SAPIENZA
UNIVERSITÀ DI ROMA

FACOLTÀ DI MEDICINA E FARMACIA

Dottorato di Ricerca in TECNOLOGIE BIOMEDICHE IN MEDICINA CLINICA

XXVIII CICLO

HLA-DQ and celiac disease: a novel risk gradient for predisposed subjects

Relatore
Prof. Fabio Midulla

Candidata
Dott.ssa Chiara Maria Trovato

Correlatore
Prof. Salvatore Oliva
Dott.ssa Monica Montuori

ANNO ACCADEMICO 2019 - 2020

INDEX:

Abstract	Page 3
Introduction	Page 5
Aims	Page 16
Materials and Methods	Page 17
Results	Page 28
Discussion	Page 35
Conclusions	Page 38
References	Page 39

ABSTRACT:

BACKGROUND

Celiac disease (CeD) is an immune-mediated systemic disorder elicited by gliadin, a prolamine found in wheat and related proteins, occurring in genetically predisposed individuals, carrying HLA-DQ2 and/or DQ8. Several heterodimers have been associated with CeD: the HLA-DQ2.5, occurring in approximately 90% of CeD patients, is encoded by the DQB1*02 and DQA1*05 alleles both *in cis* and *in trans*; the HLA-DQ2.2 heterodimer, encoded by the DQB1*02 and DQA1*02 allele; and the HLA-DQ8 heterodimer, encoded by the DQB1*03:02 and DQA1*03 alleles.

AIMS

We aimed: 1) To characterize HLA polymorphisms; 2) to confirm the association between HLA class II genes with CeD; 3) to assess their role in CeD genetic susceptibility by re-defining the current risk pyramid of genetically predisposed individuals.

METHODS

We included celiac children diagnosed based on European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) criteria and controls. The control group included healthy Italian individuals and affected family-based controls (AF-BAC). All individuals were typed for DRB1, DQA1, and DQB1 genes by sequence-specific primer–polymerase chain reaction (SSP-PCR). Disease risks are expressed as 1:N, where N is the number of individuals among which one patient is present. Considering a disease prevalence of 1:100 in the general population, for each HLA-DQ category, N is calculated as a percentage of controls with that particular HLA-DQ status multiplied by 100 and divided by percentage of patients with the same DQ typing.

RESULTS: We included 778 CeD patients (M:277= F:511), and 551 controls (292 healthy Italian individuals and 259 affected family-based controls). In conclusion, the new risk pyramid shows an increased risk for all haplotypes; it is reasonable to assert that there is an increased CeD risk in presence of DQ*B102 either as a single copy or combination with other alleles.

INTRODUCTION:

Celiac disease (CeD) is an autoimmune systemic disorder elicited by gluten in genetically susceptible (HLA DQ2 and/or DQ8) subjects. It is characterized by the presence of a combination of gluten-dependent clinical manifestations or it may be discovered in asymptomatic individuals through screening methods (high-risk groups, first-degree relatives, and related diseases). CeD-specific auto-antibodies (anti-transglutaminase Ig-A (TGA-IgA); anti-endomysium (EMA), antibodies against deamidated forms of gliadin peptides (DGP)), HLA-DQ2 or HLA-DQ8 haplotypes and enteropathy are the lead actors of the disease [1]. CeD can be considered a major public health issue, as its prevalence is increasing in different geographic areas, such as Asian countries [2]. A recent meta-analysis showed that the pooled global seroprevalence of CeD is 1.4% [3], with diverse data from different continents: prevalence is shown to be highest in Europe and in Oceania and very low in Southern America.

Differently from other autoimmune diseases, celiac disease is, to date, the only condition whose environmental trigger has been identified – i.e. gluten; however, several other environmental co-factors have been implicated in its development, including viral infections (Reovirus, Rotavirus, Norovirus, Adenovirus and Astrovirus) [4-6], gut barrier dysfunctions [7-9], changes of the gut microbiota [10], delivery methods [11-15], breastfeeding [16] and timing of gluten introduction in the diet [17]. It is currently impossible to identify the main factor which, together with gluten intake, leads to loss of tolerance and activation of innate and adaptive immunity – responsible for the presence of antibodies and for the mucosal damage pathognomonic of CeD [18].

The prevalence of the disease is increased among patients with anemia or autoimmune diseases, with short stature, or with syndromes such as Down, Turner, or Williams syndrome [19]. Moreover, CeD clusters within families,

with a prevalence among first-degree relatives ranging from 2.8% to 17.2% in different series [20] .

PATHOGENIC AND GENETIC ASPECTS:

A hereditary component for CeD has been identified by several studies showing that up to 7.5%–15% of first-degree relatives are affected [21, 22]: additionally, studies in monozygotic twins showed concordance rates of 50%–80%, decreasing up to 10% in dizygotic twins [23, 24]. However, the genetic background does not automatically equate to the presence of disease: in fact, CeD is a complex genetic disorder which develops in genetically predisposes individuals when exposed to non-genetic factors [25].

Our understanding of pathogenesis of CeD was dramatically changed by the discovery of a genetic background involving the HLA, with the first evidence of the association between HLA genes and CeD being reported almost three decades ago [26-29]. In 1983, for the first time, DC3 (a prior definition of HLA-DQ2) appeared to be the HLA specificity primarily associated to CeD [30].

This association has been largely confirmed by Margaritte-Jeannin et al [31] in 2004 when they demonstrated the highest frequency of HLA-DQ2 haplotypes in CeD patients.

Several genes, or loci, on the short arm of chromosome 6 define the HLA complex: extreme polymorphism has been reported for these loci, as an ever-increasing amount of alleles is being characterized [32]. The CeD-related HLA loci belong to the class II at the DQ locus, and include the HLA-DQA1 and DQB1 loci, respectively coding for a- and b-chain proteins. These chains bind on the surface of antigen-presenting cells, forming heterodimers which are defined as HLA-DQ molecules. HLA class II genes,

and more in detail HLA-DQ2 and DQ8, are the currently considered the strongest and best-defined genetic risk factors for CeD.

Furthermore, independently of its binding to HLA molecules, gluten is able to activate both the CeD8+ T-cells in the lamina propria (responsible for the proliferation of intra-epithelial lymphocytes, IELs) and the antigen presenting cells (APCs), therefore multiplying the effect.

Any individual presents up to 4 different heterodimers, as there are 2 alleles for each of the DQA1 and DQB1 loci. Heterodimers are defined as encoded in cis if inherited by the same parent, and therefore resulting from the binding of DQA1 and DQB1 alleles on the same chromosome, or encoded in trans when each allele comes from a different parent. Both in-cis and in-trans heterodimers result in functional molecules binding deamidated gliadin [33]. In the general population without a diagnosis of CeD, the prevalence of HLA-DQ2 and/or HLA-DQ8 is approximately 40%: however, as expected from the genetic association, this percentage increases among patients with first-degree relatives affected with CeD [34].

The activation of gluten-reactive CeD4+ T-cells in the gut is mediated by the DQ2 and DQ8 heterodimers, therefore explaining the reasons for such close association. More in detail, APCs express the disease-associated HLA-DQ molecules which specifically bind gluten-derived peptides modified by the enzyme tissue-transglutaminase (tTG); subsequently, the APCs present these peptides to intestinal T cells, eliciting a T response which triggers production of antibodies and secretion of pro-inflammatory cytokines. This response mechanism results in mucosal atrophy and is involved in the clinical manifestations of CeD [18].

Several heterodimers have been associated with CeD: the HLA-DQ2.5, occurring in approximately 90% of CeD patients [35, 36], is encoded by the DQB1*02 and DQA1*05 alleles both in cis and in trans; the HLA-DQ2.2

heterodimer, encoded by the DQB1*02 and DQA1*02 allele [37, 38]; and the HLA-DQ8 heterodimer, encoded by the DQB1*03:02 and DQA1*03 alleles. These three different allotypes have very different risks of disease, but the clinical presentation of CeD is the same.

Recent evidence shows that the molecular surface of antigen-binding clefts of HLA-DQ2.5 and HLA-DQ2.2 are very similar, but difference in the nature of the peptides presented translates to difference in responding T cell repertoires and the nature of engagement of the respective antigen-presenting molecules, which seems associated with differing disease penetrance [39].

Different studies demonstrated that T cell clones restricted by the three different allotypes recognize different and separates sets of gluten epitopes[38, 40]. Different HLA molecules bind and present gluten peptides with glutamate residues at different positions, namely P1 and/or P9 for HLA-DQ8 and P4, P6 or P7 for HLA-DQ2.2 and HLA-DQ2.5. HLA-DQ2 and HLA-DQ8 are highly homologous; however, HLA-DQ2.2 also uses a P3 anchor, with high selectivity for serine or threonine, which might contribute to different risk profiles.[39]

The HLA region alone accounts for approximately 40% of the disease heritability [41, 42], meaning that other genes are involved in CeD susceptibility. Many candidate loci have been studied but only three chromosomal regions, 5q31-q33 (CELIAC2), 2q33 (CELIAC3) and 19p13.1 (CELIAC4), have been officially recognized as predisposing genetic factors. CELIAC2 (5q31-33) was the region that first reached a genome-wide level of significance in an European metanalysis [43] and can be considered a confirmed linked region [44]. It was initially discovered in a Italian cohort [45, 46] and later confirmed by Swedish/Norwegian data [47].

