Genome Search in Celiac Disease

Luigi Greco,[1] Gino Corazza,[2] Marie-Claude Babron,[3,*] Fabienne Clot,[4,*]
Marie-Claude Fulchignoni-Lataud,[4,*] Selvaggia Percopo,[1,*] Patrizia Zavattari,[5,*]
Faouzi Bouguerra,[6] Colette Dib,[7] Roberto Tosi,[8] Riccardo Troncone,[1] Alessandro Ventura,[10]
Wilma Mantavoni,[11] Giuseppe Magazzù,[14] Rosanna Gatti,[15] Rosanna Lazzari,[12]
Annamaria Giunta,[16] Francesco Perri,[18] Giuseppe Iacono,[19] Ettore Cardi,[9] Stefano de Virgiliis,[21]
Francesco Cataldo,[20] Gianluigi De Angelis,[22] Salvatore Musumeci,[23] Roberto Ferrari,[24]
Fiorella Balli,[25] Maria-Teresa Bardella,[24] Umberto Volta,[13] Carlo Catassi,[26] Giuliano Torre,[27]
Jean-François Eliaou,[28] Jean-Louis Serre,[4] and Françoise Clerget-Darpoux[3]

[1]Department of Pediatrics, University Federico II, Naples; [2]Medicina Interna, L'Aquila, Italy; [3]Institut Nationale de la Santé et de la Recherche Médicale, Unit 155, Paris; [4]Laboratoire de Cytogénétique et Génétique Moléculaire Humaine, University of Versailles, Versailles; [5]Genetica Umana, Dipartimento di Scienze Mediche, Novara, Italy; [6]Unité de Génétique Epidémiologique, Hôpital d'Enfants and Institut National de la Santé Publique, Tunis; [7]Généthon, Evry; [8]Biologia Cellulare, Consiglio Nazionale delle Richerche, and [9]Department of Pediatrics, University of Roma La Sapienza, Rome; [10]Department of Pediatrics, University of Pisa, Pisa; [11]Laboratorio di Analisi, Ospedale Maggiore Bologna; [12]Department of Pediatrics, University of Bologna, and [13]Il Istituto di Semeiotica Medica, Bologna; [14]Department of Pediatrics, University of Messina, Messina; [15]Istituto Gaslini, Genoa; [16]Department of Pediatrics, University of Milan, and [17]Istituto di Scienze Mediche, Milan; [18]Casa Soccorso della Sofferenza, San Giovanni Rotondo, Italy; [19]Department of Pediatrics "Di Cristina" and [20]Department of Pediatrics "Lancia di Brolo," University of Palermo, Palermo; [21]Istituto di Clinica e Biologia dell'Età Evolutiva, Cagliari, Italy; [22]Department of Pediatrics, University of Parma, Parma; [23]Department of Pediatrics, University of Catania, Catania; [24]Department of Pediatrics, University of Florence, Florence; [25]Department of Pediatrics, University of Modena, Modena; [26]Department of Pediatrics, University of Ancona, Ancona; [27]Istituto Burlo Garofolo, Trieste; and [28]Institut Nationale Santé Étude et Recherche Médicale, Unit 291, Montpellier, France

Summary

Celiac disease (CD), a malabsorption disorder of the small intestine, results from ingestion of gluten. The HLA risk factors involved in CD are well known but do not explain the entire genetic susceptibility. To determine the localization of other genetic risk factors, a systematic screening of the genome has been undertaken. The typing information of 281 markers on 110 affected sib pairs and their parents was used to test linkage. Systematic linkage analysis was first performed on 39 pairs in which both sibs had a symptomatic form of CD. Replication of the regions of interest was then carried out on 71 pairs in which one sib had a symptomatic form and the other a silent form of CD. In addition to the HLA loci, our study suggests that a risk factor in 5qter is involved in both forms of CD (symptomatic and silent). Furthermore, a factor on 11qter possibly differentiates the two forms. In contrast, none of the regions recently published was confirmed by the present screening.

