JGH Open Open Access An open access journal of gastroenterology and hepatology

JGHF Journal of Gastroenterology and Hepatology Foundation

doi:10.1002/jgh3.12651

ORIGINAL ARTICLE

Human Leukocyte Antigen DQ (HLA-DQ) genotypes and haplotypes and their association with phenotype in patients with celiac disease in India

Balakrishnan S Ramakrishna,* Giriprasad Venugopal,* Alka Singh,† Srinivasan Pugazhendhi,‡ Sangitanjan Dutta,§ Vineet Ahuja and Govind K Makharia

*Institute of Gastroenterology, SRM Institutes for Medical Science, Chennai, †Department of Gastroenterology and Human Nutrition, All India Institute for Medical Sciences, New Delhi, §Department of Medicine, Guwahati Medical College, Guwahati, India and ‡Department of Cancer Biology, Kansas University Medical Center, Kansas, Kankas, USA

Key words
celiac disease, diarrhea and malabsorption, gastroenterology, genetics, intestinal disorders.

Accepted for publication 21 August 2021.

Correspondence
Balakrishnan S Ramakrishna, SRM Institutes for Medical Science, 1 Jawaharlal Nehru Road, Vadapalani, Tamil Nadu, Chennai 600026, India.
Email: wurama@hotmail.com

Declaration of conflict of interest: None of the authors has any conflict of interest to disclose.

Funding support: Department of BiotechnologyBT/MED/12/SP17783/2016

Abstract

Background and Aim: Human Leukocyte Antigen DQ (HLA-DQ) genotypes play a permissive role in the genesis of celiac disease (CeD). In this case–control study, we used next-generation sequencing to determine HLA-DQA1 and ~DQB1 genotypes and haplotypes associated with CeD in Indian patients.

Methods: HLA-DQA1 and ~DQB1 loci were amplified, using long-range polymerase chain reaction (PCR), from DNA of 259 patients with symptomatic CeD (160 typical and 99 atypical), 45 asymptomatic CeD, 96 potential CeD, and 300 healthy adults. Amplicons were fragmented and sequenced on the Illumina platform, and alleles and haplotypes were assigned by matching against the HLA-international ImMunoGeneTics (IMGT) database.

Results: HLA-DQA1*05:01 (odds ratio [OR] 8.39, 95% confidence interval [CI] 5.64–12.47) and HLA-DQB1*02:01 (OR 8.59, 95% CI 5.75–12.83) were the genotypes that showed a risk association with symptomatic CeD. Among the haplotypes, HLA-DQA1*05:01 ~ HLA-DQB1*02:01 (OR 8.56, 95% CI 5.67–13.19) showed a strong risk association with symptomatic CeD. When comparing symptomatic CeD with subclinical forms (asymptomatic and potential) CeD, HLA-DQA1*05:01 ~ HLA-DQB1*02:01 (OR 2.34, 95% CI 1.61–3.43) was significantly associated with risk of symptomatic disease. The strength of association between the HLA-DQA1*05:01 ~ HLA-DQB1*02:01 haplotype and the CeD phenotype showed a gradient in the order typical > atypical > asymptomatic > potential CeD. Genotypes consistent with expression of HLA DQ2 and/or 8 were noted in 128 (80%) typical, 73 atypical (74%), 27 (60%) asymptomatic, and 52 (54%) potential CeD participants.

Conclusion: HLA-DQA1*05:01 ~ HLA-DQB1*02:01 (haplotype DQ2.5) showed a very strong risk association with symptomatic CeD in Indian patients. The strength of association showed a gradient of increase from potential to typical CeD, coinciding with a phenotypic change in the celiac iceberg.

Introduction

Celiac disease (CeD), characterized by gluten intolerance, has a strong genetic influence, occurring only in individuals with a specific HLA background. Gluten peptides, produced by digestion in the intestine, are presented to CD4 positive-T cells by Human Leukocyte Antigen DQ (HLA-DQ) proteins on the surface of antigen-presenting cells. The DQ2 protein is heterodimeric and composed of alpha- and beta-subunits, the nature of which is determined by the sequence of the genes at HLA-DQA1 and -DQB1 loci. It is believed that CeD occurs only in individuals expressing HLA-DQ2 and/or -DQ8 antigens on the surface of antigen-presenting cells.[1] In wheat-eating populations, the prevalence of CeD is proportional to the population prevalence of HLA-DQ2 and -DQ8.[2] It is believed that CeD does not occur in the absence of HLA-DQ2 or -DQ8 expression.[3] Testing for HLA-DQ2 and -DQ8 is included as a second-line investigation in the diagnostic algorithm for CeD.[4]

While the HLA-DQ type was previously determined serologically using antibodies, it is now customary to genotype HLA-DQB1 or the HLA-DQA1 ~ DQB1 loci and use that information to assign the serotype by imputation. HLA-DQ heterodimer (HLA-DQA1 ~ DQB1) typing has been used to assess CeD risk in a large sample of patients with CeD.[5] The

