Am. J. Hum. Genet. 70:51–59, 2002

Genomewide Linkage Analysis of Celiac Disease in Finnish Families

Jianjun Liu,[1] Suh-Hang Juo,[1] Päivi Holopainen,[6] Joseph Terwilliger,[1,5] Xiaomei Tong,[1] Adina Grunn,[1] Miguel Brito,[1] Peter Green,[3] Kirsi Mustalahti,[7] Markku Mäki,[7] T. Conrad Gilliam,[1,2,4] and Jukka Partanen[6]

[1]Columbia Genome Center and Departments of [2]Psychiatry, [3]Medicine, and [4]Genetics and Development, Columbia University, and [5]New York State Psychiatric Institute, New York; [6]Department of Tissue Typing, Finnish Red Cross Blood Transfusion Service, Helsinki; and [7]Medical School, University of Tampere and Tampere University Hospital, Tampere, Finland

Celiac disease (CD), or gluten-sensitive enteropathy, is a common multifactorial disorder resulting from intolerance to cereal prolamins. The only established genetic susceptibility factor is HLA-DQ, which appears to explain only part of the overall genetic risk. We performed a genomewide scan of CD in 60 Finnish families. In addition to strong evidence for linkage to the HLA region at 6p21.3 ($Z_{max} > 5$), suggestive evidence for linkage was found for six other chromosomal regions—1p36, 4p15, 5q31, 7q21, 9p21-23, and 16q12. We further analyzed the three most convincing regions—4p15, 5q31, and 7q21—by evaluation of dense marker arrays across each region and by analysis of an additional 38 families. Although multipoint analysis with dense markers provided supportive evidence (multipoint LOD scores 3.25 at 4p15, 1.49 at 5q31, and 1.04 at 7q21) for the initial findings, the additional 38 families did not strengthen evidence for linkage. The role that HLA-DQ plays was studied in more detail by analysis of DQB1 alleles in all 98 families. All but one patient carried one or two HLA-DQ risk alleles, and 65% of HLA-DQ2 carriers were affected. Our study indicates that the HLA region harbors a predominant CD-susceptibility locus in these Finnish families.

Introduction

Celiac disease (CD [MIM 212750]) is a multifactorial disorder of the small intestine and results from intolerance to prolamins from wheat, barley, and rye. CD can have different manifestations, ranging from overt clinical cases to milder silent cases (Mäki and Collin 1997). The prevalence of CD has, until recently, been estimated to be >1/1,000. However, recent data from various populations in Europe and the United States (Not et al. 1998) suggest a much higher prevalence, of 0.4%–1%. Development of CD is known to be influenced by both environmental (gluten) and genetic (HLA) factors. Disease prevalence among first-degree relatives of affected individuals has been reported to be ~10% (Ellis 1981; Mäki et al. 1991), and the concordance rate among MZ twins has been reported to be >70% (Polanco et al. 1981; Hervonen et al. 2000; Greco et al., in press).

Although the pathogenic mechanism of CD is not known, it has been clear for some time that the immune system is involved. CD susceptibility is strongly associated with the risk alleles DQA1*05 and DQB1*02, which encode the heterodimeric HLA-DQ2 molecule (in *cis* in the DR3 haplotype and in *trans* in the DR5/7 heterozygotes). To date, >90% of patients with CD in most populations that were tested carry the DQ2 heterodimer (Spurkland et al. 1990). In addition, the DQ8 molecule, which is encoded by the DQA1*03 and DQB1*0302 alleles, has been shown to associate with CD in cases in which DQ2 is not present (Spurkland et al. 1992; Polvi et al. 1998). In summary, a key role that DQ2 and DQ8 play in disease pathogenesis has been described (Sollid 2000) that establishes the HLA-DQA1 and HLA-DQB1 loci as the principal CD-associated genetic determinants in the HLA region.

Despite its strong association with CD, HLA-DQ likely explains only part of the genetic susceptibility. Evidence for the existence of other genetic factors comes mainly from three sources. First, although almost all patients carry the DQ risk molecules, ⩽22% of individuals in the general Finnish population also carry the DQ2 risk molecules (Partanen and Westman 1997) and still remain unaffected despite cereal intake; therefore, only a small portion (~5%) of individuals positive for DQ2 subsequently develop CD, suggesting that, although certain DQ alleles are necessary for the development of disease, other genetic or environmental risk factors must also exist. Second, although, for CD, the concordance rate among MZ twins is >70%, the concordance rate among DQ-risk-allele–identical siblings is

Received September 21, 2001; accepted for publication October 26, 2001; electronically published November 19, 2001.