Introduction

Celiac disease (CD) is a permanent immunomediated intolerance to the gluten-composing proteins of wheat, barley, and rye. The population prevalence in the Western world is between 0.2% and 0.5% (Catassi et al. 1994). The prevalence probably represents an underestimate, because not all cases of CD are symptomatic. Indeed, a variety of clinical features result from gluten consumption in genetically predisposed individuals, ranging from absence of significant symptoms to life-threatening health complaints (Troncone et al. 1996). A unifying classification of the clinical features was proposed by the European Society of Paediatric Gastroenterology and Nutrition (ESPGAN) in 1970 (Meeuwisse 1970) and has since been updated (Walker-Smith et al. 1990). Malabsorption with chronic diarrhea, iron deficiency and growth retardation are often the main symptoms. At the histological level, the small bowel presents various degrees of mucosal lesions and intraepithelial lymphocyte invasion (Marsh 1992). Enteropathy-associated T-cell lymphoma and small bowel, pharyngeal, or esophageal cancer can develop. Extraintestinal manifestations of CD, such as neurological symptoms, epilepsy, posterior cerebral calcifications, dermatitis herpetiformis, and various associated or subsequent disorders, are sometimes present (Mäki and Collin 1997). The autoimmune process leads to CD-specific antigliadin, antireticulin, and antiendomysium antibodies.

CD has a multigenic and multifactorial origin: the

existence of a genetic predisposition is supported by a 10% prevalence in first-degree relatives and a 80%–100% concordance in MZ twins, compared with 20% in DZ twins (Sollid and Thorsby 1993). A strong predisposing factor lies in the HLA class II region. Eighty percent to 95% of celiac patients carry the same DQ heterodimer, DQA1*0501 DQB1*0201, encoded in *cis,* in DR3 individuals, or in *trans,* in DR5/DR7 heterozygous individuals (Sollid et al. 1989). All of the remaining cases carry a DR heterodimer encoded by the genes DRA1 and DRB4 (Bouguerra et al. 1997; Clerget-Darpoux et al. 1997). However, only part of the familial aggregation is explained by the HLA component (Risch 1987; Petronzelli et al. 1997), which suggests the presence of additional non-HLA genetic risk factors. To find other factors of genetic susceptibility, we collected a large sample of CD-affected sib pairs with living parents, and we screened the whole genome by testing linkage with polymorphic genetic markers.

Families, Material, and Methods

Family Sample

A sample of Italian sib pairs was ascertained by identification of one sib who had a symptomatic form of CD, according to Società Italiana di Gastroenterologia ed Epatologia Pediatrica criteria. The other sib could present one of two clinical forms of CD: a symptomatic form characterized by villous atrophy, antiendomysium antibodies, and clinical symptoms (diarrhea, growth failure, or vomiting) or a silent form characterized by villous atrophy and antiendomysium antibodies but without clinical symptoms. Each sib was diagnosed according to the criteria set by the ESPGAN (Walker-Smith et al. 1990). Families in which patients were not diagnosed through a small-bowel biopsy were not considered for admission. Ethical approval was granted by the National Scientific Committee of the Italian Celiac Society. Informed consent was obtained from all members of the family under study.

One hundred five families with living parents and either two affected sibs (98 families) or three affected sibs (7 families) were thus recruited. Note that, after verification of Mendelian segregation for each marker, two families were excluded from the sample, giving rise to 110 independent affected sib pairs. They are distinguished according to their clinical-symptomatic or silent-form status: S1 pairs are those in which both sibs have a symptomatic form (39 pairs), and S2 pairs are those in which one sib has a symptomatic form and the other a silent form (71 pairs).

