Trends in biological and genetic markers of gluten-related disorders research

Tendências em marcadores biológicos e genéticos de pesquisas sobre distúrbios relacionados ao glúten

ABSTRACT

This study aimed to evaluate the distribution and progress of the scientific production on biological and genetic markers of gluten-related disorders (BGM-GRD), supporting by a scientometric analysis. Emerging trends were mapped using CiteSpace. 1,050 studies were retrieved from Web of Science (1945-2022), of which 452 were eligible. Early 90s, research started to rise significantly until now. Celiac disease (CD) was the most studied. Studies were performed in patients with 53 different diseases or conditions, mainly type 1 diabetes mellitus (24.1%). Scientific production increased by 2.4 times from 2001-2021, while the citations numbers quintupled, which reflects the high quality of research (H-index of 56). Italy, USA and Spain stood out in productivity; USA was the core of the network of the largest and oldest cluster of countries. Australia, Germany, Israel and Poland stood out in quality. Eighteen from 817 institutions produced 46.7% of studies. Recently, the main interests are in the effect of environmental on metabolic pathways, determination of the prevalence and diagnosis of CD, mainly in children, as well as risk assessment and validation of biomarkers for CD diagnosis, with a focus on non-invasive methods.

Keywords: Biomarkers; Celiac disease; Dermatitis herpetiformis; Gluten sensitivity
RESUMO
Este estudo teve como objetivo avaliar a distribuição e progresso da produção científica sobre marcadores biológicos e genéticos de distúrbios relacionados ao glúten (MBG-DRG), apoiado pela análise cienciométrica. Tendências emergentes foram mapeadas usando o CiteSpace. 1.050 estudos foram recuperados da Web of Science (1945-2022), dos quais 452 foram elegíveis. No início dos anos 90, a pesquisa iniciou e aumentou significativamente até hoje. Doença celíaca (DC) foi a mais estudada. Estudos foram conduzidos com pacientes com 53 doenças ou condições diferentes, principalmente diabetes mellitus tipo 1 (24,1%). A produção científica aumentou 2,4 vezes entre 2001 e 2021, enquanto o número de citações quintuplicou, o que reflete a elevada qualidade da pesquisa (índice-H de 56). Itália, EUA e Espanha destacaram-se em produtividade; EUA foram o centro da rede de colaboração do maior e mais antigo cluster de países. Austrália, Alemanha, Israel e Polônia destacaram-se em qualidade. Dez oito das 817 instituições produziram 46,7% dos estudos. Recentemente, os principais interesses estão no efeito do ambiente nas vias metabólicas, na determinação de prevalência e no diagnóstico da DC, principalmente em crianças, bem como na avaliação de risco e validação de biomarcadores para diagnóstico de DC, com foco em métodos não invasivos.
Palavras-chave: Biomarcadores; Doença celíaca; Dermatite Hepertiformes; Sensibilidade ao glutén

INTRODUCTION

Increasing intake of wheat and its derivatives, as well as the increase in the gluten content of wheat varieties, has contributed to raising the diagnosis and/or prevalence of gluten-related disorders (GRD) (Taraghikhah et al., 2020). This trend has been also associated with a better understanding of the clinical conditions of these disorders, coupled with the development of new diagnosis techniques and the increase in scientific knowledge, which results in a greater dissemination of information about GRD among the population (Auricchio & Troncone, 2021).

GRD consist of a set of heterogeneous disorders with intestinal and/or extraintestinal manifestations caused by gluten intake and that improve after excluding gluten from the diet. GRD include celiac disease (CD), dermatitis herpetiformis (DH), gluten ataxia (GA), wheat allergy (WA), and non-celiac gluten sensitivity (NCGS). The first three are of autoimmune etiology; the fourth is an allergic condition, while the last is a disease that leads to intestinal and extra-intestinal manifestations that are not mediated by an immune or allergic response to gluten (Asri & Rostami-Nejad, 2022; Cabanillas, 2020; Taraghikhah et al., 2020).

Celiac disease (CD) is a serious autoimmune disease triggered by the consumption of gluten in the diet of genetically predisposed people. CD is characterized by inflammation and atrophy of the small intestine villi and, therefore, can be associated with several manifestations, including chronic and persistent intestinal symptoms, atypical and extra-intestinal symptoms, or even absence of them. (Perrotta & Guerrieri, 2022; Smithson, Siegelman, Oki, Maxwell, & Leffler, 2021).

Gluten ataxia is a neurological manifestation induced by gluten, with CD being its main cause. The primary target in the nervous system is intestinal axis tissue transglutaminase 2
(tTG2), although tTG6 may also be a causative antigen. In general, GA is investigated in diagnosed celiac patients and may not necessarily be associated with gastrointestinal symptoms (Hadjivassiliou, Sanders, & Aeschlimann, 2015; Osman et al., 2021).

Another immune reaction to gluten that has been linked to CD is dermatitis herpetiformis, a rare but persistent blistering cutaneous disease that presents granular deposits of IgA along the dermal-epidermal junction. This GRD is strongly associated with the presence of HLA-B8 and HLA-DR3 antigens, although different autoantibodies have been detected in small numbers in patients with DH (Nguyen & Kim, 2021).

NCGS refers to a non-autoimmune-allergic reaction to gluten that leads to intestinal and extra-intestinal manifestations and can cause a wide variety of symptoms, including weight loss, headache, muscle pain, fatigue, malaise, abdominal pain, diarrhea, recurrent oral ulceration, and depression. Although the primary cause of CD is known and very specific, since it is an inflammatory response to gluten, NCGS may not only be related to the ingestion of gluten, and its causes are still unknown. Unlike CD, the development of NCGS has also been related to the ingestion of other components of wheat, including poorly absorbed short-chain carbohydrates (FODMAPs) and amylase/trypsin inhibitors (ATIs) from wheat (Barbaro, Cremon, Wrona, et al., 2020; Catassi, 2015; Catassi et al., 2015).

