

Candidate gene regions and genetic heterogeneity in gluten sensitivity

P Holopainen, K Mustalahti, P Uimari, P Collin, M Mäki, J Partanen

Abstract

Background—Gluten sensitivity is a common multifactorial disorder, manifested in the small intestine or on the skin as typical coeliac disease or dermatitis herpetiformis, respectively. The only established genetic risk factor is HLA DQ2.

Aims—We tested genetic linkage of previously reported chromosomal loci 5q and 11q in Finnish families with gluten sensitivity. We also tested if genetic linkage to candidate loci on 5q, 11q, 2q33, and HLA DQ differed with respect to clinical manifestations or sex.

Subjects—We studied 102 Finnish families with affected sibpairs. For heterogeneity analysis, families were divided into subgroups according to sex and the presence of dermatitis herpetiformis, the skin manifestation of gluten sensitivity.

Methods—Non-parametric linkage between microsatellite markers and disease was tested. Linkage heterogeneity between subgroups was tested using the M test. The transmission/disequilibrium test and association analysis were performed.

Results—Evidence of linkage to 11q (MLS 1.37), but not to 5q, was found in the entire dataset of 102 families. Heterogeneity between subgroups was suggested: families with only the intestinal disease showed linkage mainly to 2q33 whereas families with dermatitis herpetiformis showed linkage to 11q and 5q, but not to 2q33. Linkage in all three non-HLA loci was strongest in families with predominantly male patients. HLA DQ2 conferred much stronger susceptibility to females than males.

Conclusions—Independent evidence for the suggested genetic linkage between 11q and gluten sensitivity was obtained. The possible linkage heterogeneity suggests genetic differences between intestinal and skin manifestations, and the gender dependent effect of HLA DQ2.

(Gut 2001;48:696-701)

Keywords: gluten sensitive enteropathy; coeliac disease; dermatitis herpetiformis; Finnish population; linkage analysis

Gluten sensitive enteropathy (MIM 212750), also known as coeliac disease, is a common multifactorial disorder characterised by malabsorption and small intestinal injury due to the sensitivity of the intestine to cereal prolamins.^{1,2} In addition to intestinal symptoms, patients can also present with various

extraintestinal symptoms. Dermatitis herpetiformis is a skin manifestation of gluten sensitivity characterised by a blistering rash and mucosal changes in the small intestine, which both respond to a gluten free diet.³ Genetic susceptibility to both manifestations is well established. Concordance between identical twins is high and more than 10% of first degree relatives of patients are affected, although even half of these patients are clinically asymptomatic.¹ To date, the only susceptibility locus confirmed in various studies is located in the major histocompatibility complex (MHC) on chromosome 6p21.3.⁴ In most populations, approximately 90% of patients have the human leucocyte antigen (HLA) class II alleles DQA1*0501 and DQB1*02 (= DQ2). DQ2 negative patients usually carry the DR4-DQ8 haplotype.^{4,5} Thus the HLA association in gluten sensitivity is simple compared with diseases such as insulin dependent diabetes or rheumatoid arthritis, in which much more heterogeneity in the HLA can be found. The simple HLA association together with the fact that the major environmental trigger, ingested gluten, is encountered by all individuals makes genetic analyses of this complex trait feasible.

HLA alone does not wholly explain genetic susceptibility as the concordance rate between identical twins (about 70%) is higher than that (30%) between HLA identical sibs.² Furthermore, only a small fraction of DQ2 positive individuals ever develop coeliac disease although they eat gluten. The HLA locus was recently estimated to account for about 30% of familial clustering of coeliac disease in Italy.⁶ Other genetic components, as yet not identified, are likely to be involved. Genome screens carried out in Irish⁷ and Italian⁸ families identified a number of potentially important non-HLA loci. These loci were however different between the two studies, and findings have not been replicated in other populations.^{9,10} Heterogeneity in genetic as well as in environmental factors is a serious problem in genetic analyses of complex diseases.¹¹ To reduce this heterogeneity, in the present study we focused on the Finnish population which has a limited gene pool and a strong founder effect.^{12,13} This population has been found to be most useful in mapping single gene defects^{12,13} and there is some evidence that it may also be informative in complex diseases.¹⁴ In addition to restricted genetic variation, we minimised phenotypic heterogeneity by subdividing families according to phenotypic features.