Address for correspondence and reprints: Dr. Jukka Partanen, FRC Blood Transfusion Service, Kivihaantie 7, 00310 Helsinki, Finland. E-mail: jukka.partanen@bts.redcross.fi

© 2002 by The American Society of Human Genetics. All rights reserved.
0002-9297/2002/7001-0007$15.00

only ~30% (Mearin et al. 1983; Petronzelli et al. 1997); the difference between these two rates indicates that other genetic factors may play a role in the development of CD. Third, several family-based studies have been conducted to estimate recurrence risk, along with the population risk (λ_s), for CD in siblings (with or without identical HLA alleles) (Rotter and Landaw 1984; Risch 1987; Petronzelli et al. 1997; Bevan et al. 1999); from such studies, the HLA-associated risk factor was estimated to account for 29%–40% of overall CD familiarity. It should be pointed out, however, that because DQ risk alleles are so common in the general population, estimates based solely on identical-by-descent (IBD) calculations are expected to bias the results.

In an attempt to identify non-HLA genetic factors related to CD, three independent genomewide-mapping studies on CD have been conducted. The first, from western Ireland, analyzed 15 families with 40 affected siblings (Zhong et al. 1996). In addition to the HLA region that demonstrated strong linkage to CD, six other regions—a secondary region on 6p, as well as 7q31, 11p11, 15q26, 19q, and 22cen—were reported to yield suggestive evidence for linkage. Subsequently, the non-HLA regions that were implicated from the Irish study were reexamined by Houlston et al. (1997) and Brett et al. (1998), by use of independent sample sets. Although the former study showed moderate evidence for linkage in the 15q26 region, the latter study did not find any evidence for linkage from these regions. A second genomewide-mapping study analyzed 110 Italian sib pairs (Greco et al. 1998) and provided suggestive evidence for linkage to chromosome 5q. From the same study, analysis of a subset of 39 sib pairs, in which both siblings manifested a symptomatic form of CD, provided suggestive evidence for linkage to chromosome 11q. Interestingly, this finding was not replicated in the remaining 71 sib pairs in which one sibling displayed symptomatic CD and the other sibling was diagnosed with a silent form of CD. Recently, Greco et al. (2001) have presented further evidence, from an independent sample set, in support of linkage to chromosome 5q. Recent candidate-gene studies performed in Finnish families likewise are consistent with the reported linkage at chromosomes 5q and 11q (Holopainen et al. 2001). A third genome scan with 16 multiplex families from the United Kingdom reported putative linked loci at 10q and 16q and supporting evidence for linkage to 6q, 11p, and 19q, three suggested regions in the first genome scan from western Ireland (King et al. 2000). By use of a candidate-gene approach, evidence is accumulating that the *CTLA4/CD28* gene region on chromosome 2q33 may, in the northern European populations, confer susceptibility to CD (Holopainen et al. 1999; Torinsson Naluai et al. 2000). For all non-HLA loci thus far reported, evidence for genetic linkage or association is moderate and inconclusive.

In the present study, a genomewide-mapping approach was used to search for non-HLA genetic factors for CD in a collection of families from Finland, a population that has been characterized by a somewhat less diverse gene pool and by the identification of founder effects for a number of single-gene diseases (de la Chapelle 1993; Peltonen et al. 1999). Perhaps more importantly, the cultural and environmental variability is, in comparison to that of many other populations, low in the Finnish population; the population is unusually cooperative for genetic studies; and there are reliable and extensive computerized medical and population records available for research purposes. The Finnish population has proved to be extremely valuable for the mapping of rare single-gene defects, although its advantages for the study of more-common diseases remain unproved. In the present study, genomewide mapping was initially performed with 60 families, each with multiple affected individuals. The three most interesting regions of the genome (other than HLA) were further investigated by the genotyping of additional markers and of an additional 38 families. The role that HLA-DQ plays in the familial aggregation of disease was further examined in the family sample by genotyping DQB1, a well-established and reliable assay for the CD-associated DQ2 molecule.

Families, Material, and Methods

Families

Families were initially recruited through the Finnish Coeliac Society, by advertising in the patients' newsletter. Of the families who volunteered, those with at least one affected sib pair were accepted for further evaluation. The earlier diagnoses were then reevaluated by scrutinization of the medical records. Healthy family members were screened for antiendomysium antibody activity, and ~10% were found to have an asymptomatic form of the disease. Diagnosis based on small-bowel biopsy while on normal gluten-containing diet was obtained for all but 17 of 256 patients, for whom no definitive evidence based on biopsy was available. Thirty-five patients had dermatitis herpetiformis, and 20 were asymptomatic patients who were found by screening of the family members. Median age at diagnosis was 37 years. All families were of apparent Finnish origin, and there was no evidence for any particular clustering in their current places of residence. Sixty families with multiple affected siblings were selected for genomewide mapping. Among these families, 39 had two affected siblings, 16 had three affected siblings, and 5 had four affected siblings. Only affected siblings and parents (when available) were ge-

notyped, except in nine families for whom one or two unaffected siblings were genotyped to increase the marker informativeness. Both parents were available in 19 families; only one parent was available in 20 families; and neither parent was available in 21 families. Altogether, 9 parents were affected, thereby resulting, together with 4 other relatives affected, in 159 patients with CD studied genomewide.