The wheat allergy (WA) is characterized by the activation of type 2 T helper cells (Th2), which can result in IgE mediated and non-IgE mediated reactions. WA mediated by immunoglobulin E (IgE) can be related to wheat inhalation (respiratory allergy) or wheat ingestion (food allergy) and occurs immediately. Thus, it causes asthma or rhinitis, and it may be identified by the presence of wheat specific IgE antibodies. On the other hand, non-IgE mediated WA is characterized by chronic eosinophilic and lymphocytic infiltration of the gastrointestinal tract (Cianferoni, 2016).

In general, the diagnosis of GRD is performed by a combination of serological tests, clinical symptoms, and histopathological evaluation. However, the variety and overlapping of typical and atypical clinical manifestations of GRD make the diagnosis very difficult. Particularly, in the case of CD, the best characterized GRD, there is a global distribution among the population and not in a specific group with a higher incidence. It is estimated that it is still highly underdiagnosed, as 75-90% of the celiac population in Western countries has not been identified, presumably due to the absence or atypical nature of symptoms (Pratesi et al., 2003; Sahin, 2021). Although the diagnosis of CD has been made with the help of serologic, molecular, and histopathological biomarkers, no accurate biomarker has been established for diagnosis and monitoring of NCGS so far (Brusca, 2015).

Genetic markers of GRD have been studied to determine the prevalence or risk of GRD. Despite the association between CD and the presence of HLA class II haplotypes (HLA-DQ2 and HLA-DQ8) is well established, CD results from the combination of effects of different gene
products, including other HLA and non-HLA genes, which are also important in the development of the disease. To date, 40 genes have been genetically associated with CD, but it is not known how these pathways transform a predisposed individual into an affected person (Ramirez-Sanchez et al., 2022; Rodríguez-Martín, Vaquero, & Vivas, 2020). Besides, although HLA-DQ2 and HLA-DQ8 haplotypes are found in approximately 50% of patients with NCGS, HLA typing cannot be used for diagnosing NCGS as is the case in biopsy-confirmed celiac patients (Cabanillas, 2020).

Due to diagnostic limitations based mainly on similar evidence or the absence of some of them, research aimed at the identification and selection of BGM-GRD have been increased in the last decades to be used in clinical practice to diagnose and monitor the progression of these disorders. Therefore, this study aims to evaluate the distribution and progress over time of the scientific production on biological and genetic markers from gluten-related disorders, supporting by a scientometric analysis.

METHODS

A scientometric analysis has been adopted to quantify the current evidence regarding BGM-GRD, which include celiac disease (CD), dermatitis herpetiformis (DH), gluten ataxia (GA), wheat allergy (WA) and non-celiac gluten sensitivity (NCGS).

Selection of the scientific production was systematically conducted following The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). Scientific articles published from 1945 to October 9th, 2022 (date of the last article consulted) were searched in the core collection of the Web of Science (WoS) database.

Keywords for “biomarkers” and “genetic markers” used in the WoS advanced search were structured vocabularies DeCS/MeSH (Health Sciences descriptors/ Medical Subject Headings), available at the U.S. National Library of Medicine (NLM).

The advanced search was performed by topics (title, abstract, and keywords) using different combinations of keywords and Boolean scripts. After multiple trials of combinations, the most complete algorithm was found and used for selection, as follows: TS= ((“gluten disorder*” or “gluten related disorder*” or “c$eliac disease” or “c$eliac sprue” or “Gluten sensitive enteropath*” or “gluten enteropath*” or “nontropical sprue” or sprue or “wheat allerg*” or “gluten allerg*” or “gluten ataxia” or “Triticum allerg*” or “dermatitis herpetiformis” or “non c$eliac gluten sensitivit*” or “gluten *sensitivit*” or “wheat *sensitivi*” or “non c$eliac gluten intolerance”) AND (biomarker* or “biologic* marker*” or “immun* marker*” or “serum marker*” or “serologic* marker*” or “histo* marker*” or “clinical marker*” or “biochemical marker*” or “laboratory marker*” or “surrogate marker*” or...
“viral marker*” or “surrogate endpoint*” or “surrogate end point*” or “surrogate marker*” or “molecular biomarker*” or “genetic* marker*” or “chromosome marker*” or “DNA marker*”).

The search was performed without any filter or limit. In addition, no linguistic or geographic restrictions were imposed. Early access, meeting abstracts, editorial material, book chapters, corrections, news items, letters, notes as well as review articles were excluded. Finally, the first three authors manually removed the duplicate records from the selected list. Additionally, Endnote was used as an automation tool in this process.

Then, the screening step was conducted based on their alignment with the proposed subject and papers out of scope were removed. Eligible articles were studies on BGM-GRD, performed in the human biological system as serum, blood, urine, feces, cells, and tissues, analyzing antibodies, proteins, RNA, DNA, and the gut microbiome. Biomarkers of gluten intake, commonly used for monitoring free-gluten diet adherence by GRD patients, were also included. On the other hand, studies with no evaluation of BGM-GRD, in vitro, cell culture, and animal studies as well as survey/questionnaires studies were not included. Bioinformatics research and studies about the development and performance evaluation of assays/tests/sensors for the detection or quantification of BGM-GRD were also excluded. Non-gluten-related gastrointestinal disorders such as ulcerative colitis, irritable bowel syndrome and Crohn’s Disease were not included.

Articles were manually separated into five subgroups of GRD (CD, DH, GA, NCGS, and WA) according to the group of patients (children/adolescents, adults, seniors), the type of marker and diseases or conditions addressed together GRD, which were represented into a word clouds figure generated by the free online Word Cloud generator (wordclouds.com).

