Are there new biomarkers of the gastroduodenal microbiota useful in the diagnosis of coeliac disease in children? A pilot study

D. Salamon1, K. Kowalska-Duplaga2, A. Krawczyk1, M. Duplaga3, A. Gurgul4 and T. Gosiewski5*

1Department of Molecular Medical Microbiology, Division of Microbiology, Faculty of Medicine, Jagiellonian University Medical College, ul. Czysta 18, 31-121 Krakow, Poland; 2Department of Pediatrics, Gastroenterology and Nutrition, Faculty of Medicine, Jagiellonian University Medical College, ul. Wielicka 265, 30-663 Krakow, Poland; 3Department of Health Promotion and e-Health, Institute of Public Health, Faculty of Health Sciences, Jagiellonian University Medical College, ul. Skawińska 8, 31-066 Krakow, Poland; 4Center for Experimental and Innovative Medicine, University of Agriculture in Krakow, ul. Rędzina 1c, 30-248, Krakow, Poland; 5Microbiome Research Laboratory, Department of Molecular Medical Microbiology, Division of Microbiology, Faculty of Medicine, Jagiellonian University Medical College, ul. Czysta 18, 31-121 Krakow, Poland; *tomasz.gosiewski@uj.edu.pl

Received 18 October 2023 | Accepted 27 March 2024 | Published online 23 April 2024 | Published in issue 30 April 2024

Abstract

The changing of microbiome could precede the development of coeliac disease (CeD). We compared the bacterial profile of microbiota of tissues collected simultaneously from the stomach and duodenum in newly diagnosed patients with CeD. Biopsies were collected from 60 children and adolescents aged 2-18 years: (1) 40 patients with CeD; (2) 20 children as control group. The evaluation of the bacterial microbiota was carried out by sequencing the V3-V4 regions of the 16S rRNA subunit, using next-generation sequencing (NGS). The composition of bacterial microbiota was correlated with clinical and blood parameters. The beta diversity analysis revealed a significant dissimilarity in the gastric samples between the CeD and control group (Bray-Curtis index, \( P = 0.008 \), and weighted UniFrac distance, \( P = 0.024 \)). At L2 (phylum level), *Campylobacterota* was only present in the stomach of the CeD group. A comparison of the abundance of bacteria between the stomach and duodenum showed significant differences in 10 OTUs (operational taxonomic units) in the control and 9 OTUs in the CeD group at L6 (genus) and in 8 OTUs and in 6 OTUs, respectively, at L7 (species). A significant correlation was observed between the genus *Novosphingobium* in stomach of CeD group and possession of the DQ2.5 and DQ8 allele, and in the duodenum – between the DQ 8 allele and the species *Blautia wexlerae*. Significant differences in selected, little-known genera of bacteria suggest their potential role as new biomarkers in the development of CeD. To fully understand the mechanism of CeD development in genetically predisposed individuals, it is necessary to take into account not only the abundance of a given genus or species of bacteria, but also the anatomical location of its occurrence.

Keywords
gastric and duodenal biopsies – DQ2.5 allele – DQ8 allele – next-generation sequencing
1 Introduction

Although genetic predisposition and exposure to gluten are considered essential triggers for the development of coeliac disease (CeD), other factors leading to a loss of immune tolerance to gluten in predisposed individuals are still under investigation. The role of the digestive tract microbiota in the development and progression of this disease is currently a very popular concept and the focus of numerous studies (Catassi et al., 2022; Leonard and Fasano, 2016; Palmieri et al., 2022; Valitutti et al., 2019; Verdu and Schuppan, 2021). Recent research findings suggest that changes in the microbiome could precede the development of CeD or accompany the development of inflammation (Catassi et al., 2022; Valitutti et al., 2019) Increasingly, it is postulated in the literature that such research should be developed in order to improve and personalise the methods of prevention and treatment, as well as to monitor the course of the disease, with the use of potential new diagnostic bacterial biomarkers (Fasano et al., 2023; Verdu and Schuppan, 2021). Using samples from various body environments (saliva, pharynx, duodenum, and stool), Arcila-Galvis et al. (1922) analysed 9 selected studies to prepare a comprehensive map of microbial biomarkers along the digestive tract for CeD patients. The composition of the microbiota was assessed on the basis of next-generation sequencing methods (NGS), primarily focusing on the V3-V4 region of 16S rRNA (Arcila-Galvis et al., 2022). While other studies have analysed the microbiota profile of the oral cavity, duodenum or large intestine (De Angelis et al., 2016; Leonard et al., 2021; Valitutti et al., 2019) none have provided results regarding the composition of the gastric microbiota in patients with CeD.

The aim of our study was to comparatively evaluate the composition of the bacterial microbiota of tissues collected simultaneously from the stomach (which to our knowledge has never been explored before) and duodenum in children with newly diagnosed CeD and the control group. We also attempted to analyse the correlation between the obtained bacterial profiles, biochemical and immunological results, clinical presentation, histopathological evaluation of duodenal biopsies, and HLA II DQ2/DQ8 allele typing in blood samples.

