1 Metataxonomic review to elucidate the role of the microbiome in 2 Celiac disease across the gastrointestinal tract.

3 Juliana Arcila^{1,2}, Viviana Loria-Kohen^{3,4}, Ana Ramírez de Molina³, Enrique
4 Carrillo de Santa Pau^{1*}, Laura Judith Marcos-Zambrano^{1*}

5 ¹Computational Biology Group, Precision Nutrition and Cancer Research Program,
6 IMDEA Food Institute, Madrid, Spain. ²Current address: Computational Epigenomics
7 Laboratory, Faculty of Medical Sciences, Newcastle University. Newcastle, UK.

8 ³Nutrition and Clinical Trials Unit, GENYAL Platform IMDEA-Food Institute, CEI
9 UAM+CSIC, 28049, Madrid, Spain. ⁴ Current address: Departamento de Nutrición y
10 Ciencia de los Alimentos, Faculty of Pharmacy, Universidad Complutense de Madrid,
11 28040 Madrid, Spain.

12
13 *** Correspondence:**

14 Enrique Carrillo de Santa Pau
15 enrique.carrillo@imdea.org
16 Laura Judith Marcos-Zambrano
17 judith.marcos@imdea.org
18

19 Abstract

20 **Background:** Dysbiosis of the microbiome has been related to the Celiac disease (CeD)
21 progress, an autoimmune disease characterised by gluten intolerance developed in
22 genetically susceptible individuals under certain environmental factors. The microbiome
23 contributes to CeD pathophysiology modulating the immune response by the action of
24 short-chain fatty acids (SCFA), affecting gut barrier integrity allowing the entrance of
25 gluten derived proteins, and degrading immunogenic peptides of gluten through
26 endoprollyl peptidase enzymes. We reviewed state of the art in taxonomic composition
27 for CeD and compiled the larger dataset of 16S prokaryotic ribosomal RNA (rRNA)
28 gene high-throughput sequencing for consensus profiling. We present for the first time

an integrative analysis of metataxonomic data from CeD patients, including samples from different body sites (saliva, pharynx, duodenum, and stool). We found the presence of coordinated changes through the gastrointestinal tract characterised by an increase in *Actinobacteria* species in the upper tract (pharynx and duodenum), and an increase in *Proteobacteria* in the lower tract (duodenum and stool), as well as site-specific changes evidencing a dysbiosis in CeD patients' microbiota. Moreover, we described the effect of adherence to a gluten-free diet (GFD) evidenced by an increase in beneficial bacteria and a decrease in some *Betaproteobacteriales* but not fully restoring CeD-related dysbiosis. We illustrate that the gut microbiota acts as an enhancer of immune response in CeD through the production of lipopolysaccharides and other bacterial components that activate the immune response and by decrease SCFA producers bacteria. Furthermore, microbial changes observed through the gastrointestinal tract of CeD patients may help manage the disease and follow-up GFD treatment.

Keywords: microbiome, celiac disease, 16S rRNA gene, Gluten-free diet.

1. Background

The gut microbiome comprises the genome of all the microorganisms that inhabit the gut, and it plays a crucial role in nutritional, metabolic, physiological and immunological processes of the body; it also produces several metabolites such as short-chain fatty acids (SCFAs), branched-chain amino acids, vitamins, and hormones, that affects several metabolic and inflammatory pathways [1] Numerous immune-mediated diseases have been linked to shifts in the gut microbiome composition, including celiac disease (CeD) [2].

53 CeD is an autoimmune disease affecting the small intestine, ranging from an
 54 intraepithelial lymphocytosis to the total atrophy of intestinal villi as a response to
 55 gluten consumption [3]. This disease is characterised by a genetic predisposition given
 56 by the alleles coding for the HLA-DQ2 and/or HLA-DQ8, the presence of antibodies
 57 against transglutaminase type 2 (TG2) and IgA and IgG anti-gluten as well as
 58 gastrointestinal (GI) symptoms when consuming gluten-containing foods [3]. The HLA
 59 variant with a higher association with CeD is HLA-DQ2, being present in more than
 60 90% of patients. About half of the remaining patients possess an HLA allele coding for
 61 HLA-DQ8. However, a small fraction of individuals at genetic risk develops CeD.
 62 Suggesting that besides genetic factors, the risk of disease can also depend on non-
 63 gluten environmental factors [4].