A scientometric analysis was carried out from the eligible articles through descriptive quantitative analysis. The data were downloaded using the WoS tool “Analyze Results”, which provided the publication numbers per WoS categories, publication years, affiliations, journals, Publishers, Funding agencies, countries, authors, languages, and research areas. In Microsoft Excel, graphics were generated regarding the relationship between these variables and the number of publications. Furthermore, the WoS Citation Report was generated to provide the H-Index and citations of eligible articles. CiteSpace 6.1.R3 (64-bit) Advanced software (2022) (Chen, 2006), a Java-based scientific visualization software package, was used for mapping out trends and gaps in current knowledge over time, producing infographics related to countries, references, institutions, and keywords.

RESULTS AND DISCUSSION

Characteristics of publication outputs
The initial advanced search retrieved 1,050 scientific studies published from 1945 to October 9th, 2022 in the WoS core collection. No duplicate records were found. Early access, meeting abstracts, editorial material, proceedings papers, corrections, news items, letters as well as review articles were excluded, totaling 311 documents. Therefore, only articles were included, remaining 739 records. The screening step of the remaining records was carried out based on the alignment of the articles with the proposed subject, and 287 studies were manually excluded, totaling 452 eligible articles for scientometric analysis.

The eligible articles about BGM-GRD primarily consisted of studies including celiac disease (91.3%). Non-celiac gluten sensitivity, dermatitis herpetiformis, wheat allergy, and gluten ataxia corresponded to 6.2%, 2.6%, 0.4%, and 0.2%, respectively.

These results agreed with the recent findings on BGM-GRD. So far, the most accurate diagnostic support involving BGM occurs in celiac patients (Gandini, Gededzha, De Maayer, Barrow, & Mayne, 2021). CD diagnosis is made based on clinical conditions and various serologic assays as well as endoscopic and histological tests, by gluten-containing diet maintenance during the diagnostic process. The gold standard for CD diagnosis is the presence of villous atrophy, crypt hyperplasia, and inflammation assessed by endoscopy with duodenal biopsy, with lesions intensity classified according to the Modified Marsh Classification system (grades 0–3) or qualitatively, plus clinical remission after a gluten-free diet (GFD). Because biopsy is an invasive method, video capsule endoscopy has been performed primarily in adults to avoid that (Caja, Maki, Kaukinen, & Lindfors, 2011). Therefore, due to the heterogeneous nature of CD and to avoid the invasive diagnosis, several serum-based antibody assays have been used for the diagnosis of CD before performing gastrointestinal endoscopies with small-bowel mucosal biopsies. Antigliadin antibodies (AGA) assays, antibodies of IgA and IgG classes found in sera of celiac patients, were the first used, but they are not recommended anymore due to their poor sensibilities and specificities. These antibodies mainly target gliadin-derived peptides. Antibodies against Endomysium (Ema, IgA and IgG) (gold standard), tissue transglutaminase (tTG, IgA and IgG), deamidated gliadin peptide (DGP, IgA and IgG) and IgA-R1 type reticulin (ARA) are the most frequently used and their sensitivities and specificities ranging from 61-98% and 90-100%, respectively, where their combination increases test performance (Caja et al., 2011; Schyum & Rumessen, 2013).

On the other hand, no biomarkers have been established for the accurate diagnosis and monitoring of other GRD to date, even though many studies have been carried out for this purpose (Catassi et al., 2015).

Dermatitis herpetiformis (DH) is a specific cutaneous manifestation of CD characterized by insoluble deposits of immunoglobulin A (IgA) in the papillary dermis. While CD prevalence has increased in the last decades, DH seems to show an opposite trend (Salmi, 2019), which justifies 66.7% of eligible articles on DH has been published up 2005 and the rest between 2007
and 2019. DH is usually diagnosed in patients with clinical and/or histopathological indicative symptoms of CD by direct immunofluorescence of skin biopsies, the gold standard used for the identification of granular IgA deposits, together with positive serology. Anti-epidermal transglutaminase antibodies (anti-eTG), present in about 50% of adult CD patients, have shown to be a primary diagnostic serology, while anti-tissue transglutaminase autoantibody (anti-TG2-IgA or anti-tTG-IgA), anti-endomysial antibodies (anti-EmA-IgA) and others may be used to support the diagnosis. A gluten-free diet (GFD) is the treatment indicated for all DH patients (Nguyen & Kim, 2021).

Wheat allergy (WA) is a GRD rarely investigated, particularly in adults. Epidemiological data for that are sparse, with prevalence estimated at 0.3-0.5% in children and less in adults (Venter et al., 2016). This scenario may justify the low level of scientific production on WA amongst the eligible studies. WA is characterized by the activation of type 2 T helper cells, which can result in reactions mediated by immunoglobulin E (IgE) and non-IgE. Of more than 20 different wheat allergens described so far, glutenin and gliadin from gluten are the most important. Indeed, after identifying clinical and serological features associated with CD and WA in 2,965 patients, Spoerl et al. (2019) concluded that WA does not seem to be associated with CD since values of wheat-specific IgE and IgA-tTG were negatively correlated in some WA-suspected patients (Spoerl et al., 2019). Usually, IgE-mediated WA is diagnosed based on clinical history, skin tests, and levels of anti-wheat IgE, as performed in both eligible studies recently published by Scibilia et al. (2019) (Scibilia et al., 2019) and Spoerl et al. (2019) (Spoerl et al., 2019). However, even IgE has been used as a WA biomarker, it is not found in all WA patients and may be present in only 50% of them (Spoerl et al., 2019).

NCGS diagnosis includes the initial exclusion of CD, WA, and other diseases based on medical history and symptoms (irritable bowel syndrome, inflammatory bowel disease) followed by a gluten-free diet for at least six weeks and a gluten challenge (“blinded” test)(Barbaro, Cremon, Morselli-Labate, et al., 2020; Catassi et al., 2015).