2 Materials and methods

Patients

Between 2019 and 2022, the study recruited children and adolescents aged 2 to 18 years from the Department of Paediatrics, Gastroenterology and Nutrition, University Children's Hospital in Krakow, Poland. Participants who underwent gastroscopy were divided into two groups. The first group consisted of 40 patients with newly diagnosed CeD according to the diagnostic criteria of the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (ESPGHAN) published in 2012 (Husby et al., 2012). The second group (control) comprised 20 children in whom CeD and other gastroenterological diseases were excluded.

The exclusion criteria for the study included individuals below 2 or over 18 years of age, recent treatment with antibiotics and/or probiotics within 30 days before tissue sample collection (Elvers et al., 2020; Sroka-Oleksiak et al., 2020; Vassour et al., 2016), confirmed acute or chronic infections and inflammatory diseases in the gastrointestinal tract (especially Heli- cobacter pylori infection and inflammatory bowel disease), obesity, thyroid disorders, diabetes, active neoplastic diseases (especially gastrointestinal), congenital and/or acquired immune deficiencies (especially selective IgA deficiency), and lack of consent to participate in the study.

All procedures performed in this study involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments and with the ethical standards of the Jagiellonian University Bioethics Committee in Krakow, Poland (decision No. 1072.6120.82.2018). Informed consent was obtained from patients’ parents or legal guardians (for all patients under 18 years of age) and, additionally, from patients themselves if they were above 16 years old.

Samples

The children underwent routine laboratory blood testing, including fasting blood glucose, total cholesterol, low-density lipoprotein (LDL), high-density lipoprotein (HDL), triglycerides (TGs), iron (Fe) levels and alanine aminotransferase (ALT). Tissue transglutaminase IgA antibodies (anti-tTG) were assessed in all enrolled children using an enzyme-linked immunosorbent assay (ELISA), with levels greater than 20 RU/ml considered abnormal. Moreover, coeliac HLA DQ2/DQ8 allele typing was performed (CeliacStrip; Operon; Cuarte de Huerva, Spain). All laboratory tests were carried out at the laboratory of the University Children's Hospital in Krakow, in accordance with the principles of Good Laboratory Practice (GLP) and using tests certified by In Vitro Diagnostics/Food and Drug Administration (IVD/FDA) intended for routine medical diagnostics and included in the clinical protocol of specialised lab-
oratories at this hospital. Double biopsy samples were taken from stomach and duodenum during gastroscopy. One pair of biopsy specimens was evaluated histopathologically to exclude inflammation and villous atrophy in the control group and to assess the severity of lesions in patients with CeD, according to the Marsh-Oberhuber classification (Oberhuber, 2000). The other pair was immediately frozen at −80 °C but for no longer than 2 months and transported under deep-freeze conditions to the Department of Molecular Medical Microbiology, Division of Microbiology, Jagiellonian University Medical College in Krakow. Subsequently, library preparation and sequencing were performed for microbiome analysis.

Next generation sequencing

To obtain high-quality sequence reads from low biomass samples, such as the stomach and duodenum, we followed a protocol described by Sroka-Oleksiak et al. (2020), i.e. we used the Nested-PCR method, which requires two primer sets and two successive PCR reactions. This procedure enabled to increase the specificity and sensitivity of the isolates amplified for sequencing (Sroka-Oleksiak et al., 2020).

Microbiological data obtained at 6 taxonomic levels (phylum [L2], class [L3], order [L4], family [L5], genus [L6] and species [L7]) were compared between groups (stomach of CeD vs stomach of control group and duodenum of CeD vs duodenum of control group). We also conducted a comparison of samples within the CeD group (stomach vs duodenum) and within the control group (stomach vs duodenum). The data obtained at the L6 and L7 were correlated with the clinical and laboratory parameters of children with CeD (glycemia, total cholesterol, LDL, HDL, TGs, ALT, creatinine, Fe levels, tTG antibodies titre, coeliac HLA DQ2/DQ8 genes, histopathological results), as well as age and body mass index (BMI).

Bioinformatic and statistical analysis

Raw sequencing reads were included in the QIIME2 analysis pipeline (Bolyen et al., 2019). In this pipeline, initial read quality control, filtering, denoising and feature table generation were performed using DADA2 software (Callahan et al., 2016). Subsequently, a phylogenetic tree was generated from the masked alignment using MAFFT software (Katoh and Toh, 2010) and FastTree software (Price et al., 2009). Alpha and beta diversity analyses were conducted through the q2-diversity plugin with 2,000 subsampled reads based on the rarefaction curve. Kruskal-Wallis or PERMANOVA tests were used to compare alpha and beta diversity parameters among groups. Taxonomic classification of reads was done using a pre-trained Naive Bayes classifier (sklearn) and the q2-feature-classifier plugin. This classifier was trained on the Weighted Silva 138 99% OTUs full-length sequences database. Finally, differential abundance analysis was conducted using MicrobiomeAnalyst software (Dhariwal et al., 2017) using edgeR approach (Robinson et al., 2009) with the Benjamini-Hochberg (false discovery rate-FDR) correction. Taxa were considered significantly altered if they had an FDR < 0.05.

