International Journal of Celiac Disease, 2017, Vol. 5, No. 3, 86-94
Available online at http://pubs.sciepub.com/ijcd/5/3/3

DOI:10.12691/ijcd-5-3-3

Celiac Disease: The Evolutionary Paradox

Kayla Morrell[1], Melissa K. Melby[2,3,*]

[1]Department of Biological Sciences, University of Delaware, Newark, United States
[2]Department of Anthropology, University of Delaware, Newark, United States
[3]College of Health Sciences, University of Delaware, Newark, United States
*Corresponding author: mmelby@udel.edu

Abstract Celiac disease (CD), an autoimmune disorder triggered by gluten ingestion, negatively affects individuals' health if left untreated. In individuals with the requisite genes, CD destroys the intestinal lining because specific peptide fragments of gluten (protein found in wheat and related grains) bind to the receptor proteins coded for by HLA-DQ2 and DQ8 genes, thereby causing an autoimmune response that damages intestinal cells and results in malnutrition. Because the disease has a genetic basis and can lead to impaired reproductive functioning and death, natural selection should lead to a decrease in CD prevalence over time. However, evidence suggesting CD increases in some populations contradicts this hypothesis, resulting in the 'Celiac Disease Evolutionary Paradox.' The worldwide average prevalence rate is around 1%, although rates of up to 5% have been observed in certain populations. Maintenance or increase in the frequency of CD-predisposing genes in certain populations suggests potential evolutionary benefits for those genes, which may exhibit antagonistic pleiotropy, such as being beneficial for infectious disease but detrimental for chronic disease such as CD. Recent dietary and environmental changes (including dietary gluten exposure, breastfeeding duration and the intestinal microbiome, and immune function) may have led to discordance with the Environment of Evolutionary Adaptedness, thus contributing to increases in CD. Consideration of potential benefits of CD-risk gene alleles, in conjunction with a deeper understanding of environmental factors exerting positive and negative selection on those alleles, may help to explain population variation in CD prevalence rates, and shed light on the complex gene-environment interactions influencing this devastating disease.

Keywords: *celiac disease, evolution, antagonistic pleiotropy, discordance, environment*

Cite This Article: Kayla Morrell, and Melissa K. Melby, "Celiac Disease: The Evolutionary Paradox." *International Journal of Celiac Disease*, vol. 5, no. 3 (2017): 86-94. doi: 10.12691/ijcd-5-3-3.

1. Introduction

Celiac disease (CD), an autoimmune disease characterized by intestinal inflammatory responses to gluten (proteins found in wheat and related grains), has been increasing worldwide, and increases are not solely due to increased detection [1]. Untreated CD impacts growth and reproduction, thus individuals with genes for CD are expected to be subject to strong selective pressures. High rates of CD may be an example of evolutionary discordance, or mismatch between genes and the contemporary environment. In the absence of dietary gluten in an environment such as the pre-agricultural Environment of Evolutionary Adaptedness (EEA), individuals would not have suffered the negative effects of CD due to lack of trigger. However, in an environment of high dietary gluten intake (eg, western diets that are heavily wheat-based), individuals with genes predisposing them to CD would be expected to have reduced fitness. Thus as dietary gluten increases, over time gene frequencies of CD are expected to decrease. However, despite detrimental effects on many aspects of health, CD prevalence appears to remain constant or increase regardless of the duration of gluten exposure in a population, resulting in the 'Celiac Disease Evolutionary Paradox' [2]. We examine antagonistic pleiotropy and potential benefits of CD genes, as well as environmental factors such as diet, breastfeeding and the microbiome that have changed since the origin of agriculture, to investigate this paradox.

2. Physiology of Celiac Disease

Celiac disease, also called gluten-sensitive enteropathy, is a chronic inflammatory disease of the small intestine that can develop in genetically-susceptible individuals at any point in their lifetime. Most CD patients carry one of two major histocompatibility class (MHC) II human leukocyte antigens (HLA) that predispose to CD, HLA-DQ2 or HLA-DQ8, which appear to be required for the development of CD [3]. The HLA complex is involved with distinguishing self-proteins from proteins produced by viruses and bacteria. In these susceptible individuals, certain triggers like exposure to gluten proteins activate the onset and symptoms of CD. However, many people who carry these genes do not develop CD. Thus, the genetic component is a necessary but not sufficient requirement of the disease.

Celiac disease manifests itself in a wide range of symptoms [4]. In children, it results in growth problems, short stature, decreased appetite with failure to gain weight, chronic constipation, abdominal bloating and pain, fatigue, and irritability. In adults, the symptoms include iron-deficiency anemia, fatigue, bone or joint pain, depression, anxiety, seizures or migraines, missed menstrual periods, infertility, or skin rashes [4]. If CD is left undiagnosed or untreated long enough, the individual can become increasingly malnourished, suffer severe vitamin deficiencies, or develop cancers relating to the digestive tract. Additionally, those with CD may develop secondary lactase deficiency, which results from injury to the bowel resulting in decreased lactase in the lining of the duodenum [5]. The symptoms are very broad and could be indicative of many different diseases, and therefore celiac diagnoses are often missed.

Gluten consumption is required to trigger the various symptoms of CD, including atrophy of the intestinal wall. When gluten is consumed in individuals who have developed the disease, it damages the intestines. Although gluten is broken down due to the activity of the enzyme pepsin in the stomach, it still exists in the intestine in relatively large peptides. This remaining gluten protein, Glia-α9, binds with high affinity to the HLA-DQ receptor proteins coded for by the HLA-DQ2 and HLA-DQ8 genes. The gliadin causes an autoimmune response of T-cells whereby they attack the villi of the small intestine. This gluten specific T-cell response is characterized by secretion of cytokines that drive local inflammation within the small intestine. The strength of the affinity between gluten and HLA-DQ2 and HLA-DQ8 explains the almost exclusive development and appearance in individuals with these genes [6].