Abbreviations used in this paper: HLA, human leucocyte antigen; MHC, major histocompatibility complex; MLS, maximum likelihood score; TDT, transmission/disequilibrium test.

Department of Tissue Typing, Finnish Red Cross Blood Transfusion Service, Helsinki, Finland
P Holopainen
J Partanen

Department of Paediatrics, Tampere University Hospital, Institute of Medical Technology, and Medical School, University of Tampere, Tampere, Finland
K Mustalahti
M Mäki

CSC-Scientific Computing Ltd, Espoo, Finland
P Uimari

Department of Medicine, Tampere University Hospital, and Medical School, University of Tampere, Tampere, Finland
P Collin

Correspondence to: J Partanen, FRC Blood Transfusion Service, Kivihaantie 7, 00310 Helsinki, Finland.
jukka.partanen@bts.redcross.fi

Accepted for publication 5 December 2000

Material and methods

FAMILIES

Families were initially recruited through the Finnish Coeliac Society by advertising in the patients' national newsletter. Voluntary families with at least one affected sibpair were accepted for further evaluation. Medical records were scrutinised and only patients with definitive disease—that is, a diagnosis based on initial small bowel biopsy or skin biopsy (for dermatitis herpetiformis) during a normal gluten containing diet—were included in the study. Healthy family members were screened for antiendomysium antibody and HLA DQ2 positivity. Approximately 10% were found to have the asymptomatic form of the disease (silent coeliac disease) and diagnosis was confirmed by biopsy. All families were of apparent Finnish origin and there was no evidence of any particular clustering in their current place of residence. The study protocol was accepted by the ethics committee of Tampere University Hospital.

The study group comprised 102 families (table 1) with at least one affected sibpair; two families also had affected sibpairs in the third generation. A total of 246 affected siblings and 111 parents were genotyped for candidate regions. All available healthy and asymptomatic siblings were HLA DQ typed. Blood samples were available from both parents in 42 families, from one parent in 27, and unavailable from both parents in 33 families. Seventeen genotyped parents were affected by coeliac disease. Median age at diagnosis was 37 years (range 2–76).

Different subgroups of families were analysed separately. The distribution of family members in the subgroups is shown in table 1. Divisions were made according to the presence or absence of dermatitis herpetiformis (DH and CD groups with 33 and 69 families, respectively), and according to the sex of the affected siblings (46 families having only female patients (FF group) and 56 with at least one affected male sibling (FM group)). Ten families in the latter subgroup had only male patients. A separate analysis of these 10 families was not meaningful because of the small sample size and they were included in the FM group.

Table 1 Distribution of family members in 102 families. The families were divided into the following subgroups (i) CD group: 69 families with only the intestinal manifestation of coeliac disease, (ii) DH group: 33 families who also had dermatitis herpetiformis patients, (iii) FM group: 56 families with both males and females as affected sibs, and (iv) FF group: 46 families with only females as affected sibs. Two of the families contained affected sibpairs in the third generation

	CD group	DH group	Total
FM group	37 families 87 siblings (40F, 47M)	19 families 54 siblings (27F, 27M)	56 families 141 siblings (67F, 74M) 62 parents (13 affected)
FF group	32 families 71 siblings	14 families 34 siblings	46 families 105 siblings 49 parents (4 affected)
Total	69 families 158 siblings 82 parents (11 affected)	33 families 88 siblings 29 parents (6 affected)	102 families 246 siblings (172F, 74M) 111 parents (17 affected)

F=female, M=male.

GENETIC MARKERS

The same microsatellite markers as were reported to show genetic linkage to coeliac disease in an Italian study⁸ were initially tested; six markers on 5q and four on 11q. As marker D11S4142 on 11q23 showed suggestive linkage with the disease and the distances between the original markers were long, four additional markers spanning a 5.6 cM region around D11S4142 were added in the present study. Genotyped markers and their distances are shown in table 2. Marker CD3D is a dinucleotide repeat within the CD3D gene (Mfd69CA). The six microsatellite markers (D2S2392, D2S116, D2S2214, CTLA4(AT)_n, D2S2189, and D2S2237) spanning a 3.3 cM region at the CD28/CTLA4 locus on chromosome 2q33 have been described previously.¹⁵ At this locus we reanalysed the data of 99 families overlapping with the family sample genotyped for 5q and 11q. Genomic DNA was amplified using fluorescent labelled primers and polymerase chain reaction products were detected by the ABI PRISM 310 Genetic Analyser. Genetic distances between the markers were taken from the maps of Genéthon and Genome Database. For monitoring the quality of genotyping, all markers were tested for the Hardy-Weinberg equilibrium and significant excess of homozygosity using the algorithms implemented in the Sib-Pair (v 0.97.5) program package¹⁶; no significant deviations were observed.