Statistical analysis was carried out with IBM SPSS Statistics 28 (Armonk, NY, USA). Since the variables reflecting biochemical parameters and BMI did not consistently follow the normal distribution, comparisons between study groups were carried out using the non-parametric Mann-Whitney U test. Differences in gender and the prevalence of HLA DQ2/DQ8 alleles between the coeliac disease and control groups were assessed with Fisher's exact test.

The dependence between the numbers of specific bacteria and quantitative clinical traits was evaluated using correlation analysis, in which traits deviating from the normal distribution were analysed using Spearman's rank correlation coefficient. A p-value of less than 0.05 was considered significant. The association between DQ gene alleles and selected microbiota was tested using Student's t-test with the JASP (https://jasp-stats.org/) statistical software.

3 Results

Characteristics of the study population

Clinical data concerning both the studied group of patients and the control group are presented in Table 1. The analysis revealed statistically significant differences between the CeD and control groups in terms of age, creatinine levels and the carriage of the HLA DQ 2.5 allele.

Based on histopathological examination of duodenal biopsies, it was found that, in the CeD group, 5 children met the diagnostic criteria for type 3a, 11 children for type 3b, 23 children for 3c and one patient for type 2 villous atrophy according to the Oberhuber-Marshall classification (Oberhuber, 2000). The mean number of intraepithelial lymphocytes per 100 enterocytes (IEL) was 56.85 (±23.91; ranging from 28 IEL to 100 IEL).
## Table 1 Clinical characteristics of the study groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeliac disease mean (SD)</th>
<th>Control group mean (SD)</th>
<th>( P )-value²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>8.7 (3.70)</td>
<td>12.6 (4.26)</td>
<td>0.002⁴</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>16.07 (2.56)</td>
<td>18.80 (4.54)</td>
<td>0.038</td>
</tr>
<tr>
<td>Glycemia</td>
<td>4.67 (1.11)</td>
<td>4.51 (0.44)</td>
<td>0.694</td>
</tr>
<tr>
<td>Total cholesterol (mmol/l)</td>
<td>3.85 (0.70)</td>
<td>3.88 (0.57)</td>
<td>0.919</td>
</tr>
<tr>
<td>HDL (mmol/l)</td>
<td>1.31 (0.35)</td>
<td>1.23 (0.24)</td>
<td>0.130</td>
</tr>
<tr>
<td>LDL (mmol/l)</td>
<td>2.26 (0.57)</td>
<td>2.35 (0.48)</td>
<td>0.712</td>
</tr>
<tr>
<td>TGs (mmol/l)</td>
<td>2.26 (0.57)</td>
<td>0.85 (0.26)</td>
<td>0.937</td>
</tr>
<tr>
<td>ALT (U/l)</td>
<td>27.89 (26.58)</td>
<td>23.62 (12.81)</td>
<td>0.392</td>
</tr>
<tr>
<td>Fe (μmol/l)</td>
<td>16.98 (7.26)</td>
<td>17.93 (6.09)</td>
<td>0.293</td>
</tr>
<tr>
<td>Creatinine (μmol/l)</td>
<td>36.92 (12.16)</td>
<td>37.54 (11.63)</td>
<td>0.002⁴</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category of variable</th>
<th>Coeliac disease n (%)</th>
<th>Control group n (%)</th>
<th>( P )-value³</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex</td>
<td>female</td>
<td>29 (72.5%)</td>
<td>15 (75.0%)</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>11 (27.5%)</td>
<td>5 (25.0%)</td>
<td></td>
</tr>
<tr>
<td>DQ2.2</td>
<td>positive</td>
<td>9 (22.5%)</td>
<td>3 (15.0%)</td>
<td>0.734</td>
</tr>
<tr>
<td></td>
<td>negative</td>
<td>31 (77.5%)</td>
<td>17 (85.0%)</td>
<td></td>
</tr>
<tr>
<td>DQ2.5</td>
<td>positive</td>
<td>33 (82.5%)</td>
<td>5 (25.0%)</td>
<td>&lt;0.001⁴</td>
</tr>
<tr>
<td></td>
<td>negative</td>
<td>7 (17.5%)</td>
<td>15 (75.0%)</td>
<td></td>
</tr>
<tr>
<td>DQ8</td>
<td>positive</td>
<td>7 (17.5%)</td>
<td>7 (35.0%)</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>negative</td>
<td>33 (82.5%)</td>
<td>13 (65.0%)</td>
<td></td>
</tr>
</tbody>
</table>

¹ ALT = alanine aminotransferase; BMI = body mass index; HDL = high-density lipoprotein cholesterol; LDL = low-density lipoprotein cholesterol; SD = standard deviation; TGs = triglycerides.