CELIAC3 (2q33) is the region that is harbouring well-characterized members of the costimulatory molecules: cytotoxic T lymphocyte-

associated antigen-4 (CTLA-4), CD28 [48] and inducible costimulator (ICOS) at 2q33 [49] and nearby to programmed death-1 (PD-1) localized at 2q37.1 It seems to take part to Lymphocyte activation and regulation. An European study suggests that the CeD risk due to 2q33 gene region is complex and may involve more than one susceptibility allele, which possibly differs from other autoimmune diseases [50]. Moreover, CELIAC3 harbours several independent loci contributing to CeD susceptibility [49].

CELIAC4 (19p13.1) was described as a susceptible locus for CeD by Van Belzen [51] in a genome-wide screen on 67 CeD families. CELIAC4 is associated with MYO9B, a gene belonging to the myosin family that binds calmodulin which contributes to the maintenance of the intestinal barrier. Based on these premises, CELIAC4 has been postulated as a risk factor of CeD [52].

In recent years, genome-wide association studies (GWAS) have identified many non-HLA genes associated with an increased risk of CeD, such as those coding for cytokines, chemokines and their receptors, cell adhesion molecules, T- and B-cell activators [53-57].

To date, in GWAS, 40 additional genetic loci have been associated with CeD; these variants are mainly small nucleotide polymorphism (SNPs) found in non-coding regions, often within enhancers, suggesting an effect on gene expression rather than causing modifications of protein functions [58], but it seems that these 40 non-HLA loci explain approximately 5% of the disease risk. These loci are represented by 57 independent SNPs, of which 29 were localized to a single gene and the remaining in intergenic regions [57].

In 2014, [59], a multicentric European study proposed to analyze these 57 variants, in order to improve the identification on potential CeD patients. The combination of HLA and non-HLA risk variants allows to classify individuals in three different subgroups (low risk: predicted risk<25%;

intermediate: 25-75% and high-risk: > 75%). This new stratification, albeit expensive, might help to make a better selection on individuals who need closer follow-up[59].

Nevertheless, non-HLA genetic contribution to CeD is weak (about 15%) and these polymorphisms are not considered in the calculation of CeD genetic risk [60-62].

HISTOLOGICAL ASPECTS:

From the description of the pathogenesis we can therefore define the principal actors of the mucosal damage by performing multiple biopsies, at least two in the bulb and four in the duodenum [63, 64].

Several facets must be considered for histological assessment of tissue samples, including the number of intraepithelial T lymphocytes (IELs, with a 25 IEL/100 enterocytes threshold being considered borderline), the presence of crypt hypertrophy (proving epithelial regeneration in the crypts with increased rate of mitosis) and the atrophy of intestinal villi (shorter villi with reduction of the villus/crypt ratio, up to complete flattening).

None of these elements, individually, appears to be pathognomonic of celiac disease, but their combination has allowed the identification of typical lesions of the disease first codified by Marsh [65] and subsequently revised by Oberhuber in 1999 [66] and by Corazza-Villanacci in 2005 [67], and as recently reaffirmed in a Italian position statement [68].

However, histological damage features a patchy distribution [69, 70] with lesions not limited to the bulb [71, 72], but also extended in caudal direction beyond the Treitz [73].

POTENTIAL CeD: Potential CeD (PCeD) is the condition related to people with a normal (Marsh grade 0) or minimally abnormal (Marsh grade 1) intestinal mucosa who are at increased risk of developing CeD, as

indicated by both positive EMA and TGA-IgA and compatible HLA [74]. Symptoms and signs of the disease are not always clinically manifest, and even when present, they can range from mild to severe. Many authors tried to understand the natural history of potential CeD, both in children and in adults [75, 76] or to design a predictive model of mucosal atrophy [77], with the help of mathematical model and machine learning alghoritms, [78].The presence of symptoms in both adults and children should be considered as the main determinant to prescribe a GFD in potential celiac disease. It is important to remember that all symptoms have to be considered important for the beginning of a GFD. There is no difference in the decision tree, in fact, if patient has gastrointestinal (diarrhea, constipation, abdominal pain) or extraintestinal manifestation (anemia, osteoporosis, migraine), as suggested by Popp and Maki in a recent review too [79]. As the timing of flattening is totally unpredictable, asymptomatic patients with PCeD should undergo a comprehensive follow-up in order to detect early symptoms and promptly start a GFD.

DIAGNOSIS:

In support of diagnosis, since the 90s several IgA and IgG antibodies have been identified for CeD, with IgA against TGA-IgA and EMA in the spotlight. Different guidelines for CeD diagnosis have been published so far by several societies, both for adults [80, 81] and children [82]. The first diagnostic criteria for CeD in children were published in 1969 by the European Society of Pediatric Gastroenterology and Nutrition (ESPGAN) [83]. At that time, three biopsies of the small bowel mucosa were necessary to establish the diagnosis of CeD: the first one when the suspicion of CeD was raised to demonstrate villous atrophy; the second one, after 1 year on a gluten-free diet (GFD), to show normalization of the intestinal mucosa; the third and last one after a gluten-challenge to prove unequivocally the

causal relationship between gluten intake and mucosal relapse. These rules were used until 1990, when criteria were revised by European Society of Pediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) [84]. Based on these revised criteria, if the clinical history and antibody findings were suggestive of the disease, a single intestinal biopsy followed by a favorable response to a GFD was sufficient to confirm the diagnosis of CeD. EMA and TGA-IgA are the well-established serological marker; indeed, they are present at the moment of the CeD diagnosis and vanish on a GFD diet, thus representing a true revolution in the field of diagnosis [19]. In 2012, the European Society of Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) allowed to diagnose CeD without the need for endoscopy and intestinal biopsy in symptomatic children with high TGA-IgA levels, i.e. ≥10-fold upper limit of normal (x ULN), EMA positivity and HLA-DQ2 and/or -DQ8 [85]. Very recently, at January 2020, this approach has also been suggested for asymptomatic children with high TGA-IgA titers (≥10 x ULN) and positive EMA, regardless of the HLA status [86]. The no-biopsy diagnostic approach is based on the strong correlation between high levels of TGA-IgA (≥10 x ULN) and a high degree of villous atrophy [87, 88], while the reliability of low positive anti-TGA IgA titers is limited [89], particularly when EMA are not detected [90]. Moreover, a recent study of our group, supports the diagnostic value of persistently low positive TGA-IgA titers (at least two determinations above the normal range) in predicting CeD in symptomatic children, regardless of EMA positivity and remarks the central role of biopsy in specific disease pattern at CeD diagnosis [91].

THERAPY:

Following confirmation of the diagnosis, a life-long gluten-free diet (GFD) is the only treatment for CeD, able to effectively switch off inflammation

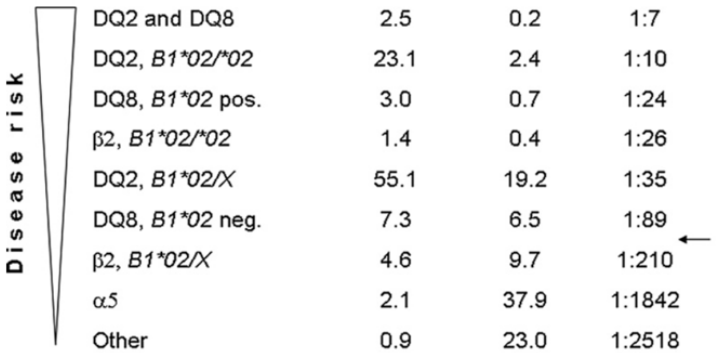
and achieve mucosal healing. Gluten-containing cereals (wheat, barley, rye, spelt, and kamut) need to be completely excluded from the diet; pure oats are safe for children and adults with CeD ([92] [93] [94]). In contrast, the main sources of complex carbohydrates in the diet of patients with CeD are rice, corn, buckwheat, amaranth, millet, sorghum, and quinoa, i.e. gluten-free cereals and pseudocereals. A strict GFD is a daunting task, and cannot be reduced to just "abstain from pasta, pizza and beer": indeed, gluten can be hidden in other foods, being used as thickening agent for gravies and coatings and in food factories as a flavoring or stabilizing agent. The chance of cross-contamination from undesirable co-cultivation and harvesting in the field, as well as inadequate storage in factories, at home and at restaurants, could pose an additional obstacle.

While many other potential, complementary treatments have been proposed in the last decades, GFD is still preferable in terms of safety; however these treatments could be used in order to prevent involuntary gluten contamination[95]. 31831308

These treatments act at different levels: some aim to decrease the immunogenic content of gluten before it reaches the intestine (caricain, latiglutenase [96], ANPEP [97]) or to prevent absorption of gluten in the gut lumen by sequestration (polymeric binders [98]) or by blocking the passage of gluten through a leaky intestinal barrier (larazotide acetate [99] and gliadin transport pathway inhibitor [100]), whereas other therapies act by modulating the immune response towards gliadin (transglutaminase inhibitors [101]), or the downstream immune activation (TIMP [102]), or even by inducing immune tolerance to gluten (immune-modulating vaccine [103]).

Based on these premises, much is currently known in regards to both the pathogenetic and clinical role of autoantibodies. Identifying subjects who will develop CeD and investigating their natural history will be primary aims for CeD research. Therefore, several studies have been published in the last few years concerning pathogenetic factors for CeD and the dynamics of autoantibodies among specific groups.

In 2009 [104] our group obtained a gradient of disease risk ranging between 1:7 and 1:2518; looking at these data, as shown in Figure 1, the highest risk value was for DQ2 and DQ8 subjects (1:7), followed by the other three categories coding for two β susceptibility chains: DQ2, B1*02/*02 (1:10), DQ8, B1*02 positive (1:24), and β2, B1*02/*02 (1:26). DQ2 subjects carrying a second DQB1 allele different from *02 or *0302 (B1*02/X) resulted in a risk of 1:35.