Microsatellite Genotyping

The affected sibs and their parents were typed, over the whole genome, by use of 254 microsatellite poly-

morphic markers, from the fluorescent panel of Généthon, with an average distance of 13 cM (Dib et al. 1996). PCR amplifications were performed in a final volume of 15 μl, containing 30 ng of genomic DNA, 0.33 μM of each primer, 0.16 mM dNTPs, 10 mM Tris pH 9, 50 mM KCl, 1.5 mM MgCl$_2$, 0.1% Triton, 0.01% gelatin, and 0.03 IU of *Taq* polymerase (Perkin Elmer) (Ginot et al. 1996). Reactions were performed by means of a "hot-start" procedure: *Taq* polymerase was added to the reaction mixtures after a first denaturation step (5 min at 96°C). This was followed by 33 cycles of increasing stringency: denaturation (40 s at 94°C), annealing from 70°C to 57°C, 30 s without elongation step. After completion of the PCR reaction, T4 DNA polymerase (0.05 UI), 3′–5′ activity (Biolab) was applied at 37°C for 30 min. PCR reaction products were pooled, and denatured aliquots were separated on 5% denaturing polyacrylamide gels. Data were collected with the 373A sequencer (ABI) and genotyped with Genescan 672 and Genotyper software (ABI).

Five regions on chromosomes 6, 7, 11, 15, and 22 were more densely typed, because some evidence of linkage had been suggested by Zhong et al. (1996). Twenty-seven markers were amplified by the "classic method" (Gyapay et al. 1994). PCR amplifications were performed in 50-μl reaction mixtures that contained 40 ng of genomic DNA, 50 pmol of each primer, 125 μM dNTPs, 10 mM Tris pH 9, 50 mM KCl, 1.5 mM MgCl$_2$, 0.1% Triton X-100, 0.001% gelatin, and 0.06 IU of *Taq* polymerase (Perkin Elmer). Amplifications were carried out by use of the hot-start procedure, with a denaturation step (5 min at 96°C), after which 35 cycles of denaturation (94°C for 40 s) and annealing (55°C for 30 s) were performed, followed by a final elongation step (2 min at 72°C). For each DNA sample, three amplification products from different markers were loaded together into single lanes of 6% polyacrylamide–8 M urea denaturing gels. After migration, DNA was transferred from the gel to a Hybond N$^+$ nylon membrane (Amersham) by a contact blotting procedure. The markers were then revealed by hybridization with the microsatellite probe, which was peroxidase labeled (Amersham) and exposed to autoradiographic X-ray films. Overall, 281 microsatellite markers were studied.

Data Analysis

Linkage was tested in accordance with the maximum-likelihood–score (MLS) approach (Risch 1990*b*). The identity-by-descent (IBD) distribution (proportion Z_2, Z_1, or Z_0 of pairs sharing 2, 1, or 0 allele(s) IBD, respectively) can be estimated, given the observations on the marker genotype. The likelihood of this IBD estimate is then compared with the one expected under no linkage ($\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$). The MLS is the decimal logarithm of the likelihood ratio. Information on several linked markers can
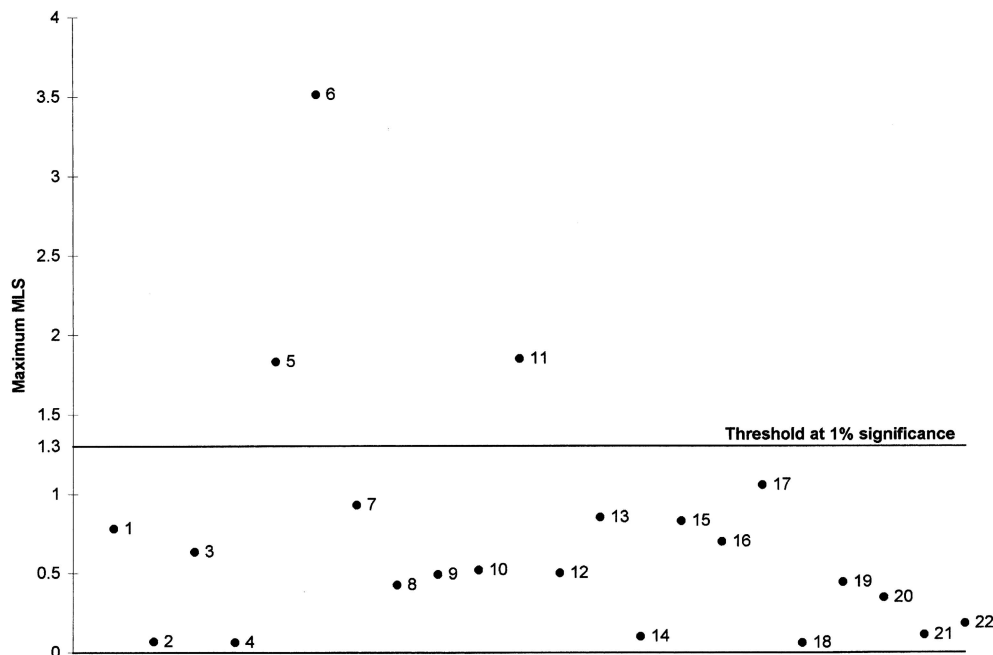
Figure 1   MLS scores obtained for each chromosome in the S1 sample (both sibs symptomatic CD).