For the follow-up study of the three most interesting regions, 38 additional families with multiple affected siblings—28 families with two affected siblings, 8 families with three affected siblings, and 2 families with four affected siblings—were studied. Both parents were available in 23 families, one parent was available in 4 families, and neither parent was available in 11 families. Altogether, 9 parents were affected, thereby resulting in 97 patients studied in these additional families.

For HLA-DQB1 alleles, all family members were genotyped in 98 families. One to seven healthy siblings were available in 54 families, thereby resulting in a total of 260 affected and 237 healthy family members genotyped.

Genotyping

A genomewide screen was performed with 352 microsatellite markers with an average marker space of 9.6 cM. All markers were screened by a semiautomatic fluorescence-labeled genotyping system (Reed et al. 1994; Ghosh et al. 1997). In brief, multiplex PCR was performed in 384-well PCR plates (Marsh) and PTC 225 thermocyclers (MJ Research) with a total volume of 10 μl containing 50 ng of genomic DNA, 0.15–0.2 mM $MgCl_2$, 0.2 mM dNTPs, and 0.5 units of *Taq* Platenum polymerase (Gibco Life Tech). On average, each multiplex PCR contains four markers with various amounts (1–50 pmol) of primer, to achieve roughly even amplification of each marker. PCR products from two to three multiplex PCR reactions were pooled and then genotyped by 377 DNA sequencers and the GENESCAN 2.0/GENOTYPER 1.1.1 software (PE Applied Biosystems). Before being imported into the genotyping database LABMAN (Adams 1994) for allele binning and Mendelization checking, the computer-generated genotypes were checked by two independent researchers who are blind to disease phenotypes.

Genotyping data were checked for Mendelian inconsistency, by the PedCheck program (O'Connell and Weeks 1998). Any marker violating the rules of Mendelian transmission was double checked by the genotyping laboratory. Ambiguous marker genotypes were deleted to reduce further problems occurring during statistical analysis. The biological relationships were also examined by the RelCheck computer program (Broman and Weber 1998). The program uses the genotypic information from autosomal markers in the genome scan and calculates the likelihoods to infer the relationship of a putative sib pair. Five relationships—MZ twins, parental/offspring, full siblings, half siblings, and unrelated individuals—were considered.

The HLA-DQB1 alleles were genotyped from all available family members by use of a DQ SSP "Low Resolution" kit (Dynal AS). Major classes of the HLA-DQB1 alleles—including the known risk alleles DQB1*02 (DQ2) and DQB1*0302 (DQ8)—could be determined. On the basis of our unpublished results, ~90% of DQB1*02 positive individuals in Finnish families with CD have the DQA1*05 DQB1*02 haplotype.

Linkage Analysis

Allele frequencies from all individuals genotyped were estimated by the GCONVERT program. Together with the likelihood-based analysis that was used, this conservative estimation of frequencies diminishes the risk of type I error due to lack of parental genotypes in part of the families (Göring and Terwilliger 2000*a*). Analyses were performed by a pseudomarker approach (Göring and Terwilliger 2000*a*, 2000*b*), which approximates a "model-free" affected-relative-pair analysis, but maintains an important property of LOD-score analysis—that is, that pedigree correlations between all relatives are considered jointly, and the pedigree is not broken into a set of all possible relative pairs. Since the actual mode of inheritance for CD is unknown, both dominant and recessive models were applied to our analyses. Two dominant models (one with high penetrance and the other with very low penetrance; neither allowing for phenocopies; and each assuming the disease allele to be infinitesimally rare, according to the pseudomarker strategy [Göring and Terwilliger 2000*b*]) were performed. Similarly, the same parameters and assumptions were used in the two recessive models. Because of com-

Table 1

Six Markers Reaching LOD Scores >1 in Genomewide Analysis of 60 Families with CD

	LOD Score Assuming			
	Dominant Inheritance		Recessive Inheritance	
Marker	High Penetrance	Low Penetrance	High Penetrance	Low Penetrance
D1S3669	1.09	0	.16	.09
D4S2639	1.25	2.11	.97	1.19
D5S816	1.72	.97	1.73	.43
D7S821	1.02	.66	.35	.40
D9S741	.65	.50	1.08	1.11
D16S3253	.21	.43	.93	1.40

Note.—Two-point analyses with four models—under the assumption of a risk allele with either high or low penetrances and either dominant or recessive inheritance—were performed.

putational restrictions, multipoint analyses were conducted by joining the meiotic information from two neighboring markers at one time (i.e., by three-point analysis). Two-point and multipoint pseudomarker analyses were performed by the FASTLINK program (Cottingham et al. 1993). In multipoint analyses, we used only the genetic models that yielded the highest LOD scores in two-point analysis.