² \( P \)-value for Mann-Whitney U test

³ \( P \)-value for Fisher’s exact test

⁴ Statistically significant \( P \) value (less than 0.05)

---

**Metagenomic sequencing**

Sequencing was performed on 60 tissue samples as follows:

- Stomach samples yielded a total of 4,029,168 paired reads, 67,152.8 reads on average per sample. The number of reads per sample ranged from a minimum of 19,906 to a maximum of 150,821, with a median of 60,110.0.

- Duodenum samples generated 4,435,481 paired reads, 75,177.64068 reads on average per sample. The number of reads per sample ranged from a minimum of 17,277 to a maximum of 780,440, with a median of 58,465.0.

The DNA sequences obtained from gastric and duodenal biopsies corresponded to a total of, respectively, 62 and 63 OTUs at the genus level (L6). At the species level (L7), 50 and 46 OTUs were detected, respectively. The reads were clustered during preliminary filtering for 97% identity, which is equivalent to sequence similarity at a given taxonomic level. Therefore, due to the possibility of errors in data interpretation, because the percentage of classified reads was below 97% at the species level (L7), the microbiota composition at L7 and correlation analyses were conducted for selected clinical data and the numbers of only those OTUs which were clearly assigned to a specific species.

**Biodiversity**

The alpha diversity profiling showed no statistical significance in sample diversity between CeD patients and controls from either the stomach or duodenum. Similarly, when comparing the diversity between the stomach and duodenum within each group, the results showed no significant variations (Table 2).

The beta diversity analysis revealed no significant differences between the gastric and duodenal samples in both groups (\( P > 0.05 \)). However, significant dissimilarities were observed in the gastric samples when comparing the CeD group to the control (Bray-Curtis index, \( P = 0.008 \), and weighted UniFrac distance, \( P = 0.024 \)) and borderline significance in the duodenal samples (Jaccard index, \( P = 0.049 \); Table 2).
<table>
<thead>
<tr>
<th>Tissue</th>
<th>Alpha diversity metrics P-values (CeD vs control)</th>
<th>Beta diversity indices i-values (CeD vs control)</th>
<th>Alfa diversity metrics P-values (stomach vs duodenum)</th>
<th>Beta diversity indices P-values (stomach vs duodenum)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Shannon</td>
<td>Evenness</td>
<td>Faith’s phylogenetic diversity</td>
</tr>
<tr>
<td>Stomach</td>
<td>0.36</td>
<td>0.56</td>
<td>0.54</td>
<td>0.43</td>
</tr>
<tr>
<td>Duodenum</td>
<td>0.06</td>
<td>0.4</td>
<td>0.13</td>
<td>0.57</td>
</tr>
</tbody>
</table>

1 Statistically significant P value (less than 0.05).

**Abundance**

We conducted a systematic assessment of the bacterial profile of the tissue samples at the L2 taxonomic level to provide an overall picture. For a more detailed analysis of the differences in the composition of the gastric and duodenal microbiota, evaluations at the L6 and L7 taxonomic levels were carried out.

Due to the large number of taxa at L6 (62 in the stomach, 63 in the duodenum) and L7 (50 in the stomach, 46 in the duodenum) and the fact that 97% of reads were classified at the species level, the description in this article refers only to a few selected, statistically significant observations. Taxa with percentages lower than 1% and/or no significant differences between the compared groups were collectively categorised as ‘other’.

*Proteobacteria* were the dominant phylum in both study groups in the stomach and duodenum. The phyla *Firmicutes* and *Actinobacteriota* were also prominent (Table 3). We found no significant differences as to the relative percentages of most bacterial phyla between groups and anatomical locations, except for the phylum *Campylobacterota*, which was only present in the stomach of children with CeD.

At the L6 level, both in the stomach and duodenum, 11 OTUs were identified which corresponded to bacterial genera and exhibited statistically significant differences in relative percentages between the patient groups (Supplementary Table S1). A comparison of the abundance of individual bacterial genera between the stomach and duodenum showed significant differences in 10 OTUs in the control group and 9 OTUs in the group of children with CeD (Supplementary Table S1).

At the L7 taxonomic level, we identified 8 OTUs in the stomach and 9 OTUs in the duodenum, corresponding to bacterial species with statistically significant differences in relative percentages between the patient groups (Supplementary Table S2). A comparison of the abundance of individual bacterial species between the stomach and duodenum showed significant differences in 8 OTUs in the control group and 6 OTUs in the group of CeD patients (Supplementary Table S2).