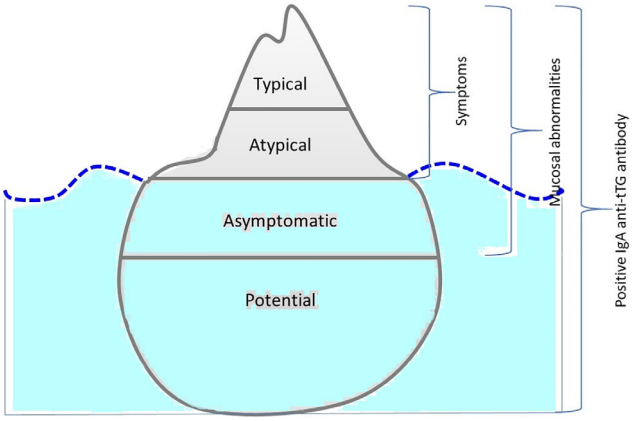

Figure 1 The celiac iceberg showing the characteristics of each phenotype of celiac disease.

highest risk for CeD was associated with HLA-DQ2.5, which is formed by the *HLA-DQA1*05:XX ~ DQB1*02:01*.[5–8] The HLA DQ2.2 heterodimer is expressed by *HLA-DQA1*02–01 ~ DQB1*02:02*, while HLA-DQ8 is expressed by *HLA-DQA1*03:01 ~ DQB1*03:02*.[5] The absence of the above three heterodimers has a high negative predictive value for the diagnosis of CeD.

CeD has been called an iceberg disease, with four-layered phenotypic forms of the disease (Fig. 1). Symptomatic patients comprise the visible portion of the iceberg, and consist of patients with classical symptoms of diarrhea who are labeled "typical" CeD as well as patients with non-classical symptoms, including stunting, metabolic bone disease, and infertility who are labeled "atypical." Patients with asymptomatic and potential CeD constitute the invisible portion of the iceberg. All CeD phenotypes have detectable IgA or IgG anti-tTG antibodies with certain exceptions. The first three groups—typical, atypical, and asymptomatic—have mucosal damage as assessed on duodenal biopsy. The first two groups—typical and atypical—have symptoms. These terms are well defined in the Oslo statements on CeD.[9]

We have previously shown a moderately high prevalence of CeD, in India.[10] In the present study, we evaluated the frequency of the HLA-DQA1 and ~ DQB1 alleles and haplotypes using next-generation sequencing (NGS) in healthy controls and in patients with typical, atypical, asymptomatic, and potential CeD.

Methods

Participants. Three hundred healthy adults aged 18 years or more, who had participated in a previous multicenter population-based study on CeD,[10] were included as healthy controls. Participants were included if they did not report a history of gastrointestinal or chronic systemic illness, and if they had a negative test for IgA antibody to tissue transglutaminase (IgA anti-tTG) (Aeskulisa Celichek tTg-A New Generation kits, Catalog No. 3503, Aesku Diagnostics Gmbh, Wendelsheim, Germany).

The symptomatic CeD cohort comprised 259 patients with symptomatic CeD recruited in the Gastroenterology outpatient clinics of the All India Institute of Medical Sciences and the SRM Institutes for Medical Science. Inclusion criteria for this cohort were an age of 18 years or older, symptoms consistent with CeD, presence of a positive test for IgA anti-tTG antibodies, the presence of histological changes of modified Marsh grade 2 or greater[11] on deep duodenal biopsy, and a clinical response to gluten-free diet. This cohort included 160 patients with typical CeD whose symptoms were predominantly gastrointestinal, and 99 patients with atypical CeD with minimal gastrointestinal symptoms who were referred from other departments for exclusion of CeD. In addition, included 45 participants with asymptomatic CeD and 96 participants with potential CeD. These latter two groups were identified during the course of previous population-based studies[10,12] on the basis of positive anti-tTG serology. Those with asymptomatic CeD had duodenal biopsy changes consistent with CeD while the potential CeD participants did not have biopsy changes.

Samples of peripheral venous blood were obtained from all participants, DNA was isolated from whole blood by salting out procedure, quantitated using fluorescence (Qubit DS DNA BR Assay Kit) on a Qubit 2.0 fluorometer (Invitrogen, Bangalore, India) or by spectrometry (NanoDrop 2000, Thermo Scientific, Wilmington, DE, USA), and stored at −20°C until analysis.