64 Studies profiling the microbiome in CeD supports an alteration in the intestinal
 65 microbiome of patients suffering from the disease [2]. The intestinal microbiome in
 66 CeD comprises an increased number of opportunistic bacteria clades while beneficial
 67 clades are decreased [5]. This phenomenon is called dysbiosis and has become subject
 68 to active and growing research; as one of the non-gluten environmental factors
 69 contributing to CeD development [6–8].

70 Gliadin is the gluten-derived protein that triggers proinflammatory cytokines in CeD
 71 patients, among them IL-6, which stimulates the growth and differentiation of T
 72 cytotoxic cells that mediate the destruction of surface epithelial cells [9–11]. Gliadin
 73 peptides present multiple proline and glutamine amino acids residues, making them
 74 resistant to proteolysis through gastric, pancreatic, and intestinal proteases. This results
 75 in incomplete digestion of gluten and the generation of high molecular weight

oligopeptides, some of which have immunogenic and toxic properties relevant to CeD.
The most studied of these peptides is 33-mer [12, 13].

Recent research has shown that peptidases from different sources can degrade gluten and gluten-derived peptides [14, 15]. In this regard, several bacteria from the human digestive tract (i.e: *Bifidobacterium* spp., *Lactobacillus* spp., *Rothia* spp) can potentially degrade gluten, and a healthy microbiome composition could modulate the symptoms of gluten-related diseases [15]. The presence of several microorganisms that produce proteases able to hydrolyse peptides rich in proline and glutamine residues has been described in the oral cavity, duodenum, and faeces [14, 15].

On the other hand, the tissular Transglutaminase (tTG) enzyme has an essential role in CeD pathogenesis by enhancing gliadin epitopes immunogenicity. This enzyme effectuates specific and organised deamination of high molecular weight derived gliadin peptides giving them an improved affinity to HLA-DQ molecules associated with CeD and forming kinetically stable HLA-DQ-peptide complexes [16]. The HLA-DQ heterodimers are HLA class II molecules expressed on the surface of antigen presenter molecules (APC). In CeD pathogenesis, HLA-DQ molecules associated with the disease play a role in presenting gliadin peptides to T CD4+ cells restricted to recognise those HLA-DQ heterodimers. These specific T cells are present in CeD patients but not in healthy people [9, 17]. Some studies have found that HLA-DQ genotype affects early gut microbiota composition, favouring the colonisation with pathogenic bacteria [18].

The gut microbiome could play a role in triggering CeD in genetically susceptible persons in several ways. First, dysbiosis could lead to an alteration in the intestinal barrier, sustained inflammation or infection can lead to deregulation in the expression of adhesion molecules at tight junctions leading to the entry of microbes and toxic

100 substances and facilitate entry of incompletely digested gliadin peptides in lamina
101 propria [2, 5, 8]. Of several proteins involved in tight junctions, disassembly of zonulin
102 has been implicated in CeD patients [19] gluten peptides and some enteric bacteria, like
103 *Escherichia coli*, can induce this protein [18].

104 Gliadin peptides and gut dysbiosis can activate innate and adaptive immune systems
105 similarly [20]. Pathogenic gram-negative bacteria activate the innate immune system by
106 activating Toll-Like Receptors (TLR-4), and CD14 complexes recognise bacterial
107 endotoxins and lipopolysaccharide and activate the innate immune system to release
108 proinflammatory cytokines [2, 20]. In patients with CeD, gluten intake is associated
109 with activating gluten-specific CD4⁺ T cells in the lamina propria and upregulation of
110 IL-15, a proinflammatory cytokine [21]. Moreover, gut microbiota can also activate
111 Th1, Th2 and Th17 mediated immune responses similar to upregulation by gliadin
112 peptides [22].

113 Finally, as discussed above, dysbiosis may also increase the amount and size of gliadin
114 peptides due to differential peptidolytic activity of the gut microbiota [14, 15]. These
115 studies suggest that the gut microbiota affects gluten digestion, intestinal permeability,
116 and the host immune system, all the mechanisms involved in the pathogenesis of CeD.