The HLA DQB1 alleles were determined from all available members of the 102 families, including healthy siblings of the affected pairs, using the Dynal AllSet SSP DQ "low resolution" kit (Dynal AS, Oslo, Norway).

DATA ANALYSIS

Genetic linkage was tested by calculating the multipoint maximum likelihood score (MLS)

Table 2 Genetic linkage to 5q and 11q in all 102 families. Multipoint maximum likelihood scores (MLS) with point-wise significance are shown

Marker	Distance (cM)	MLS	p Value
D5S410		0.52	0.09
D5S422	8.6	0.56	0.08
D5S2032	4.8	0.35	0.14
D5S425	11.9	0.03	0.44
D5S2069	3.4	0.56	0.08
D5S2111	5.8	0.88	0.03
D11S898	10.9	0.32	0.16
D11S4111	2.5	0.75	0.05
D11S4142	2.0	1.37	0.01
D11S976	1.0	0.90	0.03
D11S4171	0.1	0.99	0.03
CD3D	15.5	0.77	0.05
D11S934	16.3	0.33	0.15
D11S910		0.37	0.13

using the Genehunter 2 package.¹⁷ Allele sharing (identical by descent) in the MLS analysis was estimated only from independent sibpairs in families with more than two affected siblings. The sharing distribution was constrained to Holmans' genetically possible triangle, with no restrictions in the mode of inheritance. Marker allele frequencies were estimated from healthy parents, except for analysis of HLA DQ alleles for which we used published values.¹⁸

Statistical significance of the observed MLS scores was estimated using the Simulate program.¹⁹ Briefly, 500 data sets with 102 families were simulated using an assumption of no linkage between a disease locus and eight linked markers on chromosome 11q. The simulated data sets were analysed in a similar way as the original data. Results from these analyses formed the empirical distribution of the MLS under the null hypothesis.

Linkage heterogeneity between subgroups was tested using the M test as χ^2 statistics $2\ln(10)[MLS(a)+MLS(b)-MLS(a+b)]$, where a and b are the subgroups compared and a+b is the total sample. The significance of the test result was assessed by randomisation. The original sample of 102 families was divided randomly into groups with sizes equal to the tested subgroups—for example, 33 and 69 families for DH and CD groups, respectively. MLS was calculated for each and the M test was performed between groups. This was repeated 10 000 times, and the significance of the original test result was obtained by the number of times the simulated χ^2 exceeded the observed value, divided by 10 000.

The transmission/disequilibrium test (TDT) was calculated by Genehunter 2 to investigate genetic linkage in the presence of allelic association. TDT compares transmitted versus non-transmitted alleles from a heterozygous parent to affected offspring. Both single and two locus TDT with nominal p values were calculated. To take into account multiple testing, the statistical significance of TDT statistics was also estimated by permutation, using 1000 replicates. TDT is a valid test of allelic association if only one affected sib per family is tested. For association analysis, index cases were tested, or when not known the affected sib was picked up randomly.

Results

CANDIDATE REGIONS 5Q AND 11Q IN THE ENTIRE STUDY GROUP

Results of linkage analyses of the 5q and 11q markers in the 102 Finnish families with gluten sensitivity are presented in table 2. On chromosome 5q, a weak nominally significant MLS score of 0.88 ($p=0.03$) was obtained for the most telomeric marker D5S2111. The TDT approach showed no evidence of linkage or association of this locus with the disease. When the same set of chromosome 11q markers as used by Greco *et al* was analysed, D11S4142 showed evidence of linkage and thus a more covering set of markers were tested. Eight markers in the region were screened and linkage was confirmed on 11q23

where D11S4142 gave the highest multipoint MLS of 1.37 ($p=0.01$). Based on simulations this corresponds to chromosomal-wise probability below 0.05. The MLS scores for all five markers in the 5.6 cM interval between D11S4111 and CD3D were statistically significant (point-wise significance $p<0.05$).