be gathered simultaneously. The MLS can be computed for different locations relative to the markers. The MLS does not depend on the marker allele frequencies when the parents are typed, which is the case in the present study.

The MAPMAKER-SIBS program (Kruglyak and Lander 1995) calculates, for a sample of affected sib pairs, the MLS and the corresponding maximum-likelihood estimate of the IBD vector.

When the test is applied for one locus, an MLS of 1.3 gives evidence for linkage with a type I error of 1% (Holmans 1993). In case of systematic screening of the genome, the type I error for such a value is much higher. The correction due to the numerous (281) points analyzed drastically reduces the power to detect involved factors. To avoid losing power while keeping a reasonable type I error, we performed a two-step analysis. In a first phase, we calculated the MLS at each marker locus for the S1 sample. Then, for each marker position at which the MLS was >1.3, we calculated the MLS in the S2 sample, not only for this marker but also for the two flanking markers. Indeed, the confidence interval is large, and the maximum MLS does not necessarily indicate the true location of the risk factor (Kruglyak and Lander 1996). The significance of the MLS test for each region tested was assessed through simulations. The maximum MLS obtained in the S2 sample for the three markers of a considered replication region is denoted by V. Simulations of segregation of the three markers under study were based on the assumption of no linkage between the disease and the markers. One thousand replicates

were generated. We then calculated the number of times the MLS for these three markers exceeded the value V. This calculation gives the type I error for this region.

In case of nonreplication in the S2 sample, two explanations could be given: either results in S1 are due to random deviations, or the two samples are heterogeneous. Therefore, for each region considered in the second step, the homogeneity of the two samples was tested. The following Q values were computed for the three markers of the region: $Q = -2\mathrm{Ln}10\{\mathrm{MLS}(S1 + S2) - [\mathrm{MLS}(S1) + \mathrm{MLS}(S2)]\}$.

The significance of the maximum Q value (Q_{max}) was assessed by bootstrap. The bootstrapping was performed on the genotypes at the three markers of the original sample (S1 + S2). The sample was then divided each time in two subsamples of 39 sib pairs (equivalent to S1) and of 71 sib pairs (equivalent to S2). Two thousand bootstraps were realized, and the corresponding value of Q_{max} was evaluated for each. The significance of the test is obtained by the number of times Q exceeds the observed Q_{max} value.