Peptides are presented by antigen-presenting cells to gluten-sensitive T-cells, which vary in HLA-DQ antigen receptors as shown in Figure 1. The main CD epitope is Glia-α9. When the gluten fragment binds to HLA-DQ2 or HLA-DQ8 receptors, it results in inflammation in individuals with CD due to T-cells [8]. This T-cell infiltration results in destruction of the small intestines by flattening the villi and causing atrophy by accelerated shedding of the epithelial cells. Epithelial cell production increases to compensate, but is insufficient and the new cells are not mature enough to sustain absorptive functions, leading to malabsorption [6].

Treatment for CD is complete avoidance of gluten. This non-medical remedy, in conjunction with CD's non-specific symptoms (eg, similar to Irritable Bowel Syndrome), means that many people self-diagnose [9]. There are also limited, if any, symptomatic differences between CD and non-celiac gluten sensitivity, with the main distinguishing factor being the genetic predisposition of CD [10]. In celiac individuals who are on a gluten-free diet, there is evidence that disturbances remain in the epithelial tissue of the small intestine [11]. Even if the individual does not consume gluten, the intestines do not always return to a completely healthy state. The gluten-specific T-cell response may also drive the antibody response. To diagnose CD, a blood test is performed for antibodies specific to CD, usually immunoglobin A (IgA) anti-tissue transglutaminase antibodies. Elevated levels of these antibodies are considered indicative of CD. High levels of anti-gliadin antibodies (AgA) may also signify CD. It is possible that detectable antibody levels precede intestinal inflammation, suggesting that the disease could be diagnosed prior to intestinal damage and associated gastrointestinal symptoms [12].

The presence of certain intestinal bacteria may act as an additional environmental factor predisposing individuals to CD. Celiac disease is associated with the overgrowth of harmful bacteria and decreased protective bacteria [13]. Patients with active CD exhibit differences in their intestinal microbiome, characterized by overall higher total bacterial counts, increased levels of harmful Gram-negative bacteria (e.g., *Bacteroides, Prevotella, E. coli*), but decreased numbers of beneficial Gram-positive bacteria (e.g., *Lactobacilli, Bifidobacteria*) [13]. Breastfeeding increases both *Bifidobacteria* and *Lactobacilli* [14], supporting the observed inverse epidemiological relationship (discussed further below) between breastfeeding and CD risk [15]. *Bifidobacteria* have been observed to reduce the severity of the toxic effects of gluten in celiac patients [16]. The ratio of beneficial to harmful bacteria was significantly lower in individuals with CD compared to healthy controls. Individuals with CD that were following a gluten-free diet had higher ratios of beneficial to harmful bacteria than those who were not following the diet, however the ratio was still lower than in healthy individuals [13]. Interestingly, putting healthy individuals on a gluten-free diet leads to a reduction in beneficial bacteria, contrary to the results in patients with CD. Even following a gluten-free diet treatment, individuals with CD were not able to fully restore their intestinal microbiome [17].

HLA-DQ proteins range from DQ1-9, with the most common shown above. They are antigen receptors on the surface of cells. HLA-DQ2 and HLA-DQ8 bind the Glia-α9 epitope more strongly, triggering an immune response resulting in inflammation in predisposed individuals.

Figure 1. Schematic of gluten binding to different HLA-DQ antigen receptors (modified from [7]). Gluten-sensitive T-cells have different HLA-DQ variants of antigen receptors, represented by the different colors in the figure below. The gray substrate is Glia-α9, the primary CD epitope

3. Genetics of Celiac Disease

Celiac disease has high heritability and is considered a polygenic disease with a complex non-Mendelian pattern of inheritance that includes major histocompatibility complex (MHC) and non-MHC genes. The MHC is located on the sixth chromosome at 6p21 and contains hundreds of genes with immunological functions [18]. Although CD is considered polygenic, most individuals with CD have either HLA-DQ2 or HLA-DQ8, thus these genes appear to be sufficient to confer disease risk, with other genes contributing to phenotypic heterogeneity [3]. Most (90%) patients with CD carry the HLA-DQ2 variant, DQ2.5 [19], which is encoded by DA1*501 and DQB1*0201. The gluten-reactive T cells recognize a diverse set of gluten epitopes, or part of the antigen recognized by specific antibodies, in the binding sites of receptors coded for by DQ2.5 or DQ8 alleles, but not other MHC class II molecules. For example, another HLA-DQ2 molecule DQ2.2 is highly homologous to D2.5, but has very low risk for CD on its own. This suggests that the variation in the alpha-chain of DQ2 has considerable effect on the risk of CD. DQ2.5 and DQ2.2 differ in CLIP1 content, which encodes a protein that links endocytic vesicles to microtubules, suggesting a difference in the peptide-binding abilities of the two molecules. Phenylalanine (DQ2.2) leads to lower binding stability for most peptide ligands [20].

Many of the additional risk genes for CD also exhibit antagonistic pleiotropy. The CD Evolutionary Paradox is likely explained by the benefit-cost ratio of the fitness benefits that come at a negative cost in high gluten environments. The cost would not be as detrimental in individuals in which the disease was never activated (ie, no exposure to gluten), allowing the genes to be passed on without negatively affecting the reproductive fitness of individuals.

More than 99% of CD cases express MHC class II molecules HLA-DQ2.5, HLA-DQ8 or HLA-DQ 2.2 [19]. About 95% of CD cases are caused by HLA-DQ2.2 and 2.5, while most of the remaining 5% of CD patients have HLA-DQ8. Although HLA-DQ2 is found in about 40% of people of European descent, only about 1% of the population develops the disease, suggesting the additional genetic and/or environmental factors are required [2]. Less than 1% of Western Europeans with CD develop it without DQ2.2, 2.5 or 8, which are considered major genes. Approximately twenty other risk locations exist on the sixth chromosome and increase the risk of developing CD as minor genes. In genome-wide association studies, other non-HLA candidate genes have been identified, many of which are shared between other immune-mediated diseases such as rheumatoid arthritis [21] and Type 1 diabetes [22].