Gluten ataxia (GA) is a rare immune-mediated neurological manifestation caused by gluten intake. The primary target in the nervous system is intestinal axis tissue transglutaminase 2 (tTG2), even though tissue transglutaminase 6 (tTG6) may also be a causative antigen (Hadjivassiliou et al., 2015; Osman et al., 2021). Overall, there is a great variation in the prevalence of serological evidence in ataxias, as for GA, which has a prevalence of 15% amongst all ataxias available. Anti-TG2 antibodies are found in up to 40% of GA patients, usually in those with enteropathy. Anti-TG6 antibodies seem to be a possible more specific marker of GA; recently, it has been found in 73% of patients with ataxia and positive antigliadin antibodies (AGA); besides, GFD adherence resulted in a reduction of TG6 and TG2 antibodies. On the other hand, anti-TG6 antibodies were presented in 32% of patients with negative tests for other serological markers of GRD, which makes diagnosis a challenge. Because of this great
variation in the prevalence of serological evidence or even its absence, the evaluation of the presence of cerebellar atrophy, usually present in up to 60% of GA patients, has been more frequently performed for GA diagnosis through magnetic resonance imaging (MRI) and spectroscopy (MRS) (Hadjivassiliou et al., 2015), that was the focus of the study carried out by Rawat et al. (2022) (Rawat et al., 2022). This eligible study showed cerebellar atrophy in all GA patients (aged from 40-60 years) with no neurological/psychiatric diseases and no medical or family history of ataxia and with weak serological evidence for GRD (66.6% positive IgA anti-AGA and 16.6% positive anti-TG6 and anti-tTG2). Most importantly, this study was the first to show cerebrum atrophy in GA patients, suggesting that altered sensers and movements in patients with GA may be associated to a reduction in brain volume. In addition, it was observed a significant correlation (r = 0.9, p = <0.05) between the concentration of important neurochemicals and the brain volumetric changes.

As shown, BGM alone are not sufficient to diagnose and monitor GRD and research on discovering of new markers keeps evolving. In the present study, serologic markers, mainly antibodies (n=317), were analyzed by 72.8% of the eligible studies, while genetic markers (17.2%) were analyzed less frequently although this approach is on the rise.

Recently studied promising markers of GRD include circulating or excreted GRD markers in plasma and feces/urine, as citrulline (additive predictive value marker for latent CD), alklyresorcinols (Hakobyan et al., 2021) and gluten immunogenic peptides (Skodje et al., 2022) (specific biomarkers of gluten consumption, used for evaluation of GFD transgressions), circulating free fatty acids (FFAs) (may configure particular lipid profile of CD metabolism in infants) (Auricchio et al., 2019), zonulin (potential biomarker for NCGS diagnosis) (Barbaro, Cremon, Morselli-Labate, et al., 2020), intestinal fatty acid binding protein (I-FABP) and CX3CL1 (Fractalkine) (promise biomarkers for CD) (Gandini et al., 2022). Search for microbiome biomarkers and microRNA signature in GRD patients were approaches from 15 and 6 eligible studies, respectively. Other eligible studies carried out a complementary analysis to better understand GRD including quantitative and phenotypic analysis of specific cells (n=24), as intraepithelial lymphocytes (Camarero et al., 2021), and gene expression of specific markers (n= 35)(Bragde, Jansson, Fredrikson, Grodzinsky, & Soderman, 2018). Imaging biomarkers have also been used to look for GRD-specific biomarkers, such as optical coherence tomography from retinal layers in CD patients (Vitiello et al., 2022).

In this study, almost a quarter (24%) of eligible studies analyzed celiac disease in children or adolescents. Indeed, studies on GRD diagnosis in children have been intensively carried out, mainly in the development of non-invasive methods and in the identification of BGM for early diagnosis of GRD because the diagnosis of these disorders in children has been a challenge (Ruemmele, 2018; Singh, Pramanik, Acharya, & Makharia, 2019).
Eligible studies were performed in patients with 53 different diseases, conditions or syndromes. Type 1 diabetes mellitus (DMT1) was the most studied disease, corresponding to 24.1% of the total, followed by Down syndrome (6.2%), Autism Spectrum Disorders (ASD) (4.8%), schizophrenia (4.1%), bone diseases (4.1%) and thyroid diseases (3.4%).

There are many studies in the literature that indicate the co-occurrence of celiac disease with many diseases, including DMT1. Both T1DM and CD are frequently co-existing autoimmune diseases and share a common genetic background related to the presence of human leukocyte antigen (HLA) genes DQ2 and DQ8, which confer a high risk for developing the coexistence of CD and T1DM (Prieto et al., 2021).

The prevalence rate of T1DM range from 1.6% to 16.4% worldwide and most cases of CD are diagnosed within 5 years of T1DM diagnosis (Pham-Short et al., 2015). In pediatric patients with T1DM, the prevalence of CD was mainly determined based on serological tests and intestinal biopsy and may vary (1.4%-24.5%), differing widely between populations. The association of both diseases in children results in several complications including gastrointestinal disorders, anemia, osteoporosis, neurological disorders and, mainly, growth disorders (Jalilian & Jalali, 2021).

Down syndrome is also frequently coexisting with CD and individuals with this chromosomal abnormality constitute groups at risk for CD, as they usually have gastrointestinal disorders and changes in their innate and adaptive immunity, which contribute to increased rates of infections, autoimmune diseases and hematological malignancies (Nowak-Oczkowska, Szaflarska-Poplawska, & Soroczynska-Wrzyszcz, 2013; Ravel, Mircher, Rebillat, Cieuta-Walti, & Megarbane, 2020).

Since GRD result in chronic inflammation and malabsorption, they can cause alterations in bone metabolism and bone mineral loss in children and adults and contribute to the development of diseases such as osteoporosis, osteopenia or osteomalacia (Di Stefano, Mengoli, Bergonzi, & Corazza, 2013; Phan & Guglielmi, 2016).

People with ASD usually present a high prevalence of neurological complications compared to neurotypicals. Therefore, since gluten can cause neurological disorders, studies on the effect of GFD on ASD signs have increased nowadays, even though there is insufficient evidence for the relationship between gluten intake and ASD symptoms (Troisi et al., 2020). A similar scenario occurs with schizophrenia, a serious mental disorder (Severance, Yolken, & Eaton, 2016).