ALLELIC ASSOCIATION AND TRANSMISSION/DISEQUILIBRIUM TEST

TDT results also showed genetic linkage in the interval between markers D11S4111 and CD3D. In single locus comparisons, allele 131 of marker D11S4171 showed a nominally significant transmission to affected siblings (TDT 7.00, $p=0.008$). In addition, statistically significant two locus TDT scores (TDT 4.45–8.89) were obtained for all four adjacent marker pairs in the D11S4111 to CD3D segment (data not shown). Haplotype D11S4171*131–CD3D*82 also showed evidence of association with the disease when only a single affected sib from each family was analysed. None of these TDT statistics however remained statistically significant when their significance was evaluated using the permutation test.

SUBDIVISIONS OF FAMILIES

To minimise possible heterogeneity in clinical manifestation, two groups with sufficient number of families were established:

- (i) 33 families with at least one patient affected by dermatitis herpetiformis (DH group) were compared with 69 families with patients affected only by the intestinal manifestation of coeliac disease (CD group);
- (ii) 46 families in which all affected siblings were females (FF group) were compared with 56 families with at least one affected male sib (FM group).

Genetic linkage to the candidate loci 5q and 11q was retested in these subgroups. In addition, we reanalysed the CD28/CTLA4 locus on chromosome 2q33 for which we have previously reported linkage to coeliac disease,¹⁵ as well as the known risk locus at HLA DQ.

CANDIDATE REGIONS IN PHENOTYPIC SUBGROUPS

The three loci on 5q, 11q, and 2q33 showed variation in magnitude of linkage between subgroups (fig 1A–C). Although analysis of the 102 families as a whole did not suggest evidence of linkage on chromosome 5q, the CD and DH groups tested separately revealed linkage to marker D5S2111 in the DH group with an MLS of 1.64. In contrast, no linkage was found in the CD group. The level of genetic linkage to the 11q markers also differed between the two subgroups. The highest MLS for D11S4142 (1.40) was obtained in the DH group whereas the MLS for the CD group was only 0.32. The findings on the 5q and 11q markers supported linkage predominantly in the DH group but the opposite effect was found on the CD28/CTLA4 candidate locus (fig 1C). Restricting analysis to the CD group strengthened the linkage to MLS 2.71 for