The reported risks of CD among first-degree relatives vary from 5-20% [18]. Being genetically predisposed is required but not sufficient for development of the disease because there are environmental factors that can delay or trigger the onset of the disease. Roughly 75-83% of monozygotic twins both develop CD and 30% of HLA identical siblings do, emphasizing the important role of the environment. However, CD is the autoimmune disorder with the highest monozygotic twin concordance [23]. Due to the complex polygenic inheritance of CD, combined with the role of the environment, CD falls into the category of a threshold trait. A threshold trait is one in which the risk of developing the disease is multifactorial, determined by genetic and environmental factors, and only observable when the risk is above the threshold value [24].

While there is great variation in clinical manifestations (ranging from no symptoms to gastrointestinal distress and symptoms, to severe intestinal damage), age of onset, and antibody levels, the presence of genes conferring risk does not correlate with which symptoms manifest. That is, each specific gene adds to the likelihood of developing CD but does not determine which symptoms appear [25].

4. Epidemiology of Celiac Disease

Worldwide prevalence of CD is roughly 1% [26]. This prevalence rate was originally observed in populations of European descent, but similar frequencies exist in Middle East, India and North African populations, with slight regional differences [27,28,29]. Epidemiological changes in the prevalence of CD are influenced not only by changes in CD-predisposing gene frequencies, but also environmental changes. The main environmental change is increased gluten in the diet, following the origin of agriculture and domestication of grains such as wheat and barley. Clinical manifestations of CD are most common in populations whose diet contains wheat, and other gluten-containing grains. Thus, CD rarely occurs among populations that rely on rice or corn as staples [30].

Environmental changes that have occurred since the origin of agriculture have changed disease epidemiology considerably [31], leading to a discordance between our contemporary environment and the EEA. Pre-agriculture, human diets contained little grain or dairy. The Neolithic Transition was characterized by domesticated animals and grains, and thus increase in gluten from wheat, rye, and barley, and lactose from milk. In the case of CD, the most relevant aspect of the environment is gluten in the diet, but factors such as breastfeeding and the microbiome and timing of gluten exposure also likely play a large role. Increased ingestion of cereals such as wheat might have influenced the development of CD less directly as well. For example, gliadin from wheat exhibits lectin activity, which has been shown to induce HLA class II molecules, likely through stimulation of interferon-γ (also induced by viral infections) [32].

While changes in environmental triggers influence the epidemiology of CD, changes in the immune system (the body's response to the environmental triggers) also influence the epidemiological pattern. Both the intestinal microbiome and breastfeeding behavior have been shown to influence the development of the immune system [33] through effects on intestinal permeability and microbiome, and may thereby moderate the response to gluten of individuals with genetic predisposition to CD. Breastfeeding may contribute to healthy gut mucosa with intact tight junctions, which generally prevent passage of inappropriate macromolecules across the intestinal lining [34]. Furthermore, breast milk, in conjunction with colonization through vaginal delivery of commensal bacteria, support the development of a healthy intestinal microbiome which in turns contribute to healthy gastrointestinal innate immune

function [35]. Rates of CD thus reflect the evolutionary discordance [36,37] between our evolutionary environment (no gluten pre-agriculture, extended breastfeeding and different microbiome) and our modern environment (diets with increasing gluten, short or no breastfeeding, and microbiome influenced by multiple factors).

While CD has a very strong genetic component, there are additional factors that influence the likelihood of developing the disease, as well as some factors that could be direct triggers, such as hormonal changes of puberty and immunological changes of pregnancy, or major stressors [23]. As is true of most autoimmune diseases, women are more likely to develop CD than men because women have stronger humoral and cellular immune systems than men, manifested by higher CD4+ T-cell counts [38]. Support for the importance of the microbiome derives from observations that children delivered by cesarean section have higher rates of developing CD, possibly because they are not exposed to the same colonizing bacteria as children delivered vaginally [14]. Timing of gluten exposure, relative to immune function development, may be important as the risk of developing CD appears higher in genetically susceptible individuals who consume more gluten early in life. However, when gluten is introduced to infants between four and six months of age while they are still breastfeeding, the onset of the disease is delayed or possibly avoided entirely. This may be in part due to breastmilk's influence on intestinal permeability and support of healthy microflora. Breastfeeding in individuals with HLA-DQ2 or DQ8 has been shown to reduce the genotype-difference in microbiota, which may explain the protective association of breast milk in CD [17]. One study found that gram-negative rod bacteria were present in the intestines of CD patients but not in healthy controls, suggesting that bacterial infection could trigger gluten intolerance in children [39]. A study in Sweden found that having three or more infectious episodes early in life increased CD risk after adjusting for infant feeding and socioeconomic status; there was found to be no significant risk due to antibiotic treatment. Additionally, there was a synergistic effect between early infections and amount of gluten consumed increasing CD risk, which was more pronounced in infants who began consuming gluten following the discontinuation of breast feeding [40]. There is no one environmental factor aside from exposure to gluten that directly causes CD, but rather each factor leads to an increased risk of developing the disease (Figure 2).

It is difficult to compare rates of CD over time or across populations because of diagnostic differences. Individuals self-diagnosing also leads to potential variation CD rates [10]. A true increase in disease rates is one in which diagnostic method remains the same and the difference in rate can be attributed solely to changes in disease prevalence. Table 1 shows the change of CD rates in response to certain diets or behavior. In Table 1, the United States values represent a true increase, relying on similar diagnostic measures, while that is less clear in the other studies.

A study was performed in Sweden to explore the effect of breastfeeding and timing of introduction of dietary gluten into the diet of children at risk for CD. Between 1984 and 1996, Sweden experienced an epidemic of CD in children younger than two years old, with an abrupt increase to 2.9%, and then a decline to 0.79% following changes in dietary patterns of infants [41]. The study found that infants who were still being breastfed at the time of introduction of gluten had a lower risk of developing CD. During the Swedish epidemic, the incidence rates increased 3-fold to a rate that was higher than any country had seen; there was then an abrupt decline back to the baseline levels of 1980s [15]. The increase in rate was preceded by an increase in the amount of gluten in early weaning foods, later introduction of gluten, and average exclusive breastfeeding of only four months. The decrease in CD incidence followed an increasing duration of exclusive breastfeeding from four months to six months in 1982, then back to four months of exclusive breastfeeding in 1996 with decreased gluten in the diet, and breastfeeding for a total of six months. Swedish infant-feeding patterns shifted from favorable in protecting against gluten-triggered CD, to unfavorable, and back to favorable [15,41]. Although this differs from guidelines to exclusively breastfeed for six months, breastfeeding while introducing gluten into the diet appears key to lowering CD rates in at-risk children.