117 A schematic representation of the microbiome's involvement in CeD suggested by the
118 current knowledge in the topic is shown in Figure 1. In the intestinal lumen are the
119 peptides resulting from the degradation of the diet components, including gluten; there
120 is also the microbiome composed of beneficial and opportunistic commensal
121 microorganisms (step 1, Fig. 1). A dysbiosis characterises CeD microbiome with an increase
122 of opportunistic bacteria (mainly from the phylum *Proteobacteria*), these bacteria compete to
123 adhere to the epithelium, and when they reach, they trigger an inflammatory immune

124 response that increases the permeability of the epithelial barrier, allowing the passage of
125 immunogenic gluten peptides to the lamina propria (step 2, Fig. 1). In there, the gliadin
126 peptides are modified by TG2, improving their junction to HLA DQ2/8 heterodimers
127 (step 3, Fig. 1). These modified peptides are delivered by APCs to T lymphocytes (step
128 4, Fig. 1), producing a chain of immune responses that ends up in the destruction of
129 enterocytes (step 5, Fig. 1) and the production of anti-gliadin and anti-TG2 antibodies
130 (step 6, Fig. 1)

131 **1.1 Gluten-free diet and microbiome:**

132 Strict adherence to a Gluten-Free Diet (GFD) and a lifelong exclusion of gluten from
133 the diet is the first-line treatment and is currently the only effective therapy for CeD
134 [23]. A GFD requires the complete exclusion of gluten, a protein complex present in
135 food products from wheat, rye, barley, oats, among other cereals. It comprises only
136 naturally gluten-free food products (e.g., legumes, fruit and vegetables, unprocessed
137 meat, fish, eggs and dairy products) and/or substitutes of wheat-based foods, specially
138 manufactured without gluten or having a gluten content lower than 20 ppm, as per
139 European legislation [24].

140 GFD has become a trend in contemporary history being associated with an increase in
141 energy and boosting health; however, evidence suggests that GFD is an unbalanced diet
142 with multiple nutritional deficiencies [25]. Gluten restricted products are often
143 associated with an inadequate nutritional value, characterised by a higher fat intake, less
144 vegetal-protein intake, and higher carbohydrate and sugar consumption [24, 25].
145 Although GFD can reduce the symptoms of CeD in most patients, it does not entirely
146 restore the gut microbiota to that of healthy individuals; moreover, it has been reported

147 that up to 30% of patients will exhibit non-responsive CeD, a condition characterised by
148 the persistent enteropathy and CeD related symptoms after one year on a GFD [26].

149 **1.2 Microbiome research and CeD**

150 Traditional methodologies like PCR-DGGE [27, 28] PCR-TTGE [29], cloning [30] and
151 culture [31, 32] for subsequent identification by Sanger sequencing were used to study
152 the microbial composition of the gastrointestinal tract of CeD patients before 16S rRNA
153 gene-based metataxonomic approach become popular. These previous studies set the
154 basis of current knowledge in CeD by revealing a lower microbial diversity in CeD patients
155 with gastrointestinal symptoms than in patients with extraintestinal symptoms suggesting for the
156 first time the implication of the microbiota in the progress of the disease [27, 28]. Moreover,
157 despite the low resolution of the techniques employed, authors were able to characterise a
158 particular microbial profile [29], or even particular strains [31, 32], related to CeD, different to
159 healthy controls, and patients undergoing GFD, and the implications of the type of feeding in
160 infants, and HLA-DQ genotype in the colonisation of some bacterial species DGGE [27, 28].

161 **1.3 New high-throughput technologies:**

162 In 2012, 16S rRNA gene high throughput sequencing was used for the first time to
163 determine the composition and temporal changes of the gut microbiota on infants at risk
164 of CeD [33]. The main advantage of the technique is that it allows comparative analysis
165 of the whole microbial community, diversity, and abundance at significant sequencing
166 depths, while traditional methodologies focus only on some members of the
167 communities [34]. The method consists of sequencing the 16S rRNA gene marker
168 amplified from total DNA extracted directly from the tissue or environment sampled
169 without previous culture [35]. The 16S prokaryotic ribosomal rRNA genes sequences
170 are reliable markers for taxonomic classification and phylogenetic analysis of bacteria.

171 This gene is ubiquitous in the kingdom and has considerable variability in its sequence
172 between different species, allowing their classification into different taxa. However,
173 currently available high throughput sequencing techniques do not allow the sequencing
174 of 16S rRNA gene amplicons spanning the gene's full length. So, the strategy used is to
175 sequence any of its nine hypervariable regions (V1-V9) or the regions between them
176 [36].