HLA high resolution genotyping using NGS. HLA locus-specific amplification was performed using NGSgo-AmpX reagents (Catalog No. 2841102 and 2841502, Genome Diagnostics BV, Utrecht, The Netherlands) and LongRange polymerase chain reaction (PCR) reagents (Catalog No. 206403, Qiagen, Hilden, Germany). HLA-DQA1 and -DQB1 genes were amplified by PCR using a kit for HLA–DQA1 and HLA-DQB1 (Catalog No. 2841102 and 2841502, Genome Diagnostics BV). This kit amplified exons 1, 2, 3, and 4 of the DQA1 gene and exons 2, 3, and 4 of DQB1 gene with amplicon sizes of 5.4 to 5.8 kb for DQA1 and 3.7 to 4.1 kb for DQB1. A total of 25 μL of the final reaction mix for DQA1 amplification contained 2.5 μL LongRange PCR buffer with Mg^{2+}, 10x, 1.25 μL dNTP mix (10 mM each), 1 μL DQA1 amplification primer, 0.4 μL LongRange PCR enzyme mix, 14.85 μL nuclease-free H_2O, and 5 μL template DNA. The PCR reaction mix for DQB1 PCR condition contained 2.5 μL LongRange PCR buffer with Mg^{2+}, 10x, 1.25 μL of dNTP mix (10 mM each), 1 μL of DQB1 amplification primer, 0.4 μL of LongRange PCR enzyme mix, 14.85 μL of nuclease-free H_2O, and 5 μL of template DNA. The PCR amplification protocol consisted of initial denaturation at 95°C for 3 min, followed by 35 cycles of denaturation, annealing, and elongation at 95°C for 15 s, 65°C for 30 s, 72°C for 6 min, respectively, and final elongation at 68°C for 10 min. The PCR products were subjected to 1% agarose gel electrophoresis for examination of amplicon size and then purified using a (QiagenPure Link PCR Purification Kit, ThermoFisher Scientific, Waltham, MA, USA). The purified amplicons were quantitated with a Qubit fluorometer and processed further to generate libraries for sequencing.

Library preparation. The amplicons were pooled together in equimolar concentration and library prepared using ~250 ng of the pooled DNA. Dual index libraries were prepared using NGSgo-LibrX kits (Catalog No. 2342805, Genome Diagnostics BV). Briefly, the amplicons were fragmented by enzymatic digestion, and subjected to end repair and tailing with dA-tail followed by ligation of adapter sequences. Adapter ligated

fragments were cleaned up using SPRI beads, and the clean fragments were indexed using limited cycle PCR to generate final libraries for paired-end sequencing.

Sequencing and analysis. The final libraries were sequenced on Illumina HiSeq X to generate 2 × 150 bp sequence reads at 100× sequencing depth. The generated sequence data were analyzed after necessary quality control for variant calling and annotation. At least 75% of the sequenced bases were of Q30 value. The raw sequence reads were aligned with a FASTA-like algorithm using NGSengine Version 2.20.2.11108. Instead of aligning the reads to a single reference, the reads were aligned to all the alleles in the library. The IMGT/HLA database (www.imgt.org) was used as the source of HLA alleles. The minimum overall alignment of each locus was set at 85%, and the maximum noise at 10%.

Homozygosity or heterozygosity was first defined, and heterozygous positions were evaluated to determine the cis-trans relationship between the bases. All alleles in the library were compared against the possible haplotypes defined above. The number of nucleotide mismatches with each allele was determined and a list of alleles with the least mismatches was generated. This list was used to generate a list of possible genotypes, and the mismatch level of the two alleles with the phased data was determined for each of the potential genotypes. The genotype with the lowest mismatch count was reported. The use of NGS permitted a high degree of resolution (6–8 digit) in the genotyping process.

Sample size calculations. Using the general population prevalence of DQ2 of 27% noted in an earlier study,[10] we calculated that a sample size of 250 patients and 250 controls would allow us to reject the null hypothesis for an odds ratio of 2, with type I error of 0.05 and study power of 90%.

Statistics. Online tools available at www.imgt.org, https://hla-net.eu/tools/, and www.allelefrequencies.net were used to calculate frequencies of the different HLA genotypes, haplotypes, and serotypes. The case–control data analyses were done using the BIGDAWG pipeline,[13] and the haplotype associations were derived from this. The primary analysis compared healthy controls and patients with symptomatic (typical and atypical) CeD. In further analyses, we compared patients with symptomatic CeD and those with subclinical (asymptomatic and potential) CeD.

Ethics. The study protocol, patient information sheet, and consent forms were approved by the Institutional Ethics Committees of the SRM Institutes for Medical Science (SIMS/IEC/003/2016) and the All India Institute of Medical Sciences (AIIMS/IEC-324/01.07.2016). All participants provided informed written consent for providing relevant health data and blood samples and for the use of their blood samples for these studies.

Results

Three hundred healthy controls (150 male) ranging in age from 18 to 68 years (median 39 years) were included in the study. The case group comprised 259 patients with symptomatic CeD (160 typical CeD and 99 atypical CeD), 45 patients with asymptomatic CeD, and 96 individuals with potential CeD. Their demographic characteristics are shown in Table 1).

HLA-DQA1 and -DQB1 alleles. Following usual convention, the HLA genotypes in the published tables are shown only at four-digit resolution. The high-resolution data (up to 8 digits in some instances) is available on request from the corresponding author. Table 2 shows the association of HLA-DQA1 and ~ DQB1 alleles with symptomatic CeD. Among the HLA-DQA1 alleles, ~DQA1*05:01 showed the only predisposing association with CeD with an odds ratio of 8.39 (P = 6.31e-33). Among the HLA-DQB1 alleles, ~DQB1*02:01 (which codes for the DQ2 antigen) was very strongly associated with CeD with an odds ratio of 8.59 (P = 3.54e-33). Another predisposing allele, ~DQB1*03:02, which codes for the DQ8 antigen, did not show a statistically significant association (odds ratio [OR] 1.60, 95% confidence interval [CI] 1.02–2.50, P = 0.054) with CeD. The latter allele is of interest as it codes for the DQ8 antigen.