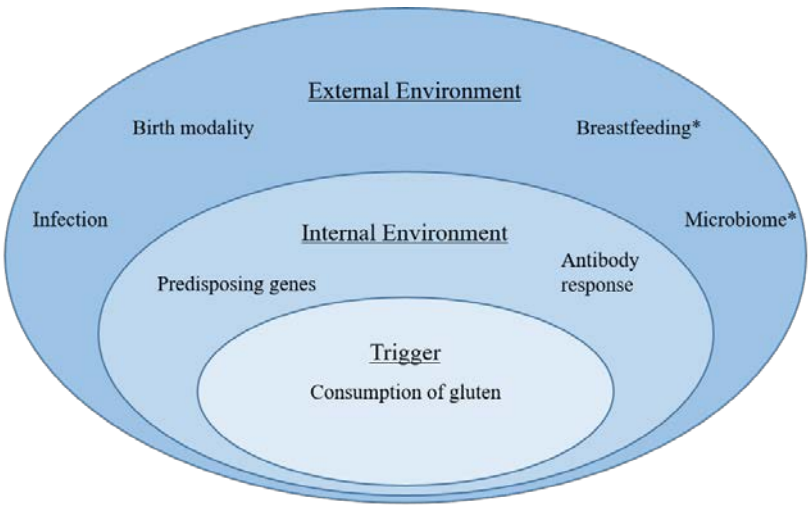

Figure 2. Schematic of levels of factors affecting celiac disease appearance and population prevalence

Table 1. Populations with Documented Changes in CD Rate

Level of Trigger	Location	Then		Now	
		Date	CD Rate	Date	CD Rate
No trigger	Burkina Faso [11]	1900	NP	2008	NP
Increased gluten trigger	United States [42]	Early 1990s	0.02%	2003	0.7%
	Ireland [43]	1700	~0.8%	2006	1.3%
	Saharawi [44]	1900	NP	2015	5.6%
Increased gluten trigger and decreased protective factors	Sweden [15]	1985-1996	2.2-2.9% in children	2013	0.79%

NP = no prevalence.

5. Evolution of Celiac Disease Risk

Given the many negative health effects of CD that impact reproductive success, there should be strong negative selection against genes that make people susceptible in an environment with gluten, and increasing exposure to dietary gluten should act as a selective pressure leading to decreased gene frequencies [2]. Thus, populations with longer exposure to gluten should have lower frequencies of CD due to negative selection depending on initial frequencies of risk alleles. However, they do not, resulting in the 'Celiac Disease Evolutionary Paradox' [2]. This paradox is supported by the co-localization of gluten consumption and celiac predisposing genes, specifically in the Middle East, with diffusion into India and Africa.

The maintenance of gene frequencies in various populations suggests that they may have protective benefits in certain environments, similar to glucose-6-phosphate dehydrogenase deficiency (G6PD) deficiency protecting against malaria, but causing hemolytic anemia and death when in the presence of pyrimidines derived from fava beans. Both G6PD deficiency and CD have no cure, but are controlled by avoiding specific foods [45]. Despite the two main genes for CD, HLA-DQ2 and HLA-DQ8 having negative consequences in the presence of gluten, the chromosomal region as a whole is selected for under balancing selection, possibly mediated by coevolution with pathogens [46]. The selection for CD genes may occur as a result of positive selection on immune-related loci, which are necessary to fight infection. Thus, CD genotypes may exhibit antagonistic pleiotropy (meaning that genes may control more than one trait where at least one of these traits is beneficial to the organism's fitness and the other detrimental), thereby aiding in immunity against some pathogens, but also increasing risk of CD (in the presence of gluten) [47]. While the genes are now deleterious in certain environments, they may have been advantageous in early (eg, gluten-free, or universal extended breastfeeding) environments. The transition from hunting and gathering to agriculture mediates susceptibility to celiac disease, shown by increase of risk genes [48].

In 1978, Simoons hypothesized that the spreading of wheat from the Fertile Crescent in the Middle East along the Mediterranean countries to northern Europe exerted a negative pressure on genes that cause CD, leading to higher HLA-B8 gene frequencies (closely linked with HLA-DQ2) in NW Europe, compared to SE Europe, where people had longer historical exposure to cereal grains [32]. However, CD is a common autoimmune disease worldwide and is increasing in areas with high gluten consumption and prevalence of celiac risk genes, as well as increasing in diagnosis [2]. This suggests: (1) rapid recent environmental changes have led to discordance between genetics and the environment, and/or (2) positive selection on CD-risk genes may have occurred in human history to minimize disease, followed by some negative selection when gluten exposure increased, but with rapid and substantial changes to practices such as birthing (cesarean vs. vaginal), infant feeding (formula vs. breastmilk), sanitation and antibiotic use, recent CD risks appear to be increasing further. Support for this comes from Sweden's reversal of CD rates by changing infant feeding practices and timing of gluten introduction (since gene frequencies are unlikely to have changes so rapidly) [49]. However, to explain in the existence and maintenance of CD-risk alleles despite documented negative health effects, there must have been some type of positive selection with benefits for reproductive fitness. Evidence for the selective advantage of several CD-predisposing variants exists and is summarized below (see also Table 2).