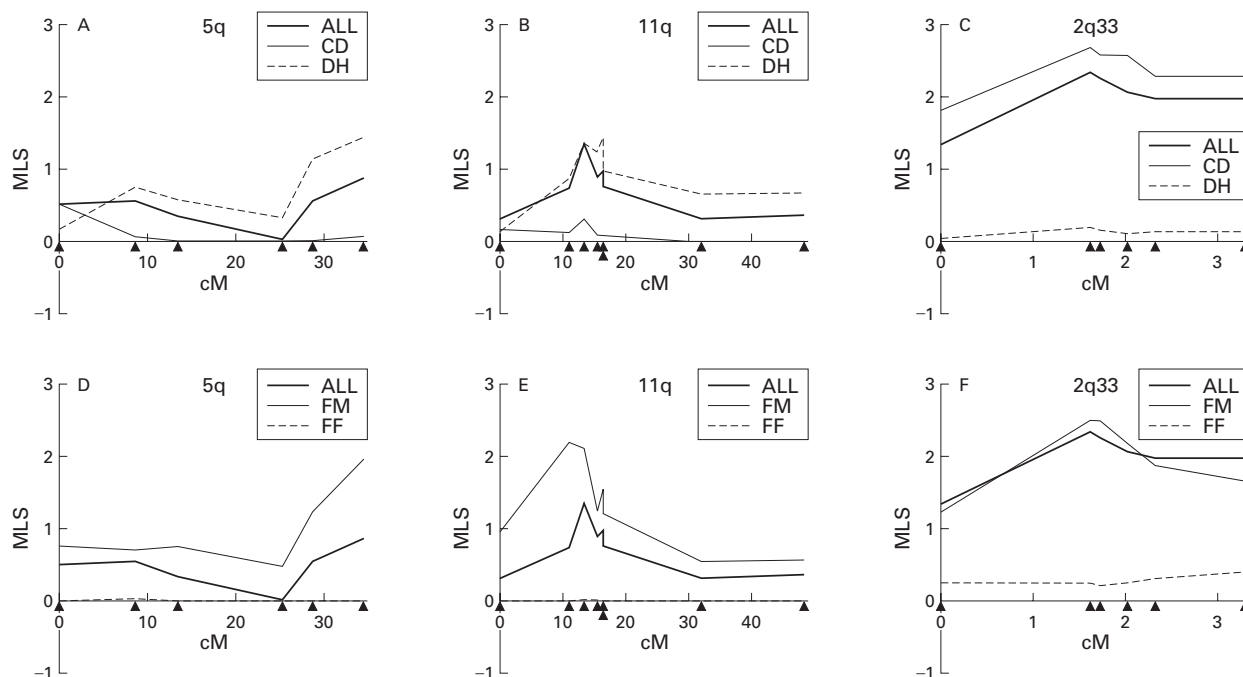


Figure 1 Maximum likelihood scores (MLS) in candidate regions 5q, 11q, and CD28/CTLA4 on 2q33 in the entire study group of 102 families and in the subgroups. Group ALL (all 102 families) was divided into group DH (dermatitis herpetiformis) and group CD (absence of dermatitis herpetiformis) (A–C) or group FM (male patients among affected siblings) and group FF (absence of male patients among affected siblings) (D–F). Marker positions are shown as triangles, from left to right: at 5q, D5S410, D5S422, D5S2032, D5S425, D5S2069, and D5S2111; at 11q, D11S898, D11S4111, D11S4142, D11S976, D11S4171, CD3D, D11S934, and D11S910; and at 2q33, D2S2392, D2S116, D2S2214, CTLA4(AT)₂₃, D2S2189, and D2S2237.

marker D2S116. The MLS of 0.21 in the DH families, on the other hand, suggested no linkage.

The degree of genetic linkage also varied when families were divided according to the sex of the affected siblings. MLS scores >2.0 were obtained for all three candidate regions in the FM group—that is, in families which included at least one male patient. Conversely, families with only female patients (FF group) showed no evidence of linkage to any of the three regions (fig 1D–F).

Linkage heterogeneity between subgroups was tested using the M test statistics for non-parametric MLS scores. Significance of the obtained χ^2 statistics were estimated by 10 000 simulations for each locus and division of families. Nominally significant ($p < 0.05$) heterogeneity was obtained only between FM and FF groups in chromosome 5q; the probability of other comparisons varied from 0.06 to 0.36 (table 3).

RISK DUE TO HLA DQ IN PHENOTYPIC SUBGROUPS

Genetic linkage to the HLA DQ locus was highly significant in the total sample of 102 families (MLS 14.3, $p = 10^{-16}$) as well as in all subgroups: an MLS of 10.9 was obtained for the CD group, 3.8 for the DH group, 6.8 for the FM group, and 8.2 for the FF group. All but 10 patients carried the known risk allele HLA DQ2, nine patients were positive only for the DQ8 risk allele, and one was negative for both DQ2 and DQ8. The other HLA DQ alleles did not differ between the CD and DH groups (data not shown), indicating no role for HLA DQ alleles other than DQ2 and DQ8 in

the determination of the clinical outcome of the disease.

To further study the role of HLA DQ in disease susceptibility between females and males, we compared the prevalence of gluten sensitivity between the HLA DQ2 risk allele positive females and males in the entire family material, including all available siblings. Sixty three per cent of all DQ2 positive family members were affected, and a statistically significant difference between the sexes was observed. Seventy one per cent (176/247) of all DQ2 positive females were affected compared with 52% (84/163) of all DQ2 positive males ($p = 0.0001$ by Fisher's exact test). It must be noted however that the overall high prevalence of gluten sensitivity in these families was inflated because

Table 3 Linkage heterogeneity test between subgroups. The highest maximum likelihood scores (MLS) in the total sample (ALL) are compared with the scores in the subgroups (FM, FF, CD, DH), and respective M statistics with significance obtained by simulation are presented

Marker	Group	MLS	M test χ^2	Simulated p value
D5S2111	FM	2.06	5.44	0.019
	FF	0.00		
	ALL	0.88		
D11S4142	FM	2.10	3.43	0.088
	FF	0.01		
	ALL	1.37		
D2S116	FM	2.54	1.96	0.360
	FF	0.25		
	ALL	2.36		
D5S2111	CD	0.06	3.76	0.057
	DH	1.64		
	ALL	0.88		
D11S4142	CD	0.32	1.61	0.265
	DH	1.40		
	ALL	1.37		
D2S116	CD	2.71	2.55	0.245
	DH	0.21		
	ALL	2.36		

multiplex families were selected. Importantly, the difference between the sexes should not be biased. The female:male ratio among patients in the families was 2.2 (185:85).

