Development of a Sequence Searchable Database of Celiac Disease-Associated Peptides and Proteins for Risk Assessment of Novel Food Proteins

Plaimein Amnuaycheewa[1], Mohamed Abdelmoteleb[2], John Wise[3], Barbara Bohle[4], Fatima Ferreira[5], Afua O. Tetteh[6], Steve L. Taylor[3] and Richard E. Goodman[3]*

[1] Department of Agro-Industrial, Food, and Environmental Technology, King Mongkut's University of Technology North Bangkok (KMUTNB), Bangkok, Thailand, [2] Botany Department, Faculty of Science, Mansoura University, Mansoura, Egypt, [3] Food Allergy Research and Resource Program (FARRP), Department of Food Science and Technology, University of Nebraska, Lincoln, NE, United States, [4] Christian Doppler Laboratory for Immunomodulation, Department of Pathophysiology and Allergy Research, Medical University of Vienna, Vienna, Austria, [5] Department of Biosciences, University of Salzburg, Salzburg, Austria, [6] BASF Corporation, Morrisville, NC, United States

Celiac disease (CeD) is an autoimmune enteropathy induced by prolamin and glutelin proteins in wheat, barley, rye, and triticale recognized by genetically restricted major histocompatibility (MHC) receptors. Patients with CeD must avoid consuming these proteins. Regulators in Europe and the United States expect an evaluation of CeD risks from proteins in genetically modified (GM) crops or novel foods for wheat-related proteins. Our database includes evidence-based causative peptides and proteins and two amino acid sequence comparison tools for CeD risk assessment. Sequence entries are based on the review of published studies of specific gluten-reactive T cell activation or intestinal epithelial toxicity. The initial database in 2012 was updated in 2018 and 2022. The current database holds 1,041 causative peptides and 76 representative proteins. The FASTA sequence comparison of 76 representative CeD proteins provides an insurance for possible unreported epitopes. Validation was conducted using protein homologs from Pooideae and non-Pooideae monocots, dicots, and non-plant proteins. Criteria for minimum percent identity and maximum *E*-scores are guidelines. Exact matches to any of the 1,041 peptides suggest risks, while FASTA alignment to the 76 CeD proteins suggests possible risks. Matched proteins should be tested further by CeD-specific CD4/8+ T cell assays or *in vivo* challenges before their use in foods.

Keywords: celiac disease, gluten, T-cell epitopes, peptide database, risk assessment, sequence comparison, Pooideae, prolamin

INTRODUCTION

Novel proteins and complex foods are being introduced into the human diet through the creation of genetically engineered organisms, by the addition of isolated proteins, and by the introduction of novel organisms with limited or no history of safe human consumption (1). Prior to marketing, novel proteins and foods containing proteins from species related to wheat should undergo a

safety evaluation to ensure safe consumption by those with CeD. The 2003 Codex Alimentarius Commission guideline called for evaluating proteins encoded by genes transferred from wheat and wheat relatives into a different species to be evaluated for potential risks of eliciting CeD as part of the overall food safety evaluation (2).

Celiac disease involves gluten-mediated inflammation that damages the epithelial lining of the upper small intestine leading to villous flattening and nutrient malabsorption. Extra-intestinal manifestations of the chronic enteropathy include short stature, osteopenia, dermatitis herpetiformis, and ataxia (3–5). Life-long, strict adherence to a gluten-free diet reverses the damage and function. The causative gluten proteins are the seed storage prolamins and glutelins, which in wheat are called gliadins and glutenins, respectively. The prolamins are called hordeins in barley and secalins in rye. Among plant-derived, dietary proteins, prolamins and glutelins contain a high amount of proline (P) and glutamine (Q) amino acids in their sequences contributing to the visco-elastic property of bread dough and conferring resistance to gastrointestinal digestion (6). Many digestion-resistant gluten peptides are found to travel across intestinal epithelium via transcellular absorption or through stimulating CXC chemokine receptor 3 (7, 8). The peptides then bind to human leukocyte antigen (HLA) DQ2.5 (DRB1*301-DQA1*0501-DQB1*0201) or DQ8 (DRB1*04-DQA1*0301-DQB1*0302) on antigen-presenting cells and are presented to activate pro-inflammatory T cells (9, 10). In Europe and the United Kingdom, more than 90% of patients with CeD express HLA-DQ2.5 and 5–10% of the patients express HLA-DQ8 (11–13). The percentage of patients with CeD patients in the United States carrying the genotypes has been estimated to be 82 and 16%, respectively (14). A multicenter study reported that nearly 0.4% of the patients carry DR5-DQ7 (DRB1*11/12-DQA1*0505-DQB1*0301) or DR7-DQ2 (DRB1*07-DQA1*0201-DQB1*0202), which can form a heterozygous DQ2.5 (DQA1*0505-DQB1*0202) binding groove (15). These MHC receptors present both native and tissue transglutaminase (TG2) deamidated peptides to the reactive cluster of differentiation 4 (CD4) T cells, and the TG2-gluten peptide complex can be a target of the activated T cells (16). Studies demonstrated that DQ2.5 preferentially binds to peptides having a 9-mer binding core with negatively charged anchors at positions 4 and 6 or 7, whereas the DQ8 allele preferentially binds peptides with negatively charged anchors at positions 1 and 9. To a lesser extent, DQ2.5 and DQ8 alleles preferentially bind to peptides with P at positions 1 and 6, respectively (17–22). Digestion-resistant gluten peptides lack polar acidic amino acids, yet the position of specific amino acids in these peptides allows TG2 deamidation at appropriately spaced Q residues, converting them to glutamic acid (E). Deamidation is important for the selection of T-cell epitopes since most of the DQ2.5 recognized epitopes are in the deamidated form (17, 20, 23–26). For example, a number of deamidated positions were reported on the 33-mer, CD4+ T cell reactive alpha2-gliadin peptide (LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) (27). The DQ8 molecule recognizes the epitopes equally in both native and deamidated forms (28). The specificity of TG2 deamidation of the peptides has not been conclusively demonstrated, but

some residues are more effectively modified in peptides with the configuration of QXP, where X represents amino acids other than P (25, 27, 29). The MHC restriction is predictive, but not definitive, since nearly 40% of the general population carries HLA-DQ2 or HLA-DQ8 genes, but only 1% of the population exhibit CeD (30). Genome-wide association studies indicated that non-HLA, immune-related genes also relate to CeD (31–33).

The adaptive immune response to gluten proteins appears to involve both helper and cytotoxic T cells. An α-gliadin peptide (p123-p132: QLIPCMDVVL) induced HLA-A2 specific CD8+ T cells isolated from biopsies of DQ2/DQ8 CeD patients to undergo maturation to express Fas ligand and secrete interferon-gamma (IFN-γ) and granzyme B (34). The CeD pathogenesis also implicates innate immunity. A 13-AA gliadin peptide (LGQQQPFPPQQPY) induced IFN-γ, tumor necrosis factor-alpha (TNF-α), and interleukin 15 (IL-15) secretion from intestinal epithelial cells, macrophages, and dendritic cells (DCs) (35, 36). IL-15 is recognized as the dominant driver in refractory patients with CeD who exhibit villous atrophy and lymphoma without recent ingestion of gluten proteins (37, 38). IL-15 induces MHC class I chain-related (MIC) expression on enterocytes and the natural killer group 2D (NKG2D) expression on the intraepithelial lymphocytes. The T cell receptor-independent MIC-NKG2D interaction leads to enterocyte apoptosis and destructive lesions (35, 39–41).