Evolutionary discordance between our contemporary environment and the EEA in which our ancestors evolved may provide additional insight into the evolution of CD. The human genome evolved in an environment characterized by lower gluten intake in a pre-agricultural environment, but also extended breastfeeding on the order of years, not months [50], leading to different intestinal microbiome composition that could affect gluten digestion and exposure. Dietary changes that have occurred too rapidly for adequate genetic adaptation, and behavioral changes related to birth, child rearing and infant feeding have contributed to the modern-day epidemic of chronic diseases, such as CD [51]. Yet the CD-risk genes likely evolved and were maintained in populations with evolutionarily more normal birth and infant feeding behaviors. Thus it is important to consider possible positive selective factors influencing the maintenance of the CD-risk gene frequencies in many populations.

Table 2. CD Risk Genes that Exhibit Antagonistic Pleiotropy

Gene	Primary alleles	Other functions
HLA-DQ2	HLA-DQ2.5, 2.2	Protective against dental caries
SH2B3		Protective against bacterial infection
HFE	C282Y	Causes hemochromatosis

First, HLA-DQ2, the gene that accounts for 95% of CD cases, has been shown to protect against dental decay [2]. When grain-based carbohydrates were introduced into the diet, overall health declined and the number of cavities increased because of increased exposure to simple sugars from breakdown of carbohydrates [31]. Dental caries result from an interaction between the host, the host's diet, and microflora on the tooth surface [52]. In an environment with higher carbohydrate consumption, individuals with the random mutation of HLA-DQ2 might have had higher survival rates and reproductive success, leading to increased gene frequencies. Thus, ironically, a mutation that likely experienced positive selection following the increase in grain-based diets, also ultimately led to increased risk of CD and associated decreased fitness, especially when combined with environmental changes involving breastfeeding and the microbiome.

Second, the SH2B3 gene is correlated with other autoimmune disorders and metabolic diseases, and carries both increased risk of CD but also fitness benefits. SH2B3 likely does not directly cause onset of CD, but adds to the risk of CD development [53]. This gene protected against an unnamed bacterial infection that swept across European populations 1200-1700 years ago [53] (thousands of years after the origin of agriculture). Individuals with this gene were better able to survive the infection, thus the gene was passed on. Molecules that contain the SH2 domain are known to modulate intermolecular interactions and to inhibit cytokine responses, which are diminished in those with the risk allele. Additionally, the protein coded for by SH2B3 is an important receptor for bacterial pathogens, and thus SH2B3 may be protective not just against the unnamed bacterial infection 1200-1700 years ago, but also current and future bacterial selective pressures [53].

Third, not only are several genes involved in CD likely pleiotropic, but many also exhibit linkage disequilibrium (non-random association) [3]. The C282Y mutation in the hemochromatosis susceptibility gene *HFE*, which leads to a disease called hemochromatosis (a buildup of iron in the body that eventually harms the pancreas, heart and liver) is also associated with CD and forms part of the extended HLA haplotype [54]. The C282Y mutation of *HFE* results in an abnormality in *HFE* protein trafficking and complete lack of cell surface expression of the protein, leading to an increase in iron absorption in the intestine, and higher hemoglobin and iron levels compared with the HLE wildtype in celiac patients [54]. The H63D mutation is the most common mutation identified with the *HFE* gene; it does not prevent cell surface association, but the physiological consequences besides hemochromatosis are unknown. Interestingly, H63D shows the highest European frequency among the Basques, which are characterized by persistence of Paleolithic iron-rich diet, lower exposure to major infectious threat, and limited genetic mixing with Celts and Vikings, who have the highest prevalence of C282Y [55], highlighting the importance of history. The origin of the C282Y mutation coincides with Neolithic farmers expanding into Central Europe and the spread of the allele could have been an adaptation to a dietary shift from hunting and gathering foods rich in iron to domesticated cereals poor in iron [56]. However, in individuals with CD who are often iron deficient, this buildup of iron allows them to absorb appropriate amounts. This decreases the infertility rates associated with CD due to lack of iron, allowing the genes for both CD and hemochromatosis to be passed on. While this mutation does not actively positively-select for CD, it mitigates the negative effect of the disease that selects against the genes, and thus selection may act on this haplotype of CD [43].

While the genes discussed above constitute the primary CD-associated genes with evidence for antagonistic pleiotropy, other genes exist that are part of other risk loci, many of which may protect against the development of CD in different populations throughout the world. MICA-A5 was found to decrease CD risk. HLA-Cw4 and DQ1 were found to be protective against the development of CD in a Spanish population. Likewise, HLA-DQB1*06 acts protectively in white Brazilians. DQA1*0101, DQA1*0201, and DQB1*0301 were protective in children from Santiago, Chile. Their distribution among populations suggest that many of them may be due to founder effects [57].

6. Historical and Cross-Cultural Considerations

To understand the current distribution of gene frequencies throughout the world and observed population variation, there is a need to consider historical and cross-cultural factors. With minimal to no gluten in diets before the origin of agriculture, there was minimal trigger in those who may have been genetically susceptible, thus the disease was unlikely to develop. Without genetic testing, it is difficult to estimate gene frequency, since expression of the CD phenotype requires both predisposing genes and gluten trigger. Three different combinations of trigger-genotype can lead to non-CD phenotypes. Before the origins of agriculture, there was no fitness cost nor little selective advantage for DQ2, and no gluten trigger, and thus no CD. But over time there was an increase in the selective advantage of DQ2 to protect against dental caries in environments with high carbohydrates. Absence of CD could also occur in populations with high CD-risk gene frequencies maintained or selected for in populations for non-CD-related reasons due to absence of gluten trigger, such as among those in Burkino Faso, or possibly even in the presence of gluten trigger as long as other evolutionary practices such as vaginal births, extended breast feeding and minimal antibiotics were common.

As wheat cultivation continued and spread from the Middle East to Europe, the amount of gluten within the plant increased only slightly [58]. While people initially selected for bigger wheat plants to maximize the grain recovery during threshing, increase of gluten within the plant itself was insignificant. However, over time plants with higher wheat content were selected for because leavened bread requires high gluten content [58]. Wheat was initially domesticated in the region of Gobleki Tepe, where Einkorn wheat is found. Varieties of Eikorn wheat are found to have less intestinal toxicity in individuals with CD than modern grain gliadins [59]. Additionally, whole wheat foods, which are increasing in popularity in the United States, have higher percentages of gluten [58]. Additionally, the Glia-α9 epitope is the one with the highest

binding affinity by receptors in CD patients and is found in greater quantities in modern wheat than ancient wheat [8].