Discussion

Genetic linkage studies of complex traits have focused on whole genome scans to identify new risk loci, or on replication studies of the observed candidate regions or functionally interesting genes. Controversial results between genome scans indicate the presence of false positive results although genuine heterogeneity between populations cannot be excluded. Candidate gene approaches have some advantages over whole genome studies. The regions are readily genotyped with more dense sets of markers, which maximises the information content and allows the use of methods utilising linkage disequilibrium. The markers can also be chosen to be identical to those in original studies, which makes comparison of results more reliable and enables pooling of different study samples, increasing significantly the power of the study.

In our study, we tested genetic linkage to previously suggested candidate regions in 102 Finnish families with gluten sensitivity. The Finnish population is characterised by a relatively narrow gene pool and strong founder effect,¹³ both of which are potentially useful in genetic mapping studies. This advantage is particularly clear in single gene defects. Although it must be remembered that susceptibility alleles for common diseases can be rather frequent in a general population, hence decreasing the potential value of the founder effect, the narrow gene pool may also be an advantage in studies of common multifactorial disorders.

Our candidate locus study provides independent support for the presence of a risk locus for gluten sensitivity on chromosome 11q; the locus was originally suggested in the genome wide screening by Greco and coworkers.⁸ Linkage in our family material was found for markers at 11q23, the highest MLS being 1.37 for D11S4142. Linkage was also supported by the nominally significant TDT and association results within the same genomic segment. This can indicate a shared ancestral risk haplotype, or fragments of it, in some families. Further studies are needed because the associations were not statistically significant after the permutation test which corrected for the multiple comparisons made. Although the region showing the best evidence for linkage to the trait does not necessarily indicate the exact location of the susceptibility gene, markers showing linkage and association at 11q23 are located near certain functionally interesting candidate genes. The genes CD3D, CD3E, and CD3G code for molecules forming the CD3 molecule which regulates signalling through the T cell receptor. The gene for the IL-10 receptor is also located in the same region. All of these genes are of interest in a disease with T cell mediated mechanisms. It is possible to speculate that the genes of the

immune system play a critical role in gluten sensitivity as strong linkage to the HLA DQ and also the CTLA4/CD28 locus at 2q33 is involved.^{15 20}

Our results also highlight the importance of stratification of patients in genetic studies. Heterogeneity between disease phenotypes and the sex of patients was studied by evaluating linkage in subgroups of families. Two groups in our sample were sufficiently large in number for meaningful analyses and divisions were reasonable from a clinical point of view: intestinal versus skin manifestations of gluten sensitivity and the predominance of female versus male patients. Most of the families with dermatitis herpetiformis also included patients with the classical intestinal form of coeliac disease. As both manifestations are dependent on gluten ingestion, a gluten free diet after the initial diagnosis of either trait effectively hinders the outcome of the other manifestation. Therefore, discordance between the affected family members could not be unequivocally assumed, and the division was made at the family level: some families had a greater tendency towards the skin manifestation and others to the intestinal disease. Although the strength of genetic linkage was found to differ between the groups, the presence of heterogeneity could not be unequivocally proved statistically. Nominally significant test statistics ($p=0.019$) were found only for chromosome 5 between groups divided by sex but as multiple divisions were made, no firm conclusions can be drawn. However, as heterogeneity in complex diseases can be generally assumed, even weak evidence for it should be reported, enabling independent replication studies to be carried out.

As expected, strong genetic linkage to the HLA DQ locus was found, and HLA DQ2 seems to be the shared risk marker in all subgroups. The chromosome 5q and 11q markers showed stronger linkage in families with dermatitis herpetiformis whereas the CTLA4/CD28 region seemed to predispose only to the intestinal manifestation. Classification of families also indicated gender dependent variation in genetic linkage. Similar findings have been reported in other autoimmune diseases such as type I diabetes and Graves' disease.²¹⁻²³ All three candidate regions showed stronger evidence of linkage in families with affected males. Autoimmune diseases are known to occur more frequently in females, and over representation of females in our study sample fits well with the reported estimates for adult patients with coeliac disease.²⁴ It was intriguing that HLA DQ2 positive females had a statistically significant higher risk of developing coeliac disease than HLA DQ2 positive males in the same families. It is tempting to speculate that the HLA linked risk factor is needed for disease onset for both sexes but non-HLA risk factors play a stronger role in males than females. This implies that identification of novel non-HLA susceptibility genes for gluten sensitivity could be more feasible in families with predominantly male patients.