Studies have shown that the prevalence of autoimmune thyroiditis is higher in patients with celiac disease and vice versa. Autoimmune thyroid diseases are characterized by lymphocytic infiltration of the thyroid parenchyma, the main ones being Hashimoto's thyroiditis and Graves' disease. Studies suggest that these diseases share a genetic background with CD and
therefore can coexist with it (Ashok, Patni, Fatima, Lamis, & Siddiqui, 2022; Lerner, Jeremias, & Matthias, 2017).

**Characteristics of publication outputs Citation Report results**

According to the WoS Citation report from eligible articles, BGM-GRD is a relatively new field of study as it was in 1984 that research on that began. Only in the early 90s, it started to rise significantly until now. Therefore, there has been a growing scientific interest in BGM-GRD as the number of publications has doubled in the last two decades and keeps growing to date (Figure 1).

**Figure 1** – Number of scientific articles on BGM-GRD published over years (bars), annual citation numbers (straight line), and annual average of citations (dotted line), from 1984 to 2021 (scientific production in 2022 (n= 18 articles) was not considered because it was still in progress at the time of analysis).

According to Figure 1, while the number of publications increased by 2.4 times from 2001 to 2021, the number and the average of citations per year have increased by 5.0 and 55 times, respectively, in the same period, which reflects the relevance of the research published in this area in the last 20 years. The most expressive production volume occurred in 2019, while the most significant drop happened in 2005 and ten years later.

This upward trend in the increase of scientific production on BGM-GRD has been related to the increasing prevalence of GRD worldwide, mainly CD and NCGS, along with diagnostic limitations of these disorders and the development of new diagnosis techniques to avoid the non-invasive ones (Singh, Pramanik, Acharya, & Makharia, 2019). Furthermore, the research in this area keeps growing because BGM are useful elements for predicting,
diagnosing, and monitoring the progression of GRD, mainly CD and NCGS. Thus, data from this study follow the rising publishing trends on biomarkers in recent decades (Choong & Tsafnat, 2012).

Overall, the annual average of citations continued to grow from 1984 to nowadays, showing the relevance of this field to GRD research. Figure 1 also highlights an increase in the quality of the scientific contribution in this approach, mainly involving NCGS and CD biomarkers. The number of citations showed an upward trend up to 2009 with some fluctuations and showed steady growth from 2009 to 2014 and from 2016 until now. 2005 was the year of the lowest citations in this field of research.

This trend is particularly related to the difficulty of accurately diagnosing GRD, mainly NCGS. Although CD pathology is considerably well understood, it is estimated that up to 95% of CD remains undiagnosed worldwide due to its diverse appearance (Pratesi et al., 2003). Likewise, NCGD diagnosis is difficult and is based on the exclusion of DC and WA and its prevalence (0.5-13.0%) may be higher than CD (0.5-1.0%) (Cárdenas-Torres, Cabrera-Chávez, Figueroa-Salcido, & Ontiveros, 2021; Molina-Infante, Santolaria, Sanders, & Fernández-Bañares, 2015). Therefore, in the last decade, many studies have been performed to identify biomarkers that can help to establish a clear process for the definitive diagnosis of individuals with suspected NCGS. However, placebo-controlled gluten challenges must be carried out for NCGS diagnosis due to the lack of reproducible and sensitive biomarkers for NCGS diagnosis so far (Cárdenas-Torres et al., 2021).

According to the WoS citation report, 7,774 articles (7,471 without self-citation) cited the 452 eligible articles (1945 to October 9th, 2022). The H-index (56), the average citation per article (27.77), and the sum of times cited (12,552) from these articles were relatively high as well as this value without self-citation (11,908). According to Hirsch (Hirsch, 2005), H-index reflects the productivity of authors based on their publication and citation records and, overall, after 20 years of research; a H-index of 20 reflects good production while 40 and 60 indicate excellent and exceptional productions, respectively. Although it is difficult to provide global references to the H-index value for a set of articles, based on this individual categorization and considering an average value, the scientific production on BGM-GRD may be considered excellent or exceptional, reflecting its relevance (Pasterkamp, Rotmans, & De Kleijn, 2007).

Most prominent authors and publications

The most cited article (Table 1), published in 2013 by Biesiekierski and collaborators, received 529 citations and the highest annual average citation (52.2), which corresponded nearly to twice the number of citations per article in that year. Ten most cited articles on BGM-GRD, published between 1984 and 2022, are shown in Table 1. Most of them (n=8) evaluated
biological and/or genetic markers from CD patients, three studies assessed NCGS, while one evaluated anti-EmA-IgA biomarkers in DH patients.

Although the number of citations is linked to several factors (Tahamtan, Afshar, & Ahamdzadeh, 2016), it has been used as an important metric of scientific research quality. However, its performance is related to publication date once citations tend to increase over time as older papers had more time to accumulate citations (Aksnes, Langfeldt, & Wouters, 2019). The time publication effect can be seen when comparing the annual average of citations from...
studies published by Ciacci et al. (2002) and Sellitto et al. (2012), which received a similar number of citations but a different annual average of citations (Table 1). From this point of view, the research from Biesiekierski et al. (2013) presents the most scientific relevance (highest annual average of citations) on BGM-GRD approach (focus on NCGS), followed by studies from Carroccio et al. (2012), Sapone et al. (2011) and De Lourdes et al. (2017). In general, as presented in Table 1, the greatest scientific contribution on BGM-GRD, mainly NCGS and CD, has occurred in the last eleven years.

Overall, in the first two decades (up to 2004), the most cited studies have focused on diagnosis markers of CD, by mainly antibodies and HLA genes (Table 1). Indeed, the CD serologic biomarkers anti-gliadin and anti-endomysial were discovered in 1983, while anti-tissue transglutaminase and anti-deamidated gliadin peptide were discovered in 1997 and 2001, respectively (Lerner, Ramesh, & Matthias, 2019). Along with HLA-DQ typing, these antibodies have been used as CD biomarkers (Brusca, 2015).