Collective evidence illustrates that the immunogenicity of gluten proteins is complicated and thus safety evaluation of novel proteins for the CeD risk is challenging. The immune responses of novel proteins and products could be evaluated in patients with CeD using oral challenges. However, such an approach is costly and would involve risks for patients. The assessment of CeD-specific T cell reactivity can be performed using T cells derived from *in vitro* cultured intestinal biopsies or from the blood of patients with CeD after oral challenges. Yet those tests would require relevant clinical infrastructure and CeD subjects following informed consent. Animal models are available alternatives but require complicated evaluation and validation. Although there are commercial enzyme-linked immunosorbent assay (ELISA) kits that are widely employed to test for the presence of gluten proteins in various food products, the available antibodies (e.g., R5 and G12) may not target all proteins or peptides that pose a risk to CeD subjects. Furthermore, those ELISA tests do not predict peptides that might be deamidated by the endogenous enzyme in the lamina propria. Since many studies have reported that the gluten peptides are confirmed to elicit CeD reactivity, our goal was to establish a searchable database of the causative peptides as well as to develop simple but effective bioinformatics assessment tools for evaluating the potential risk of new dietary proteins for those with CeD. This publication describes the literature review used to identify the causative epitopes and proteins and describes the evaluation process used to validate the database, computer search routines, and risk assessment criteria. The practical utility of the database and the bioinformatics tools is then discussed.

MATERIALS AND METHODS

Literature Review and Collection of CeD-Associated Peptides

The PubMed literature database (http://www.ncbi.nlm.nih.gov/pubmed) was searched for relevant publications using the keywords "celiac" and "coeliac" to identify studies investigating proteins and peptides capable of eliciting CeD pathogenesis. Native and deamidated peptides with documented evidence of stimulating CD4+ T cells restricted to MHC class II molecule DQ2.5, DQ2.2, DQ8, or DQ9 that lead to T cell proliferation with greater than a 2-fold stimulatory index or release of IFN-γ were collected. In addition to the immunogenic peptides, those with rigorous evidence of eliciting toxic reactions in the intestines of patients with CeD were collected as toxic peptides. Published toxic properties reported included one or more of the following indications: reduction in epithelial brush border alkaline phosphatase activity; increased intestinal permeability; reduction in enterocyte surface cell height (ECH) or reduction in villus height to crypt depth ratio (VH:CD); expression of epithelial apoptotic mediator ligand HLA-E molecule; maturation and migration of macrophage, DC, and CD4+ T cells to the lamina propria; or expression of inflammatory cytokines IFN-γ, TNF-α, and IL-15 (35, 36, 42–50). The peptides obtained from the selected publications were reviewed by five peer review panel members of the AllergenOnline.org database to confirm the selection rationale prior to constructing the database.

Construction of the Database (Version 1, 2012) and the Bioinformatics Searching and Comparison Tools

The collected CeD peptides from literature searches were compared with the non-redundant National Center for Biotechnology Information (NCBI) protein database using the Basic Local Alignment Search Tool (BLAST) to identify identical peptides and the source proteins. Default BLASTP search parameters were used with an Expect threshold (E-score) of 10, matrix selection of BLOSUM62, gap costs of 11 for existence, and 1 for extension. The BLAST results showed that 425 native peptides from the 1,016 identified peptides had identity matches with 147 prolamins and glutelins of the Pooideae grass subfamily. The 147 proteins were then evaluated using the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) multiple sequence alignment program ClustalW2. Identical sequences were removed and 68 non-redundant sequences were collected as representatives for CeD-associated proteins. Bread wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) are the major sources of the CeD reactive prolamins, which account for about 63% (43 out of 68 proteins) and 16% (11 out of 68), respectively (**Table 1**).

The 1,016 identified CeD peptides and the 68 representative CeD proteins with the NCBI protein accession numbers were loaded into a MySQL relational database management system. The peptides with CeD-associated evidence and publication links

TABLE 1 | Statistics of the AllergenOnline.org CeD peptide and protein database version construction and inclusion characteristics.

		Version 1 (released in 2012)	Version 2 (released in 2018)
References	Number of publication references*	68	**72**
	Publication year of references	1984–2012	1984–**2017**
Peptides	Number of peptides	1,016	**1,013**
	Number of native peptides	464	**465**
	Number of deamidated peptides	552	**548**
	Number of immunogenic peptides	998	**1,004**
	Number of CD4+ T cell reactive peptides	997	**1,003**
	Number of CD8+ T cell reactive peptides	1	1
	Number of toxic peptides (without T cell reactivity)	18	**9**
	Length of peptides (AA)	8–55	**9**–55
	Averaged length of peptides (AA)	16 ± 4	16 ± 4
Proteins	Number of proteins	68	**72**
	Number of proteins in *Triticum aestivum*	43	43
	Number of synthetic constructs in *Triticum aestivum*	1	1
	Number of proteins in *Triticum monococcum*	2	2
	Number of proteins in *Hordeum vulgare*	11	**12**
	Number of proteins in *Secale cereale*	6	6
	Number of proteins in *Avena sativa*	3	**6**
	Number of proteins in *Avena nuda*	2	2
	Length of proteins (AA)	20–800	20–800

*Changes between the two database versions are in bold font. * Complete list of the 72 reference publications with the PubMed links is available on the database.*

were available in the browse function of the database. Query sequences entered by database users in the search window can be compared with CeD peptides by the exact identity match tool. The 68 representative CeD source proteins can be viewed in the browse function, and the sequences of query proteins can be compared for identity scores to each of the 68 representative CeD proteins by the full-length FASTA3 sequence alignment tool, version 35.04 (51). The peptide and protein database sections with complete references from 68 publications and the two sequence comparison tools were available for public use at http://www.allergenonline.org/celiachome.shtml from January 2012 until November 2017.

Update of the Database (Version 2, 2018)

In 2017, an additional literature review was conducted and affirmed by six AllergenOnline.org peer review panel members. A total of 34 previously included peptides were removed from the database as they were less than nine AA long and were considered too short to effectively bind MHC and activate T cells. The core nine-AA peptides listed in the 2017 European Food Safety Authority (EFSA) guidance on allergenicity assessment of genetically modified plants and their predicted deamidated forms were added to the database (52, 53). Four additional publications

were added to the database as references. One barley and three oat prolamins were identified through BLASTP and ClustalW2 bringing the representative proteins in the database to 72. The updated database version 2 was made public in October 2017 with a full explanation in January 2018 (**Supplementary Tables 1, 3**).

Testing the Database to Define Criteria for Potential Risks for Eliciting CeD

Testing of version 1 of the database was conducted in 2012 and version 2 in 2018. Proteins tested included prolamins and glutelins from known CeD-associated species (e.g., wheat, barley, rye, and oat) and homologous proteins from sources outside of Pooideae that have a history of safe use for subjects with CeD (e.g., maize, sorghum, coix (adlay), millet, rice, and teff). The analyses were conducted using query sequences identified from the NCBI protein database from Pooideae sources and non-Pooideae sources using keywords: gluten, glutelin, glutenin, prolamin, prolamine, gliadin, hordein, secalin, avenin, zein, kafirin, coixin, canein, and pennisetin. In addition, each of the representative CeD protein sequences was searched against the non-redundant NCBI protein database by BLASTP using the Expect threshold of 10 and with the exclusion of the Pooideae proteins (NCBI taxonomic identifier: 147368), but

TABLE 2 | FASTA sequence identity scores and alignments of the representative prolamin-like protein groups clustered by source organism types that were tested with the AllergenOnline.org CeD database version 1.