177 On the other hand, shotgun metagenomics is based on the untargeted sequencing of
178 genetic material present in the sample. It allows the taxonomic analysis and functional
179 profiling of all types of microorganisms present in the microbiome (bacteria, viruses,
180 fungi and protozoa). Moreover, it allows identifying species, strains, and even Single
181 Nucleotide Polymorphisms (SNP) [37].

182 We performed a scoping review of the literature employing high-throughput sequencing
183 of the 16S rRNA gene marker to analyse the microbiome profile in CeD. Also, we
184 performed a merged-data analysis combining the multiple datasets of 16S rRNA gene
185 microbiome sequencing available in public databases following the same methodology
186 and leveraging proper metadata to get a consensus profiling of the microbiome in the
187 disease. We compile datasets from different body sides and extensive metadata,
188 considering the influence of GFD over the gut microbiome, trying to find microbial
189 biomarkers for CeD not only in the duodenum but also in less invasive samples as
190 saliva, stool, and oropharynx exudates.

191 **2. Methods**

192 **2.1 Literature search, identification, and selection of relevant studies.**

193 This study follows the scoping review methodology for searching and assessment of the
194 relevant studies [38]. This methodology adopts the following workflow (1)

195 identification of a research question; (2) identification of relevant studies; (3) study
196 selection; (4) charting the data; (5) collecting, summarising, and reporting the results,
197 for this, we conducted a systematic and comprehensive literature search of the studies in
198 years from 2010 to August 2020; that carried out metagenomic sequencing of the 16S
199 rRNA gene for microbiomes of CeD patients, including those on a GFD as well as on an
200 unrestricted diet. We searched on PubMed, Google Scholar, and SCOPUS. However, no
201 additional studies to those in PubMed were found in Google Scholar or SCOPUS. The
202 terms used for the search, the automatic filters, and the statistics results are summarised
203 in Table S1; Figure S1 summarises the methodology used for the scoping review. The
204 studies retrieved with the previous strategy undergo manual curation to exclude non-
205 CeD related studies and studies using non-high-throughput 16S rRNA gene sequencing,
206 which escaped the automatic filters.

207 **2.2 Merged-data analysis of publicly available microbiome data sets.**

208 Out of the total selected studies for the literature review, those included in the merged-
209 data analysis met the following criteria: (a) Bacterial 16S rRNA gene sequenced from
210 total DNA using high-throughput sequencing, (b) Data must be available in fastq format
211 in one of the publicly accessible databases NCBI Sequence Read Archive [SRA] or EBI
212 European Nucleotide Archive [ENA]), (c) Metadata on the sequenced biological
213 samples must be available and must include tissue of origin, information on the type of
214 diet (with or without gluten), and clinical classification (case or control).

215 **2.3 Amplicon Sequence Variant detection and taxonomic assignment.**

216 All data from the selected studies were available in the European Nucleotide Archive
217 (ENA) database. Raw 16S rRNA gene sequencing data sets from each study were
218 downloaded from that database. The quality of the sequencing was examined using

219 FastQC v.11.9 software [39], and the primers used in the PCR amplification of
220 hypervariable regions were removed using the Cutadapt v.2.9 software [40]. For data
221 obtained by paired-end sequencing, each pair of reads was joined by overlapping
222 assembling using the software FLASH v.1.2.11 [41]. Then, each study was processed
223 individually for the construction of the ASV count table for each sample, using the
224 software DADA2 v.1.15 [42]. Briefly, DADA2 was used for quality filtering of the
225 sequences, detecting exact amplicon sequence variants (ASV), removing chimeras, and
226 finally, the taxonomic assignment of ASVs with SILVA SSU v.138 database [43]. The
227 computer code was developed with the R and "bash shell" programming environments
228 for GNU / Linux. All the workflow and the specific criteria for each step in the analysis
229 for each dataset are available on GitHub [https://github.com/juearcilaga/Assesing-](https://github.com/juearcilaga/Assesing-microbiome-profiles-of-celiac-disease-patients.git)
230 [microbiome-profiles-of-celiac-disease-patients.git](https://github.com/juearcilaga/Assesing-microbiome-profiles-of-celiac-disease-patients.git).