Risch (1990a) introduced a measure of the contribution of a genetic risk factor, denoted λ. Given the marker observation, exclusion of those areas of the genome that are unlikely to contain a genetic risk factor corresponding to a given relative risk λ^* is possible by comparing the likelihood of $\lambda = 1$ to the likelihood of λ^*. When presence of a genetic factor with a relative risk $\geqslant \lambda^*$ has been excluded, the existence of a risk locus in the region is still possible, but its associated risk must be $<\lambda^*$.
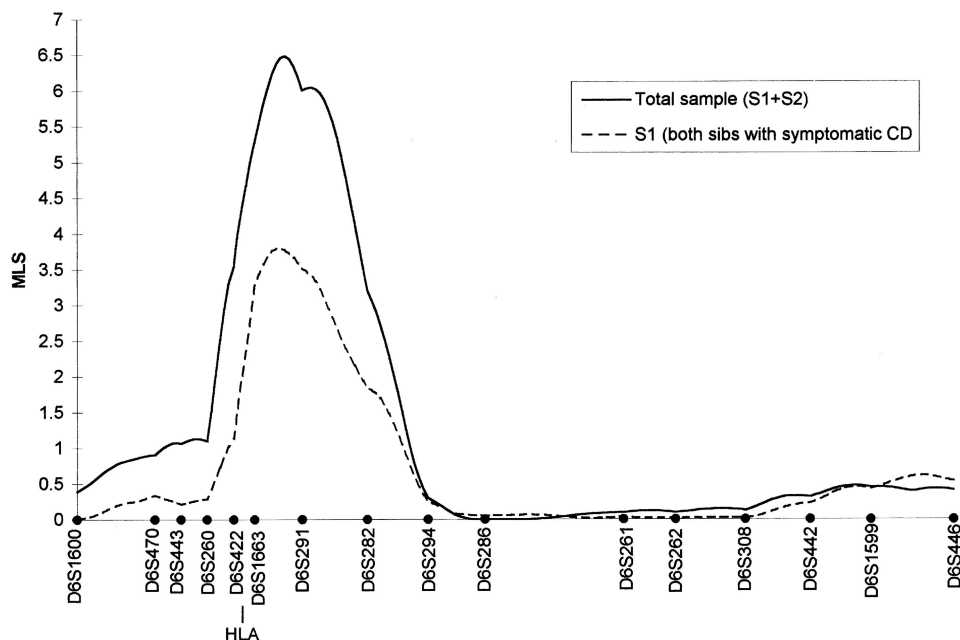
Figure 2 Distribution of the MLS scores for chromosome 6. Solid line: total sample (S1 + S2); broken line: S1 sample (both sibs symptomatic CD). The dots on the abscissa indicate the positions and names of the markers used in the analysis.

Results

Systematic Screening

In the S1 sample, the MLS values were computed for each chromosome, at each marker position, taking into account the genotype information available on all the chromosome markers. Figure 1 displays, for each chromosome, the maximum of these MLS values. Note that the type I errors that correspond to an MLS value between markers are not the same as those at the exact marker locations. Eichenbaum-Voline et al. (1997) showed that they are greater between markers. Consequently, only values at marker locations are interpreted in this study.

The MLS in the S1 sample exceeds 1.3 for three regions: 6p21 (MLS = 3.5), 5qter (MLS = 1.8), and 11qter (MLS = 1.8). The MLS curves (dashed line) for these three chromosomes are shown in figures 2, 3, and 4, respectively. The highest MLS is obtained for the 6p21 (HLA) region, which corresponds to the effect of the well-known HLA component.

Figure 2 displays the MLS results in the HLA region for S1 (dashed curve) and the global sample (solid curve). The maximum MLS are 3.5 and 6.0, respectively. There is no evidence of heterogeneity between S1 and S2. The global IBD estimate is $Z_2 = .50$; $Z_1 = .41$; $Z_0 = .09$.

The first step of the analysis suggests the existence of other factors in two new regions. Therefore, in the S2

sample, the MLS were computed only for three markers in these two regions.

On the region 5qter, the three markers used for replication in the S2 sample were D5S2069, D5S2111, and D5S2006. An MLS value of 1.0 was obtained for D5S2006 (with a *P* value, assessed by simulations, of .04). There is no evidence for heterogeneity between S1 and S2 for the 5qter markers ($Q_{max} = 3.6$; $P = .18$). The MLS value on the global sample (S1 + S2) is 2.0 (solid line in fig. 3) at the terminal marker with an IBD estimate $Z_2 = .42$; $Z_1 = .39$; $Z_0 = .19$.