	Patients%	Controls%	Risk	F	M
DQ2 and DQ8	2.5	0.2	1:7	1:7	1:8
DQ2, *B1*02/*02*	23.1	2.4	1:10	1:8	1:13
DQ8, *B1*02* pos.	3.0	0.7	1:24	1:16	1:52
β2, *B1*02/*02*	1.4	0.4	1:26	1:27	1:26
DQ2, *B1*02/X*	55.1	19.2	1:35	1:26	1:54
DQ8, *B1*02* neg.	7.3	6.5	1:89	1:62	1:157
β2, *B1*02/X*	4.6	9.7	1:210 ←	1:211	1:208
α5	2.1	37.9	1:1842	1:8327	1:1027
Other	0.9	23.0	1:2518	1:2530	1:2497

Figure 1: Risk- gradient. Risks were evaluated from unrounded case-control DQ frequencies, considering a disease prevalence of 1:100 (arrow) and a female: male gender ratio of 1.8. DQ2 =DQA1*05 and DQB1*02; DQ8 =DQA1*03 and DQB1*0302; beta2 =DQB1*02 in the absence of DQA1*05 (DQ2.2); alfa5 = DQA1*05 in the absence of DQB1*02. X, different from DQB1*02 or DQB1*0302. Adapted from Megiorni et al, 2009 [104].

AIMS

The aim of the present study is:

1) To characterize HLA polymorphisms and distinguish the different roles of single haplotype (by discriminating the DQ2.5 from the other susceptible haplotype)

2) to confirm the association between HLA class II genes with CeD;

3) to assess their role in CeD genetic susceptibility by re-defining the current risk pyramid of genetically predisposed individuals

MATERIALS AND METHODS

A. Patients

We enrolled all CeD patients who performed HLA testing at the moment of CeD diagnosis and their nuclear families (parents and siblings) from June 2005 to June 2020. All patients and their families were enrolled in the Pediatric Gastroenterology and Liver Unit, Sapienza University of Rome, Italy and all diagnoses of CeD were based on the current criteria of the European Society for Pediatric Gastroenterology, Hepatology and Nutrition (ESPGHAN) published in 1990, 2012 and 2020 previously descripted.

Nuclear families performed HLA testing after CeD patient's diagnosis within the project of family screening.

All relatives were tested for total IgA and TGA-IgA; in selected individuals EMA antibodies were measured and small-intestinal biopsy was performed, where necessary.

B. Control subjects

The control data set included healthy Italian individuals and affected family-based controls (AF- BAC). The affected family-based sample was established by selecting the chromosomes not transmitted from parents to affected children [105].

The AF-BAC method was introduced in 1981 [106]. In families ascertained for the presence of an affected child, the parental marker alleles not transmitted to the affected child are used as population (control) alleles. This matched design for patient (parental transmitted) and "control" (parental non-transmitted) marker alleles avoids ethnic confounding in the case of a stratified population. [107, 108]. This method has been extended in 1989 to nuclear pedigrees ascertained for the presence of at least two

affected sibs by using the alleles from both parents that are never transmitted to either sib in the affected sib pair, as the "control" population [109, 110].

Moreover, with the AF-BAC method, the effects of population stratification, migration, and admixture on the expected values of the family marker allele frequencies are considered.

Both healthy controls and affected family-based controls samples were in Hardy-Weinberg equilibrium at the HLA-DR and -DQ loci and the comparison of allele and genotype frequencies between the two groups did not show significant differences.

C. Methods

1. Genotyping

All the following methods are performed in collaboration with Dr. Barbara Mora and Prof. Antonio Pizzuti of the Unit of Genetics, of "Sapienza – University of Rome" Policlinico Umberto I.

1.1. DNA extraction

DNA is extracted with the salt-chloroform method starting from fresh or frozen blood, using the following protocol (modified from Müllenback et al., 1989):

1) Add 45 ml of a hypotonic solution (0.32 M saccarose, 10 mM Tris-HCl pH 7.5; 5 mM MgCl2; 1% Triton 100X) to 5 ml of blood
2) Gently shake the pellet, then put in ice for 10'
3) Centrifuge (10', 4000 rpm, 4° C)
4) Store the pellet and wash it once or twice with a Salt-EDTA solution (0,075 M NaCl e 0,025 M EDTA pH8)
5) Centrifuge after every wash (10', 3000 rpm, room temperature)
6) Add 5 ml SE solution, 40 µl Proteinase K (25 mg/ml) and 500 µl SDS 10% to the pellet
7) Incubate overnight at 37° C, or 55° C for three hours
8) Add pre-heated NaCl to the pellet (reaching 1.5 M concentration, starting from a 5 M solution)
9) Add an even amount of chloroform and mix for 45' in a rotor at room temperature
10) Centrifuge (2000 rpm, 10', room temperature)
11) Extract supernatant and add an even amount of isopropanol
12) Mix for 5'
13) Centrifuge (2000 rpm, 10', 4° C)

14) Wash the pellet with 1.5 ml 70% ethanol and incubate at -80° C for at least 1h
15) Centrifuge (9000 rpm, 30', 4° C)
16) Discard supernatant, then add an adequate volume of TE solution (10 mM Tris-HCl; 1 mM EDTA)

1. Spectrophotometer reading

The solution obtained from the last step in the DNA extraction process is read using a spectrophotometer with two different wave lengths (260 nm and 280 nm), corresponding to the DNA and protein peak absorption, respectively. A 260 nm/280 nm ratio between 1.6 and 2.2 is considered adequate for a pure solution. As a reference to measure the DNA concentration of the sample, a 50 μg/ml solution has a 260 nm absorption equal to 1. The value obtained from the 260 nm reading is then multiplied by a factor or 50 and by the dilution performed; the resulting concentration is expressed in μg/ml.

1.2 Polymerase Chain Reaction (PCR)

PCR is an enzymatic method for amplification of specific DNA sequences relying on the Taq polymerase and on specific short sequences of nucleotides (primers) to make new strands of DNA. As the newly synthesized chain of DNA can be used as a template for further copies, at every cycle an exponential number of DNA copies is generated. Each cycle consists of three different step, occurring at different temperature thresholds: denaturation occurs at 96° C, annealing between 55° C and 65° C, and extension at 72° C. Denaturation is needed in order to produce single-strand DNA chains acting as templates for following steps; during annealing, primers bind to the single-strand DNA chain; lastly, during extension, the DNA polymerase extends the primers by adding new

nucleotides to the chain. The most commonly used DNA polymerase is called Taq polymerase, after the heat-tolerant bacterium from which it was isolated (Thermus aquaticus).

Besides the Taq polymerase, PCR requires a double-stranded DNA chain, a couple of primers delimiting the sequence (one for the 5′ strand, and one for the 3′ strand), the four triphosphate nucleotides (dATP, dCTP, dGTP, dTTP), a buffer containing inorganic ions for the enzyme action. At the end of the PCR, in order to assess the result of DNA amplification and its quality, a gel electrophoresis is usually run on a small quota of the product by using ethidium bromide 0.5 µl/ml. Ethidium bromide binds by intercalating between DNA base pairs; when the agarose gel is exposed to UV light, DNA bands become visible, and by measuring their mobility relative to a molecular weight standard the base pair length can be identified.

1.3 HLA typing

During the last 40 years, HLA typing has been performed by using serological methods; however, the introduction of PCR has introduced a revolution in HLA genes study and several techniques based on DNA amplification via PCR are now available. Compared to serological methods, PCR techniques can be used in order to better define the polymorphisms, mostly for the following two reasons: some nucleotide differences between alleles do not result in aminoacidic changes; and serological methods cannot be used for identification of sequences which do not define an epitope, or for sequences which define an epitope for which no specific antibody is available.

1.4 PCR-SSP (PCR-Sequence Specific Primers)

The PCR-SSP method is based on the grounds that primers identical to a "target" DNA sequence are more efficacious in amplification of specific alleles or groups of alleles compared to partially complementary primers, especially in the presence of a "mismatch" on the 3'.

At the end of the polymerization process, amplified DNA fragments are separated by running a gel electrophoresis and examined under UV light following exposure to ethidium bromide. The presence or absence of DNA fragments with different, specific molecular weight is the key to understanding the results of the PCR. However, during amplification, several factors can impair the reliability of the reaction (pipetting errors, low DNA quality, presence of inhibitors, and so on). In order to assess the integrity and purity of the reaction, an internal control consisting of a primer couple for a known gene is generally used. In this case, positive reaction for an allele (or an allele group) appears in the gel as a DNA fragment (band) between the internal control and the primer bands. The first application of this technique for HLA typing has been used DRB1 (Olerup et al., 1992).

1.5 HLA-DRB1, -DQA1 e -DQB1 gene typing

In this study, the class II HLA genes has been performed by using the method just described for all subjects involved by using commercially available kits. The kit consists of 8-vial strips, with each vial containing a different primer mix in liquid phase: two or three primer pairs with different specificity are used for each vial, allowing for amplification of different DNA sequences at the same time. Primers are designed and mixed in order to be compatible in the same PCR without the risk of creating dimers. Each vial also contains the material for amplification of a "housekeeping" gene, such as human β-globin, as an internal control for assessment of the reaction quality.