Individuals in modern Burkina Faso have a diet very similar to that of populations 10,000 years ago in that they consume very little wheat, subsisting on meat, fruits and vegetables. There are no instances of CD due to low amounts of trigger [11]. The Saharawi of Algeria consumed a diet very similar to those in Burkina Faso, until they were forced to become refugees and consume a diet very high in gluten [44]. Celiac disease rates in this population increased to 5.6% from zero in just a century due to this abrupt change in diet [60] (Figure 3). Additionally, the population has one of the highest frequencies of HLA-DQ2 likely due to the founder effect or that they have not yet begun to experience the negative selection exerted by gluten. However, in other populations with a similar HLA-DQ2 gene frequency, CD rates of around 1% are observed [61], suggesting other mediating environmental factors.

Prior to agriculture when there was less gluten in the diet, people were less likely to develop the disease. In an environment with no gluten trigger, there would have been no selective pressure against these genes because populations were not developing the disease and suffering the negative effects. As more gluten was introduced to the diet, more people may have developed the disease, but CD-risk genes did not disappear, perhaps due to positive selection on CD-predisposing genes (antagonistic pleiotropy) and buffering of CD risk through behavioral and environmental factors.

As discussed above in the Epidemiology of Celiac Disease section, Swedish patterns of CD highlight important behavioral factors that influence disease rates, including breastfeeding, and timing and amount of gluten-containing weaning foods (Figure 3). Intestinal microbiome and effects on immune function likely also play a role, thus expanding the role of diverse cultural practices and environments on the etiology of CD.

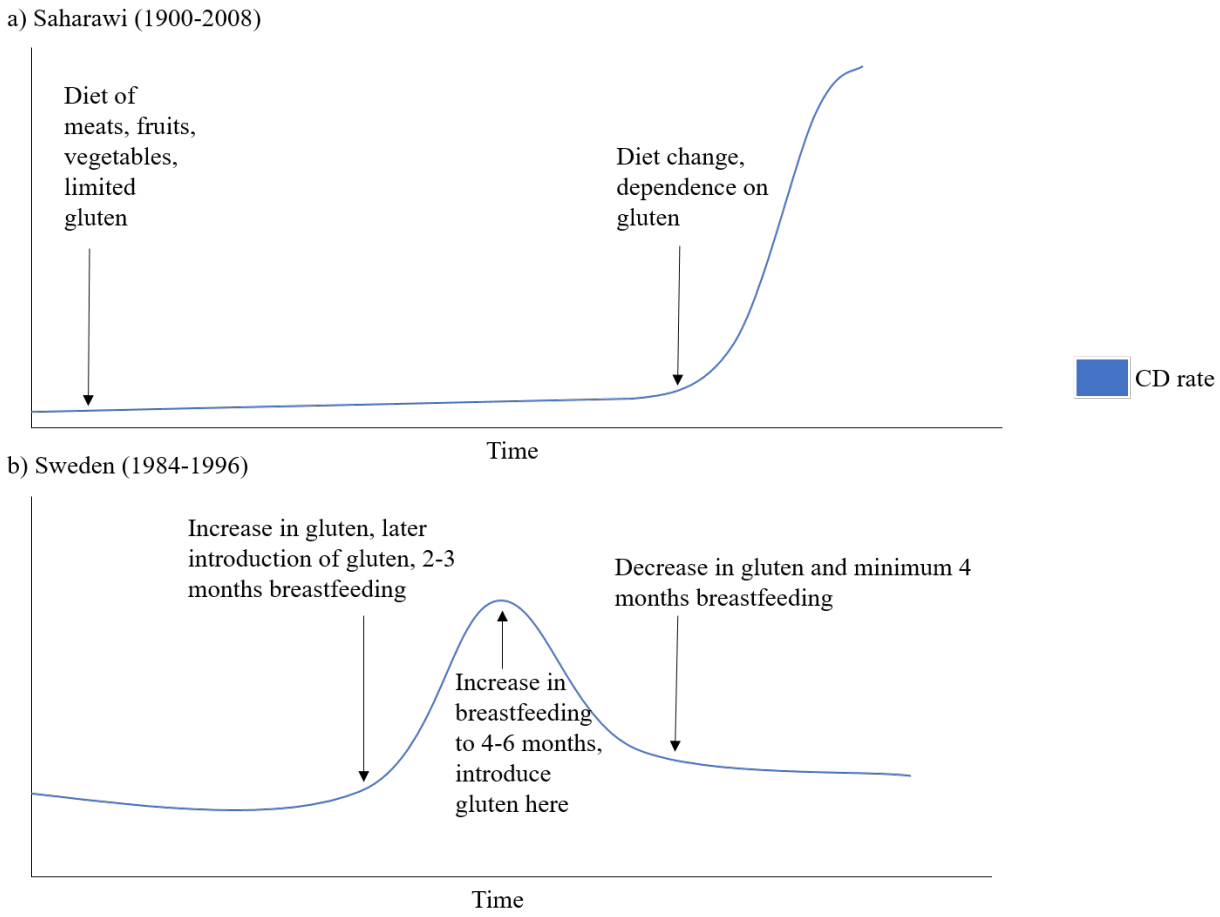

Figure 3. Schematic of environmental factors influencing changes in CD rates over time. Changes in CD rates may be caused by changes in environmental factors, such as breastfeeding duration timing of gluten introduction, and amount of gluten exposure

7. Conclusions

It is important to understand the many factors influencing the phenotypic expression of CD. It may be tempting to think that gene therapy might cure CD, but gene therapy only works for diseases caused by a single gene with no environmental factors. Before attempting to find alternative therapeutics to minimize the development of CD, it is important to understand the positive benefits that maintain CD in different populations, the environmental factors that may decrease or delay CD's appearance, and the environmental and developmental interventions that could minimize the risk of CD while maximizing the fitness benefits associated with CD-predisposing alleles. As shown in Table 2, certain genes for CD exhibit antagonistic pleiotropy and may contribute to maintenance of CD-risk genes in populations due to health and fitness benefits. These benefits may have been offset by cultural and dietary environments that have changed with increasing gluten exposure and changes to

birthing and breastfeeding practices, intestinal microbiome and immune function.