HLA-DQ heterodimer haplotypes. Specific heterodimer haplotypes that is combinations of HLA-DQA1 and -DQB1 alleles, and their association with CeD are shown in Table 3. The haplotype *HLA-DQA1*05:01 ~ DQB1*02:01* (colloquially termed haplotype DQ2.5) showed a strong association with CeD with an odds ratio of 8.56 (P < 2.22e-16). Among the two other haplotypes that are reported to be associated with CeD in other parts of the world, *HLA-DQA1*02:01 ~ DQB1*02:02* (colloquially termed DQ2.2), and *HLA-DQA1*03:01 ~ DQB1*03:02* (colloquially termed DQ8), did not show an association with CeD in this study (P=0.252 and 0.062), respectively.

Haplotype-phenotype association in CeD. Table 4 compares haplotype distribution between patients with symptomatic CeD and individuals with subclinical CeD (asymptomatic and potential) picked up by community screening. *HLA-DQA1*05:01 ~ DQB1*02:01* was significantly associated with symptomatic CeD compared to subclinical CeD. On the other hand, *HLA-DQA1*06:01 ~ DQB1*03:01*, *HLA-DQA1*01:02 ~ DQB1*06:01*, *HLA-DQA1*01:03 ~ DQB1*06:01*, and *HLA-DQA1*01:02 ~ DQB1*06:04* showed significant protective associations with symptomatic CeD when compared to subclinical CeD.

CeD is an iceberg disease ranging from potential CeD to typical CeD. Table 5 compares the association of each of the four phenotypes of CeD with HLA-DQ heterodimer haplotypes. As seen in Table 5 and in Figure 2, the strength of association of *HLA-DQA1*05:01 ~ DQB1*02:01* (also known as haplotype DQ2.5) with CeD phenotype showed a gradient in the direction typical > atypical > asymptomatic > potential. Figure 2 contrasts the haplotype-phenotype association observed in the case of DQ2.5 with the lack of association in the case of DQ2.2 and DQ8, which are the other haplotypes that are generally associated with CeD genesis.

HLA-DQ2 and/or DQ8 antigen expression. The HLA-DQ antigen makeup of the individual participants was derived by imputation from the DQB1 genotype. Genotypes consistent with expression of HLA DQ2 and/or 8 were noted in 128 of 160 (80%) patients with typical CeD, 73 of 99 patients (74%) with atypical CeD, 27 of 45 (60%) individuals with asymptomatic CeD, and 52 of 96 (54%) participants with potential CeD.

Table 1 Demographic characteristics of participants

		Celiac disease			
		Symptomatic (259)			
	Controls	Typical	Atypical	Asymptomatic	Potential
Number	300	160	99	45	96
Age	18–68	18–63	18–75	18–65	18–75
Sex (male:female)	150:150	102:58	69:30	33:12	50:46
Biopsy					
Marsh 2		12	13	9	—
Marsh 3a		30	30	11	—
Marsh 3b		57	26	13	—
Marsh 3c		61	30	12	—

Table 2 HLA-DQA1 and DQB1 allele associations with symptomatic celiac disease

	Frequency			95% confidence interval		
Allele	Cases	Controls	Odds ratio	Lower	Upper	*P* value
HLA-DQA1 alleles						
DQA1*01:01	0.0135	0.0833	0.1507	0.0677	0.3354	1.24 e-07
DQA1*01:02	0.0830	0.1333	0.5884	0.3980	0.8699	0.0117
DQA1*01:03	0.1448	0.1717	0.8169	0.5909	1.1294	0.2870
DQA1*01:04	0.0714	0.1500	0.4359	0.2915	0.6519	1.11 e-04
DQA1*01:05	0.0328	0.0700	0.4508	0.2534	0.8021	0.0117
DQA1*02:01	0.1390	0.1433	0.9649	0.6883	1.3526	0.8636
DQA1*03:01	0.0907	0.0617	1.5184	0.9703	2.3761	0.0926
DQA1*05:01	0.3282	0.0550	8.3924	5.6488	12.4716	6.31 e-33
HLA-DQB1 alleles						
DQB1*02:01	0.3263	0.0533	8.5953	5.7576	12.8315	3.44 e-33
DQB1*02:02	0.1120	0.0933	1.2248	0.8313	1.8048	0.3225
DQB1*03:01	0.0676	0.1167	0.5487	0.3590	0.8384	0.0086
DQB1*03:02	0.0927	0.0600	1.6000	1.0211	2.5071	0.0547
DQB1*05:01	0.0444	0.1500	0.2633	0.1638	0.4231	8.44 e-09
DQB1*05:03	0.0734	0.1550	0.4316	0.2900	0.6422	6.23 e-05
DQB1*06:01	0.0888	0.1583	0.5181	0.3565	0.7528	0.0011
DQB1*06:03	0.0695	0.0433	1.6489	0.9815	2.7701	0.0758

Two hundred fifty-nine patients with symptomatic celiac disease were compared with 300 healthy controls. Comparisons were made using the BIGDAWG program and *P* values were corrected for multiple comparisons.