This is the process for DNA amplification using this kit:

1. Put 5 μl primer mix from each primer vial into the corresponding empty vial of each strip, for each sample requiring amplification
2. For each sample, put in an Eppendorf vial 225 μl Reagent A and 135 μl Reagent B (master mix)
3. Add 45 μl DNA (diluted up to the concentration of 110 ng/μl) to the master mix
4. Shake the Eppendorf vials
5. Put 45 μl of the resulting mixture in each one of the 8 vials filled with primer mix
6. Firmly close each vial with the cap
7. Shake the strip.

Amplification has been performed by using an Applied Biosystem 2720 thermal cycler (table 1).

Temperature	Time	Cycles
95°C	5 min	1
95°C	30 sec	5
61°C	30 sec	
72°C	45 sec	
95°C	30 sec	30
58°C	30 sec	
72°C	45 sec	
72°C	7 min	1
15°C	HOLD	-

Table 1: Amplification process

After the first 5 cycles, the temperature for annealing is reduced (from 61° C to 58° C) in order to improve the output of each cycle, as the primer-target sequence complex is more stable at lower temperatures. As previously stated, amplified DNA sequences are separated by electrophoresis gel (120 Volt for 20') and examined under UV light following exposure to ethidium bromide (table 2).

	Gel Solution (30 ml)	Buffer Solution (10 ml)
Agarose	0.6 gr	_
TBE 5X	6 ml	2 ml
H_2O	24 ml	8 ml
Ethidium Bromide	1.5 µl	0.5 µl

Table 2: agarose gel preparation

Results of the PCR are assessed by examining control and reaction bands, using the scheme in table 3, with lighter sequences running longer distances than heavier ones. By comparing the migration distance of each reaction band to the control band (341bp, β-globin) and to the marker band, it is possible to identify the different alleles.

In CeD patients, as well as in healthy controls, HLA haplotypes have been estimated based on the strong linkage disequilibrium between the DRB1, DQA1 e DQB1 genes. The molecular typing of the DRB1, DQA1 e DQB1 alleles and the definitions used for the DR-DQ haplotypes in the remainder of the study are reported in table 4.

As an example, the DRB1*03-DQA1*05-DQB1*02 combination would define the DR3-DQ2 haplotype.

	Allele	Molecular weight
1	DQA1*01	225 bp
	DRB1*07	129 bp
2	DQB1*04	207 bp
	DQA1*02:01	143 bp
3	DQA1*03	196 bp
	DQB1*03:01, *03:03	Mix
4	DQA1*05	209 bp
	DQB1*03:02	121 bp
5	DRB1*03	221 bp
	DQA1*06	121 bp
6	DRB1*04	260 bp
	DQB1*02	140 bp
7	DRB1*11	182 bp
	DQB1*03:05	132 bp
8	DQB1*03:01, *03:04	102 bp

Table 3: HLA typing scheme for PCR.

DR-DQ haplotypes	DRB1	DQA1	DQB1
DR3-DQ2	*03	*05	*02
DR11-DQ7	*11	*05	*03:01
DR7-DQ2	*07	*02:01	*02
DR4-DQ8	*04	*03	*03:02

Table 4: HLA haplotypes associated with CeD.

2. Antibody measurements

2.1 Dosage of serum IgA

Serum IgA is detected by Nephelometry. It is performed by measuring the scattered light at an angle from the sample being measured and it is widely used in clinical laboratories because it is relatively easily automated. It is based on the principle that a dilute suspension of small particles will scatter light (usually a laser) passed through it rather than simply absorbing it. The amount of scatter is determined by collecting the light at an angle (usually at 30 and 90 degrees). The amount of light scatter is measured and compared to the amount of scatter from known mixtures. The amount of the unknown is determined from a standard curve.

2.2 Dosage of anti-transglutaminase IgA

Anti-tissue TransGlutaminase (TGA-IgA) antibodies are measured by an ELISA method using a commercially available kit (Eurohospital, Trieste, Italy). Reference values are as follows: Negative: < 9.00 U/mL, Uncertain: 9.00 – 16.00 U/mL; Positive: > 16.00 U/mL.

All dosages are performed in the Immunopathology Laboratory of "Sapienza – University of Rome" Policlinico Umberto I, directed by Prof. Fabrizio Mainiero.

3. Upper gastrointestinal endoscopy and biopsies

All upper gastrointestinal endoscopy were performed under general anesthesia or deep sedation by an expert pediatric endoscopist. Targeted or

random mucosal biopsies were obtained (4 in duodenum and 2 in bulb) according to mucosal appearance. All specimens were adequately oriented and graded according to the Marsh–Oberhuber (MO) classification by an expert and dedicated pathologist **.**

4. Statistics

Statistical significance was calculated by Fisher's exact test using 2 X 2 contingency tables. Values of $p < 0.05$ were considered significant.

Disease risks are expressed as 1:N, where N is the number of individuals among which one patient is present. Considering a disease prevalence of 1:100 in the general population, for each HLA-DQ category, N is calculated as a percentage of controls with that particular HLA-DQ status multiplied by 100 and divided by percentage of patients with the same DQ typing.

The risks in female and male subjects were calculated using the DQ frequencies previously reported in the two genders and considering the 1.8:1 female to male ratio.

For calculation of odds-ratio in non-DQ2.5 patients we elaborated a calculation with the exclusion of DQ2.5, as suggested by literature [111].

The odds ratio in at risk-population was calculated with the construction of two for two table by OpenEpi, Version 3.

RESULTS

We enrolled 778 (F 502) celiac patients in our study group and 551 controls that included 292 healthy Italian individuals and 259 affected family-based controls. Of these, it was selected chromosomes not transmitted from parents to affected children, as suggested by literature [105].

1. HLA distribution

Table 5 shows HLA distribution in our study cohort (Controls and CeD patients).

	Controls (N= 551)	**CeD (N=778)**
DQ2/DQ8	1 (0.2%)	25 (3.2%)
DQ2/DQ2,β2	13 (2.4%)	177 (22.8%)
DQ8/β2	4 (0.7%)	34 (4.4%)
β2/β2	2 (0.3%)	12 (1.5%)
DQ2/x	106 (19.2%)	408 (52.4%)
DQ8/x	36 (6.5%)	64 (8.2%)
β2	53 (9.6%)	48 (6.2%)
α5	209 (37.9%)	8 (1.0%)
Other	127 (23%)	2 (0.3%)

Table 5: HLA distribution.

In details, we can observe the distribution of single haplotypes in CeD patients in Table 6.

	n	%
DQ2.5 cis	237	30.46%
DQ2.5, DQ2.5	177	22.75%
DQ2.5 trans	171	21.98%
DQ8	64	8.22%
DQ2.2	48	6.17%
DQ8/DQ2.2	34	4.37%
DQ8/DQ2.5	25	3.21%
DQ2.5/DQ2.2	12	1.54%
DQ2neg/DQ8neg	10	1.29%

Table 6: HLA haplotypes in CeD patients.

As shown, DQ2.5 (HLA-DQA1*05- DQB1*02) *in cis* appears to be the most common haplotype associated with CeD in our study cohort. This haplotype is present in 30.46% of our celiac patients.

Among 778 HLA-DQ genotyped CeD patients, 704 (90.5%) were proven to be carriers of at least one copy of the HLA-DQB1*02 allele, whereas 74 carry different allelic variants at both HLA-DQB1 loci (HLA-DQB1*0302 or others). Therefore, the HLA-DQB1*02 carrier frequency appears to be 90.5%, meaning that only 9.5% of CeD population is completely lacking this allelic variant.

2. Female:Male ratio

Data from literature shows a cumulative Female:Male ratio of 2:1 in CeD. Figure 2 illustrates the F:M ratio of different haplotypes. It is possible to observe that different haplotypes show different female:male ratio.