231 **2.4 Data merging filtering and normalisation.**

232 As a result of the variant detection process, two tables were obtained per each study.
233 The first one with the abundance of each variant (ASV-count table) and the second one,
234 with the taxonomic assignment of ASVs in six taxonomy ranks (Kingdom, Phylum,
235 Class, Order, Family, and Genus). Tables were combined using the phyloseq_merge
236 command. With phyloseq taxa_glom command, taxa were agglomerated at the genus
237 level, to avoid ASV duplication bias. All tables, along with raw and processed data from
238 this study, are available in FigShare.
239 https://figshare.com/projects/Analisis_del_microbioma_en_Enfermedad_Celaca/82547

240 **2.5 Statistical analysis.**

241 All the tests were carried out using the Phyloseq and Vegan v.2.5.6. [44] packages of the
242 R programming environment. For the representation of statistical significance in the

243 graphs, the following equivalences were used: $p < '***' \rightarrow 0.001$, $'**' \rightarrow 0.01$, $'*' \rightarrow 0.05$,
 244 $'.' \rightarrow 0.1$, $' ' \rightarrow 1$. The bar graphs and the scatter graphs were obtained in the R programming
 245 environment using the ggplot2 package [45].

246 **2.6 Estimation of the biological diversity and composition of the microbiome.**

247 The Phyloseq v.1.3 package [46] from R v.3.6.3 [47] was used to estimate rarefaction
 248 curves for each sample. Rarefaction curves were useful to determine the minimum
 249 sequencing depth for reaching the saturation of observed species.

250 The diversity and richness of each sample were estimated by Shannon, Simpson, and
 251 Chao1 indexes using the phyloseq package. The Shannon and Simpson indexes were
 252 estimated after normalising the data using the Centered log-ratio method in the
 253 Compositions package [48], while the Chao1 index was estimated on the raw data. The
 254 normality of the data was evaluated using the Shapiro-Wilk test. The Wilcoxon-Mann-
 255 Whitney test was used to compare diversity means between two groups and the
 256 Kruskal-Wallis test for more than two groups. The homoscedasticity of the variances
 257 was calculated using Levene's test. Since the Shannon index data did not meet the
 258 criteria of homoscedasticity or normality, the data were adjusted to a normal distribution
 259 by square root transformation and subsequently analysed with the Welch test.

260 Also, a principal coordinate analysis (PCoA) was performed on weighted Unifrac
 261 distances to display whether the data was grouped by any of the variables included in
 262 the metadata (case/control and tissue sampled, 16S rRNA gene region, age, gluten / free
 263 diet, and sequencing technology). Before PCoA data analysis, taxa with less than five
 264 counts in raw ASV-count tables or being present in just one sample were discarded.

265 **2.7 Differential abundance, association, and LDA for the discovery of microbial** 266 **markers on CeD.**

267 Data for each tissue was subsetted from raw ASV-count tables. The taxa present in less
268 than 10% of samples or having less than ten counts were discarded. The CLR
269 transformation was performed on raw ASV-count tables before differential abundance
270 analysis with RNAseq methods and Correlation Analysis. Before the LDA Effect Size
271 (LEfSe) analysis, taxa with less than ten counts in raw ASV-count tables were
272 discarded. Both data rarefaction and the Total Sum Scaling (TSS) normalisation were
273 performed on raw ASV-count tables.

274 The difference in taxa abundances between cases and controls for each sampled tissue
275 was evaluated with DESeq2 software v.1.26.0 [49]. Differences with adjusted *p*-value
276 < 0.01 and Fold change < 3 were considered significant. Association among the taxa at
277 the genus level and CeD were studied by Spearman-Rank correlation analysis. Also, a
278 LEfSe analysis was performed, and genera were considered significant if adjusted *p*-
279 value < 0.1 and LDA > 2.

280 **2.8 Inference of the microbiome metabolic potential**

281 The software PICRUST2 [50], which infers the genes encoded in studied the taxa
282 genomes, was used to predict the microbial communities' functional and metabolic
283 capacities present in each sample. The results consist of tables of genes, metabolic
284 pathways, and enzymes abundance in each sample. Data were normalised by using the
285 CLR transformation. The differential abundance of metabolic pathways and genes,
286 between cases and controls, was estimated using DESeq2. Subsequently, a targeted
287 search for differentially abundant genes whose impact on the celiac disease's
288 pathogenesis is of interest was carried out.