**Correlation analysis**

In the stomach but also in the duodenum, positive and negative correlations with selected laboratory and clinical data were observed at L6 and L7 levels between some taxa and: BMI, total cholesterol, LDL, HDL, TGs, Fe levels, ALT and anti-tTG titre (Supplementary Tables S3 and S4). Additionally, at L6, bacteria (in biopsies from the stomach and duodenum) exhibited correlations with creatinine concentration and (in the duodenal samples) with IEL (Supplementary Table S3). At L7, a relationship was observed between selected taxa and creatinine level, specifically in the duodenum (Supplementary Table S4).
Table 3: Comparison of the relative percentage of bacteria at the phylum level (L2) between groups: CeD and control, in the stomach and in the duodenum

<table>
<thead>
<tr>
<th>L2 (phylum)</th>
<th>Stomach</th>
<th>Duodenum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CeD [%]</td>
<td>Control [%]</td>
</tr>
<tr>
<td>Actinobacteriota</td>
<td>10.63</td>
<td>8.47</td>
</tr>
<tr>
<td>Bacteroidota</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Campylobacterota</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td>0.44</td>
<td>0.62</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>24.25</td>
<td>32.24</td>
</tr>
<tr>
<td>Fusobacteriota</td>
<td>0.47</td>
<td>0.53</td>
</tr>
<tr>
<td>Not assigned</td>
<td>0.76</td>
<td>1.90</td>
</tr>
<tr>
<td>Patescibacteria</td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>62.60</td>
<td>55.75</td>
</tr>
</tbody>
</table>

*Statistically significant P_{FDR} value (less than 0.05).

Analysis of HLA DQ2/DQ8 allele occurrence in relation to gastric bacterial profiles in the CeD group showed a significant correlation only with bacteria belonging to the genus Novosphingobium: possession of the DQ 2.5 allele resulted in lower abundance, while possession of the DQ 8 allele led to a greater abundance of these bacteria (Figure 1). In the duodenum, a positive association was observed between the presence of the DQ 8 allele and the abundance of bacteria belonging to the genus Blautia, but a statistically significant relationship was observed only for the species Blautia wexlerae (Figure 1). No statistically significant correlations were observed for the remaining parameters.

4 Discussion

Data from many studies indicate that particular changes in the composition of the gut microbiota may precede the onset of CeD in genetically susceptible individuals. The alterations in the microbiota profile and resulting changes in specific metabolic pathways could potentially lead to a shift from gluten tolerance to an immune response to gluten (Leonard and Fasano, 2016). In our study, we focused on assessing the composition of the tissue microbiota, not only in the duodenal but also in the gastric mucus of young patients with newly diagnosed CeD which, to the best of our knowledge, has never been done before. This study was based on the analysis of the genetic material of microorganisms using NGS.

Biodiversity analysis did not show statistically significant differences concerning alpha diversity among the studied samples (Table 2). Similar results were obtained in three other studies (Federica et al., 2022; Leonard et al., 2021; Palmieri et al., 2022). In contrast, a comparison of beta diversity between the two studied groups seems to indicate a significant difference (but only in single metrics) in the number and abundance of bacteria in gastric samples (Bray-Curtis index) and duodenal samples (Jaccard index) between CeD patients and the control group. The phylogenetic relationship (weighted UniFrac distance) between the microorganisms found in the gastric samples of the two study groups was also statistically significant (Table 2), indicating a low phylogenetic affinity between these groups.

Comparison of bacterial profiles at the highest taxonomical level (phylum – L2) revealed the presence of the Campylobacterota phylum only in patients with CeD (in the stomach) (Table 3). Campylobacterota, previously known as Epsilonproteobacteria, is now classified as a phylum consisting of a large group of mostly Gram-negative, spiral-shaped motile bacteria, including Sulfospirillum spp., Helicobacter spp., and Campylobacter spp., among others. H. pylori is the most well-known and the most prevalent Helicobacter species in the human stomach (Matos et al., 2021). Although H. pylori remains one of the bacterial species that is most likely to trigger autoimmunity, various studies have either failed to define its causal role, as Yue et al. (2022) indicated in their meta-analysis, or did not show a significant association between H. pylori and CeD, as shown by the cross-sectional study by Basyigit et al. (2017). Our study involved children in whom the presence of H. pylori colonisation was excluded, and the sequencing results did not indicate the presence of its genetic material in the samples. In contrast, at L6, we found a significant relative percentage of bacteria belonging to the...
Paediatric microbiota in coeliac disease

Figure 1 Significant relationship between the occurrence of specific HLA DQ alleles and *Novosphingobium* abundance in the stomach.

![Graph](image.png)

**a** Relationship between the occurrence of specific HLA DQ alleles and *Novosphingobium* abundance in the stomach

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>df</th>
<th>p</th>
<th>Test</th>
<th>Statistic</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>0.814</td>
<td>38.000</td>
<td>0.421</td>
<td>Student</td>
<td>-0.799</td>
<td>38.000</td>
<td>0.429</td>
</tr>
<tr>
<td>Welch</td>
<td>0.925</td>
<td>10.146</td>
<td>0.377</td>
<td>Welch</td>
<td>-0.917</td>
<td>10.289</td>
<td>0.380</td>
</tr>
<tr>
<td>Mann-Whitney</td>
<td>154.000</td>
<td></td>
<td>0.020</td>
<td>Mann-Whitney</td>
<td>78.000</td>
<td></td>
<td>0.034</td>
</tr>
</tbody>
</table>

![Graph](image.png)

**b** Relationship between the occurrence of specific HLA DQ alleles and *Blautia* abundance in the duodenum

1 – selected DQ allele absent, 2 – selected DQ allele present,
A – statistically significant *p* value (less than 0.05)