Discussion

In the present study, we document the HLA-DQA1 and ~ DQB1 genotypes in healthy adults and in patients with CeD in India. In this population, the *HLA-DQA1*05:01 ~ DQB1*02:01* (DQ2.5) haplotype showed a very strong association with CeD. We identified a gradient of association for this haplotype with the phenotype of CeD as the phenotype goes from the typical form to the disease to the individuals who are classified as having potential CeD. Twenty percent of patients with typical CeD lacked the HLA-DQ makeup that allows expression of the HLA antigens DQ2 or DQ8. These findings have implications for diagnosis and practice.

The HLA system accounts for 40% of the heritable influence on CeD, and the Human Genome Organization (HUGO) Gene Nomenclature Committee has named the HLA-DQA1 and -DQB1 class II HLA genes as CELIAC1 locus.[6] HLA typing has significantly advanced recently due to the introduction of NGS resulting in high resolution genotyping at these loci. This is the first report of detailed genotyping at the HLA-DQA1 and DQB1 loci in healthy controls and patients with CeD and is of interest because NGS provided a complete genotype profile at these loci unlike earlier studies where DQ2 and DQ8 alone were genotyped, but not the other DQ determinants. The method used allowed for detection of new alleles, and allowed assignment of the homozygosity and heterozygosity of CeD risk haplotype, genotype, and serotype for each participant.

The prevalence of CeD in a population is dependent both on the amount of wheat consumed and on the prevalence of HLA Class II DQ2 and/or DQ8 genotypes.[14,15] Thus, the prevalence of CeD is greater than 1 % in Europe and North America where approximately one-third of the general population has

Table 3 HLA-DQ heterodimer haplotype associations of symptomatic celiac disease

	Frequency			95% CI		
DQA1-DQB1 haplotype	Cases	Controls	Odds ratio	Lower	Upper	*P* value
01:01 ~ 05:01	0.0115	0.0766	0.14	0.05	0.34	2.808 e-07
01:01 ~ 05:02	0.0019	0.0066	0.29	0.01	2.94	0.23864
01:02 ~ 05:02	0.0289	0.05	0.57	0.28	1.11	0.076877
01:02 ~ 06:01	0.0193	0.0433	0.44	0.19	0.95	0.024112
01:02 ~ 06:02	0.0231	0.0083	2.84	0.92	10.33	0.042251
01:02 ~ 06:04	0.0057	0.0166	0.35	0.06	1.35	0.092353
01:02 ~ 06:09	0.0019	0.0133	0.14	0	1.08	0.033935
01:03 ~ 05:03	0.0057	0.015	0.38	0.07	1.55	0.13833
01:03 ~ 06:01	0.0694	0.115	0.58	0.37	0.9	0.010017
01:03 ~ 06:03	0.0675	0.0416	1.67	0.96	2.96	0.052979
01:04 ~ 05:03	0.0675	0.14	0.45	0.29	0.69	0.00010087
01:05 ~ 05:01	0.0328	0.07	0.45	0.24	0.83	0.0058808
02:01 ~ 02:02	0.1100	0.09	1.26	0.83	1.9	0.25292
02:01 ~ 03:03	0.0289	0.0533	0.53	0.26	1.03	0.044545
03:01 ~ 03:02	0.0888	0.06	1.53	0.95	2.49	0.062291
03:02 ~ 03:03	0.0077	0.0033	2.34	0.33	25.93	0.31371
04:01 ~ 04:02	0.0077	0.005	1.56	0.26	10.67	0.56077
05:01 ~ 02:01	0.3243	0.0533	8.56	5.67	13.19	<2.22 e-16
05:05 ~ 03.01	0.0405	0.0383	1.07	0.55	2.04	0.83742
05:09 ~ 03:01	0.0193	0.0266	0.72	0.29	1.71	0.42234
06:01 ~ 03:01	0.0057	0.0416	0.13	0.03	0.45	0.00013576
Binned			0.86	0.42	1.69	0.63126

Two hundred fifty-nine patients with symptomatic celiac disease compared with 300 healthy controls, using the BIGDAWG program. *P* values were corrected for multiple comparisons.