On the region 11qter, the three markers used for replication in the S2 sample are D11S4142, D11S934, and D11S910. In contrast to the 5qter region, evidence for linkage was not replicated in the S2 sample for the 11qter region (MLS = 0 for the three markers). In addition, there is evidence of heterogeneity between S1 and S2 ($Q_{max} = 6.14$; $P = .03$). This means that either the results for S1 cannot be interpreted as evidence for linkage or there exists, in 11qter, a risk factor that differentiates the symptomatic and silent forms of CD.

Nonreplication of Regions Already Published

In a sample of Irish sib pairs, Zhong et al. (1996) found evidence for linkage for five new chromosome locations apart from HLA: 6p23, 7q31, 11p11, 15q26, and 22cen. Table 1 provides, for these five locations, the MLS of Zhong et al., our MLS on the global sample (S1
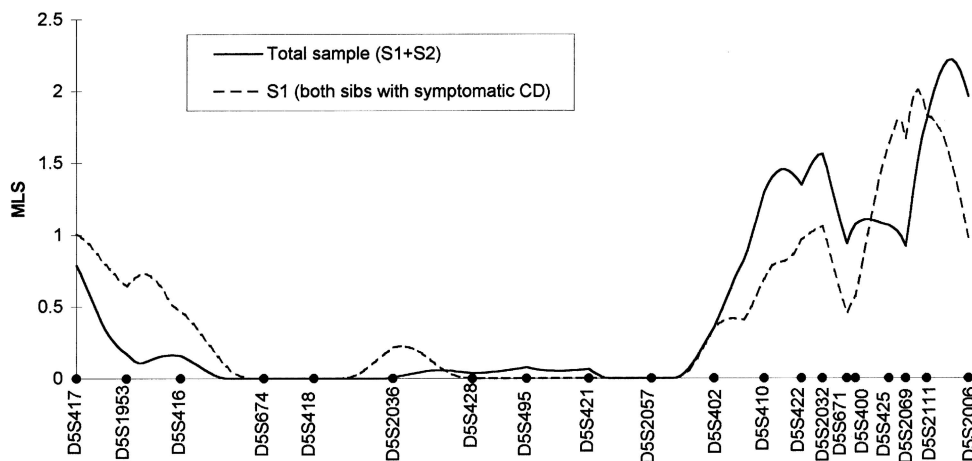
Figure 3   Distribution of the MLS scores for chromosome 5. Solid line: total sample (S1 + S2); broken line: S1 sample (both sibs symptomatic CD). The dots on the abscissa indicate the positions and names of the markers used in the analysis.

+ S2), and the λ value for which the likelihood ratio is below the traditional criterion of -2. For the last four locations, the presence of a risk factor with a relative risk >1.8 is excluded in our study. For the region 6p23, the MLS of 1.0 at D6S443 corresponds, in our data, to what is expected for a risk factor located in HLA. Indeed, given the well-known HLA determinism and the corresponding HLA IBD vector (0.50, 0.41, 0.09), it is possible to calculate the expected IBD vector at D6S443 (28 cM from HLA). The observed IBD at D6S443 (0.32, 0.49, 0.19) is very close to the expected one (0.35, 0.47, 0.18).

Although our study does not confirm the regions proposed by Zhong et al., note that, in contrast to the marker in 5qter for which we obtained a MLS of 2.0 in S1, the MLS in Zhong et al. was 0.9 (fig.1*a* in their article). This is not at all inconsistent with our suggestion of a susceptibility locus in this region.

In addition to our linkage analysis, it was possible to estimate the typing error rate through the additional analysis of 15 MZ twins. The rate was estimated to be 1%. Our study also allowed more accurate estimation

Table 1

Genomic Regions of Interest (Zhong et al. 1996): Multipoint MLS Values and Our λ Exclusion Values Corresponding to a Likelihood Score <-2

| | MULTIPOINT MLS | | |
REGION	Zhong et al. (1996)	Present Study	λ
6p23	4.66	1.0	…
7q31	2.99	0	1.6
11p11	3.92	0	1.8
15q26	2.12	.5	1.8
22cen	2.69	0	1.6

of the allelic frequencies of Généthon panel-D microsatellites, through genotyping of all 236 parents (472 alleles). These data are available on the Généthon web server (http://www.genethon.fr/projets/genfluo/COELIAQUE).