Results

Two-Point Genomewide Analysis of 60 Families

The only highly significant finding over all models tested was a LOD score >5 for marker D6s1281, which maps adjacent to the HLA gene region. Six other markers—D1s3669, D4s2639, D5s816, D7s821, D9s741, and D16s3253—yielded LOD scores >1 (table 1). The maximum LOD scores over the four tested models were 1.09 for D1s3669 and 1.02 for D7s821, under the assumption of a high-penetrance gene with dominant inheritance, and 2.11 for D4s2639, under the assumption of a dominant, low-penetrance gene. The best LOD scores under the assumption of a recessive inheritance were 1.73 for D5s816, for a high-penetrance gene, and 1.11 for D9s741 and 1.40 for D16s3253, assuming a low-penetrance gene.

Follow-Up Analysis with Additional Markers and 38 Families

Three of the linked regions—chromosomes 4p15, 5q31, and 7q21—were selected for follow-up analysis based on statistical evidence for linkage to CD and/or based on independent evidence for linkage in earlier genome scans. In the follow-up study, additional adjacent markers spanning the implicated genetic regions were genotyped in the original 60 families and in an additional 38 Finnish families. The follow-up linkage analyses were performed based on the same model that yielded the best LOD scores in the genome scan. The results of the follow-up analyses are summarized in table 2. Two-point analysis of the original 60 families with the new markers provided consistent evidence for linkage for some of the adjacent markers. The multipoint analysis with dense markers gave stronger evidence for linkage in the 4p region, with a LOD score of 3.25, and consistent evidence for linkage in the chromosomes 5q and 7q, with LOD scores of 1.49 and 1.04, respectively. Linkage analysis of the 38 new families alone did not provide supporting evidence for any of the three regions (data not shown), which is not surprising considering the smaller sample. Linkage analysis of the combined 98 families did not markedly strengthen evidence for linkage in any of three regions.

Table 2

Two-Point and Multipoint LOD Scores of the Follow-Up Analysis of Three Regions

Marker[a]	Map Position[b] (cM)	Initial 60 Families		All 98 Families[c]	
		Two-Point LOD Score	Multipoint LOD Score[d]	Two-Point LOD Score	Multipoint LOD Score[d]
D4s2960	29.1	0		0	
D4s2639	29.8	2.11	1.05	1.33	.63
D4s2633	33.4	2.14	3.25	1.78	2.40
D4s3017	35.6	.68	1.38	.19	.69
D4s2397	42.7	.78	1.61	.37	.34
D5s642	134.7	.08		.09	
D5s1995	138.0	.33	.68	.25	.42
D5s2115	138.6	.43	.55	.36	.29
D5s396	139.3	.34	.46	.15	.10
D5s458	139.3	.33	.55	.17	.27
D5s816	139.3	1.72	1.49	1.81	1.00
D5s500	140.7	1.56	1.50	1.62	1.52
D5s1480	147.5	.46	1.20	.21	1.11
D7s479	109.1	.0		.03	
D7s821	109.1	1.01	.39	.84	1.02
D7s2480	111.3	.96	1.04	.73	1.30
D7s1799	113.9	.51	.89	.63	.83

[a] Underlined markers are those used in the initial genomewide analysis.

[b] Map distances are based on the genetic map from Marshfield Genetic Laboratory (Center for Medical Genetics, Marshfield Medical Research Foundation).

[c] Ninety-nine families were analyzed for the markers from 5q region.

[d] Three-point analysis was performed using the current marker and that above it.

Table 3

CD Prevalence in Relation to HLA Risk Status among 497 Family Members in 98 Families

DQ Status	No. Affected/No. Unaffected (CD Prevalence)
DQ2+:	
Heterozygous	175/117 (60%)
Homozygous	77/20 (79%)
Overall	252/137 (65%)
DQ2−/DQ8−	1/75 (1%)
DQ2−/DQ8+	7/25 (22%)
Overall	260/237 (52%)

Known Risk Alleles at HLA-DQB1, in 98 Families

HLA-DQB1 alleles were determined, in all available family members in the 98 families. The DQB1*02 allele forms part of the DQ2 heterodimer, a well-known risk factor for CD. Altogether, 497 individuals—of whom 260 were affected and 237 were unaffected—were genotyped. The distribution of the DQ risk alleles in our family sample was as expected for CD. Only one affected individual did not carry either a DQ2 (DQB1*02) or a DQ8 (DQB1*0302) risk allele. DQ2 was found in all but three of the families. Disease prevalence in relation to the DQ risk status is shown in table 3. Of the 389 DQ2-positive family members, 252 were affected, thereby resulting in a disease prevalence of 65% for all DQ2 carriers—without any correction for ascertainment. Difference in prevalence was observed between individuals carrying one copy of DQ2 and individuals carrying two copies of DQ2 (with prevalences of 60% and 79%, respectively). In contrast, prevalence did not differ between DQ2 heterozygous persons carrying either DQ8 or nonrisk DQ alleles (data not shown). Although it is interesting, the significance of these observations is not clear, because our family ascertainment plan, designed to maximize the power for detection of linkage, obviously biases estimation of prevalence.