	Group	Number of proteins searched from NCBI	Contain exact CeD-associated peptides	Three best FASTA identity score results		
				Alignment overlap length (Length of the representative CeD protein)	% Identity to the CeD protein	*E*-score
I	Prolamins in Pooideae with CeD peptides	2,104*	Yes	827 (827)	100	2.8e−179
				287 (290)	100	7.8eb−81
				842 (838)	98.1	1.4e−195
	Prolamins in Pooideae without CeD peptides	562*	No	20 (20)	95	2.9e−05
				187 (288)	98.4	2.7e−45
				290 (288)	79.3	3.5e−63
II	Prolamins and prolamin-like proteins in Chloridoideae, Ehrhartoideae, and Panicoideae	1,059*‡	No	54 (52)	40.7	6.7
				12 (20)	66.7	1.9
				268 (360)	41	3.5e−17
III	Prolamin-like proteins in Dicotyledons	1,050*	No	68 (68)	33.8	2.3
				10 (20)	60	8.8
				121 (648)	30.6	1.8e−06
IV	Unrelated proteins (animals, fungi and microbes)	48△	No	29 (29)	58.6	3.8
				11 (20)	72.7	5.8e−03
				437 (439)	41.2	8.7e−25

* Proteins were identified from the NCBI protein database using keywords: gluten, glutelin, glutenin, prolamin, prolamine, gliadin, hordein, secalin, avenin, zein, kafirin, coixin, canein, and pennisetin.

‡ 35 proteins were obtained by BLAST, which searched the 68 representative celiac proteins against the NCBI Protein-Protein (non-redundant sequences) database with the exclusion of Pooideae (taxid: 147368) that had close to 45% identity over 100 AA and an E-score close to 1e-14. None had a direct CeD peptide match.

△ proteins were obtained by BLAST searches with the 68 representative celiac proteins against the NCBI Protein-Protein (non-redundant sequences) database with the exclusion of Pooideae (taxid: 147368).

TABLE 3 | Repeat of the FASTA sequence identity scores and alignments of the larger representative prolamin-like protein groups clustered by source organism types that were tested with the AllergenOnline.org CeD database version 2.

	Group	Number of proteins searched from NCBI	Contain exact CeD-associated peptides	Three best FASTA identity score results		
				Alignment overlap length (Length of the representative CeD protein)	% Identity to the CeD protein	*E*-score
I	Prolamins in Pooideae with CeD peptides	4,623*	Yes	828 (828)	100	1.0e−177
				439 (290)	100	1.6e−165
				455 (455)	100	8.4e−153
	Prolamins in Pooideae without CeD peptides	1,163*	No	291 (288)	98.6	3.7e−09
				264 (279)	98.9	1.1e−73
				266 (269)	98.5	3.6e−68
II	Prolamins and prolamin-like proteins in non-Pooideae monocots	1,755[*‡]	No	292 (250)	37.3	3.6e−09
				168 (181)	40.5	9.1e−09
				222 (222)	37.4	2.4e−08
III	Prolamin-like proteins in Dicotyledons	4,724*	No	305 (838)	32.1	1.6e−04
				372 (439)	28.8	9.5e−04
				253 (290)	29.2	9.3e−03

* Proteins were identified from the NCBI protein database using keywords: gluten, glutelin, glutenin, prolamin, prolamine, gliadin, hordein, secalin, avenin, zein, kafirin, coixin, canein, and pennisetin.

‡ 44 proteins were obtained by BLAST, which searched the 72 representative celiac proteins against the NCBI Protein-Protein (non-redundant sequences) database with the exclusion of Pooideae (taxid: 147368), that are close to significance based on percent identity and E-scores close to or below 1e-14. None contained a CeD peptide match, however one had 45–50% identity to four wheat glutens.

excluding patented proteins. The 2012 results were compiled and sorted into four groups: (1) 2,666 prolamins from the Pooideae subfamily that may be considered possibly unsafe for patients with CeD; (2) 1,059 prolamins and prolamin related proteins from the grass subfamilies of Chloridoideae, Ehrhartoideae, and Panicoideae, sources that are considered to be safe for individuals with CeD; (3) 1,050 prolamin-like proteins from the dicotyledon class that are considered to be safe for patients with CeD; and (4) 48 unrelated proteins, obtained from the BLAST search from sources considered safe for patients with CeD (**Table 2**). Each sequence of the four groups was manually tested against the CeD database using both the exact peptide match and FASTA3 tools and the results (exact match hits and FASTA sequence homology scores [percent identity score, alignment overlap length, and *E*-score]) were recorded. The evaluation of the FASTA3 alignment scores was used to set the criteria for minimum percent identity and maximum *E*-scores that suggest risks of CeD. In 2018, similar NCBI searches were conducted, and version 2 database was tested using (1) 5,786 prolamins from the Pooideae subfamily; (2) 1,755 prolamins and prolamin related proteins from the grass subfamilies of Chloridoideae, Ehrhartoideae, and Panicoideae; and (3) 4,724 prolamin-like proteins from the dicotyledon class (**Table 3**). The results were used to validate the previously proposed criteria.

Database Tests Using Hypothetical Alanine-Substituted Alpha-Gliadin

Evaluation of the utility of the FASTA3 algorithm was tested using a sequence of α-gliadin of *Triticum aestivum* (NCBI accession

number: CAB76964). This protein is one of the 72 representative proteins in the database and it contains 53 overlapping CeD-associated peptides identified with the exact sequence match tool. The sequence was altered *in silico* by substitutions in amino acid sequence to eliminate 53 exact peptide matches by amino acid substitutions of 13 alanine (A) residues in place of 12 Q and one tyrosine (Y) residues. For the second *in silico* modification, 11 theoretical substitutions were made with the addition of A in place of three serine (S), two glycine (G), four lysine (L), and one P and one Q amino acid residues. The two *in silico* modified alpha-gliadin sequences were evaluated using both the exact peptide match and FASTA3 tools. The exact match tool did not identify any possible risks, but the results with FASTA3 allowed us to consider criteria for meaningful alignments.

RESULTS

The CeD-Associated Peptide and Protein Database

In reviewing publications of CeD-associated peptides, broad differences were noted in specificity, sensitivity, and severity of reactions (34). For example, pure oat products that were not contaminated by wheat, barley, or rye, were reported to be well-tolerated by the majority of the CeD consumers (54, 55). Oats has been considered to be safe for CeD consumers by the United States Food and Drug Administration (US FDA). However, avenin-reactive T cells that mediate the intestinal inflammation typical of CeD were identified in several patients with CeD (56–59), and the adequacy of a number of the oat