Discussion

Our study confirms the implication of HLA as a risk factor for CD. No heterogeneity could be found between the S1 and S2 samples. This result shows that the HLA component is not a factor that differentiates the symptomatic and silent forms of CD.

Beyond the well-known factor in the HLA region (6p21), our study suggests that a risk factor in the terminal portion of chromosome 5 (5qter) is involved in both symptomatic and silent forms of CD. The results of Zhong et al. (1996) for the same region provide additional evidence for a risk factor in this region.
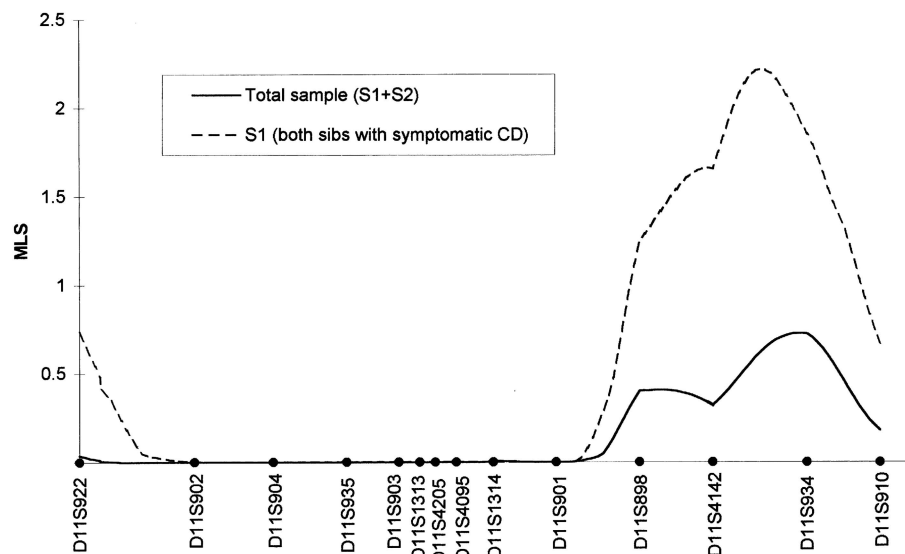
For the 11qter region, our data cannot discriminate between the two hypotheses: presence of a susceptibility factor that is involved in the symptomatic forms of CD only, or S1 as the result of chance.

In a sample of Irish sib pairs, Zhong et al. (1996) found some evidence for linkage for five new chromosome locations apart from HLA: 6p23, 7q31, 11p11, 15q26, and 22cen. In a recent study, Houlston et al. (1997) found no significant evidence in favor of linkage to CD for these regions, except for chromosome 15. In our study, the presence of a risk factor with a relative risk >1.8 in the regions 15q26 and 11p11, and 1.6 in the regions 7q31 and 22cdn, was excluded. On the assumption of the presence of risk factors in these regions with relative risks <1.8, MLS values as high as those obtained by Zhong et al., in a sample of 45 dependent

Figure 4 Distribution of the MLS scores for chromosome 11. Solid line: total sample (S1 + S2); broken line: S1 sample (both sibs symptomatic CD). The dots on the abscissa indicate the positions and names of the markers used in the analysis.

sib pairs, are not expected. Two explanations may be given to explain this discrepancy: either the genetic component of CD is different in the Irish and Italian populations, or the high MLS values obtained by Zhong et al. are due to the particular familial structure they studied. Indeed, among their 45 sib pairs, 31 belong to just three nuclear families. Babron et al. (1997) showed that, when parents are untyped, inclusion of a large number of sib pairs that belong to a few nuclear families may lead to artificially high values of MLS. In particular, it may explain the second MLS peak on chromosome 6 in 6p23. A replication study in a different sample is planned, to confirm the presence of a risk factor in 5qter and to demonstrate that a risk factor in 11qter differentiates symptomatic and silent forms of CD.