Later (2004-2017), the scientific community has sought to better understand GRD, mainly NCGS and CD, in the same degree of interest but from different perspectives (Table 1).

For NCGS, studies are mainly focusing on discovery of accurate biomarkers that could be useful to its diagnosis and its differentiation from other GRD, mainly CD, since its pathogenesis is not completely understood. Besides, research on the discovery of components rather than gluten that could trigger this disease and their effect on potential biomarkers for NCGS have received more attention nowadays, evidenced by the greatest number of citations of article published by Biesiekierski and collaborators (2013), from an Australian research group (Table 1). Studies have been shown that grain components including fermentable oligo-, di-, monosaccharides, and polyols (FODMAP) as well as amylase/ trypsin inhibitors (ATIs) found in all gluten-containing cereals may also trigger NCGS. Thus, the effectiveness of different dietary interventions including gluten-free and low-FODMAP diets on GRD patients’ metabolism has been recently evaluated since the effects on the immune system remain unclear (Cárdenas-Torres et al., 2021).

Mapping research areas, Journals and Publishers

Among the 45 research areas from the Web of Science, most scientific knowledge about BGM-GRD has been concentrated in Gastroenterology Hepatology (40.9%), followed by Pediatrics (13.5%), Nutrition Dietetics (9.7%), General Internal Medicine (9.3%) and Immunology (9.1%) areas; 82.5% of published articles were included in these research areas. The other research areas each covered less than 4.2% of the scientific production (n=452 articles). Recently, similar ranking of research fields was demonstrated for general studies on celiac disease (Demir & Comba, 2020).
Knowledge on BGM-GRD was widely distributed in 221 scientific journals worldwide. The top nine journals (that covered at least 2.0% of knowledge about BGM-GRD) covered a total of 28.4% of the publications between 1984 and 2022: American Journal of Gastroenterology (3.1%), Digestive Diseases and Sciences (3.1%), European Journal of Gastroenterology Hepatology (2.9%), Gut (2.9%), Journal of Clinical Gastroenterology (2.9%), Scandinavian Journal of Gastroenterology (2.7%), Plos One (2.4%) and Digestive and Liver Disease (2.0%). Journal of Pediatric Gastroenterology and Nutrition (JPGN) was the most prominent scientific journal in this area and covered 6.4% of the world scientific production on BGM-GRD. Indeed, this journal has ranked within the top four positions of scientific production on CD studies in the last decade (Clarance & Raja, 2021; Demir & Comba, 2020).

Early diagnosis of CD in children is extremely important to avoid long-term complications and the onset of associated diseases, including malnutrition, delayed puberty, growth problems, and weight loss. In general, it is usually difficult because just a minority of celiac patients present classical symptoms and older children have either minimal or atypical symptoms (Sahin, 2021). This scenario may be constantly stimulating the publication of scientific articles in the JPGN.

More than half (52.8%) of research on BGM-GRD were equally published by Lippincott Williams & Wilkins (n=63), Elsevier (n=62), Wiley (n=57) and Springer Nature (n=57), from a total of 90 publishers. According to Nishikawa-Pacher (2022), Springer, Taylor and Francis, Elsevier and Wiley are the largest scientific publishers, which comprised 3,763, 2,912, 2,674, and 1,691 journals, respectively, in 2022 (Nishikawa-Pacher, 2022). Lippincott Williams & Wilkins ranked in the 16th position, with 375 Journals (Nishikawa-Pacher, 2022).

Countries cooperation

Studies about BGM-GRD have been published by researchers from 59 different countries, most of them from Italy (21.2% of the total scientific production), The United States of America (15.3%), Spain (10.8%), Sweden (8.9%), England (6.0%) and Finland (5.8%), that accounted for 68.0% of word production.

Overall, according to WoS data set, scientific research in these countries is driven by the support of government funding agencies. Overall, 15.7% of research on BGM-GRD was funded by USA government agencies including the United States Department of Health and Human Services (HHS), the National Institutes of Health (NIH), and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). In addition, 5.0% of the research received financial resources from European Commission, a politically independent institution that represents the interests from European Union. Swedish Research Council, a government agency
from Sweden, funded 2.2% of the eligible studies. The rest of the 378 agencies has supported less than 2% of all scientific research.

Even though more than half of the research is carried out in the USA and in specific institutions from Europe, some countries with a significantly low volume of scientific production on BGM-GRD have stood out in the quality of research with publications of great relevance; they include Australia (2.0%), Germany (3.3%), Israel (2.4%) and Poland (2.9%). Studies from these countries were among the ten most cited articles (Table I).

The evolution of the network formed among the countries (nodes) throughout the last decades showed the most important groupings among countries that conduct research activities on BGM-GRD, among the 23 clusters identified (Figure 2). The clusters present homogeneity and robustness, as indicated by the high weighted mean silhouette ($S=0.862$), which shows a great cohesion among clusters. Modularity $Q$ of 0.4635 means that the network is reasonably divided into clusters coupled in a medium intensity (Chen, 2006).

**Figure 2** – Cluster view of evolving networks (lines) among countries (nodes) that conduct research activities on BGM-GRD (1984 -2022). Red filled node represents the citation burst. Nodes with purple halos present high betweenness centrality. Bigger nodes indicate higher levels of published articles.

According to Figure 2, USA formed the most consolidated and oldest cluster (1984-2022) on BGM-GRD studies and continues to expand its network with other countries around the globe today. The US appears as the core of the network of the largest country cluster (cluster #0) on BGM-GRD, collaborating with 16 other countries, mainly England and Sweden. In addition, the US showed the strongest betweenness centrality (thickest purple halo), connecting
large research groups on BGM-GRD, mainly from Europe. This country presented the unique citation burst (strength of 4.54) from 2013 to 2017, represented by the red-filled node, which attracted an extraordinary degree of attention from the scientific community in this field of study in this period. After Italy, the US was the second most cited cluster member.