Table 4 HLA-DQ heterodimer haplotypes

	Frequency		Odds ratio	95% CI		
DQA1-DQB1 haplotype	Cases	Controls	Odds ratio	Lower	Upper	*P* value
01:01 ~ 05:01	0.0115	0.0248	0.46	0.13	1.62	0.15707
01:02 ~ 05:02	0.0289	0.0460	0.62	0.27	1.43	0.20753
01:02 ~ 06:01	0.0193	0.0496	0.38	0.15	0.93	0.015245
01:02 ~ 06:02	0.0231	0.0177	1.31	0.43	4.81	0.61054
01:02 ~ 06:04	0.0057	0.0212	0.27	0.04	1.27	0.047257
01:03 ~ 05:03	0.0057	0.0070	0.82	0.09	9.82	0.82352
01:03 ~ 06:01	0.0694	0.1347	0.48	0.29	0.8	0.002394
01:03 ~ 06:03	0.0675	0.0602	1.13	0.6	2.19	0.68971
01:04 ~ 05:03	0.0675	0.0815	0.82	0.46	1.48	0.46591
01:05 ~ 05:01	0.0328	0.0531	0.6	0.28	1.32	0.16006
02:01 ~ 02:02	0.1100	0.0886	1.27	0.76	2.18	0.3407
02:01 ~ 03:03	0.0289	0.0354	0.81	0.34	2.05	0.6135
03:01 ~ 03:02	0.0888	0.0638	1.43	0.79	2.67	0.21354
03:02 ~ 03:03	0.0077	0.0	Inf	0.5	Inf	0.097918
04:01 ~ 04:02	0.0077	0.0106	0.72	0.12	4.98	0.6722
05:01 ~ 02:01	0.3243	0.1702	2.34	1.61	3.43	2.7229 e-06
05:05 ~ 03.01	0.0405	0.0212	1.94	0.75	5.95	0.14945
05:09 ~ 03:01	0.0193	0.0248	0.77	0.26	2.42	0.60516
06:01 ~ 03:01	0.0057	0.0567	0.1	0.02	0.34	6.1523 e-06
Binned			1.03	0.42	2.66	0.9451

Comparison of 259 patients with CeD and 141 participants with subclinical (asymptomatic or potential) celiac disease. These comparisons were obtained using the BIGDAWG program.

Table 5 Haplotype-phenotype association in celiac disease

DQA1 ~ DQB1 haplotype		Celiac disease phenotype			
		Typical	Atypical	Asymptomatic	Potential
01:01 ~ 05:01	OR (95% CI)	0.08 (0.01–0.29)	0.25 (0.06–0.69)		0.26 (0.07–0.72)
	P value	4.77e-06	0.0044		0.0056
01:03 ~ 06:01	OR (95% CI)	0.54 (0.31–0.91)			
	P value	0.0163			
01:03 ~ 06:03	OR (95% CI)	2.03 (1.11–3.74)			
	P value	0.0124			
01:04 ~ 05:03	OR (95% CI)	0.41 (0.23–0.69)	0.5 (0.26–0.91)		0.45 (0.22–0.83)
	P value	0.0004	0.0174		0.0078
01:05 ~ 05:01	OR (95% CI)	0.38 (0.16–0.82)			
	P value	0.0082			
02:01 ~ 03:03	OR (95% CI)	0.34 (0.11–0.84)			
	P value	0.0120			
05:01 ~ 02:01	OR (95% CI)	9.82 (6.34–15.49)	6.66 (4.04–11.05)	5.74 (2.99–10.83)	2.78 (1.54–4.96)
	P value	<2.22e-16	<2.22 e-16	3.07e-10	0.0001
06:01 ~ 03:01	OR (95% CI)	0.07 (0–0.45)	0.23 (0.03–0.96)		
	P value	0.0007	0.0331		

Only significant associations are shown. All other associations were not significant. OR, CI, and *P* values are shown, compared to healthy controls.
CI, confidence intervals; OR, odds ratios.

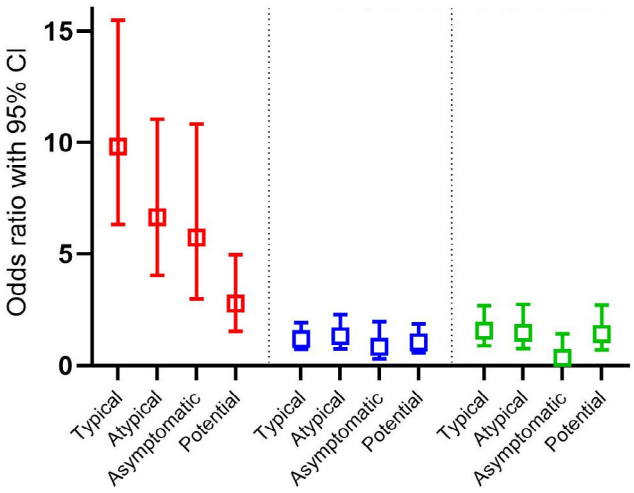

Figure 2 Association of haplotype with phenotype in celiac disease. Haplotype DQ2.5 showed a significant gradient across the phenotypes, while two other haplotypes—DQ2.2 and DQ8—did not show a similar trend. The Center box is the odds ratio and the two whiskers represent the 95% confidence intervals. (▭), DQA1*05:01~DQB1*02:01; (▭), DQA1*02:01~DQB*1*02:02; (▭), DQA1*03:01~DQB1*03:02.

these genotypes, while prevalence is very low in Japan and the Far East where these genotypes are infrequent, and virtually nil in Burkina Faso where people do not have these genotypes or eat wheat.[14] In our earlier study using low resolution genotyping, we identified DQB1*02 in 26.7%, DQB1*0302 in 8.3%, and both in 0.3% of 573 healthy adult Indians, and showed that there was no difference in prevalence of these genotypes in residents of different regions in India.[10] This present study returned a similar figure of the population prevalence of HLA DQ2/DQ8 of 30% in healthy Indian adults, which explains the ~1% population prevalence of CeD in the wheat-eating areas of India. Going farther than earlier studies, we now have high resolution genotyping and identification of all the other DQA1 and DQB1 genotypes in the healthy population.