Acknowledgments

References

Babron MC, Truy F, Eichenbaum-Voline S, Clerget-Darpoux F (1997) Behavior of the maximum likelihood score when many affected sibpairs are issued from a few untyped parents. Genet Epidemiol 14:A4

Bouguerra F, Babron MC, Eliaou JF, Debbabi A, Clot J, Khaldi F, Clerget-Darpoux F (1997) Synergistic effect of two HLA heterodimers in the susceptibility of Celiac disease in Tunisia. Genet Epidemiol 14:413–422

Catassi C, Ratsch IM, Fabiani E, Rossini M, Bordicchia F, Candela F, Coppa GV, et al (1994) Celiac disease in the year 2000: exploring the iceberg. Lancet 343:200–203

Clerget-Darpoux F, Babron MC, Bouguerra F, Clot F, Djilali-Saiah I, Khaldi F, Debbabi A, et al (1997) Synergistic effect of two HLA heterodimers in celiac disease. In: Charron D (ed) Genetic diversity of HLA functional and medical implication. Vol 2. Edition Dufour Krief, Paris, pp 718–719

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature 380:152–154

Eichenbaum-Voline S, Génin E, Babron MC, Margaritte-Jeannin P, Prum B, Clerget-Darpoux F. Caution in the interpretation of MLS. Genet Epidemiol 14:1079–1083

Ginot F, Bordelais I, Nguyen S, Gyapay G (1996) Correction of some genotyping errors in automated fluorescent micro-

satellite analysis by enzymatic removal of one base over-hangs. Nucleic Acids Res 24:540–541

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millas-seau P, Marc S, et al (1994) The 1993–1994 Généthon hu-man genetic linkage map. Nat Genet 7:246–249

Holmans P (1993) Asymptotic properties of affected-sib-pair linkage analysis. Am J Hum Genet 52:362–374

Houlston RS, Tomlinson IPM, Ford D, Seal S, Marossy AM, Ferguson A, Holmes GKT, et al (1997) Linkage analysis of candidate regions for coeliac disease genes. Hum Mol Genet 6:1335–1339

Kruglyak L, Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 57:439–454

——— (1996) Limits on fine mapping of complex traits. Am J Hum Genet 58:1092–1093.

Mäki M, Collin P (1997) Coeliac disease. Lancet 349: 1755–1759

Marsh M (1992) Gluten, major histocompatibility complex, and the small intestine. Gastroenterology 102:330–354

Meeuwisse G (1970) Diagnostic criteria in coeliac disease. Acta Paediatr Scand 59:461–463

Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mar-iani P, Appollonio I, et al (1997) Genetic contribution of the HLA region to the familial clustering of coeliac disease. Ann Hum Genet 61:307–317

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1–14

——— (1990a) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:223–228

——— (1990b) Linkage strategies for genetically complex traits. III. The effects of markers polymorphism on analysis of affected relative pairs. Am J Hum Genet 46:242–253

Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E (1989) Evidence for a primary association of celiac disease to a particular HLA-DQ alpha-beta heterodimer. J Exp Med 169:345–350

Sollid LM, Thorsby E (1993) HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. Gastro-enterology 105:910–922

Troncone R, Greco L, Auricchio S (1996) Gluten sensitive enteropathy. Pediatr Clin North Am 43:355–373

Walker-Smith JA, Guandalini S, Schmitz J, Shmerling DH, Vi-sakorpi JK (1990) Revised criteria for diagnosis of coeliac disease. Arch Dis Child 65:909–911

Zhong F, McCombs C, Olson JM, Elston RC, Stevens FM, McCarthy CF, Michalsky JP (1996) An autosomal screen for genes that predispose to celiac disease in the western counties of Ireland. Nat Genet 14:329–333