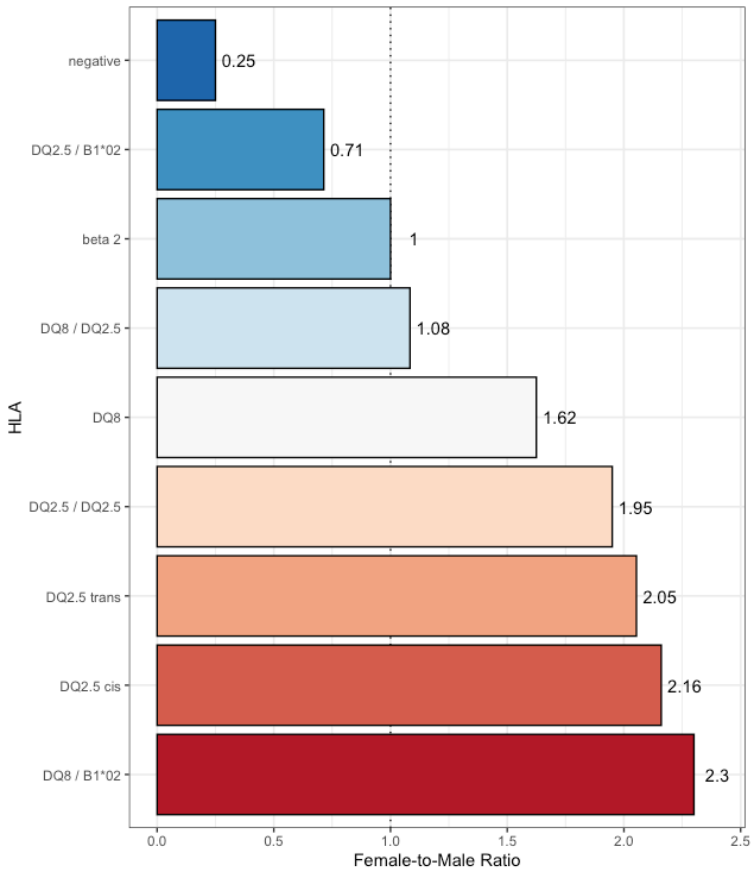
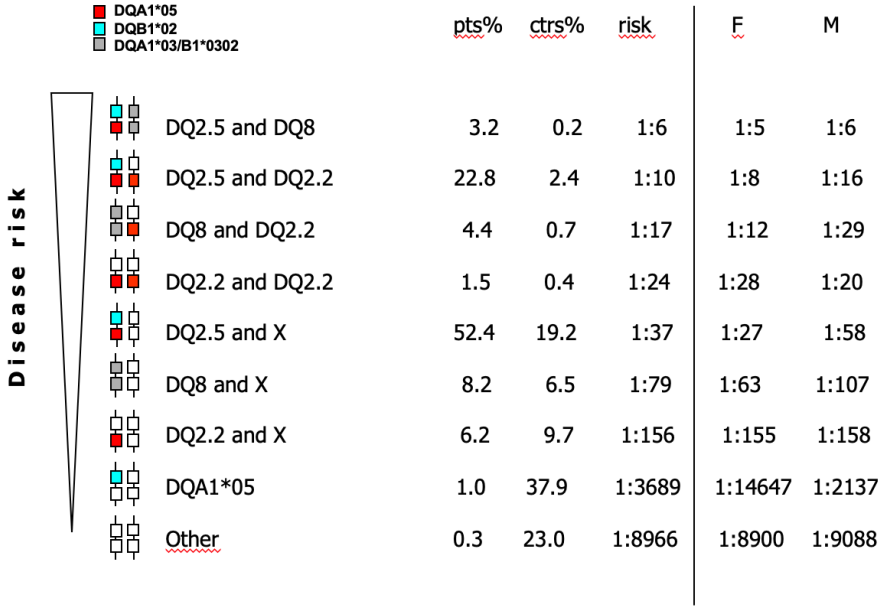


Figure 2: F:M ratio in CeD population.

It's interesting to highlight that negative haplotypes have a 1:4 F:M ratio, whereas the presence of DQ2.2 is represented with 1:1 ratio. The other haplotypes are represented with a ratio comparable with data from literature.

3. Risk pyramid

Figure 3 shows the new risk pyramid with combined risk for male and female, and risk according to gender, in a separate column.



	pts%	ctrs%	risk	F	M
DQ2.5 and DQ8	3.2	0.2	1:6	1:5	1:6
DQ2.5 and DQ2.2	22.8	2.4	1:10	1:8	1:16
DQ8 and DQ2.2	4.4	0.7	1:17	1:12	1:29
DQ2.2 and DQ2.2	1.5	0.4	1:24	1:28	1:20
DQ2.5 and X	52.4	19.2	1:37	1:27	1:58
DQ8 and X	8.2	6.5	1:79	1:63	1:107
DQ2.2 and X	6.2	9.7	1:156	1:155	1:158
DQA1*05	1.0	37.9	1:3689	1:14647	1:2137
Other	0.3	23.0	1:8966	1:8900	1:9088

Figure 3: New-risk pyramid.

In addition, Figure 4 shows the old and the new risk pyramid in comparison

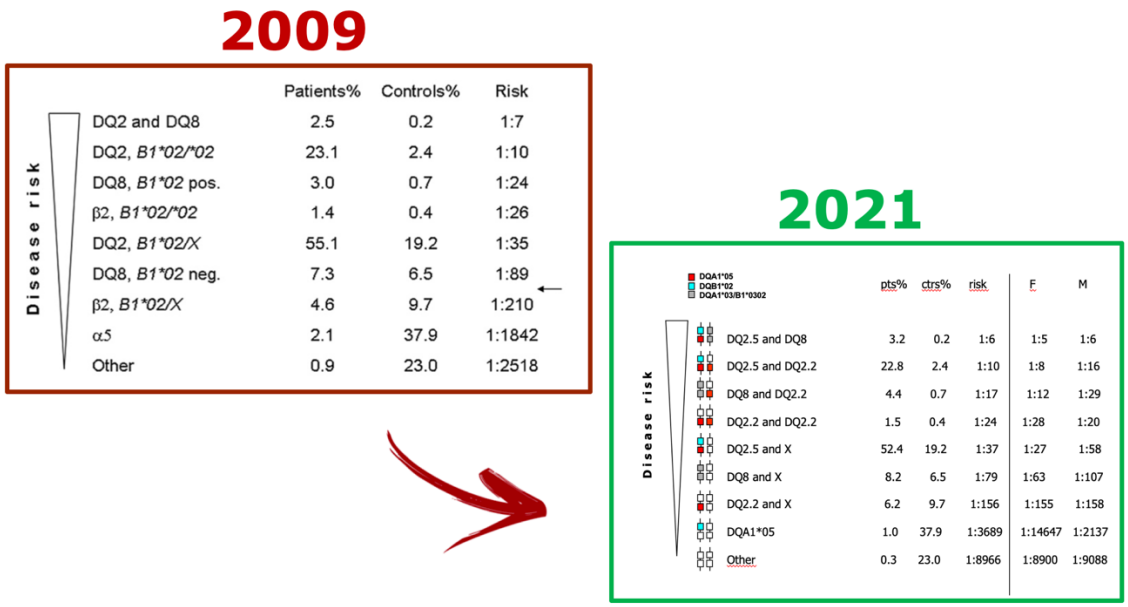


Figure 4: Old and the new risk pyramid in comparison

4. Odds Ratio

The following table (table 7) shows DQ2/DQ8 frequencies and Odds Ratio in Control and CeD patients. Haplotypes have been considered as combined and this Odds Ratio can be considered as a cumulative risk.

	Controls (N= 551)	**CeD (N=778)**	**Odds Ratio**	**C.I. 95%**	**p**
β2	55 (10.0%)	60 (7.7%)	0.75	0.51-1.11	n.s.
α5	209 (37.9%)	8 (1.0%)	0.02	0.008-0.03	<0.001
DQ2	120 (21.8%)	610 (78.4%)	13.04	10.01-16.99	<0.001
DQ2/DQ8	160 (29.0%)	708 (91.0%)	24.72	18.19-33-59	<0.001
DQ2/DQ8/β2	215 (39.0%)	768 (98.7%)	120.02	62.85-229.19	<0.001
DQ2/DQ8/β2/α5	424 (76.9)	776 (99.7%)	116.22	28.61-472.13	<0.001
DQ8	41 (7.4%)	123 (15.8%)	2.34	1.61-3.39	<0.001

Table 7: Cumulative OR

In table 8, Odds Ratio of single haplotypes versus "others" has been calculated. We can observe higher OR in patients with DQ2.5/DQ8 followed by the presence of two copies of DQ B1*02 (one of these, DQ2.5 *in cis*). In this table DQ2.2/DQ2,2 and DQ8/- do not reach significance despite the presence of a positive OR.

	Controls (N= 551)	CeD (N=778)	Odds Ratio	C.I. 95%	p
DQ2.5 and DQ8	1 (0.2%)	25 (3.2%)	18.26	2.46-135.17	0.045
DQ2.5 and DQ2.2	13 (2.4%)	177 (22.8%)	12.18	6.85-21.6	<0.001
DQ8 and DQ2.2	4 (0.7%)	34 (4.4%)	6.24	2.204-17.71	0.0006
DQ2.2 and DQ2.2	2 (0.3%)	12 (1.5%)	4.30	0.95-19.29	n.s.
DQ2.5 and X	106 (19.2%)	408 (52.4%)	4.62	3.58-5.96	<0.001
DQ8 and X	36 (6.5%)	64 (8.2%)	1.28	0.83-1.95	n.s
DQ2.2 and X	53 (9.6%)	48 (6.2%)	0.62	0.41-0.92	0.02
DQ A1*05	209 (37.9%)	8 (1.0%)	0.02	0.01-0.00	<0.001
other	127 (23%)	2 (0.3%)	0.008	0.002-0.03	<0.0001

Table 8: OR of single haplotypes versus other

As is well known, the presence of the positive dimer DQ A1*05-DQ B1*02 is able to hide all other risk alleles. As suggested by literature [111], Odds Ratio of single haplotypes with exclusion of positive dimer DQ A1*05-DQ B1*02 is calculated (table 9). In this table, it is possible to observe that DQ8 and DQ2.2 can be considered as risk alleles with a positive OR when compared to no-risk alleles DQ A1*05 and "others". This table confirms the protective role of DQ A1*05 when it is present in a single copy and of the other haplotypes different from DQ2, DQ8 and DQ B1*02 allele.