Community-based studies suggest that specific HLA-DQB1 and ~DQA1 haplotypes confer approximately 5-fold higher risk for CeD when compared to the absence of these haplotypes.[14,16–19] Among the healthy controls, the DQ2.5, DQ2.2, and DQ8.1 haplotypes were found in 11, 20, and 11%, respectively. In the present study, the presence of the DQ2.5 haplotype conferred a markedly increased risk of symptomatic CeD. The DQ8 haplotype showed a borderline association with CeD in this population, while the DQ2.2 haplotype was not associated with CeD in this population. While the DQ haplotype has been associated with CeD in many populations, there has been controversy about the DQ2.2 haplotype, with some early studies identifying a linkage of the haplotype with CeD. A recent analysis of a large number of samples received in a clinical laboratory confirms the absence of a linkage between DQ2.2 and CeD.[5]

In the present study, nearly 20% of patients with typical and atypical CeD lacked the genes associated with expression of HLA-DQ2 or DQ8. Previous small studies from India have reported absence of these two DQ antigens in 6–16% of patients with CeD.[20–22] Small studies in other populations indicate the absence of DQ2 and/or DQ8 in 12–35% of CeD patients.[23–25] Studies now indicate that other HLA-DQ antigens, including DQ7 and DQ9, may confer risk to CeD.[26,27]

Interestingly, there was a gradient of association of CeD phenotype with the DQ2.5 haplotype, with the association progressively weakening from typical to potential CeD. Other authors have described a high frequency of low-risk HLA class II genotypes in latent (potential) CeD.[28] In addition, in this case–

control study, certain haplotypes had a statistically significant protective association with CeD. The role of protective haplotypes in CeD has not been previously discussed and the concept may require further exploration.

These findings have implications in the diagnostic algorithm for CeD. The value of HLA DQ testing in the diagnostic algorithm is related to its negative predictive value.[29] Where the situation warrants HLA DQ testing to exclude the diagnosis, it is necessary to keep in mind that some individuals with CeD may not express either DQ2 or DQ8 and that the negative predictive value of DQ testing may vary in different populations.

In conclusion, the HLA-DQ heterodimer haplotype DQ2.5 was strongly associated with CeD in this population and showed a diminishing gradient going toward the bottom of the celiac iceberg. Haplotype DQ8 showed a trend to associate with CeD, while DQ2.2 was clearly not associated with CeD in this population.

Acknowledgments

We acknowledge the support of Kamal Chettri, Anil K Verma, and Balamurugan Ramadass for their participation in the recruitment of patients; Meenakshi Sharma, Ashu Miyan, Anita Soni, and Prince Kumar for assistance in data maintenance and sample storage. The study was funded by a Consortium grant from the Department of Biotechnology (BT/MED/12/SP17783/2016).