Populations undergoing rapid changes in dietary gluten and CD rates may hold clues to the origins of CD and future research should examine the evolutionary and environmental factors influencing the development of CD. For example, the Saharawi of Northern Africa both exhibited rapidly increasing CD rates following the introduction of gluten as a staple of the diet. Since environmental conditions such as length of breastfeeding with respect to timing of gluten introduction(as observed in Sweden) may play an important role in determining CD prevalence, studies of populations with dramatic shifts from non-gluten to gluten-based diets should ensure that data are collected on birth, breastfeeding and weaning practices, as well as infectious disease exposure and factors influencing the microbiome since they moderate the gene-environment interactions resulting in the phenotypic expression of CD.

Prospective studies could be performed, particularly in refugee populations whose diets often change rapidly, although such studies must also control for significant stressors and changes in other important variables such as breastfeeding duration, timing of gluten-containing weaning foods, infectious disease, and intestinal microflora composition. Examination of CD prevalence rates in populations with varying prevalence of HLA-DQ2 gene frequencies and varying duration of gluten exposure, modified by timing and amount of gluten exposure, breastfeeding, microbiome and immune function, will shed light on the complex gene-environment interactions between major and minor genes influencing this devastating disease.

Acknowledgements

We gratefully acknowledge Karen Rosenberg, Wenda Trevathan, Patricia DeLeon, and Dunia Tonob for critical feedback and comments on earlier drafts.

References

[1] Lohi S, Mustalahti K, Kaukinen K, et al. Increasing prevalence of coeliac disease over time. Aliment Pharmacol Ther. 2007; 26: 1217-25.

[2] Lionetti E and Catassi C. Co-localization of gluten consumption and HLA-DQ2 and -DQ8 genotypes, a clue to the history of celiac disease. Digestive and Liver Disease. 2014; 46: 1057-1063.

[3] Karell K, Louka AS, Moodie SJ, 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: 469-477.

[4] Gheller-Rigoni A, Celiac Disease: celiac sprue, gluten-sensitive enteropathy. Clin Med & Res. 2004; 2: 71-72.

[5] Basso MS, Luciano R, Ferretti F et al. Association Between Celiac Disease and Primary Lactase Deficiency. Eur J Clin Nutr. 2012; 66: 1364-1365.

[6] Koning F. Pathophysiology of Celiac Disease. J Pediatr gastroenterl Nutr. July 2014; 59: S1-S4.

[7] Celiac Disease: Genetics and Nutrition. http://www.nchpeg.org/nutrition/index.php?option=com_content&view=article&id=459&Itemid=56&showall=1. [Accessed 6 Oct 2016].

[8] van den Broeck H. Presence of celiac disease epitopes in modern and old hexaploid wheat varieties: wheat breeding may have contributed to increased prevalence of celiac disease. Theor Appl Genet. 2010; 121: 1527-1539.

[9] Copelton D. "You don't need a prescription to go gluten-free": the scientific self-diagnosis of celiac disease. Soc Sci Med. 2009; 69: 623-631.

[10] Stevenson JC, Maki C, Bruna S, et al. Community based pilot study of diagnostic paths to the gluten free diet. Int J Celiac Disease. 2015; 3: 17-24.

[11] Dubois PC. Translational Mini-Review Series on the Immunogenetics of Gut Disease: Immunogenetics of coeliac disease. Clin Exp Immunol. 2008; 153: 162-173.

[12] Hoffenberg EJ, Bao F, Eisenbarth GS et al. Transglutaminase antibodies in children with a genetic risk for celiac disease. J Pediatr. 2000; 137: 356-360.

[13] Nylund L. The microbiota as a component of the celiac disease and non-celiac gluten sensitivity. Clin Nutr Exp. 2016; 6: 17-24.

[14] Chen J. Development of intestinal bifidobacteria and lactobacilli in breast-fed neonates. Clin Nutr (Scot) 2007; 26: 559-566.

[15] Ivarsson A. Breast-feeding protects against celiac disease. Am J Clin Nutr. 2002; 75: 914-921.

[16] Marasco G. Gut Microbiota and Celiac Disease. Dig Dis Sci. 2016; 61: 1461-1472.

[17] Beres NJ and Sziksz E. Role of the Microbiome in Celiac Disease. Internat J Celiac Disease. 2014; 2: 150-3.

[18] Dieli-Crimi R, Cénit MC, and Núñez C. The genetics of celiac disease: A comprehensive review of clinical implications. J Autoimmun. 2015; 64: 26-41.

[19] Megiorni F PA, HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. J Biomed Sci. 2012; 19: 88.

[20] Fallang LE, Bergseng E, Hotta K, 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: 1096-101.

[21] Zhernakova A, Stahl EA, Trynka G et al. Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci. PLoS Genetics. 2011; 7: e1002004.

[22] Parkes M, Cortes A, van Heel DA,Brown MA, Genetic insights into common pathways and complex relationships among immune-mediated diseases. 2013; 14: 661-73.

[23] Wang L, Human autoimmune diseases: a comprehensive update. J Intern Med. 2015; 278: 369-395.

[24] Hartl DL and Jones EW. Essential genetics: a genomics perspective. Jones & Bartlett Publishers. 2006.

[25] Popat S, Hearle N, Hogber L, et al. Variation in the CTLA4/CD28 gene region confers an increased risk of coeliac disease. Ann Hum Genet. 2002; 66: 125-137.

[26] Mustalahti K, Catassi C, Reunanen A, et al. The prevalence of celiac disease in Europe: Results of a centralized, international mass screening project. SANN Annals of Medicine. 2010; 42: 587-595.

[27] Lionetti E. Celiac disease from a global perspective. Bailliere's Best Practice & Research: Clinical Gastroenterology. 2015; 29: 365-379.