Although linkage between chromosome 6p markers proximal to HLA and CD was definitive (LOD score >5) in the 60 families, the strength of linkage was also tested directly by assaying DQB1 alleles in the combined 98 families. Pseudomarker analysis was performed using the most obvious dominant inheritance model on the basis of DQ2 characteristics in the family sample—penetrances of 60% and 79%, for individuals who are heterozygous or homozygous carriers, respectively, for risk alleles; phenocopy rate of 0%; and a risk allele frequency of 13% in the population (Partanen and Westman 1997). As expected, highly significant evidence for linkage was observed (maximum LOD score 19.6).

Discussion

Genomewide linkage analysis was performed with 60 Finnish families with CD, and subsequent analysis of the three most interesting findings was performed, together with analysis of 38 additional families, with a denser marker array across the implicated regions. Finland has proved to be a particularly interesting study population for genetic analyses. Importantly, in the Finnish population, cultural and environmental heterogeneity is lower than it is in many other populations, and medical records are reliable and comparable. The Finnish population has a relatively homogeneous gene pool, and many studies have identified strong founder effects. Together, these characteristics have facilitated successful mapping of single-gene defects (Peltonen et al. 1999). The level of linkage disequilibrium in the general population may not, however, be significantly different from that in other populations (Laan and Paabo 1997; Eaves et al. 2000), and, in general, it is not clear whether the Finnish population offers a significant advantage in the search of complex-disease alleles. It is certainly true that, in smaller, more isolated populations, there would be a reduction of environmental and genetic (both locus and allelic) heterogeneity; however, what remains to be seen is whether this reduction in heterogeneity will be sufficient, in populations as large as that of Finland, to make feasible the identification of risk factors for complex traits. In Finland, the HLA genetic component of CD is rather simple; ~90% of patients carry the HLA-DQ2 risk molecule, and the remaining ~10% usually carry the HLA-DQ8 risk molecule. Also, most individuals in Finland constantly ingest cereals—that is, they encounter the major environmental trigger for the disease.

In view of this background of population homogeneity and founder effect, it is interesting that our study produced only moderate evidence for a non-HLA genetic component to CD. The only definitive genetic locus identified was HLA, which appears to be the major susceptibility locus, which carries the known CD risk genes DQA1 and DQB1. Six additional regions showed suggestive evidence for linkage; the most impressive region around locus D4s2623 produced a multipoint LOD score of 2.4 in the 98 families. This may result from a higher level of locus heterogeneity for non-HLA risk factors between, as well as within, the populations—even in the relatively homogeneous Finnish population. The role that non-HLA loci play in disease susceptibility may also be much more complex than that played by HLA, perhaps featuring epistatic interactions between multiple loci and possibly including diverse environmental factors.

Two-point LOD scores of 1.02–2.11 were found for non-HLA regions on chromosomes 1p36, 4p15, 5q31,

Table 4

Summary of Evidence for Linkage, for CD

	Evidence from Study (Sample)														
	Zhong et al. (1996) (40 Sib Pairs)			Greco et al. (1998) (37 or 110 Sib Pairs)			Greco et al. (2001) (182 Sib Pairs[a])			King et al. (2000) (16 Families)			Present Study (60 or 98 Families)			
Chromosome	Locus	Position (cM)	MMLS[b]	Locus	Position (cM)	MMLS[b]	Locus	Position (cM)	MMLS[b]	Locus	Position (cM)	MMLOD[c]	Locus	Position (cM)	MLS[d]	MMLS[e]
1p													D1s3669	37	1.90	
4p													D4s2633	33	2.11	3.25 (2.40)
5q	D5s211–D5s498	183–185	.92	D5s2111–D5s2006	188–198	2.0	D5s640	153	2.92				D5s816	139	1.73	1.49 (1.00)
6p	D6s259–D6s469	28–30	4.66													
6q										D6s460	90	1.2				
7q	D7s480–D7s1809	126–128	2.99										D7s2480	111	1.02	1.04 (1.30)
9p													D9s741	43	1.08	
10q										D10s597	129	1.9				
11p	D11s871–D11s1385	58	3.92							D11s935	46	1.5				
11q				D11s4142–D11s934	111–126	1.8										
15q	D15s207–D15s107	102–109	2.12													
16q										D16s3091	111	1.9	D16s3253	72	1.40	
19q	D19s418	92	1.84							D19s210	100	.9				
22(cen)	D22s420	4	2.69													

[a] Overlaps with the sample in an earlier study by Greco et al. (1998).

[b] MMLS = multipoint maximum LOD score; determined by MAPMAKER/SIBS.

[c] MMLOD = multipoint maximum LOD score; determined by VITESSE/MFLINK programs.

[d] MLS = two-point LOD scores; determined by pseudomarker analysis, by FASTLINK, of 60 families.

[e] Determined by pseudomarker analysis, by FASTLINK, of 60 or 98 (in parentheses) families.