289 **3. Results**

290 **3.1 Literature search, identification, and selection of relevant studies.**

291 We found 19 original research papers matching our criteria; we decided to group the
 292 studies into five categories according to the final aim and the main findings as follow:
 293 (1) Gut microbiome of children at-risk of CeD before the onset of the disease (2) Gut
 294 microbiome of CeD patients in comparison with healthy controls, and other groups of
 295 interest, (3) Influence of GFD on the microbiome of healthy people (gut) and CeD and
 296 Non-Celiac Gluten Sensitivity patients (saliva and gut) (4) Upper gastrointestinal tract
 297 microbiome in CeD and (5) Treatments: dietary interventions, prebiotics and hookworm
 298 infections. Table S2 summarise the main findings of the select papers.

299 **3.1.1 Gut microbiome of children at risk of CeD before the onset of the** 300 **disease.**

301 We found four research in which a longitudinal study of the microbiome profile in stool
 302 samples of children at-risk of CeD before the disease's onset was performed. Some of
 303 these studies focused on identifying microbiome composition at an early age in at-risk
 304 infants that could be predictive of CeD development [33, 51, 52], the effect of the time of
 305 first gluten exposure [33] and other environmental factors [53]; such as delivery mode,
 306 antibiotic exposure and infant feeding type on microbial gut composition and/or CeD
 307 development.

308 Sellitto et al. (2012) were the pioneers in using the Pyrosequencing of the 16S rRNA
 309 gene to study infant faecal microbiome before CeD onset. Their research aimed to know
 310 if gut microbiota trajectory in early life may predict the development of celiac disease
 311 and if first-time gluten exposure could affect the time of CeD onset. The authors found
 312 that the microbiota composition was highly different among at-risk infants before 18

313 months of age but converged at 24 months. In all children, the microbiota was
314 characterised by a low abundance of members of the phylum *Bacteroidetes*. Infants
315 were breastfed until six months of age. From 6 to 12 months, eight children followed a
316 GFD and the other eight children followed a gluten-containing diet. From 12 to 24
317 months, all the children followed a gluten-unrestricted diet. At 12 months of age, more
318 children in the group of early exposure to gluten presented antigliadin antibodies. Only
319 one out of all children belonging to the group of early exposure to gluten developed
320 CeD. The children who developed CeD showed a reduction in bacterial richness from 6
321 to 10 months. However, no definitive answer can be given to the original questions
322 because only one child is insufficient to make statistical analysis and achieve valid
323 conclusions [33].

324 Six years later than Sellitto et al. (2012) study was published, Rintala et al. 2018
325 assessed the same question. May gut microbiota trajectory in early life be a predictor of
326 celiac disease development? But, increasing sample size and follow-up time. They
327 studied changes in microbial communities of 27 infants genetically at risk of CeD from
328 9 to 12 months of age. Unlike Sellitto et al., they did not test the effect of first-time
329 gluten exposure. The introduction of gluten was initiated from 4.4 to 5.7 months of age
330 in all cases, similar to the early exposure time in Sellitto et al. The follow-up time of this
331 study was more extensive than in Sellitto et al. Nine children (all girls) out of the 27
332 developed CeD at the median age of 3.5 years (range 2.6–4.2 years). These observations
333 would not have been made in the followed time studied by Sellitto et al. (0-24 months).
334 Rintala et al. study showed a significant increase in children's microbiota alpha diversity
335 in general between 9 and 12 months of age. Still, no differences among healthy controls
336 and the group who develop CeD were detected. The study concludes that the individuals
337 who develop CeD, do not already in the early infancy, have a distinct faecal microbiota

338 composition compared to other infants with risk-HLA-haplotype, suggesting that the
339 onset of CeD is more likely a consequence of a solid external trigger rather than gradual
340 development due to peculiarly vulnerable gut microbiota [52].