safety evaluation studies remains controversial (60, 61). Since our goal is to include all known prolamin and glutelin peptides with scientific evidence of CeD induction to ensure that all CeD individuals are protected by our bioinformatics tools, the reported T-cell reactive avenin peptides were identified as risky for some CeD consumers, and 47 oat-derived peptides were included in the database.
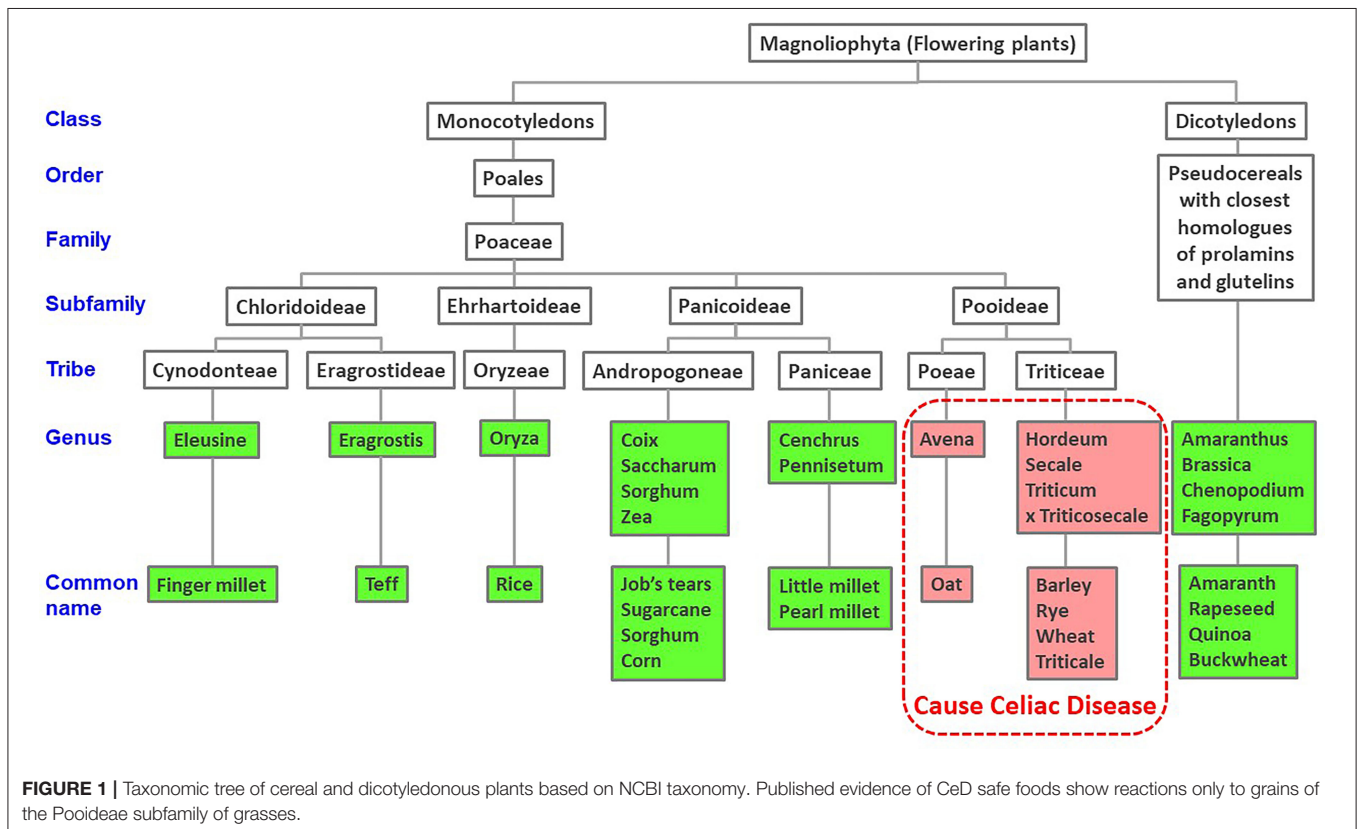
Statistics of the database versions 1 and 2 are summarized in **Table 1**. Overall, 68 relevant publications published between November 1984 and October 2012 were selected to collect 1,016 overlapping prolamin and glutelin peptides 8–55 AA long. Most peptides included in the database are immunogenic. From the 1,016 collected peptides, 997 are CD4+ T cell reactive while only one is CD8+ T cell reactive. This highlights the role of T cell-mediated inflammation in CeD pathogenesis. In addition, more than half of the prolamin and glutelin peptides are post-translationally deamidated. Of the 997 CD4+ T cell reactive peptides, 445 were in native sequences and 552 were in predicted deamidated sequences.

We note that the core nine-AA peptide of the collected T cell epitopes is different for HLA-DQ2 versus HLA-DQ8, and that the identity of immunogenic peptides relevant in HLA-DQ8 is less frequently noted than those of DQ2 epitopes. Possibly some immunogenic peptides may not have been reported yet, and this database should be updated periodically to ensure accuracy. From the 1,016 collected peptides, 18 elicited pathological effects to the intestine without evidence of specific T cell activation. Most of the collected toxic peptides appeared to trigger innate immune responses, yet some of their sequences overlap the immunogenic peptides. Version 2 of the database comprises a total of 1,013 causative peptides and 72 representative CeD-associated proteins (**Table 1**).

The Bioinformatics Tools for CeD Risk Assessment

BLAST searches indicated that all the CeD-associated peptides are found only in the prolamin and glutelin storage proteins of the Pooideae subfamily or in predicted deamidation products of those sequences. There were no other cereals outside of Pooideae that are likely to elicit CeD including corn, rice, sorghum, or millets (**Figure 1**). Our recommendation for the database users is that any query sequences found to contain even one of the known 1,013 peptides could be a risk for some individuals susceptible to CeD. Those proteins should be tested further to evaluate risks with MHC II-restricted CD4+ T cells from subjects with CeD or relevant toxicity before being introduced into a gluten-free food. We recognized that approximately 20% (562 of 2,666 for the first FASTA analysis and 1,163 of 5,786 for the second FASTA analysis) of the gluten-like proteins identified from Pooideae do not contain any of the known CeD reactive peptides (**Table 2**). Those proteins might be safe for CeD consumers; however since some T-cell reactive or toxic peptides may remain undiscovered, testing would be a conservative choice (62). Therefore, we recommend using the full-length FASTA sequence alignment tool to identify query sequences that may lack an exact peptide match to the 1,013

FIGURE 1 | Taxonomic tree of cereal and dicotyledonous plants based on NCBI taxonomy. Published evidence of CeD safe foods show reactions only to grains of the Pooideae subfamily of grasses.

peptides, but may include previously undefined CeD reactive peptides to confer an added layer of confidence.

To reinforce the utility of the FASTA alignment tool, *in silico* amino acid substitution in positions of exact CeD peptides of a CeD-associated α-gliadin (NCBI accession number: CAB76964) was conducted. The substitutions were made so that each of the known 53 overlapping CeD-associated peptides were no longer native and they are not identical to CeD peptides (**Figures 2A–C**). When the modified sequences were searched with the full FASTA3 sequence alignment tool, they showed > 95.5% identity full-length alignment to the original α-gliadin with *E*-scores smaller than 1.1e-78. These conservative substituted sequences might still be recognizable by the DQ2 or 8 restricted T-cells of patients with CeD. Without laboratory or clinical evidence of safety, it is prudent to flag these two sequences as probable risk factors for CeD. It is clear when using the full FASTA3 sequence alignment that careful evaluation of matching data is required since the query sequence can align with segments of the 72 representative CeD protein sequences in regions that do not harbor antigenic CeD determinants (**Figure 2A**, AA 98–219 and 255–290).

There are many gluten-like proteins in other grass subfamilies outside of Pooideae and some in dicotyledonous plants that have a clear history of safe consumption by those with CeD. A large number of these protein sequences were collected and used for testing with the first FASTA analysis in 2012 to provide identity scores, alignment overlap lengths, and *E*-scores that were used to set limits to differentiate conservative safety guidelines that are useful to identify possibly risky sequences (**Table 2**).

The alignment results indicated that the 562 Pooideae prolamin sequences lacking any exact match to the known CeD-associated peptides could have high identity FASTA alignments up to 98.4% over half-sequence length (187/288) and with an *E*-score of 2.7e-45 and can be up to 79.3% identical for a full-length (290/288) alignment with an *E*-score of 3.5e-63 to the representative CeD proteins. In contrast, a number of query sequences in non-Pooideae grass subfamilies (group II) were found to align over their full length with representative CeD proteins, and none were more than 43% identical with the representative CeD proteins (**Table 2**). Nearly all the query sequences in group II represent very short alignments with the representative CeD proteins and with the minimum *E*-score of 3.5e-17. In addition, full-length alignment comparison analyses of the prolamin-like sequences from dicotyledon class (group III) resulted in even lower identity scores and larger *E*-score values, while short overlaps (10/20) had up to 60% identities with *E*-scores as large as 8.8 (**Table 2**). Finally, 48 protein sequences from animals, fungi, or bacteria (group IV) were compared with the database by full-length alignment, and the FASTA results indicated that these 48 proteins could produce full-length (437/439) alignments with up to 41.2% identity and with the smallest *E*-score of 8.7e-25. The maximum identity was 72.7% over the half-sequence lengths (11/20) alignment, having a minimum *E*-score of 5.8e-03 (**Table 2**). The second FASTA analysis was performed in 2018 using version 2 of the database, and a total of 12,265 identified sequences led to the consistent results with those obtained from the first FASTA analysis (**Table 3**). In that analysis, of the 1,163 Pooideae prolamins lacking the exact peptide match (group I),

FIGURE 2 | (A) Amino acid sequence alignments of an α-gliadin (NCBI accession number: CAB76964) with 53 overlapping CeD-associated peptides identified with the exact sequence match tool; **(B)** full FASTA sequence alignment results with homology scores of the α-gliadin theoretically substituted with 13 alanine residues; **(C)** full FASTA sequence alignment results with homology scores of the α-gliadin theoretically substituted with 11 alanine residues.

the best FASTA alignment can be up to 98.9% identical over a full-length (264/279) alignment with an E-score of 1.4e-73 to the representative CeD proteins. Of the 1,755 query sequences from non-Pooideae grass subfamilies (group II) and 4,724 query sequences from dicotyledons class (group III), the best is a full-length FASTA alignment (168/181) of 40.5% identical with an E-score of 9.1e-09 to the representative CeD proteins.