7q21, 9p21-23, and 16q12. The highest two-point LOD score, 2.11, was generated on chromosome 4p. Regions on chromosomes 5q and 7q were close to previously implicated CD regions in independent genome scans described elsewhere (Zhong et al. 1996; Greco et al. 1998). These three regions were reanalyzed with dense marker coverage, in the original 60 families and an additional 38 families. Although multipoint analysis with dense markers yielded stronger evidence for linkage on chromosome 4p and consistent evidence for linkage on chromosomes 5q and 7q, analysis of the combined 98 families did not bolster evidence for linkage in any of the three regions. Such a finding may reflect the underlining genetic complexity (i.e., genes of small effect or complex gene-gene or gene-environmental interaction, together with nonallelic heterogeneity), or they may indicate false-positive findings. Distinction between these interpretations is the challenge posed to those who undertake complex genetic studies.

Table 4 shows the findings from this study compared with those of other previously reported linkage studies. Evidence from this study, for linkage to 5q31, maps 14 cM centromeric to the location in the Italian population described elsewhere (Greco et al. 2001). Multiple studies—including the Irish study (Zhong et al. 1996), our candidate-gene study with Finnish families (Holopainen et al. 2001), and the original genome screen reported by Greco et al. (1998)—have found hints of linkage to the 5q telomere, ~50 cM from 5q31. The 5q region is of particular relevance to immune-mediated diseases, since it harbors a cytokine-gene cluster that is involved in the T-helper-cell subset 2 type of immune response. The present study provided moderate evidence for linkage at chromosome 7q, which is ~17 cM centromeric to the locus implicated in the study by Zhong et al. (1996). It is also notable that the pathogenic deletions described for Williams syndrome, a rare disease that has been reported in association with CD (Pankau et al. 1993; Pittschieler et al. 1993; Chiaravalloti et al. 1995; Papadatou et al. 1996), span the elastin gene at 7q11.23. However, no association to the elastin-gene marker was found in a recent study (Grillo et al. 2000).

It is notable that, in the present study, no evidence for linkage to candidate regions 2q33 (*CTLA4/CD28* gene region) and 11q23, which were implicated in candidate-gene studies of 100 Finnish families (60 of whom overlap with those in the present study), was detected. Previous studies showed a nonparametric-linkage score of 2.5 for locus D2s116 and a maximum-likelihood score of 1.37 for locus D11s4142 (Holopainen et al. 1999, 2001). By comparison, denser marker sets were used in the candidate-gene studies, relative to the whole-genome scan. Several comparative studies provide examples of putative linkages that are undetected by genome scans at marker-interval spacing of $\geqslant$10 cM (Aita et al. 1999; Liu et al. 2001*a*, 2001*b*), although it is remains unclear whether the putative linkage findings from dense-marker analysis in the candidate-gene studies, or from our genome scan, are significant.

Linkage to HLA-DQ was extremely strong in this sample of Finnish families with CD. All but one patient carried the known risk factor DQ2 or DQ8. The prevalence of CD was 65% among all DQ2-positive family members, 60% for individuals carrying one copy of DQ2, and 79% for individuals carrying two copies of DQ2. The relevance of DQ2 dose effect in CD could not be evaluated in our sample, however, because of strong ascertainment bias, nonindependence of family members, and high frequency of DQ risk alleles. Methods for estimating the role that risk alleles play in disease familiarity are not, in general, applicable to common alleles but, specifically, are not applicable in the presence of strong ascertainment bias, as was the case in this study. In our sample, all but one patient carried either the DQ2 or DQ8 risk molecule and thus rendered estimation of IBD sharing meaningless, since many individuals were identical-by-state for the "disease" allele without being IBD. Estimates based solely on IBD thus will systematically underestimate the effect of this locus.

In addition to the need to continue the search for non-HLA genetic components to CD, it will be important, in future studies, to explore the relative contribution of HLA-related risk alleles to CD susceptibility in Finnish families. As the HLA gene region harbors a large number of genes involved in the immune response, HLA-linked non-DQ genes may also contribute to CD susceptibility. If not all DQ2 haplotypes confer equal risk to CD, then the prevalence of the putative high-risk DQ2 haplotypes would be higher in families with patients with CD. Indeed, recently, Lie et al. (1999), as well as our group (Karell et al., in press), reported evidence for differences in susceptibility between DQ2-positive haplotypes. Despite weak evidence for HLA-unlinked genes, the present genome scan provides a starting point for a further search for CD-susceptibility alleles in the Finnish population.

Acknowledgments

The Coeliac Disease Study Group (P.H., K.M., M.M., and J.P.) is supported by the Sigrid Juselius Foundation, the University of Helsinki, the Emil Aaltonen Foundation, the Maud Kuistila Memorial Foundation, and the Medical Research Fund of Tampere University Hospital. This study has been partially funded by the Commission of the European Communities (specific RTD program "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"). The present study does not necessarily reflect the views of the Commission and

in no way anticipates the Commission's future policy in this area. Most importantly, we thank the families who have participated in this study.