References

1 Sollid LM, Jabri B. Celiac disease and transglutaminase 2: a model for posttranslational modification of antigens and HLA association in the pathogenesis of autoimmune disorders. *Curr. Opin. Immunol.* 2020; **23**: 732–8.
2 Gujral N, Freeman HJ, Thomson AB. Celiac disease: prevalence, diagnosis, pathogenesis and treatment. *World J. Gastroenterol.* 2012; **18**: 6036–59.
3 Poddighe D, Rebuffi C, De Silvestri A, Capittini C. Carrier frequency of HLA-DQB1*02 allele in patients affected with celiac disease: a systematic review assessing the potential rationale of a targeted allelic genotyping as a first-line screening. *World J. Gastroenterol.* 2020; **26**: 1365–81.
4 Husby S, Koletzko S, Korponay-Szabó I *et al*. European Society Paediatric Gastroenterology, Hepatology and Nutrition Guidelines for diagnosing coeliac disease 2020. *J. Pediatr. Gastroenterol. Nutr.* 2020; **70**: 141–56.
5 Choung RS, Mills JR, Snyder MR, Murray JA, Gandhi MJ. Celiac disease risk stratification based on HLA-DQ heterodimer (HLA-DQA1 ~ DQB1) typing in a large cohort of adults with suspected celiac disease. *Hum. Immunol.* 2020; **81**: 59–64.
6 Megiorni F, Pizzuti A. HLA-DQA1 and HLA-DQB1 in celiac disease predisposition: practical implications of the HLA molecular typing. *J. Biomed. Sci.* 2012; **19**: 88.
7 Wolters VM, Wijmenga C. Genetic background of celiac disease and its clinical implications. *Am. J. Gastroenterol.* 2008; **103**: 190–5.
8 De Silvestri A, Capittini C, Poddighe D *et al*. HLA-DQ genetics in children with celiac disease: a meta-analysis suggesting a two-step genetic screening procedure starting with HLA-DQ β chains. *Pediatr. Res.* 2018; **83**: 564–72.
9 Ludvigsson JF, Leffler DA, Bai JC *et al*. The Oslo definitions for coeliac disease and related terms. *Gut*. 2013; **62**: 43–52.
10 Ramakrishna BS, Makharia GK, Chetri K *et al*. Prevalence of adult celiac disease in India: regional variations and associations. *Am. J. Gastroenterol.* 2016; **111**: 115–23.
11 Singh A, Verma AK, Das P *et al*. Non-immunological biomarkers for assessment of villous abnormalities in patients with celiac disease. *J. Gastroenterol. Hepatol.* 2020; **35**: 438–45.
12 Makharia GK, Verma AK, Amarchand R *et al*. Prevalence of celiac disease in the northern part of India: a community based study. *J. Gastroenterol. Hepatol.* 2011; **26**: 894–900.
13 Pappas DJ, Marin W, Hollenbach JA, Mack SJ. Bridging Immuno-Genomic Data Analysis Workflow Gaps (BIGDAWG): an integrated case-control analysis pipeline. *Hum. Immunol.* 2016; **77**: 283–7.
14 Abadie V, Sollid LM, Barreiro LB, Jabri B. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. *Annu. Rev. Immunol.* 2011; **29**: 493–525.
15 Ramakrishna BS. Celiac disease: can we avert the impending epidemic in India? *Indian J. Med. Res.* 2011; **133**: 5–8.
16 Martínez-Ojinaga E, Molina M, Polanco I, Urcelay E, Núñez C. HLA-DQ distribution and risk assessment of celiac disease in a Spanish center. *Rev. Esp. Enferm. Dig.* 2018; **110**: 421–6.
17 Fernández-Fernández S, Borrell B, Cilleruelo ML *et al*. Prevalence of celiac disease in a long-term study of a Spanish at-genetic-risk cohort from the general population. *J. Pediatr. Gastroenterol. Nutr.* 2019; **68**: 364–70.
18 Leonard MM, Serena G, Sturgeon C, Fasano A. Genetics and celiac disease: the importance of screening. *Expert Rev. Gastroenterol. Hepatol.* 2015; **9**: 209–15.
19 Liu E, Lee HS, Agardh D. Risk of celiac disease according to HLA haplotype and country. *N. Engl. J. Med.* 2014; **371**: 1074.
20 Verma AK, Singh A, Gatti S *et al*. Validation of a novel single-drop rapid human leukocyte antigen-DQ2/-DQ8 typing method to identify subjects susceptible to celiac disease. *JGH Open*. 2018; **2**: 311–16.
21 Amarapurkar DN, Somani VS, Shah AS, Kankonkar SR. HLA - DQ genotyping in celiac disease in western India. *Trop. Gastroenterol.* 2015; **36**: 174–8.
22 Agrawal S, Gupta A, Yachha SK, Müller-Myhsok B, Mehrotra P, Agarwal SS. Association of human leucocyte-DR and DQ antigens in coeliac disease: a family study. *J. Gastroenterol. Hepatol.* 2000; **15**: 771–4.
23 Murad H, Jazairi B, Khansaa I, Olabi D, Khouri L. HLA-DQ2 and -DQ8 genotype frequency in Syrian celiac disease children: HLA-DQ relative risks evaluation. *BMC Gastroenterol.* 2018; **18**: 70.
24 Fallah H, Akbari MT, Mirzajani S *et al*. Association between HLA alleles and risk of celiac disease in Iranian patients. *Hum. Antibodies*. 2020; **28**: 123–9.
25 Özgenel ŞM, Temel T, Üsküdar Teke H, Yıldız P, Korkmaz H, Özakyol A. HLA-DQ2/DQ8 frequency in adult patients with celiac disease, their first-degree relatives, and normal population in Turkey. *Turk. J. Gastroenterol.* 2019; **30**: 321–5.
26 Tinto N, Cola A, Piscopo C *et al*. High frequency of haplotype HLA-DQ7 in celiac disease patients from South Italy: retrospective evaluation of 5,535 subjects at risk of celiac disease. *PLoS One*. 2015; **10**: e0138324.
27 Bodd M, Tollefsen S, Bergseng E, Lundin KE, Sollid LM. Evidence that HLA-DQ9 confers risk to celiac disease by presence of DQ9-restricted gluten-specific T cells. *Hum. Immunol.* 2012; **73**: 376–81.
28 Schirru E, Jores RD, Cicotto L *et al*. High frequency of low-risk human leukocyte antigen class II genotypes in latent celiac disease. *Hum. Immunol.* 2011; **72**: 179–82.
29 Husby S, Murray JA, Katzka DA. AGA clinical practice update on diagnosis and monitoring of celiac disease - changing utility of serology and histologic measures: expert review. *Gastroenterology*. 2019; **156**: 885–9.