In summary, the exact sequence match searches and the full FASTA sequence alignment searches indicated that the tested proteins in groups II, III, and IV from both the 2012 and 2018 analyses, are unlikely to implicate any of these proteins in CeD pathogenesis. The proteins found to contain no known CeD-associated peptides and when searched with FASTA comparison with the representative CeD proteins do not have significant identity matches and have relatively large E-scores. The percent identity describes the proportion of amino acids that are identical in alignment taking into consideration sequence and spacing. The E-score is a parameter describing the expected number of hits when searching a database of a particular size based on a log scale. Small numbers (e.g., $E = 1e-20$ or less) indicate likely significant matches.

DISCUSSIONS AND CONCLUSION

Several approaches can be undertaken to evaluate the CeD-associated immune response to novel foods including *in vitro* T cell studies using lines and clones and T cells obtained from patients after oral challenges and *in vivo* schemes using specific animal models. However, our database and algorithms aim at providing simple yet effective screening tools with low rates of false positive and false negative results. Identity to even one of the known CeD-associated peptides indicates a potential risk, and thus our proposed exact peptide sequence match tool is the most definitive comparison. Among the existing gluten-associated databases, the AllergenOnline.org CeD database contains the largest number of identified CeD-associated sequences (63, 64), and this allows higher quality and more versatile comparison. Since it has been a while since 2018, an expedited update was conducted in April 2022 to include newly identified CeD-associated peptides (65) and subsequently it will be given a rigorous evaluation as described for the 2018 version. Six more formal nine-AA CeD-relevant T cell epitopes, 22 CD4+ T cell reactive peptides, and four representative protein sequences were included bringing a total of 1,041 peptides and 76 protein sequences in our database. The full FASTA sequence alignment algorithm appears to be a useful tool to identify proteins with possible CeD risks while recognizing that the knowledge of the CeD-associated peptides is incomplete. The sequence comparison provides an additional assessment and compensates for the lack of identification of all CeD-associated peptides or cases of mutation that might remove exact match sequences, but possibly not diminish the CeD reactivity.

Upon evaluating the 4,823 (first analysis) and the 12,265 (second analysis) sequences of proteins with known risks of CeD and homologs from outside of Pooideae considered to be safe for CeD consumers, we compared alignment scores and characteristics of the two sequence groups. When compared with the representative CeD-associated proteins, all sequences from the latter group fall into one of the following alignment characteristics: having <45% sequence identity; having an E-score >1e-14; and aligned with many gaps or aligned with less than a full 100-AA overlap. These did not align with regions harboring the antigenic determinants. Although a single bioinformatics threshold cannot be attained, these four major observations together signify a less likely risk of CeD and are useful for careful evaluation of the FASTA alignment result. As a result, any query sequences identified with the FASTA comparison outside these alignment characteristics should not be defined as CeD safe without further verification. In fact, higher percent identity, lower E-scores, or longer the full alignments suggest a higher probability of sequence identity to the known CeD-associated proteins and thus appear to be of potential risk for eliciting CeD that must be critically evaluated further for the safe use for individuals with CeD.
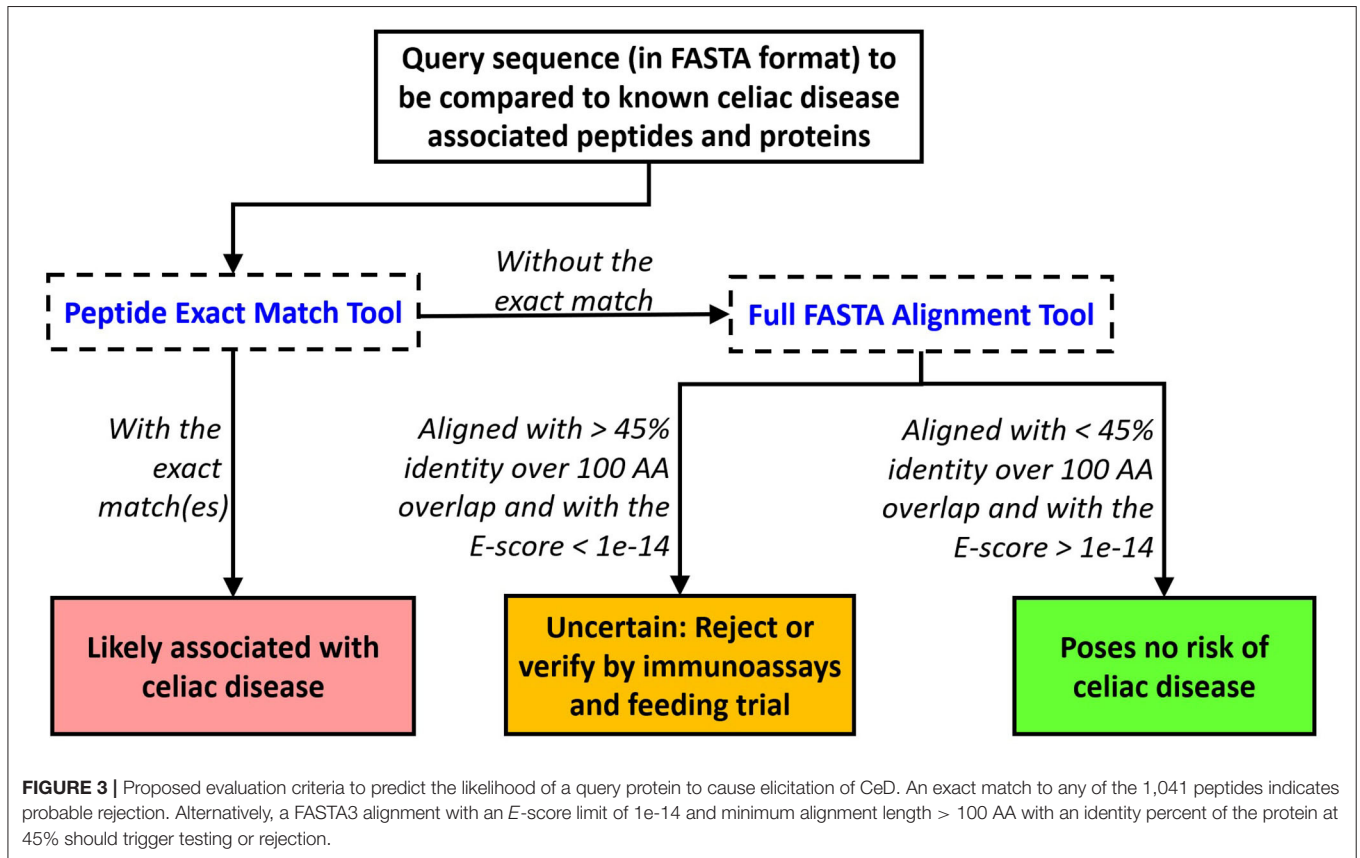
Our bioinformatics scheme and evaluation criteria for assessing novel food proteins for eliciting CeD are depicted in **Figure 3**. First, each query sequence should be screened for the presence of any of the 1,041 CeD-associated peptides using the peptide exact match tool. An exact match to any of the known CeD-associated peptides indicates a probable risk. Query sequences without matches to the known CeD-associated peptides are then evaluated to identity high-scoring matches to the 76 representative CeD proteins using the full FASTA alignment tool. Sequence alignments to any of the representative CeD proteins with at least 45% identity and an E-score <1e-14 over 100 AA alignment suggest that they may harbor antigenic determinants. The significance of matches by either method can be verified by testing using the appropriate specific T cell (66).