	Controls (N= 431)	CeD (N=168)	Odds Ratio	C.I. 95%	P
DQ8 and DQ2.2	4 (0.9%)	34 (20.23%)	26.76	9.32-76.81	<0.0001
DQ8 and X	36 (8.3%)	64 (38.1%)	6.75	4.25-10.71	<0.001
DQ2.2 and X	53 (12.3%)	48 (28.57%)	2.85	1.83-4.43	<0.0001
DQ A1*05	209 (48.5%)	8 (4.8%)	0.05	0.02-0.11	<0.0001
Other	127 (29.5%)	2 (1.2%)	0.02	0.007-0.11	<0.0001
DQ2.2 and DQ2.2	2 (0.4%)	12 (7.14%)	16.5	3.65-74.55	0.0003

Table 9: OR of single haplotypes versus other after the exclusion of positive dimer (DQ A1*05-DQ B1*02)

DISCUSSION

This study characterized HLA polymorphisms in patients with celiac disease and re-defined the risk pyramid of genetically predisposed individuals. From our study cohort, 90.5% of CeD patients' carries at least one copy of the HLA-DQB1*02, which overlaps the frequency of this allele as described by Cabrera [112] in 2018. A high frequency of HLA-DQB1*02 in CeD population, both in association with other haplotypes and in a single copy, has recently been reported by Poddighe (94.94%) [113] and previously by Margaritte-Jeannin [31] in 2004. No gene dose effect has been observed regarding symptoms, age at diagnosis, mucosal atrophy and association with other autoimmune disease, as is also confirmed by a recent meta-analysis [114]. In accordance with what was described by Margaritte-Jeannin [31], DQ2.5 haplotypes (HLA DQA1*05-DQB1*02) have the highest frequency in our cohort, especially DQ2.5 (HLA-DQA1*05-DQB1*02) *in cis* results the most common haplotype followed by a double dose of DQ2.5 and DQ2.5 (HLA-DQA1*05-DQB1*02) *in trans.* Haplotype HLA-DQ8 is present in 15.8% of our celiac population, and only 8.10% among these subjects has this allele as the only risk allele, in a slightly higher prevalence than previously reported [115, 116]. The frequency of associated haplotypes for CeD is not comparable with the risk and the odds ratio of single haplotypes. In our study, indeed, differently from what has previously been reported [31, 36, 117, 118], DQ2.5/DQ8 is the haplotype with the highest risk and odds ratio to develop celiac disease. The presence of DQ2.5/DQ8 shows a risk of 1:6, the odds ratio of cumulative haplotypes is 24.72 ($p< 0.001$) and when the odds ratio is calculated for the presence of both haplotypes versus the others it is 18.26 ($p=0.045$). Differently from previous reports [31, 36, 117, 118] but according to our previous data [104], this result could be explained with the different distribution of haplotypes from North to South in Europe

[119], and Italian data could be more similar to Spanish than to Scandinavian ones [120].

Comparing the new and the old risk pyramid [104], it is possible to observe an increased risk for all haplotypes, especially for DQ2.2 (single "dose" of DQB1*02). It is indeed possible to observe an increase in the risk of developing CeD when DQ2.2 binds to a DQ2.5, a DQ8, and another copy of DQ2.2. More in detail, the presence of a single DQ2.2 carries a 1:156 risk of CeD, which results protective (OR 0.62) when considering all the haplotypes together but becomes a risk factor (OR 2.85) when investigated among the negative (DQ2.5) dimer.

In 2017, De Silvestri [121] drew a risk pyramid where the highest risk was associated with DQ2.5/DQ2.5 haplotypes. Our results show a high risk to develop CeD in presence of DQ2.5/DQ2.5 (1:10,) with an odds ratio of 12.18, comparably to previous reports from Brazil [35].

Analyzing the risk of single haplotypes, it can be easily noticed that DQ8/X and DQ2.2/DQ2.2 and DQ2.2 do not seem to increase the risk. As suggested by literature [111], in order to observe the real risk for this single haplotypes it would be necessary to exclude patients with DQ2.5 and perform a new calculation. In this way, we can recognize the OR of DQ8/X, DQ2.2/DQ2.2 and DQ2.2 (6.75, 16.5 and 2.85, respectively) for the risk to develop CeD. From the analysis of these data it is possible to underline the importance of DQ2.2, especially when associated with the other haplotypes (DQ2.5 or DQ8). The presence of DQ2 with DQ2.2 still carries a 1:10 risk, while the risk for DQ8 with DQ2.2 has increased from 1:24 to 1:17.

Furthermore, it is possible to observe that F:M ratio in our population appears to differ by single haplotypes compared to the cumulative F:M ratio described (2:1) [122]. Indeed, as reported in literature [119], a higher prevalence of males among celiac patients negative for DQ2 and/or DQ8 has been detected. In our study group, we identified a 1:4 F:M ratio for the

presence of other haplotypes and a 1:1 F:M ratio for DQ2.2 subjects. The reason of this difference is not clear, but it could be interesting to hypothesize a role of sexual hormones, or the presence of regulatory genes located of the Y-chromosome on different haplotypes in development of CeD. Concerning the classical haplotype, the well-described F:M ratio is maintained.

Based on our results and on the presence of the DQB1*02 chain (90.5%) in our study population, we can confirm recent findings from literature [123] [124] suggesting, for subjects with a suspected CeD diagnosis, a two-step typing in which searching for the B chain would be the first step, followed by the investigation of DQ8 alleles. Additionally, a new quick test for identification of the beta chains only (DQB1*02 and DQB1*03:02) has recently been developed and could potentially become a useful tool for risk assessment both in subjects with suspected CeD and in the general population, by identifying, although not stratifying, genetic risk factors [125, 126].

The present approach, however, cannot be considered as acceptable for the typing of at-risk subjects (e.g. relatives, or subjects with associated syndromes or comorbidities) as it would not allow a real estimate of the risk of developing CeD and would result in inadequate follow-up, as previously suggested by Murray [117]. Therefore, the complete high-resolution typing of DQ2 and DQ8 becomes necessary considering the new risk pyramid, strengthened by different risk and odds ratios for single haplotype, and even more so in the context of specific at-risk categories.

It is also important to stress that comparing genetic data from different populations is difficult, since every genetic variant has a different prevalence in different settings, and therefore risk calculation needs to be corrected based on the haplotype prevalence itself.

CONCLUSIONS

In conclusion, the new risk pyramid shows an increased risk for all haplotypes; it is reasonable to assert that there is an increased CeD risk in presence of DQ*B102 allele either as a single copy or combination with other alleles. The awareness of the specific single haplotype allows a real estimate of the risk of developing CeD and could be a milestone for personalizing follow-up.

REFERENCES

1. Ludvigsson, J.F., et al., *The Oslo definitions for coeliac disease and related terms.* Gut, 2013. **62**(1): p. 43-52.
2. Catassi, C., S. Gatti, and A. Fasano, *The new epidemiology of celiac disease.* J Pediatr Gastroenterol Nutr, 2014. **59 Suppl 1**: p. S7-9.
3. Singh, P., et al., *Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis.* Clin Gastroenterol Hepatol, 2018. **16**(6): p. 823-836 e2.
4. Kagnoff, M.F., et al., *Possible role for a human adenovirus in the pathogenesis of celiac disease.* J Exp Med, 1984. **160**(5): p. 1544-57.
5. Stene, L.C., et al., *Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study.* Am J Gastroenterol, 2006. **101**(10): p. 2333-40.
6. Zanoni, G., et al., *In celiac disease, a subset of autoantibodies against transglutaminase binds toll-like receptor 4 and induces activation of monocytes.* PLoS Med, 2006. **3**(9): p. e358.
7. Fasano, A., et al., *Vibrio cholerae produces a second enterotoxin, which affects intestinal tight junctions.* Proc Natl Acad Sci U S A, 1991. **88**(12): p. 5242-6.
8. Paterson, B.M., et al., *The safety, tolerance, pharmacokinetic and pharmacodynamic effects of single doses of AT-1001 in coeliac disease subjects: a proof of concept study.* Aliment Pharmacol Ther, 2007. **26**(5): p. 757-66.
9. Tripathi, A., et al., *Identification of human zonulin, a physiological modulator of tight junctions, as prehaptoglobin-2.* Proc Natl Acad Sci U S A, 2009. **106**(39): p. 16799-804.
10. Valitutti, F., S. Cucchiara, and A. Fasano, *Celiac Disease and the Microbiome.* Nutrients, 2019. **11**(10).

11. Roberts, S.E., et al., *Perinatal risk factors and coeliac disease in children and young adults: a record linkage study.* Aliment Pharmacol Ther, 2009. **29**(2): p. 222-31.
12. Decker, E., et al., *Cesarean delivery is associated with celiac disease but not inflammatory bowel disease in children.* Pediatrics, 2010. **125**(6): p. e1433-40.
13. Marild, K., et al., *Pregnancy outcome and risk of celiac disease in offspring: a nationwide case-control study.* Gastroenterology, 2012. **142**(1): p. 39-45 e3.
14. Lionetti, E., et al., *Mode of Delivery and Risk of Celiac Disease: Risk of Celiac Disease and Age at Gluten Introduction Cohort Study.* J Pediatr, 2017. **184**: p. 81-86 e2.
15. Koletzko, S., et al., *Cesarean Section on the Risk of Celiac Disease in the Offspring: The Teddy Study.* J Pediatr Gastroenterol Nutr, 2018. **66**(3): p. 417-424.
16. Girbovan, A., et al., *Is the evidence of breast feeding protection against coeliac disease real?* Allergol Immunopathol (Madr), 2017. **45**(6): p. 616-618.
17. Lionetti, E., et al., *Introduction of gluten, HLA status, and the risk of celiac disease in children.* N Engl J Med, 2014. **371**(14): p. 1295-303.
18. Jabri, B. and L.M. Sollid, *Tissue-mediated control of immunopathology in coeliac disease.* Nat Rev Immunol, 2009. **9**(12): p. 858-70.
19. Hill, I.D., et al., *Guideline for the diagnosis and treatment of celiac disease in children: recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition.* J Pediatr Gastroenterol Nutr, 2005. **40**(1): p. 1-19.
20. Esteve, M., et al., *Spectrum of gluten-sensitive enteropathy in first-degree relatives of patients with coeliac disease: clinical relevance of lymphocytic enteritis.* Gut, 2006. **55**(12): p. 1739-45.

