JL, Michalski JP, McCombs CC, et al. Evidence for a dominant gene mechanism underlying coeliac disease in the West of Ireland. *Genet Epidemiol* 1991;8:13-27.
- 7 Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 1990;46:229-41.
- 8 Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 1996;59:983-9.
- 9 Lander E, Kruglak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;11:241-7.
- 10 Risch N, Merikangas K. The future of genetic studies of complex human disease. *Science* 1996;273:1516-17.
- 11 Camp NJ. Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 1997;61:1424-30.
- 12 Mylotte M, Egan-Mitchell B, McCarthy CF, et al. Incidence of coeliac disease in the West of Ireland. *BMJ* 1973;ii:703-5.
- 13 Logan RFA, Rifkind EA, Busuttill A, et al. Prevalence and "incidence" of celiac disease in Edinburgh and the Lothian region of Scotland. *Gastroenterology* 1986;90:334-42.
- 14 Midhagen G, Jarnerot G, Kraaz W. Adult coeliac disease within a defined geographic area in Sweden: a study of prevalence and associated diseases. *Scand J Gastroenterol* 1988;23:1000-4.
- 15 Greco L, Tozzi AE, Mayer M, et al. Unchanging clinical picture of coeliac disease presentation in Campania, Italy. *Eur J Pediatr* 1989;148:610-13.
- 16 Logan RFA, Tucker G, Rifkind EA, et al. Changes in clinical features of coeliac disease in adults in Edinburgh and the Lothians 1960-79. *BMJ* 1983;286:95-7.
- 17 Catassi C, Fabiani E, Ratsch IM, et al. The coeliac iceberg in Italy. A multicentre antigliadin antibodies screening for coeliac disease in school-age subjects. *Acta Paediatr Suppl* 1996;412:29-35.
- 18 Catassi C, Ratsch IM, Fabiani E, et al. High prevalence of undiagnosed coeliac disease in 5280 Italian students screened by antigliadin antibodies. *Acta Paediatr* 1995;84:672-6.
- 19 Grodzinsky E. Screening for coeliac disease in apparently healthy blood donors. *Acta Paediatr Suppl* 1996;412:36-8.
- 20 Not T, Horvath K, Hill ID, et al. Celiac disease risk in the USA: high prevalence of antiendomysium antibodies in healthy blood donors. *Scand J Gastroenterol* 1998;33:494-8.
- 21 Petronzelli F, Bonamico, Ferrante P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 1997;61:307-17.
- 22 Bevan S, Popat S, Ferguson A, et al. Contribution of the MHC region to the familial risk of coeliac disease. *J Med Genet* 1999 (in press).
- 23 Zhong F, McCombs CC, Olson JM, et al. An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. *Nat Genet* 1996;14:329-33.
- 24 Greco L, Corazza G, Babron MC, et al. Genome search in celiac disease. *Am J Hum Genet* 1998;62:669-75.
- 25 Houlston RS, Tomlinson IP, Ford D, et al. Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997;6:1335-9.
- 26 Muller-Myhsok B, Abel L. Genetic analysis of complex diseases. *Science* 1997;275:1328-9.
- 27 Risch N, Merikangas K. Genetic analysis of complex diseases. *Science* 1997;275:1329-30.