Likewise, Italy, the country with the highest number of citations and published articles as well as high betweenness centrality, was the leading country in the second largest cluster (#1), in cooperation with French, Iran, and eight other countries.

Even though with less productivity, Netherlands and Germany, countries with high betweenness centrality, took part in the third largest cluster (#2) composed by eleven members, including Estonia, Canada, Argentina, and other countries. This grouping has timidly expanded collaborations over time.

Spain, the third most cited country with high betweenness centrality, was the core country from the fourth cluster, in a network with Denmark, Ireland, Cuba, and others six members.

This scenario reflects the greater influence of these countries on BGM-GRD research, as can be seen by their highest level of scientific production (bigger nodes from Figure 2) and agrees with previous scientometric studies that demonstrated the leadership of the US and Italy in scientific research on different aspects of CD, which was probably motivated by the higher availability of financial resources in these countries (Bansal, Gupta, & Bansal, 2017; Clarance & Raja, 2021).

Institution collaboration network

Almost half (46.7%) of the world’s scientific production on BGM-GRD was concentrated in 18 institutions among the 817 institutions around the world that have been engaged in this research field. From that, five are in USA and the rest in Europe (Finland, Sweden, Italy, French, Spain, and England).

Tampere University, Tampere University Hospital and University of Helsinki (Finland), Linkoping University (Sweden) and University of Naples Federico II (Italy) have been the most productive institutions in this field of study so far; these institutions together covered 17.6% of world production on BGM-GRD. Overall, these institutions have ranked among the top eight positions as the most productive in studies involving different aspects of CD (Clarance & Raja, 2021; Demir & Comba, 2020). The others institutions covered less than 3.0% of world's scientific production.

Psychiatry (#0), Immunology (#1), Pathology (#2), Pediatrics (#3), Genetics and Heredity (#4) were the fields of studies (cluster labels) from the largest clusters of institutions that conduct research on BGM-GRD.
Psychiatry (cluster #0) was the field with the largest cluster of institutions, from a total of 192 clusters identified. Because ingestion of gluten and other components have been related to several neurological and psychiatric disorders or syndromes including ataxia, schizophrenia, peripheral neuropathy, autism spectrum disorders, depression, anxiety, and hallucinations (so-called gluten psychosis) (Catassi, 2015; Catassi et al., 2015), studies on the effects of gluten-free and low-FODMAP diets on symptoms remission has been increasing.

In addition, cluster analysis identified one citation burst in Tampere University, Finland, (Strength of 3.88), from 2008 to 2014, a period in which its scientific production in this field of study attracted extraordinary attention from the scientific community. This institution has stood out in research on GRD with focus on Immunology (cluster #1). Timeline view of institutions’ clusters showed that the University of Tampere and the University Hospital of Tampere were pioneers in studies on BGM-GRD and these institutions continue to expand their collaboration network to date in several areas, mainly Immunology, Psychiatry and Pediatrics.

Trends and knowledge gaps

Keywords analysis of studies on BGM-GRD allowed finding the most researched topics over time as well as their evolutionary trends in the related research field. Thus, hotspots in this area of study could be identified by the frequency of the keywords and the similarity between them (network), obtained from co-word analysis using CiteSpace (Chen, 2006).

The keywords most frequently cited (bigger diamonds) in publications (1984-2022) as well as the clusters of keywords are shown in Figure 3. These groups of keywords presented good cohesion (S= 0.7307) and reasonable splitting quality of network-coupled clusters (Q=0.3805). As expected, keywords were grouped mainly in topics from Immunology, Gastroenterology, Psychiatry, and Pediatrics, the largest clusters (Figure 3).
**Figure 3** – Most frequent keywords (nodes) cited on BGM-GRD from 1984 to 2022. Keyword clusters were labeled with subject categories. Nodes (diamonds) of the largest cluster are colored in red. Higher numbers (#) indicate smaller clusters. Concentric diamonds represent how many citations were made to the keywords in corresponding years. Red filled diamonds represent citation bursts. Nodes with purple halos present high betweenness centrality.

![Image of Figure 3](CLIUM.ORG)

Source: Authors (2023)

Top 21 keywords of studies on BGM-GRD with the strongest citation bursts (red-filled diamonds in Figure 3) highlighted the change of trends on this field of study (Figure 4).