21. Wessels, M.M.S., et al., *Towards an individual screening strategy for first-degree relatives of celiac patients.* Eur J Pediatr, 2018. **177**(11): p. 1585-1592.

22. Singh, P., et al., *Risk of Celiac Disease in the First- and Second-Degree Relatives of Patients With Celiac Disease: A Systematic Review and Meta-Analysis.* Am J Gastroenterol, 2015. **110**(11): p. 1539-48.

23. Nistico, L., et al., *Concordance, disease progression, and heritability of coeliac disease in Italian twins.* Gut, 2006. **55**(6): p. 803-8.

24. Kuja-Halkola, R., et al., *Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins.* Gut, 2016. **65**(11): p. 1793-1798.

25. Brown, N.K., et al., *A Clinician's Guide to Celiac Disease HLA Genetics.* Am J Gastroenterol, 2019. **114**(10): p. 1587-1592.

26. Betuel, H., et al., *Adult celiac disease associated with HLA-DRw3 and -DRw7.* Tissue Antigens, 1980. **15**(3): p. 231-8.

27. DeMarchi, M., et al., *Two HLA-D and DR alleles are associated with coeliac disease.* Tissue Antigens, 1979. **14**(4): p. 309-16.

28. Keuning, J.J., et al., *HLA-DW3 associated with coeliac disease.* Lancet, 1976. **1**(7958): p. 506-8.

29. Solheim, B.G., et al., *Strong association between an HLA-Dw3 associated B cell alloantigen and dermatitis herpetiformis.* Tissue Antigens, 1977. **10**(2): p. 114-8.

30. Tosi, R., et al., *Evidence that celiac disease is primarily associated with a DC locus allelic specificity.* Clin Immunol Immunopathol, 1983. **28**(3): p. 395-404.

31. Margaritte-Jeannin, P., et al., *HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on Coeliac Disease.* Tissue Antigens, 2004. **63**(6): p. 562-7.

32. Mungall, A.J., et al., *The DNA sequence and analysis of human chromosome 6.* Nature, 2003. **425**(6960): p. 805-11.

33. Habig, D.F., et al., *Donor-specific antibody to trans-encoded donor HLA-DQ heterodimer.* Hum Immunol, 2015. **76**(8): p. 587-90.

34. Cecilio, L.A. and M.W. Bonatto, *The prevalence of HLA DQ2 and DQ8 in patients with celiac disease, in family and in general population.* Arq Bras Cir Dig, 2015. **28**(3): p. 183-5.

35. Almeida, L.M., et al., *Presence of DQ2.2 Associated with DQ2.5 Increases the Risk for Celiac Disease.* Autoimmune Dis, 2016. **2016**: p. 5409653.

36. Liu, E., et al., *Risk of pediatric celiac disease according to HLA haplotype and country.* N Engl J Med, 2014. **371**(1): p. 42-9.

37. Fallang, L.E., et al., *Differences in the risk of celiac disease associated with HLA-DQ2.5 or HLA-DQ2.2 are related to sustained gluten antigen presentation.* Nat Immunol, 2009. **10**(10): p. 1096-101.

38. Bodd, M., et al., *T-cell response to gluten in patients with HLA-DQ2.2 reveals requirement of peptide-MHC stability in celiac disease.* Gastroenterology, 2012. **142**(3): p. 552-61.

39. Ting, Y.T., et al., *A molecular basis for the T cell response in HLA-DQ2.2 mediated celiac disease.* Proc Natl Acad Sci U S A, 2020. **117**(6): p. 3063-3073.

40. Tollefsen, S., et al., *HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease.* J Clin Invest, 2006. **116**(8): p. 2226-36.

41. Schuppan, D., Y. Junker, and D. Barisani, *Celiac disease: from pathogenesis to novel therapies.* Gastroenterology, 2009. **137**(6): p. 1912-33.

42. Wolters, V.M. and C. Wijmenga, *Genetic background of celiac disease and its clinical implications.* Am J Gastroenterol, 2008. **103**(1): p. 190-5.

43. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.* Nat Genet, 1995. **11**(3): p. 241-7.

44. Amundsen, S.S., et al., *A comprehensive screen for SNP associations on chromosome region 5q31-33 in Swedish/Norwegian celiac disease families.* Eur J Hum Genet, 2007. **15**(9): p. 980-7.

45. Greco, L., et al., *Genome search in celiac disease.* Am J Hum Genet, 1998. **62**(3): p. 669-75.

46. Latiano, A., et al., *Analysis of candidate genes on chromosomes 5q and 19p in celiac disease.* J Pediatr Gastroenterol Nutr, 2007. **45**(2): p. 180-6.

47. Babron, M.C., et al., *Meta and pooled analysis of European coeliac disease data.* Eur J Hum Genet, 2003. **11**(11): p. 828-34.

48. Ling, V., et al., *Assembly and annotation of human chromosome 2q33 sequence containing the CD28, CTLA4, and ICOS gene cluster: analysis by computational, comparative, and microarray approaches.* Genomics, 2001. **78**(3): p. 155-68.

49. Amundsen, S.S., et al., *Genetic analysis of the CD28/CTLA4/ICOS (CELIAC3) region in coeliac disease.* Tissue Antigens, 2004. **64**(5): p. 593-9.

50. Holopainen, P., et al., *Candidate gene region 2q33 in European families with coeliac disease.* Tissue Antigens, 2004. **63**(3): p. 212-22.

51. Van Belzen, M.J., et al., *A major non-HLA locus in celiac disease maps to chromosome 19.* Gastroenterology, 2003. **125**(4): p. 1032-41.

52. Monsuur, A.J., et al., *Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect.* Nat Genet, 2005. **37**(12): p. 1341-4.

53. Garner, C.P., et al., *Replication of celiac disease UK genome-wide association study results in a US population.* Hum Mol Genet, 2009. **18**(21): p. 4219-25.

54. Romanos, J., et al., *Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease.* Gastroenterology, 2009. **137**(3): p. 834-40, 840 e1-3.

55. Einarsdottir, E., et al., *Genome-wide analysis of extended pedigrees confirms IL2-IL21 linkage and shows additional regions of interest potentially influencing coeliac disease risk.* Tissue Antigens, 2011. **78**(6): p. 428-37.

56. Plaza-Izurieta, L., et al., *Revisiting genome wide association studies (GWAS) in coeliac disease: replication study in Spanish population and expression analysis of candidate genes.* J Med Genet, 2011. **48**(7): p. 493-6.

57. Trynka, G., et al., *Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.* Nat Genet, 2011. **43**(12): p. 1193-201.

58. Withoff, S., et al., *Understanding Celiac Disease by Genomics.* Trends Genet, 2016. **32**(5): p. 295-308.

59. Romanos, J., et al., *Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants.* Gut, 2014. **63**(3): p. 415-22.

60. Tack, G.J., et al., *The spectrum of celiac disease: epidemiology, clinical aspects and treatment.* Nat Rev Gastroenterol Hepatol, 2010. **7**(4): p. 204-13.

61. Kumar, V., C. Wijmenga, and S. Withoff, *From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases.* Semin Immunopathol, 2012. **34**(4): p. 567-80.

62. Megiorni, F. and A. Pizzuti, *HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing*. J Biomed Sci, 2012. **19**: p. 88.

63. Green, P.H., *Celiac disease: how many biopsies for diagnosis?* Gastrointest Endosc, 2008. **67**(7): p. 1088-90.

64. Dai, Y., et al., *Celiac Disease: Updates on Pathology and Differential Diagnosis.* Adv Anat Pathol, 2019. **26**(5): p. 292-312.

65. Marsh, M.N., *Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue').* Gastroenterology, 1992. **102**(1): p. 330-54.

66. Oberhuber, G., G. Granditsch, and H. Vogelsang, *The histopathology of coeliac disease: time for a standardized report scheme for pathologists.* Eur J Gastroenterol Hepatol, 1999. **11**(10): p. 1185-94.

67. Corazza, G.R. and V. Villanacci, *Coeliac disease.* J Clin Pathol, 2005. **58**(6): p. 573-4.

68. Villanacci, V., et al., *Histopathology of Celiac Disease. Position Statements of the Italian Group of Gastrointestinal Pathologists (GIPAD-SIAPEC).* Transl Med UniSa, 2020. **23**: p. 28-36.

69. Scott, B.B. and M.S. Losowsky, *Patchiness and duodenal-jejunal variation of the mucosal abnormality in coeliac disease and dermatitis herpetiformis.* Gut, 1976. **17**(12): p. 984-92.

70. Ravelli, A., et al., *How patchy is patchy villous atrophy?: distribution pattern of histological lesions in the duodenum of children with celiac disease.* Am J Gastroenterol, 2010. **105**(9): p. 2103-10.

71. Bonamico, M., et al., *Patchy villous atrophy of the duodenum in childhood celiac disease.* J Pediatr Gastroenterol Nutr, 2004. **38**(2): p. 204-7.

72. McCarty, T.R., et al., *Efficacy of duodenal bulb biopsy for diagnosis of celiac disease: a systematic review and meta-analysis.* Endosc Int Open, 2018. **6**(11): p. E1369-E1378.

73. Valitutti, F., et al., *Mapping histologic patchiness of celiac disease by push enteroscopy.* Gastrointest Endosc, 2014. **79**(1): p. 95-100.

74. Trovato, C.M., et al., *The Challenge of Treatment in Potential Celiac Disease.* Gastroenterol Res Pract, 2019. **2019**: p. 8974751.