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>df</th>
<th>p</th>
<th>Test</th>
<th>Statistic</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>-2.252</td>
<td>38.000</td>
<td>0.030</td>
<td>Student</td>
<td>-2.488</td>
<td>38.000</td>
<td>0.017</td>
</tr>
<tr>
<td>Welch</td>
<td>-1.026</td>
<td>6.042</td>
<td>0.344</td>
<td>Welch</td>
<td>-1.101</td>
<td>6.012</td>
<td>0.313</td>
</tr>
<tr>
<td>Mann-Whitney</td>
<td>80.000</td>
<td></td>
<td>0.075</td>
<td>Mann-Whitney</td>
<td>74.000</td>
<td></td>
<td>0.014</td>
</tr>
</tbody>
</table>

**Figure 1** Significant relationship between the occurrence of specific HLA DQ alleles in children with coeliac disease and bacterial abundance in: (A) stomach, (B) duodenum.

*Helicobacter* genus (significantly higher in the stomach than in the duodenum) in patients with CeD but not in the control group. Presumably, this is a group of other gastric, spiral-shaped non-*Helicobacter pylori* helicobacters (NPHH) which are associated with gastrointestinal disorders. Some of the known NPHH species, such as *Helicobacter suis*, *Helicobacter felis*, *Helicobacter bizzozeronii*, *Helicobacter salomonis* and *Helicobacter helmanii*, have been detected in the human stomach and may be associated with antral erosions, gastritis, duodenal ulcers, and even gastric MALT lymphoma (Matos et al., 2021; Taillieu et al., 2023). While we were unable to identify any of these specific species, the presence of the *Helicobacter* genus in the stomach of patients with CeD, but not in the control group, could be considered as one of the candidates for a microbial biomarker triggering CeD.

At the genus level (L6), we identified several other significant differences in bacterial abundance between the CeD and control groups. Particularly noteworthy is the genus *Novosphingobium*, which was exclusively found in the stomach of the CeD group and was absent...
in the control (Supplementary Table S1). As in the case of *Helicobacter*, bacteria belonging to the genus *Novosphingobium* were present in patients with CeD, especially in the stomach, while in the control group, it was absent. Although it was not possible to identify bacterial species of this genus in this study, at the L6, correlations with clinical data showed a significant negative (rho −0.31, *P* = 0.047) relationship between *Novosphingobium* and HDL in the stomach (Supplementary Table S3) and anti-tTG (rho −0.324, *P* = 0.042) in the duodenum, and a positive correlation with creatinine (rho 0.34, *P* = 0.034) in the duodenum of patients with CeD (Supplementary Table S4). We observed an intriguing significant relationship between the genus *Novosphingobium* and the presence of selected HLA DQ genes: a negative correlation with DQ2.5 (rho = 0.029, Mann-Whitney U test) and a positive correlation with DQ8 (rho = 0.034, Mann-Whitney U test) (Figure 1). The genus *Novosphingobium*, classified in the class Alphaproteobacteria and family Erythrobacteraceae, comprises Gram-negative, rod-shaped non-spore-forming bacteria widely distributed in various environments, such as toxic chemical-contaminated soil and fresh and sea water (Y. Liu *et al.*, 2021). However, some species may be associated with the development of autoimmune diseases like primary biliary cirrhosis (PBC). In a review about alphaproteobacteria causing such diseases, Mohammed and Mattner cited reports associating *Novosphingobium aromaticivorans* infection with PBC. The authors proposed molecular mimicry between these bacteria and mammalian proteins as a possible mechanism. This relationship between mammalian and bacterial structures can trigger immunological cross-reactions leading to tissue pathology and autoimmunity (Mohammed and Mattner, 2009). Moreover, PBC frequently coexists with CeD (Kingham and Parker, 1998). A similar mechanism has also been described by Galipeau and Verdu in their analysis, citing the study of Petersen *et al.* who demonstrated that DQ2.5-restricted gliadin epitopes shared high similarity with peptides from *Pseudomonas fluorescens* and *Pseudomonas aeruginosa*. They postulated that cross-reactivity could occur between other bacterial or viral peptides and gliadin, potentially initiating or maintaining CeD (Galipeau and Verdu, 2022). Thus, it is also possible that both bacteria belonging to the genus *Novosphingobium* and other microorganisms exert similar effects on the development of autoimmune diseases in predisposed individuals, including CeD.

In our study, we observed a noteworthy relationship between low HDL and high creatinine levels, the occurrence of DQ8 allele, and the percentage of bacteria of the genus *Novosphingobium* in the stomach of CeD children. These findings suggest that these microorganisms could be involved in the disruption of homeostasis in humans, and it is possible that a specific species within this genus, distinct from *N. aromaticivorans*, could play a role in the development of CeD. Interestingly, the correlation between the abundance of *Novosphingobium* genus in the duodenum and the level of anti-tTG antibodies is negative, suggesting a potentially beneficial effect of these microorganisms on the disease course in children with CeD. Such a hypothesis aligns with reports presented in the research review by Caputo *et al.* (2010), in which one of the possible treatment strategies, instead of gluten free diet (GFD), involves using prolyl oligopeptidases (large cytosolic enzymes that belong to the class of serine peptidases) to eliminate harmful gluten peptides from the coeliac diet, thereby neutralising the immunogenic effect of the gluten epitopes. Notably, such enzymes are produced by, among others, *Novosphingobium capsulatum* (homotypic synonym: *Sphingomonas capsulata*) (Caputo *et al.*, 2010).