Electronic-Database Information

The accession number and URLs for data in this article are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://research.marshfieldclinic.org/genetics/

David Duffy's QIMR Homepage, http://www2.qimr.edu.au/davidD/ (for GCONVERT program)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CD [MIM 212750])

References

Adams P (1994) LABMAN and LINKMAN: a data management system specifically designed for genome searches of complex diseases. Genet Epidemiol 11:87–98

Aita VM, Liu J, Knowles JA, Terwilliger JD, Baltazar R, Grunn A, Loth JE, Kanyas K, Lerer B, Endicott J, Wang Z, Penchaszadeh G, Gilliam TC, Baron M (1999) A comprehensive linkage analysis of chromosome 21q22 supports prior evidence for a putative bipolar affective disorder locus. Am J Hum Genet 64:210–217

Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, Ferguson A, et al (1999) Contribution of the MHC region to the familial risk of coeliac disease. J Med Genet 36:687–690

Brett PM, Yiannakou JY, Morris MA, Bronson SR, Mathew C, Curtis D, Ciclitira PJ (1998) A pedigree-based linkage study of coeliac disease: failure to replicate previous positive findings. Ann Hum Genet 62:25–32

Broman KW, Weber JL (1998) Estimation of pairwise relationships in the presence of genotyping errors. Am J Hum Genet 63:1563–1564

Chiaravalloti G, Rossomando V, Quinti S, Assanta N, Ughi C, Ceccarelli M (1995) Williams-Beuren syndrome and celiac disease. Minerva Pediatr 47:43–46

Cottingham RW Jr, Idury RM, Schäffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

Eaves IA, Merriman TR, Barber RA, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Cucca F, Todd JA (2000) The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. Nat Genet 25:320–323

Ellis A (1981) Coeliac disease: previous family studies. In: McConnel RB (ed) The genetics of coeliac disease. MTB, Lancaster, pp 197–200

Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Eldridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers: FUSION (Finland-U.S. Investigation of NIDDM Genetics) Study Group. Genome Res 7:165–178

Göring HHH, Terwilliger JD (2000*a*) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. Am J Hum Genet 66:1298–1309

Göring HHH, Terwilliger JD (2000*b*) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. Am J Hum Genet 66:1310–1327

Greco L, Babron MC, Corazza GR, Percopo S, Sica R, Clot F, Fulchignoni-Lataud MC, Zavattari P, Momigliano-Richiardi P, Casari G, Gasparini P, Tosi R, Mantovani V, De Virgiliis S, Iacono G, D'Alfonso A, Selinger-Leneman H, Lemainque A, Serre JL, Clerget-Darpoux F (2001) Existence of a genetic risk factor on chromosome 5q in Italian coeliac disease families. Ann Hum Genet 65:35–41

Greco L, Corazza G, Babron MC, Clot F, Fulchignoni-Lataud MC, Percopo S, Zavattari P, et al (1998) Genome search in celiac disease. Am J Hum Genet 62:669–675

Greco L, Romino R, Coto I, Cosmo ND, Percopo S, Maglio M, Paparo F, Gasperi V, Limongelli MG, Cotichini R, D'Agate C, Tinto N, Sacchetti L, Tosi R, Stazi MA. The first large population-based twin study of coeliac disease. Gut (in press)

Grillo R, Petronzelli F, Mora B, Bonamico M, Mazzilli MC (2000) Search for coeliac disease susceptibility loci on 7q11.23 candidate region: absence of association with the ELN17 microsatellite marker. Hum Hered 50:180–183

Hervonen K, Karell K, Holopainen P, Collin P, Partanen J, Reunala T (2000) Concordance of dermatitis herpetiformis and celiac disease in monozygous twins. J Invest Dermatol 115:990–993

Holopainen P, Arvas M, Sistonen P, Mustalahti K, Collin P, Mäki M, Partanen J (1999) CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to celiac disease: a linkage and family-based association study. Tissue Antigens 53:470–475

Holopainen P, Mustalahti K, Uimari P, Collin P, Mäki M, Partanen J (2001) Candidate gene regions and genetic heterogeneity in gluten sensitivity. Gut 48:696–701

Houlston RS, Tomlinson IP, Ford D, Seal S, Marossy AM, Ferguson A, Holmes GK, Hosie KB, Howdle PD, Jewell DP, Godkin A, Kerr GD, Kumar P, Logan RF, Love AH, Johnston S, Marsh MN, Mitton S, O'Donoghue D, Roberts A, Walker-Smith JA, Stratton MF (1997) Linkage analysis of candidate regions for coeliac disease genes. Hum Mol Genet 6:1335–1339

Karell K, Holopainen P, Mustalahti K, Collin P, Mäki M, Partanen J. Not all HLA DR3 DQ2 haplotypes confer equal susceptibility to coeliac disease: transmission analysis in families. Scand J Gastroenterol (in press)

King AL, Yiannakou JY, Brett PM, Curtis D, Morris MA, Dearlove AM, Rhodes M, Rosen-Bronson S, Mathew C, Ellis HJ, Ciclitira PJ (2000) A genome-wide family-based linkage study of coeliac disease. Ann Hum Genet 64:479–490