[28] Malekzadeh R. Coeliac disease in developing countries: Middle East, India and North Africa. Bailliere's Best Practice & Research: Clinical Gastroenterology. 2005; 19: 351-358.

[29] Dubé C, Rostom A, Sy R et al. The prevalence of celiac disease in average-risk and at-risk Western European populations: A systematic review. Gastroenterology. 2005; 128: S57-S67.

[30] Cataldo F. Celiac disease in the developing countries: a new and challenging public health problem. World Journal of Gastroenterology. 2007; 13: 2153-2159.

[31] Cohen MN and Armelagos GJ (eds). Paleopathology at the origins of agriculture. University Press of Florida. 2013.

[32] Cordain L. Cereal Grains: Humanity's Double-Edged Sword. World Rev Nutr Diet. 1999; 19-23.

[33] Fasano A, Catassi C, and Gatti S. The New Epidemiology of Celiac Disease, J Pediatr Gastroenterol Nutr. 2014; 59: Suppl 1:S7-9.

[34] Visser J, Rozing J, Sapone A et al. Tight Junctions, Intestinal Permeability, and Autoimm,unity: Celiac Disease and Type 1 Diabetes Paradigms. Ann N Y Acad Sci. 2009; 1165: 195-205.

[35] Jakaitis BM and Denning PW. Human breast milk and the gastrointestinal innate immune system. Clin Perinatol. 2014; 41: 423-35.

[36] Eaton SB and Konner, M. Paleolithic Nutrition. A consideration of its nature and current implications. N Engl J Med. 1985; 312: 283-9.

[37] Eaton SB, Konner M, Shostak M. Stone agers in the fast lane: chronic degenerative diseases in evolutionary perspective. Am J Med. 1988; 84: 739-49.

[38] Kivity S. Can we explain the higher prevalence of autoimmune disease in women? Expert Rev Clin Immunol. 2010; 6: 691-694.

[39] Sams A and Hawks J. Celiac disease as a model for the evolution of multifactorial disease in humans. Human Biology. 2014; 86: 19-36.

[40] Myléus A , Hernell O, Gothefors L, et al. Early infections are associated with increased risk for celiac disease: an incident case-referent study. BMC Pediatrics. 2012; 12: 194.

[41] Ivarsson A, Myleus A, Norstrom F, et al. Prevalence of childhood celiac disease and changes in infant feeding. Pediatrics. 2013; 131: e687-94.

[42] Fasano A, Berti I, Gerarduzzi T, et al. Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. Arch Intern Med. 2003; 163: 286-292.

[43] Whittington C. The C282Y mutation may have been positively selected as it mitigates the infertility of celiac disease. Med Hypotheses. 2006; 66: 769-772.

[44] Ercolini D. From an imbalance to a new imbalance: Italian-style gluten-free diet alters the salivary microbiota and metabolome of African celiac children. Scientific Reports. 2015; 5: 18571.

[45] Mager J, Glaser G, Razin A et al. Metabolic effects of pyrimidines derived from fava bean glycosides on human erythrocytes deficient in glucose-6-phosphate dehydrogenase. Biochem Biophys Res Commun. 1965; 20: 235-240.

[46] Augusto D and Petzl-Erler ML. *KIR* and *HLA* under pressure: evidences of coevolution across worldwide populations. Human Genetics. Sept 2015; 134: 929-940.

[47] Sams A and Hawks J. Patterns of Population Differentiation and Natural Selection on the Celiac Disease Background Risk Network. PLoS ONE. 2013; 8: e70564.

[48] Crespi BJ. The emergence of human-evolutionary medical genomics. Evolutionary Applications. 2011; 4: 292-314.

[49] Olsson C, Stenlund H, Hö rnell A, et al. Regional variation in celiac disease risk within Sweden revealed by the nationwide prospective incidence register. Acta Paediatrica. 2008; 98: 337-342.

[50] Stuart-Macadam P and Dettwyler KA. Breastfeeding: biocultural perspectives. Routledge. 1995.

[51] Konner M. Paleolithic nutrition: twenty-five years later. Nutr Clin Pract. 2010. 25: 594-602.

[52] Valarini N. Association of Dental Caries with HLA Class II Allele in Brazilian Adolescents. Caries Res. 2012; 46: 530-535.

[53] Zhernakova A, Elbers C, Ferwerda B et al. Evolutionary and Functional Analysis of Celiac Risk Loci Reveals SH2B3 as a Protective Factor against Bacterial Infection. Amer J Hum Genet. 2010; 86: 970-977.

[54] Butterworth JR, Cooper BT, Rosenberg WMC, et al. The role of hemochromatosis susceptibility gene mutations in protecting against iron deficiency in celiac disease Gastroenterology. 2002; 123: 444-449.

[55] Bauduer F. C282Y/H63D hemochromatosis mutations and microevolution: Speculations concerning the Basque population. HOMO - J Comp Human Biology. 2017; 68: 38-41.

[56] Heath KM, Axton JH, McCullough JM, et al. The evolutionary adaptation of the C282Y mutation to culture and climate during the European Neolithic. Am J Phys Anthropol. 2016; 160: 86-101.

[57] Aaron L. The last two millennias echo-catastrophes are the driving forces for the potential genetic advantage mechanisms in celiac disease. Med Hypotheses. 2011; 77: 773-776.

[58] Kasarda D. Can an increase in celiac disease be attributed to an increase in the gluten content of wheat as a consequence of wheat breeding? J Agric Food Chem. 2013; 61: 1155-1159.

[59] Freeman HJ The Neolithic revolution and subsequent emergence of the celiac affection. Int J Celiac Disease 2013; 1: 19-22.

[60] Ratsch IM and Catassi C. Coeliac disease: a potentially treatable health problem of Saharawi refugee children. Bull World Health Organ. 2001; 79: 541-545.

[61] Catassi C. The distribution of DQ genes in the Saharawi population provides only a partial explanation for the high celiac disease prevalence. Tissue Antigens. 2001; 58: 402-406.