We also observed a significant negative correlation (rho −0.034, *P* = 0.034) between IEL in the duodenum and the relative percentage of bacteria belonging to the genus *Blautia* (Supplementary Table S4). The abundance of these microorganisms, particularly the species *B. wexlerae*, was significantly higher in the duodenum of CeD children compared to the control group, and interestingly, these microorganisms were also more abundant in the duodenum than in the stomach of the same patients (Supplementary Tables S1 and S2). In addition, the relative percentage of *Blautia* positively correlated with the presence of the HLA DQ 8 gene, as in the case of *Novosphingobium* bacteria. Specifically, a statistically significant relationship (*P* = 0.011, Mann-Whitney U test) was found between the HLA DQ 8 gene and the species *B. wexlerae* (Figure 1). Some *Clostridium* and *Ruminococcus* species have been reclassified as *Blautia* based on phenotypic and phylogenetic analyses. *Blautia* is a genus that comprises obligately anaerobic, non-sporulating, coccobacillus-shaped Gram-positive bacteria, commonly found in the feces and intestines of mammals. Within this genus, *B. wexlerae* and *B. luti* have been identified as the most abundant species, especially in the human intestine. Although these bacteria are believed to possess probiotic properties, their role in health and disease has not yet been fully explained (X. Liu *et al.*, 2021). *Blautia* members are associated with the production of short-chain fatty acids (SCFAs) and antimicrobial peptides. Additionally, these bacteria
have been found to be decreased in the microbiota of individuals with obesity or diabetes (Benítez-Páez et al., 2020a; Maturana and Cárdenas, 2021). Studies analysing the composition of the intestinal microbiota in people with CeD have yielded various results. Shi T. et al., in their examination of faecal samples from adults with newly diagnosed CeD, reported a significantly lower abundance of bacteria from the genus Blautia compared to the control group (Shi et al., 2022). Similarly, Palmieri et al. observed a reduction in Blautia bacteria abundance in faecal samples of coeliac patients who were in remission and had followed a GFD for at least 2 years, when compared to the control group (Palmieri et al., 2022). Additionally, Sample et al. analysed stool samples from children with CeD and found a unique microbiota signature in these patients when placed on a GFD for one year, including significantly lower abundance of Blautia bacteria, compared to the control group (Sample et al., 2021). Moreover, according to Hansen et al., a low-gluten diet resulted in a decrease in B. wexlerae abundance compared to a high-gluten diet (Hansen et al., 2018). In our study, we observed a significantly higher relative percentage of bacteria from the genus Blautia in the duodenum of patients with CeD compared to children from the control group. This finding can be attributed to two factors. Firstly, our assessment focused on the microbiota of tissues from the upper digestive tract rather than faecal samples. Secondly, the CeD group in our study consisted of newly diagnosed individuals who had not yet excluded gluten from their diet, which may explain the increased abundance of this bacterium species. On the other hand, a negative correlation with IEL suggests a potential anti-inflammatory effect of these bacteria. It seems to confirm the supposition of other researchers regarding the role of the genus Blautia, and particularly the species B. wexlerae, in fat metabolism, obesity reduction and anti-inflammatory effects (Benítez-Páez et al., 2020; Hosomi et al., 2022).

The negative correlation between the relative percentage of bacteria from the Novosphingobium genus and the presence of the HLA DQ2.5 gene, as well as the positive relationship between these microorganisms and bacteria from the Blautia genus with the HLA DQ8 gene, may suggest that certain species or strains of the mentioned genera may exert distinct effects, beneficial or harmful, depending on the particular HLA allele possessed by coeliac patients and on the significant abundance of these bacteria in either the stomach or duodenum. To fully comprehend the underlying mechanisms and functional roles of these microbial associations, further research is warranted.