75. Biagi, F., et al., *Prevalence and natural history of potential celiac disease in adult patients.* Scand J Gastroenterol, 2013. **48**(5): p. 537-42.

76. Lionetti, E., et al., *Prevalence and natural history of potential celiac disease in at-family-risk infants prospectively investigated from birth.* J Pediatr, 2012. **161**(5): p. 908-14.

77. Auricchio, R., et al., *Progression of Celiac Disease in Children With Antibodies Against Tissue Transglutaminase and Normal Duodenal Architecture.* Gastroenterology, 2019. **157**(2): p. 413-420 e3.

78. Piccialli, F., et al., *Precision medicine and machine learning towards the prediction of the outcome of potential celiac disease.* Sci Rep, 2021. **11**(1): p. 5683.

79. Popp, A. and M. Maki, *Gluten-Induced Extra-Intestinal Manifestations in Potential Celiac Disease-Celiac Trait.* Nutrients, 2019. **11**(2).

80. Rubio-Tapia, A., et al., *ACG clinical guidelines: diagnosis and management of celiac disease.* Am J Gastroenterol, 2013. **108**(5): p. 656-76; quiz 677.

81. Bai, J.C., et al., *World Gastroenterology Organisation global guidelines on celiac disease.* J Clin Gastroenterol, 2013. **47**(2): p. 121-6.

82. Murch, S., et al., *Joint BSPGHAN and Coeliac UK guidelines for the diagnosis and management of coeliac disease in children.* Arch Dis Child, 2013. **98**(10): p. 806-11.

83. Meeuwisse, G., *Diagnostic criteria in coeliac disease.* Acta Paediatr. Scand., 1970. **59**: p. 461-463.

84. *Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition.* Arch Dis Child, 1990. **65**(8): p. 909-11.

85. Husby, S., et al., *European Society for Pediatric Gastroenterology, Hepatology, and Nutrition guidelines for the diagnosis of coeliac disease.* J Pediatr Gastroenterol Nutr, 2012. **54**(1): p. 136-60.

86. Husby, S., et al., *European Society Paediatric Gastroenterology, Hepatology and Nutrition Guidelines for Diagnosing Coeliac Disease 2020.* J Pediatr Gastroenterol Nutr, 2020. **70**(1): p. 141-156.

87. Singh, P., et al., *Titers of anti-tissue transglutaminase antibody correlate well with severity of villous abnormalities in celiac disease.* J Clin Gastroenterol, 2015. **49**(3): p. 212-7.

88. Trovato, C.M., et al., *Are ESPGHAN "biopsy-sparing" guidelines for celiac disease also suitable for asymptomatic patients?* Am J Gastroenterol, 2015. **110**(10): p. 1485-9.

89. Werkstetter, K.J., et al., *Accuracy in Diagnosis of Celiac Disease Without Biopsies in Clinical Practice.* Gastroenterology, 2017. **153**(4): p. 924-935.

90. Salmi, T.T., et al., *Endomysial antibody-negative coeliac disease: clinical characteristics and intestinal autoantibody deposits.* Gut, 2006. **55**(12): p. 1746-53.

91. Trovato, C.M., et al., *Diagnostic Value of Persistently Low Positive TGA-IgA Titers in Symptomatic chIldren with Suspected Celiac Disease.* Journal of Pediatric Gastroenterology & Nutrition, 2021. **Publish Ahead of Print**.

92. Peraaho, M., et al., *Oats can diversify a gluten-free diet in celiac disease and dermatitis herpetiformis.* J Am Diet Assoc, 2004. **104**(7): p. 1148-50.

93. Gatti, S., et al., *Oats in the diet of children with celiac disease: preliminary results of a double-blind, randomized, placebo-controlled multicenter Italian study.* Nutrients, 2013. **5**(11): p. 4653-64.

94. Koskinen, O., et al., *Oats do not induce systemic or mucosal autoantibody response in children with coeliac disease.* J Pediatr Gastroenterol Nutr, 2009. **48**(5): p. 559-65.

95. Caio, G., et al., *Therapeutic options for coeliac disease: What else beyond gluten-free diet?* Dig Liver Dis, 2020. **52**(2): p. 130-137.
96. Gass, J., et al., *Combination enzyme therapy for gastric digestion of dietary gluten in patients with celiac sprue.* Gastroenterology, 2007. **133**(2): p. 472-80.
97. Mitea, C., et al., *Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease.* Gut, 2008. **57**(1): p. 25-32.
98. McCarville, J.L., et al., *BL-7010 demonstrates specific binding to gliadin and reduces gluten-associated pathology in a chronic mouse model of gliadin sensitivity.* PLoS One, 2014. **9**(11): p. e109972.
99. Khaleghi, S., et al., *The potential utility of tight junction regulation in celiac disease: focus on larazotide acetate.* Therap Adv Gastroenterol, 2016. **9**(1): p. 37-49.
100. Heyman, M., et al., *Intestinal permeability in coeliac disease: insight into mechanisms and relevance to pathogenesis.* Gut, 2012. **61**(9): p. 1355-64.
101. Sulic, A.M., et al., *Transglutaminase as a therapeutic target for celiac disease.* Expert Opin Ther Targets, 2015. **19**(3): p. 335-48.
102. Alhassan, E., et al., *Novel Nondietary Therapies for Celiac Disease.* Cell Mol Gastroenterol Hepatol, 2019. **8**(3): p. 335-345.
103. Anderson, R.P. and B. Jabri, *Vaccine against autoimmune disease: antigen-specific immunotherapy.* Curr Opin Immunol, 2013. **25**(3): p. 410-7.
104. Megiorni, F., et al., *HLA-DQ and risk gradient for celiac disease.* Hum Immunol, 2009. **70**(1): p. 55-9.
105. Thomson, G., *Mapping disease genes: family-based association studies.* Am J Hum Genet, 1995. **57**(2): p. 487-98.
106. Falk, C.T., N.R. Mendell, and P. Rubinstein, *Effect of population associations and reduced penetrance on observed and expected genotype*

frequencies in a simple genetic model: application to HLA and insulin dependent diabetes mellitus. Ann Hum Genet, 1983. **47**(2): p. 161-5.

107. Khoury, M.J., *Case-parental control method in the search for disease-susceptibility genes.* Am J Hum Genet, 1994. **55**(2): p. 414-5.

108. Archimbaud, E., et al., *Prospective genetically randomized comparison between intensive postinduction chemotherapy and bone marrow transplantation in adults with newly diagnosed acute myeloid leukemia.* J Clin Oncol, 1994. **12**(2): p. 262-7.

109. Field, L.L., *Genes predisposing to IDDM in multiplex families.* Genet Epidemiol, 1989. **6**(1): p. 101-6.

110. Thomson, G., et al., *HLA and insulin gene associations with IDDM.* Genet Epidemiol, 1989. **6**(1): p. 155-60.

111. Polvi, A., et al., *HLA-DQ2-negative celiac disease in Finland and Spain.* Hum Immunol, 1998. **59**(3): p. 169-75.

112. Cabrera, C.M., I.M. Mendez-Lopez, and A. Caballero, *Risk variation in celiac disease in a population from Southern Spain: evaluating the influence of the DQB1*02:02 allele frequency.* Scand J Gastroenterol, 2018. **53**(3): p. 266-272.

113. Poddighe, D., et al., *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**(12): p. 1365-1381.

114. Bajor, J., et al., *Classical celiac disease is more frequent with a double dose of HLA-DQB1*02: A systematic review with meta-analysis.* PLoS One, 2019. **14**(2): p. e0212329.

115. van Heel, D.A., et al., *Genetics in coeliac disease.* Best Pract Res Clin Gastroenterol, 2005. **19**(3): p. 323-39.

116. Karell, K., et al., *HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European*

Genetics Cluster on Celiac Disease. Hum Immunol, 2003. **64**(4): p. 469-77.

117. Murray, J.A., et al., *HLA DQ gene dosage and risk and severity of celiac disease.* Clin Gastroenterol Hepatol, 2007. **5**(12): p. 1406-12.

118. Stankovic, B., et al., *HLA genotyping in pediatric celiac disease patients.* Bosn J Basic Med Sci, 2014. **14**(3): p. 171-6.

119. Tinto, N., 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**(9): p. e0138324.

120. Lionetti, E. and C. Catassi, *Co-localization of gluten consumption and HLA-DQ2 and -DQ8 genotypes, a clue to the history of celiac disease.* Dig Liver Dis, 2014. **46**(12): p. 1057-63.

121. De Silvestri, A., et al., *HLA-DQ genetics in children with celiac disease: a meta-analysis suggesting a two-step genetic screening procedure starting with HLA-DQ beta chains.* Pediatr Res, 2018. **83**(3): p. 564-572.

122. Megiorni, F., et al., *HLA-DQ and susceptibility to celiac disease: evidence for gender differences and parent-of-origin effects.* Am J Gastroenterol, 2008. **103**(4): p. 997-1003.

123. Poddighe, D., et al., *HLA-DQB1*02 allele in children with celiac disease: Potential usefulness for screening strategies.* Int J Immunogenet, 2019. **46**(5): p. 342-345.

124. Capittini, C., et al., *Relevance of HLA-DQB1*02 Allele in the Genetic Predisposition of Children with Celiac Disease: Additional Cues from a Meta-Analysis.* Medicina (Kaunas), 2019. **55**(5).

125. Verma, A.K., 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**(6): p. 311-316.

126. Gatti, S., et al., *Increased Prevalence of Celiac Disease in School-age Children in Italy.* Clin Gastroenterol Hepatol, 2020. **18**(3): p. 596-603.