In 2017, the EFSA Panel on Genetically Modified Organisms (GMO) published guidance on the allergenicity assessment of genetically modified plants stating that any new protein expressed in a GMO must be evaluated for safety to CeD consumers. The guideline was referred to our CeD database but also suggested sequence evaluation with the 4 AA (Q/E-X1-P-X2) motifs (52). Song et al. (67) however demonstrated that the Q-X1-P-X2 motif searches for potential CeD risk yielded poor selectivity and suggested including our FASTA comparison evaluation to improve the risk assessment efficiency (67).

New varieties of crops are being developed as new food sources. Perennial grain crops offer potentially sustainable new food sources. A wheatgrass *Thinopyrum intermedium*, also in the Pooideae subfamily, has been under development for 30 years as a possible alternative to annual wheat (*Triticum* sp.) (68). Our preliminary analysis however indicated that 31 predicted probable proteins from the *T. intermedium* genome contain multiple known CeD-associated peptides. It is highly likely that consumption of *T. intermedium* grain would also trigger CeD in some individuals. The predicted *T. intermedium* proteins also match more than 30 known wheat allergen sequences in the AllergenOnline database version 21, suggesting a risk for consumers with a wheat allergy.

In summary, our CeD peptide and protein database with bioinformatics tools have been robustly evaluated by us and

FIGURE 3 | Proposed evaluation criteria to predict the likelihood of a query protein to cause elicitation of CeD. An exact match to any of the 1,041 peptides indicates probable rejection. Alternatively, a FASTA3 alignment with an *E*-score limit of 1e-14 and minimum alignment length > 100 AA with an identity percent of the protein at 45% should trigger testing or rejection.

by other users in the past 10 years and proven to offer an effective screening system for the identification and analysis of CeD-associated peptides and proteins for a thorough food safety evaluation (67, 69–77). The curated database and tools are available for public use free of charge at http://www.allergenonline.org/celiachome.shtml, and an update and validation of the risk evaluation criteria will be continued by our peer review panel members.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

PA, ST, and RG: conceptualization and writing—review and editing. PA, JW, and RG: methodology. JW: software. PA, AT, BB, FF, ST, and RG: validation. PA and MA: formal analysis, investigation, and writing—original draft preparation. PA and RG: resources, data curation, and visualization. RG: supervision and project administration. ST and RG: funding acquisition. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/falgy.2022.900573/full#supplementary-material

REFERENCES

1. van Putten MC, Frewer LJ, Gilissen LJWJ, Gremmen B, Peijnenburg AACM, Wichers HJ. Novel foods and food allergies: a review of the issues. *Trends Food Sci Technol.* (2006) 17:289 doi: 10.1016/j.tifs.2005.11.010

2. Codex Alimentarius Commission. In: *Alinorm 03/34: Joint FAO/WHO Food Standard Programme, Codex Alimentarious Commission, Twenty-Fifth Session, Rome, Italy, 30 June−5 July, 2003. Appendix III, Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants, and Appendix IV, Annex on the assessment of possible allergenicity.* Twenty-Fifth Session (FA); Joint FAO/WHO Food Standards Programme. (2003). p. 47–60.

3. Rubio-Tapia A, Ludvigsson JF, Brantner TL, Murray JA, Everhart JE. The prevalence of celiac disease in the United States. *Am J Gastroenterol.* (2012) 107:1538–44; quiz 1537, 1545. doi: 10.1038/ajg.2012.219

4. Singh P, Arora A, Strand TA, Leffler DA, Catassi C, Green PH, et al. Global prevalence of celiac disease: systematic review and meta-analysis. *Clin Gastroenterol Hepatol.* (2018) 16:823–36.e2. doi: 10.1016/j.cgh.2017.06.037

5. Cukrowska B, Sowinska A, Bierla JB, Czarnowska E, Rybak A, Grzybowska-Chlebowczyk U. Intestinal epithelium, intraepithelial lymphocytes and the gut microbiota - key players in the pathogenesis of celiac disease. *World J Gastroenterol.* (2017) 23:7505–18. doi: 10.3748/wjg.v23.i42.7505

6. Di Sabatino A, Corazza GR. Coeliac disease. *Lancet.* (2009) 373:1480–43. doi: 10.1016/S0140-6736(09)60254-3

7. Groschwitz KR, Hogan SP. Intestinal barrier function: molecular regulation and disease pathogenesis. *J Allergy Clin Immunol.* (2009) 124:3–20. doi: 10.1016/j.jaci.2009.05.038

8. Lammers KM, Lu R, Brownley J, Lu B, Gerard C, Thomas K, et al. Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3. *Gastroenterology.* (2008) 135:194–204.e3. doi: 10.1053/j.gastro.2008.03.023

9. Fasano A. Zonulin and its regulation of intestinal barrier function: the biological door to inflammation, autoimmunity, and cancer. *Physiol Rev.* (2011) 91:151–75. doi: 10.1152/physrev.00003.2008

10. Perez-Gregorio MR, Días R, Mateus N, de Freitas V. Identification and characterization of proteolytically resistant gluten-derived peptides. *Food Funct.* (2018) 9:1726–35. doi: 10.1039/C7FO02027A

11. Polvi A, Arranz E, Fernandez-Arquero M, Collin P, Mäki M, Sanz A, et al. HLA-DQ2-negative celiac disease in Finland and Spain. *Hum Immunol.* (1998) 59:169–75. doi: 10.1016/S0198-8859(98)00008-1

12. Romanos J, van Diemen CC, Nolte IM, Trynka G, Zhernakova A, Fu J, et al. Analysis of HLA and non-HLA alleles can identify individuals at high risk for celiac disease. *Gastroenterology.* (2009) 137:834–40.e3. doi: 10.1053/j.gastro.2009.05.040

13. Sollid LM. The roles of MHC class II genes and post-translational modification in celiac disease. *Immunogenetics.* (2017) 69:605–16. doi: 10.1007/s00251-017-0985-7

14. Fasano A, Berti I, Gerarduzzi T, Not T, Colletti RB, Drago S, 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) 163:286–92. doi: 10.1001/archinte.163.3.286

15. Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, 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) 64:469–77. doi: 10.1016/S0198-8859 (03)00027-2

16. du Pré MF, Blazevski J, Dewan AE, Stamnaes J, Kanduri C, Sandve GK, et al. B cell tolerance and antibody production to the celiac disease autoantigen transglutaminase 2. *J Exp Med.* (2020) 217:e20190860. doi: 10.1084/jem.20190860

17. Tripathi A, Lammers KM, Goldblum S, Shea-Donohue T, Netzel-Arnett S, Buzza MS, et al. Identification of human zonulin, a physiological modulator of tight junctions, as prehaptoglobin-2. *Proc Natl Acad Sci USA.* (2009) 106:16799–804. doi: 10.1073/pnas.0906773106

18. Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanović S, Eriksen JE, et al. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. *Eur J Immunol.* (1996) 26:2764–72. doi: 10.1002/eji.1830261132

19. van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Koning F. Peptide binding characteristics of the coeliac disease-associated DQ (alpha1*0501, beta1*0201) molecule. *Immunogenetics.* (1996) 44:246–53. doi: 10.1007/BF02602553

20. Kim CY, Quarsten H, Bergseng E, Khosla C, Sollid LM. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. *Proc Natl Acad Sci USA.* (2004) 101:4175–9. doi: 10.1073/pnas.0306885101

21. Kwok WW, Domeier ML, Raymond FC, Byers P, Nepom GT. Allele-specific motifs characterize HLA-DQ interactions with a diabetes-associated peptide derived from glutamic acid decarboxylase. *J Immunol.* (1996) 156:2171–7.