Laan M, Paabo S (1997) Demographic history and linkage disequilibrium in human populations. Nat Genet 17:435–438

Lie BA, Sollid LM, Ascher H, Ek J, Akselsen HE, Ronningen KS, Thorsby E, Undlien DE (1999) A gene telomeric of the HLA class I region is involved in predisposition to both type 1 diabetes and coeliac disease. Tissue Antigens 54:162–168

Liu J, Juo SH, Terwilliger JD, Grunn A, Tong X, Brito M, Loth JE, Kanyas K, Lerer B, Endicott J, Penchaszadeh G, Gilliam TC, Baron M (2001*a*) A follow-up linkage study supports evidence for a bipolar affective disorder locus on chromosome 21q22. Am J Med Genet 105:189–194

Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, Geschwind D, Lord C, Iversen P, Hoh J, Ott J, Gilliam TC (2001*b*) A genomewide screen for autism susceptibility loci. Am J Hum Genet 69:327–340

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

Mäki M, Holm K, Lipsanen V, Hallstrom O, Viander M, Collin P, Savilahti E, Koskimies S (1991) Serological markers and HLA genes among healthy first-degree relatives of patients with coeliac disease. Lancet 338:1350–1353

Mearin ML, Biemond I, Pena AS, Polanco I, Vazquez C, Schreuder GT, de Vries RR, van Rood JJ (1983) HLA-DR phenotypes in Spanish coeliac children: their contribution to the understanding of the genetics of the disease. Gut 24: 532–537

Not T, Horvath K, Hill ID, Partanen J, Hammed A, Magazzu G, Fasano A (1998) Celiac disease risk in the USA: high prevalence of antiendomysium antibodies in healthy blood donors. Scand J Gastroenterol 33:494–498

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Pankau R, Partsch CJ, Gosch A, Santer R (1993) Williams-Beuren syndrome in combination with celiac disease. Monatsschr Kinderheilkd 141:577–580

Papadatou B, Digilio C, Ferretti F, Colistro F, Bella S, Orru M, Castro M, Giannotti A (1996) Celiac disease in Williams' syndrome. Paper presented at 7th International Symposium on Coeliac Disease. Tampere, September 5–7

Partanen J, Westman P (1997) Caucasian Finnish normal. In: Terasaki PI, Gjertson DW (eds) HLA 1997. UCLA Tissue Typing Laboratory, Los Angeles, pp 214–215

Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. Hum Mol Genet 8:1913–1923

Petronzelli F, Bonamico M, Ferrante P, Grillo R, Mora B, Mariani P, Apollonio I, Gemme G, Mazzilli MC (1997) Genetic contribution of the HLA region to the familial clustering of coeliac disease. Ann Hum Genet 61:307–317

Pittschieler K, Morini G, Crepaz R (1993) Williams syndrome and coeliac disease. Acta Paediatr 82:iv

Polanco I, Biemond I, van Leeuwen A, Schreuder I, Kahn PM, Guerrero J, D'Amaro J, Vasquez C, van Rood JJ, Pena AS (1981) Gluten sensitive enteropathy in Spain: genetic and environmental factors. In: McConnel RB, Lancaster RB (eds) The genetics of coeliac disease. MTP, pp 211–231

Polvi A, Arranz E, Fernandez-Arquero M, Collin P, Mäki M, Sanz A, Calvo C, Maluenda C, Westman P, de la Concha EG, Partanen J (1998) HLA-DQ2-negative celiac disease in Finland and Spain. Hum Immunol 59:169–175

Reed PW, Davies JL, Copeman JB, Bennett ST, Palmer SM, Pritchard LE, Gough SC, Kawaguchi Y, Cordell HJ, Balfour KM, Jenkins SC, Powell EE, Vignal A, Todd JA (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi- automated genome mapping. Nat Genet 7:390–395

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

Rotter JI, Landaw EM (1984) Measuring the genetic contribution of a single locus to a multilocus disease. Clin Genet 26:529–542

Sollid LM (2000) Molecular basis of celiac disease. Annu Rev Immunol 18:53–81

Spurkland A, Sollid LM, Polanco I, Vartdal F, Thorsby E (1992) HLA-DR and -DQ genotypes of celiac disease patients serologically typed to be non-DR3 or non-DR5/7. Hum Immunol 35:188–192

Spurkland A, Sollid LM, Ronningen KS, Bosnes V, Ek J, Vartdal F, Thorsby E (1990) Susceptibility to develop celiac disease is primarily associated with HLA-DQ alleles. Hum Immunol 29:157–165

Torinsson Naluai Å, Nilsson S, Samuelsson L, Gudjónsdóttir AH, Ascher H, Ek J, Hallberg B, Kristiansson B, Martinsson T, Nerman O, Sollid LM, Wahlström J (2000) The CTLA4/CD28 gene region on chromosome 2q33 confers susceptibility to celiac disease in a way possibly distinct from that of type 1 diabetes and other chronic inflammatory disorders. Tissue Antigens 56:350–355

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