This study was partially funded by the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources", QLRT-1999-00037, "Evaluation of the prevalence of the coeliac disease and its genetic components in the European population". It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area. The Coeliac Disease Study Group is also supported by the Sigrid Juselius Foundation, University of Helsinki, Emil Aaltonen Foundation, Maud Kuusitila Memorial Foundation, and the Medical Research Fund of Tampere University Hospital.

- 1 Mäki M, Collin P. Coeliac disease. *Lancet* 1997;**349**:1755–9.
- 2 Strober W. Gluten sensitive enteropathy. In: King RA, Roter JL, Motulsky AG, eds. *The genetic basis of common diseases*. Oxford: Oxford University Press, 1992:279–304.
- 3 Reunala T, Kosnai I, Karpai S, et al. Dermatitis herpetiformis: jejunal findings and skin response to gluten-free diet. *Arch Dis Child* 1984;**59**:517–22.
- 4 Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* 1993;**105**:910–22.
- 5 Polvi A, Arranz E, Fernández-Arquero M, et al. HLA-DQ2-negative celiac disease in Finland and Spain. *Hum Immunol* 1998;**59**:169–75.
- 6 Petronzelli F, Bonamico M, Ferrante P, et al. Genetic contribution of the HLA region to the familial clustering of coeliac disease. *Ann Hum Genet* 1997;**61**:307–17.
- 7 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.
- 8 Greco L, Corazza G, Babron MC, et al. Genome search in celiac disease. *Am J Hum Genet* 1998;**62**:669–75.
- 9 Houlston RS, Tomlinson IP, Ford D, et al. Linkage analysis of candidate regions for coeliac disease genes. *Hum Mol Genet* 1997;**6**:1335–39.
- 10 Brett PM, Yiannakou JY, Morris M-A, et al. A pedigree-based linkage study of coeliac disease: failure to replicate previous positive findings. *Ann Hum Genet* 1998;**62**:25–32.
- 11 Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality. *Curr Opin Biotechnol* 1998;**9**:578–94.
- 12 de la Chapelle A. Disease gene mapping in isolated human populations: the example of Finland. *J Med Genet* 1993;**30**:857–65.
- 13 Peltonen L, Jalanko A, Varilo T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 1999;**10**:1913–23.
- 14 Kuokkanen S, Gschwend M, Rioux JD, et al. Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am J Hum Genet* 1997;**61**:1379–87.
- 15 Holopainen P, Arvas M, Sistonen P, et al. CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease. A linkage and family-based association study. *Tissue Antigens* 1999;**53**:470–5.
- 16 Duffy DL. Sib-pair: a program for non-parametric linkage/association analysis. *Am J Hum Genet* 1997;**61**(suppl):A197.
- 17 Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1995;**57**:439–54.
- 18 Partanen J, Westman P. Caucasian Finnish normal. In: Terasaki PI, Gjertson DW, eds. *HLA 1997*. Los Angeles: UCLA Tissue Typing Laboratory, 1997:214–15.
- 19 Ott J, Terwilliger JD. Assessing the evidence for linkage in psychiatric genetics. In: Mendlewicz J, Hippus H, eds. *Genetic research in psychiatry*. Berlin: Springer-Verlag, 1992:245–9.
- 20 Djilali-Saiah I, Schmitz J, Harfouch-Hammoud E, et al. CTLA-4 gene polymorphism is associated with predisposition to coeliac disease. *Gut* 1998;**43**:187–9.
- 21 Morahan G, Huang D, Tait BD, et al. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science* 1996;**272**:1811–13.
- 22 Paterson AD, Petronis A. Sex of affected sibpairs and genetic linkage to type 1 diabetes. *Am J Med Genet* 1999;**84**:15–19.
- 23 Vaidya B, Imrie H, Perros P, et al. The cytotoxic T lymphocyte antigen-4 is a major Graves' disease locus. *Hum Mol Genet* 1999;**8**:1195–9.
- 24 Logan RF. Epidemiology of coeliac disease. In: Marsh MN. *Coeliac disease*. Oxford: Blackwell Scientific Publications, 1992:192–214.

Electronic database information

Genéthon, <http://www.genethon.fr>
 Genome Database, <http://www.gdb.org>
 Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim>.
 Sib-pair program, <http://www2.qimr.edu.au/davidD>