22. Henderson KN, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T, et al. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity.* (2007) 27:23–34. doi: 10.1016/j.immuni.2007.05.015

23. van de Wal Y, Kooy Y, van Veelen P, Peña S, Mearin L, Papadopoulos G, et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *J Immunol.* (1998) 161:1585–8.

24. Arentz-Hansen H, Korner R, Molberg Ø, Quarsten H, Vader W, Kooy YM, et al. The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J Exp Med.* (2000) 191:603–12. doi: 10.1084/jem.191.4.603

25. Vader LW, de Ru A, van der Wal Y, Kooy YM, Benckhuijsen W, Mearin ML, et al. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. *J Exp Med.* (2002) 195:643–49. doi: 10.1084/jem.20012028

26. Stepniak D, Vader LW, Kooy Y, van Veelen PA, Moustakas A, Papandreou NA, et al. T-cell recognition of HLA-DQ2-bound gluten peptides can be influenced by an N-terminal proline at p-1. *Immunogenetics.* (2005) 57:8–15. doi: 10.1007/s00251-005-0780-8

27. Dørum S, Arntzen MØ, Qiao SW, Holm A, Koehler CJ, Thiede B, et al. The preferred substrates for transglutaminase 2 in a complex wheat gluten digest are peptide fragments harboring celiac disease T-cell epitopes. *PLoS ONE.* (2010) 5:e14056. doi: 10.1371/journal.pone.0014056

28. Hovhannisyan Z, Weiss A, Martin A, Wiesner M, Tollefsen S, Yoshida K, et al. The role of HLA-DQ8 beta57 polymorphism in the anti-gluten T-cell response in coeliac disease. *Nature.* (2008) 456:534–38. doi: 10.1038/nature07524

29. Fleckenstein B, Molberg Ø, Qiao SW, Schmid DG, von der Mülbe F, Elgstøen K, et al. Gliadin T cell epitope selection by tissue transglutaminase in celiac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. *J Biol Chem.* (2002) 277:34109–16. doi: 10.1074/jbc.M204521200

30. Jabri B, Sollid LM. T cells in celiac disease. *J Immunol.* (2017) 198:3005–14. doi: 10.4049/jimmunol.1601693

31. Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet.* (2008) 40:395–402. doi: 10.1038/ng.102

32. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet.* (2010) 42:295–302. doi: 10.1038/ng.543

33. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet.* (2011) 43:1193–201. doi: 10.1038/ng.998

34. Gianfrani C, Troncone R, Mugione P, Cosentini E, De Pascale M, Faruolo C, et al. Celiac disease association with CD8+ T cell responses: identification of a novel gliadin-derived HLA-A2-restricted epitope. *J Immunol.* (2003) 170:2719–26. doi: 10.4049/jimmunol.170.5.2719

35. Londei M, Ciacci C, Ricciardelli I, Vacca L, Quaratino S, Maiuri L. Gliadin as a stimulator of innate responses in celiac disease. *Mol Immunol.* (2005) 42:913–8. doi: 10.1016/j.molimm.2004.12.005

36. Jabri B, Sollid LM. Tissue-mediated control of immunopathology in coeliac disease. *Nat Rev Immunol.* (2009) 9:858–70. doi: 10.1038/nri2670

37. Mention JJ, Ben Ahmed M, Bègue B, Barbe U, Verkarre V, Asnafi V, et al. Interleukin 15: a key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease. *Gastroenterology.* (2003) 125:730–45. doi: 10.1016/S0016-5085(03)01047-3

38. Malamut G, El Machhour R, Montcuquet N, Martin-Lanneree S, Dusanter-Fourt I, Verkarre V, et al. IL-15 triggers an antiapoptotic pathway in human intraepithelial lymphocytes that is a potential new target in celiac

disease-associated inflammation and lymphomagenesis. *J Clin Invest.* (2010) 120:2131–43. doi: 10.1172/JCI41344

39. Roberts AI, Lee L, Schwarz E, Groh V, Spies T, Ebert EC, et al. NKG2D receptors induced by IL-15 costimulate CD28-negative effector CTL in the tissue microenvironment. *J Immunol.* (2001) 167:5527–30. doi: 10.4049/jimmunol.167.10.5527

40. Meresse B, Chen Z, Ciszewski C, Tretiakova M, Bhagat G, Krausz TN, et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. *Immunity.* (2004) 21:357–66. doi: 10.1016/j.immuni.2004.06.020

41. Tang F, Chen Z, Ciszewski C, Setty M, Solus J, Tretiakova M, et al. Cytosolic PLA2 is required for CTL-mediated immunopathology of celiac disease via NKG2D and IL-15. *J Exp Med.* (2009) 206:707–19. doi: 10.1084/jem.20071887

42. Auricchio S, De Ritis G, De Vincenzi M, Occorsio P, Silano V. Effects of gliadin-derived peptides from bread and durum wheats on small intestine cultures from rat fetus and coeliac children. *Pediatr Res.* (1982) 16:1004–10. doi: 10.1203/00006450-198212000-00006

43. Barone MV, Zanzi D, Maglio M, Nanayakkara M, Santagata S, Lania G, et al. Gliadin-mediated proliferation and innate immune activation in celiac disease are due to alterations in vesicular trafficking. *PLoS ONE.* (2011) 6:e17039. doi: 10.1371/journal.pone.0017039

44. Caputo I, Barone MV, Lepretti M, Martucciello S, Nista I, Troncone R, et al. Celiac anti-tissue transglutaminase antibodies interfere with the uptake of alpha gliadin peptide 31-43 but not of peptide 57-68 by epithelial cells. *Biochim Biophys Acta.* (2010) 1802:17–27. doi: 10.1016/j.bbadis.2010.05.010

45. de Ritis G, Auricchio S, Jones HW, Lew EJ, Bernardin JE, Kasarda DD. *In vitro* (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. *Gastroenterology.* (1988) 94:41–9. doi: 10.1016/0016-5085(88)90607-5

46. Sturgess R, Day P, Ellis HJ, Lundin KE, Gjertsen HA, Kontakou M, et al. Wheat peptide challenge in coeliac disease. *Lancet.* (1994) 343:758–61. doi: 10.1016/S0140-6736(94)91837-6

47. Mantzaris G, Jewell DP. *In vivo* toxicity of a synthetic dodecapeptide from a gliadin in patients with coeliac disease. *Scand J Gastroenterol.* (1991) 26:392–8. doi: 10.3109/00365529108996500

48. Wieser H, Belitz HD, Idar D, Ashkenazi A. Coeliac activity of the gliadin peptides CT-1 and CT-2. *Z. Lebensm Unters Forsch.* (1986) 182:115–7. doi: 10.1007/BF01454241

49. Biagi F, Ellis HJ, Parnell ND, Shidrawi RG, Thomas PD, O'Reilly N, et al. A non-toxic analogue of a coeliac-activating gliadin peptide: a basis for immunomodulation? *Aliment Pharmacol Ther.* (1999) 13:945–50. doi: 10.1046/j.1365-2036.1999.00512.x