The abundance of bacteria from the Turicibacter genus showed significant differences in our study, with higher levels observed in the duodenum ($P < 0.001$) and lower levels in the stomach compared to the duodenum ($P < 0.01$) in CeD group, as compared to the control group (Supplementary Table S1). Nevertheless, a significant (negative) correlation between this genus and total cholesterol (rho $-0.36$, $P = 0.022$) and LDL (rho $-0.40$, $P = 0.01$) concentrations was found in the stomach (Supplementary Table S3). Turicibacter belongs to the Firmicutes phylum and is classified as Gram-positive, rod-shaped bacteria commonly found in the human faecal microbiota (Lynch et al., 2023a; Rahman et al., 2023). Interestingly, analyses of the intestinal microbiota done by Rodriguez-Diaz et al. (2022) showed higher abundances of Turicibacter in patients with non-alcoholic fatty liver disease (NAFLD), coexisting with significant liver fibrosis, compared to NAFLD patients without or with mild liver fibrosis (Rodriguez-Diaz et al., 2022). Furthermore, in a mouse model, Li et al. (2020) observed a significantly higher abundance of, among others, the Turicibacter bacteria in the non-alcoholic steatohepatitis (NASH) group compared to the non-alcoholic fatty liver (NAFL) group, suggesting its potential as a distinguishing factor between the two conditions (Li et al., 2020). Previous studies on the human gut microbiota have shown that Turicibacter has the ability to modify bile acids and lipid metabolism of the host. The effect of diverse Turicibacter isolates on host metabolites, including lipids and bile acids, was presented by Lynch et al. Conflicting reports exist, with some indicating positive and others negative correlations of Turicibacter bacteria with dietary fats and host adiposity. The authors of the cited work hypothesised that the type of dependence is related to the properties of specific bacterial strains and that the presence of bile modifying genes from these isolates influences host physiology. They found that these genes are sufficient to modify host lipid and cholesterol states (Lynch et al., 2023). Based on our clinical data and correlation results, it could be hypothesised that the stomach of children with CeD may be colonised by Turicibacter strains which have a beneficial effect on lipid metabolism and adipose tissue. However, considering the scarcity of studies on the presence of these bacteria in the upper digestive tract of children with CeD and their possible impact on lipid metabolism, liver function and the course of this disease, further comprehensive research is needed. The above findings suggest that this bacterial taxon could also serve as one of the biomarkers for CeD.
5 Conclusions

Our work indicates a significant presence of microorganisms belonging to little-known genera in the upper digestive tract of children with newly diagnosed CeD. To our knowledge, this is the first analysis involving a comprehensive assessment of bacterial profiles of both gastric and duodenal biopsies obtained from children with CeD. The results concerning the composition of the gastric microbiota and the correlation of selected genera with clinical data and specific HLA DQ genes (especially DQ 8) are particularly interesting. The analysis of the data from this study suggests that in order to fully understand the mechanism of CeD development in genetically predisposed individuals, it is necessary to take into account not only the abundance of a given genus or species of bacteria, but also the anatomical location of its occurrence. The function and correlation with the HLA gene allele may depend on whether it is present in the stomach or duodenum and in what percentage. One limitation of this study was the small number of samples, particularly in the control group. Moreover, this group was not fully representative of the general population of patients, as it was selected among patients undergoing upper endoscopy for gastrointestinal ailments that were later ruled out. Furthermore, the abundance of bacteria in these ecological niches is typically much poorer than in the lower digestive tract, as noted by Constante et al. (2022) in their work. The isolated low concentration of bacterial genetic material required additional amplification procedures (nested PCR) and ultimately limited a more precise assessment of the microbiota at the species level (L7). Due to the above-mentioned circumstances, direct comparison of our results with observations from other researchers may prove challenging, especially considering that they did not test gastric samples. However, our study highlights significant differences in selected genera of bacteria, such as Helicobacter, Novosphingobium, Blautia and Turicibacter, suggesting their potential role as new biomarkers in the development of CeD in children and adolescents. This study can also serve as a basis for further, more detailed research analysing the functions of specific species and strains within these genera of microorganisms.

Supplementary material

Supplementary material is available online at: https://doi.org/10.6084/m9.figshare.25562211

Table S1. Statistically significant ($P_{\text{FDR}}$-value < 0.05) differences in the relative percentage of bacteria at the genus level between the CeD and control groups (in the stomach and duodenum) and between two locations of the upper digestive tract (in the CeD and control groups).

Table S2. Statistically significant ($P_{\text{FDR}}$-value < 0.05) differences in the relative percentage of bacteria at the species level between the CeD and control groups (in the stomach and duodenum) and between two locations of the upper digestive tract (in the CeD and control groups).

Table S3. Correlations between clinical data and the presence of bacteria in the stomach of patients with celiac disease at the genus (L6) and species (L7) levels.

Table S4. Correlations between clinical data and the presence of bacteria in the duodenum of patients with celiac disease at the genus (L6) and species (L7) levels.

Acknowledgements

This study was funded by the National Science Centre in Poland in frame of the project: SONATA Bis7 – 2017/26/E/NZ5/00266. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Authors’ contribution

Conceptualisation, TG, DS, KK-D, AK, MD and AG; methodology, TG, HK, MAI and YS; software, MK and AG; validation, HK and YT; analysis of data, TG, DS, KK-D, AK, MK and AG; writing-original draft preparation, DS; writing-review and editing, TG, KK-D, AK, MK and AG; visualisation, DS and AG; supervision, TG, KK-D; funding acquisition, TG. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Data availability

All data relevant to the study are included in the article or uploaded as supplementary information. Fastq files are available: https://portalwiedzy.cm-uj.krakow.pl/info
References


Adherence to gluten-free diet restores alpha diversity in celiac people but the microbiome composition is different to healthy people. Nutrients 14: 2452. https://doi.org/10.3390/nu14122452