**Figure 4** – Timeline from the most frequent keywords cited on BGM-GRD from 1984 to 2022.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Year</th>
<th>Strength</th>
<th>Begin</th>
<th>End</th>
<th>1984 - 2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>childhood</td>
<td>1993</td>
<td>3.29</td>
<td>1993</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>antiretinois antibody</td>
<td>1993</td>
<td>3.01</td>
<td>1993</td>
<td>2003</td>
<td></td>
</tr>
<tr>
<td>coeliac disease</td>
<td>1996</td>
<td>9.3</td>
<td>1996</td>
<td>2005</td>
<td></td>
</tr>
<tr>
<td>serological marker</td>
<td>1997</td>
<td>6.42</td>
<td>1997</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>tissue transglutaminase</td>
<td>2000</td>
<td>7.91</td>
<td>2000</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>multicenter</td>
<td>2004</td>
<td>3.85</td>
<td>2004</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>population</td>
<td>2000</td>
<td>4.29</td>
<td>2000</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>symptoms</td>
<td>2012</td>
<td>3.8</td>
<td>2012</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>management</td>
<td>2013</td>
<td>7.23</td>
<td>2013</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>guideline</td>
<td>2014</td>
<td>7.5</td>
<td>2014</td>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>biomarker</td>
<td>2011</td>
<td>3.79</td>
<td>2014</td>
<td>2019</td>
<td></td>
</tr>
<tr>
<td>risk</td>
<td>2008</td>
<td>9.23</td>
<td>2015</td>
<td>2020</td>
<td></td>
</tr>
<tr>
<td>activation</td>
<td>2012</td>
<td>4.69</td>
<td>2017</td>
<td>2022</td>
<td></td>
</tr>
<tr>
<td>t cell</td>
<td>2012</td>
<td>6.33</td>
<td>2018</td>
<td>2022</td>
<td></td>
</tr>
<tr>
<td>expression</td>
<td>2005</td>
<td>4.62</td>
<td>2018</td>
<td>2022</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors (2023)
Along with *dermatitis herpetiformis* (1991), *anti-gliadin antibody* presented the first citation burst on BGM-GRD. Antigliadin antibodies (AGA) mainly target gliadin-derived peptides, the main proteins of gluten, and are found in sera of CD patients although may be found in patients with other gastrointestinal diseases. *Antiendomysial antibodies* (EmA; IgA class autoantibodies directed against endomysium) and *anti-tissue transglutaminase 2* (anti-tTG2; autoantibodies of class IgA and IgG produced by tTG2-specific B cells) are AGA of the IgA and IgG classes. AGA assays were the first used for CD diagnosis, but they are not usually used today because of their low diagnostic accuracy and the introduction of newer, more accurate serological tests (Caja et al., 2011; Schyum & Rumessen, 2013). This may explain why the keyword *anti-gliadin antibody* showed a citation explosion in the first decades of studies on MBG-DRG, which lasted 13 years, but is currently not highlighted. During this period, the use of *antiendomysial antibodies* (1991; 12 years) for diagnosing CD was also highly relevant, which was reflected in its long duration of citation explosion (12 years). This trend was previously discussed and is present among the ten most cited publications in the BGM-GRD (Table 1).

The keywords cited in the first two decades (1991-2004) presented citation bursts varying from 1991 to 2012. CD was the GRD that received most attention in this period, with the strongest citation burst (strength of 9.49) (Figure 4). The keywords *antigliadin antibody* (1991), *antiendomysial antibody* (1991), *celiac disease* (1996), *gliadin* (1993), *serological marker, prevalence* (1996), *population* (2000), and *tissue transglutaminase* (2000) indicated greater attention to antibodies as serologic biomarkers for the diagnosis of CD, with a focus on determining the prevalence of CD in specific populations. The keyword *childhood* (burst from 1993 to 2004) is likely related to the importance of early diagnosis of CD in childhood to minimizing its long-term health impacts, along with the challenges related to CD diagnosis in pediatric patients. *Dermatitis herpetiformis* (1991) also present more attention in the first years of studies on GRD biomarkers, with citation burst lasting seven years (1991 to 1998). Due to the reduction in the DH prevalence (Salmi, 2019), studies that assess this disorder currently show a downward trend. These issues were addressed by older clusters (Figure 3) and were considered by two of the top ten most cited articles published in the first decade, performed by Chorzelski et al. (1984) and Maki et al. (1991) (Table 1).

From 2005 to now, there was an expressive increase in the interest of scientific community on estimating the risk (2008) of GRD in different populations as well as establishing *guidelines* (2014) for accurate diagnosis and *management* (2013) of them, mainly from NCGS, based on nutritional strategies, especially gluten-free diet. Articles published by Carrocio et al. (2012), Confino-Cohen et al. (2012), Ciacci et al. (2002) and de Loudes et al. (2017)(Table 1), addressed these topics.
The explosion of citations of the keyword biomarker occurs recently, in 2011, and has continued until now. Thus, the strongest citation bursts from keywords expression and symptom may indicate that much attention has been made to studies evaluating the effect of dietary interventions on gene expression of GRD markers as well as in the reduction/remission of GRD symptoms. This approach can be seen in the study of Sapone and colleagues, a collaboration group from Italy and USA, and one of the 10 most cited articles published in the last decade (Table 1). Since gluten can cause neurologic disorders, overall, gluten-free diet (GFD) is the most efficient treatment for CD and, likewise, related to improvements in neuropathologies including ASD, schizophrenia, anxiety, depression, and in gut microbiome health. Modification of microbiome may be related to the disease’s remission since abundance and diversity of resident gut microbiota is modified with gluten withdrawal (Severance et al., 2016).

The burst of citations for the keywords T cells and activation started in 2017 and remain to date. Celiac disease is a human T cell-mediated autoimmune-like disorder caused by exposure to dietary gluten in genetically predisposed individuals. CD4+ T cells mistakenly recognize gluten peptides derived from incomplete digestion as harmful pathogens and become activated, triggering the immune response. Once CD4+ T cells become activated, they can produce different subsets of helper T cells. In DC, Th1, the most common type of helper T cell produced, produces cytokines that recruit other immune cells, resulting in tissue damage (Jabri & Sollid, 2017). Thus, the development of strategies for reducing T cells activation may reduce or inhibit immune response and have been received much attention nowadays.

In the last 39 years, investigations of new BMG-GRD are in remarkable expansion and celiac disease was the most studied disorder. The main interests of recent studies were in the effect of environmental components on metabolic pathways from CD patients or individuals genetically predisposed to CD, determination of the prevalence and diagnosis of CD in children as well as risk assessment and validation of biomarkers (proteome, transcriptome, micro-RNA, metabolome, and microbiome) for CD diagnosis thought noninvasive methods, as those present in urine and saliva. Studies have also focused on the influence of diet (mainly GFD) on gut microbiome and its impact on immune response tolerance in genetically susceptible individuals to CD.

Investigations of new biomarkers and genetic markers of gluten-related disorders are in remarkable expansion, with a focus on diagnosis, management, treatment monitoring, as well as risk assessment and prevalence of these diseases, especially in children.
REFERÊNCIAS


Pasterkamp, G., Rotmans, J. I., & De Kleijn, D. V. P. (2007). Scientometrics Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles., 70, 153–165.