50. Maiuri L, Troncone R, Mayer M, Coletta S, Picarelli A, De Vincenzi M, *In vitro* activities of A-gliadin-related synthetic peptides: damaging effect on the atrophic coeliac mucosa and activation of mucosal immune response in the treated coeliac mucosa. *Scand J Gastroenterol.* (1996) 31:247–53. doi: 10.3109/00365529609004874

51. Pearson WR. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol.* (2000) 132:185–219. doi: 10.1385/1-59259-192-2:185

52. EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), Naegeli H, Birch AN, Casacuberta J, De Schrijver A, et al. Guidance on allergenicity assessment of genetically modified plants. *EFSA J.* (2017) 15:4862. doi: 10.2903/j.efsa.2017.4862

53. Sollid LM, Qiao SW, Anderson RP, Gianfrani, C, Koning F. Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics.* (2012) 64:455–60. doi: 10.1007/s00251-012-0599-z

54. Picarelli A, Di Tola M, Sabbatella L, Gabrielli F, Di Cello T, Anania MC, et al. Immunologic evidence of no harmful effect of oats in celiac disease. *Am J Clin Nutr.* (2001) 74:137–40. doi: 10.1093/ajcn/74.1.137

55. Rashid M, Butzner D, Burrows V, Zarkadas M, Case S, Molloy M, et al. Consumption of pure oats by individuals with celiac disease: a position statement by the Canadian Celiac Association. *Can J Gastroenterol.* (2007) 21:649–51. doi: 10.1155/2007/340591

56. Vader LW, Stepniak DT, Bunnik EM, Kooy YMC, De Haan W, Drijfhout JW, et al. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. *Gastroenterology.* (2003) 125:1105–13. doi: 10.1016/S0016-5085(03)01204-6

57. Arentz-Hansen H, Fleckenstein B, Molberg Ø, Scott H, Koning F, Jung G, et al. The molecular basis for oat intolerance in patients with celiac disease. *PLoS Med.* (2004) 1:e1. doi: 10.1371/journal.pmed.0010001

58. Real A, Comino I, de Lorenzo L, Merchán F, Gil-Humanes J, Giménez MJ, et al. Molecular and immunological characterization of gluten proteins isolated from oat cultivars that differ in toxicity for celiac disease. *PLoS ONE.* (2012) 7:e48365. doi: 10.1371/journal.pone.0048365

59. Hardy MY, Tye-Din JA, Stewart JA, Schmitz F, Dudek NL, Hanchapola I, et al. Ingestion of oats and barley in patients with celiac disease mobilizes cross-reactive T cells activated by avenin peptides and immuno-dominant hordein peptides. *J Autoimmun.* (2015) 56:56–65. doi: 10.1016/j.jaut.2014.10.003

60. Pinto-Sánchez MI, Causada-Calo N, Bercik P, Ford AC, Murray JA, Armstrong D, et al. Safety of adding oats to a gluten-free diet for patients with celiac disease: systematic review and meta-analysis of clinical and observational studies. *Gastroenterology.* (2017) 153:395–409.e3. doi: 10.1053/j.gastro.2017.04.009

61. Ciacci C, Ciclitira P, Hadjivassiliou M, Kaukinen K, Ludvigsson JF, McGough N, et al. The gluten-free diet and its current application in coeliac disease and dermatitis herpetiformis. *United European Gastroenterol J.* (2015) 3:121–35. doi: 10.1177/2050640614559263

62. Koning F, Gilissen L, Wijmenga C. Gluten: a two-edged sword. immunopathogenesis of celiac disease. *Springer Semin Immunopathol.* (2005) 27:217–32. doi: 10.1007/s00281-005-0203-9

63. Juhász A, Haraszi R, Maulis C. ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families. *Database.* (2015) 2015:1–16. doi: 10.1093/database/bav100

64. Bromilow S, Gethings LA, Buckley M, Bromley M, Shewry PR, Langridge JI, et al. A curated gluten protein sequence database to support development of proteomics methods for determination of gluten in gluten-free foods. *J Proteomics.* (2017) 163:67–75. doi: 10.1016/j.jprot.2017.03.026

65. Sollid LM, Tye-Din JA, Qiao S-W, Anderson RP, Cianfrani C, Koning F. Update 2020: nomenclature and listing of celiac disease-relevant gluten epitopes recognized by CD4+ T cells. *Immunogenetics.* (2020). 72:85–8. doi: 10.1007/s00251-019-01141-w

66. Raki M, Fallang LE, Brottveit M, Berseng E, Quarsten H, Lundin KE, Sollid LM. Tetramer visualization of gut-homing gluten-specific T cells in the peripheral blood of celiac disease patients. *Proc Natl Acad Sci USA.* (2007) 104:2831–6. doi: 10.1073/pnas.0608610104

67. Song P, Podevin N, Mirsky H, Anderson J, Delaney B, Mathesius C, et al. Q-X1-P-X2 motif search for potential celiac disease risk has poor selectivity. *Regul Toxicol Pharmacol.* (2018) 99:233–7. doi: 10.1016/j.yrtph.2018.09.021

68. DeHaan L, Larson S, Lopez-Marques RL, Wenkel S, Gao C, Palmgren M. *Trends Plant Sci.* (2020) 25:525–37. doi: 10.1016/j.tplants.2020.02.004

69. Röckendorf N, Meckelein B, Scherf KA, Schalk K, Koehler P, Frey A. Identification of novel antibody-reactive detection sites for comprehensive gluten monitoring. *PLoS ONE.* (2017) 12:e0181566. doi: 10.1371/journal.pone.0181566

70. Pilolli R, Gadaleta A, Mamone G, Nigro D, De Angelis E, Montemurro N, et al. Scouting for naturally low-toxicity wheat genotypes by a multidisciplinary approach. *Sci Rep.* (2019) 9:1646. doi: 10.1038/s41598-018-38845-8

71. Maurer-Stroh S, Krutz NL, Kern PS, Gunalan V, Nguyen MN, Limviphuvadh V, et al. AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics.* (2019) 35:3020–7. doi: 10.1093/bioinformatics/btz029

72. Fiedler KL, Cao W, Zhang L, Naziemiec M, Bedford B, Yin L, et al. Detection of gluten in a pilot-scale barley-based beer produced with and without a prolyl endopeptidase enzyme. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess.* (2019) 36:1151–62. doi: 10.1080/19440049.2019.1616830

73. Li H, Bose U, Stockwell S, Howitt CA, Colgrave M. Assessing the utility of multiplexed liquid chromatography-mass spectrometry for gluten detection in Australian breakfast food products. *Molecules.* (2019) 24:3665. doi: 10.3390/molecules24203665

74. Pilolli R, Gadaleta A, Di Stasio L, Lamonaca A, De Angelis E, Nigro D, et al. A comprehensive peptidomic approach to characterize the protein profile of

selected Durum wheat genotypes: implication for coeliac disease and wheat allergy. *Nutrients.* (2019) 11:2321. doi: 10.3390/nu11102321

75. Cao W, Baumert JL, Downs ML. Compositional and immunogenic evaluation of fractionated wheat beers using mass spectrometry. *Food Chem.* (2020) 333:127379. doi: 10.1016/j.foodchem.2020. 127379

76. Daly M, Bromilow SN, Nitride C, Shewry PR, Gethings LA, Mills ENC. Mapping coeliac toxic motifs in the prolamin seed storage proteins of barley, rye, and oats using a curated sequence database. *Front Nutr.* (2020) 7:87. doi: 10.3389/fnut.2020. 00087

77. Pilolli R, De Angelis M, Lamonaca A, De Angelis E, Rizzello CG, Siragusa S, et al. Prototype gluten-free breads from processed Durum wheat: use of monovarietal flours and implications for gluten detoxification strategies. *Nutrients.* (2020) 12:3824. doi: 10.3390/nu121 23824