341 Another study by Olivares et al. (2018) aimed to answer the same question about the
342 predictive power of microbial communities' changes at an early age of infants
343 genetically at risk of developing CeD. With a similar sample size and time of gluten
344 exposure, samples were taken earlier in the baby's life. They studied the microbiota of
345 20 children from 4 to 6 months of age. The introduction of gluten was initiated from 6
346 months of age onwards in all cases. Nine out of the twenty children developed CeD
347 among 16 and 40 months of age and one at 82 months. Unlike Rintala et al. they found
348 changes among microbiota of cases vs controls. The study showed a significant increase
349 in microbiota alpha diversity of healthy control children (characterised by increases in
350 *Firmicutes* families) between the ages of 4 and 6 months, which did not occur in the
351 child group who developed CeD. An increased relative abundance of *Bifidobacterium*
352 *longum* was associated with control children, while increased proportions of
353 *Bifidobacterium breve* and *Enterococcus spp.* were associated with CeD development.
354 Authors conclude that the diversity of microbiome increases among 4 and 6 months of
355 age in children who remained healthy but not in those who develop CeD [51].

356 Finally, Leonard et al. (2020) attempted to characterise the changes in microbial
357 communities of infants genetically at risk of CeD. However, unlike the studies
358 mentioned above, they did not perform a follow-up, and they did not evaluate the
359 predictive power of the microbiota composition because their question was different:
360 Do environmental factors like delivery mode, antibiotic exposure and infant feeding
361 type impact on the microbial composition of gut in at-risk infants?. Authors found that
362 genetic risk to develop CeD was associated with decreased abundance of *Streptococcus*

363 and *Coprococcus* and decreased abundance of *Veillonella*, *Parabacteroides* and
 364 *Clostridium perfringens* at 4-6 months of age. Also, with increased abundance of
 365 *Bacteroides* and *Enterococcus* species at 0 months. Cesarean section delivery was
 366 associated with a decreased abundance of *Bacteroides vulgatus* and *Bacteroides*
 367 *dorei* and folate biosynthesis pathway. With an increased abundance of
 368 hydroxyphenyl acetic acid, alterations implicated in immune system dysfunction
 369 and inflammatory conditions. They found differences in the microbiota regarding
 370 exposure to breastmilk and formula and between antibiotic exposure (as an
 371 environmental risk factor). Despite the study provides insights into taxonomic and
 372 functional shifts in the developing gut microbiota of infants at risk of CeD linking
 373 genetic and environmental risk factors to detrimental immunomodulatory and
 374 inflammatory effects, it is unclear whether they indeed contribute to the future
 375 development of CeD [53].

376 In summary, these studies evidence changes in the microbiome of infants with CeD,
 377 particularly modifications in diversity, demonstrating the importance of the inclusion of
 378 gluten in diet.

379 **3.1.2 Gut microbiome of CeD patients in comparison with healthy controls** 380 **and other groups of interest.**

381 We reviewed five papers in which a case-control study of the microbiome profile in
 382 stool and/or duodenal samples of CeD patients was carried out.

383 In 2013, Cheng et al. performed a study analysing the duodenal microbiota from newly
 384 diagnosed children with CeD. They did not find significant differences in the abundance
 385 of taxa between groups, neither at phylum nor at the genus level. However, using
 386 Random Forest (RF) algorithm with a subpopulation of eight genus-like taxa, they were
 387 able to separate samples between CeD and healthy controls. Taxa included *Prevotella*

388 *melaninoenica*, *Haemophilus* sp. and *Serratia* sp. augmented in CeD, whereas
 389 *Prevotella oralis*, *Proteus* sp., *Clostridium stercorarium*, *Ruminococcus bromii*, and
 390 *Papillibacter cinnamivorans* were augmented in controls. This classification system had
 391 an error rate of 31.6% [54] This study showed that bacteria from *Proteobacteria* were
 392 elevated in CeD and butyrate producers *Firmicutes* were diminished despite the
 393 sequencing technology used.

394 A different approach was conducted by Pellegrini et al. (2017). They analysed the
 395 duodenal microbiota of CeD and patients with Diabetes Type 1 (T1D), trying to find
 396 differences in the inflammatory state compared with healthy controls. The authors did
 397 not found differences in the diversity measures across the studied groups. However,
 398 they found differences in the inflammatory status of the patients studied according to
 399 the type of pathology and some relevant changes in the microbiota. In this regard,
 400 authors showed that CeD patients share some similitudes in the microbial profile with
 401 T1D *i.e.*, reduction in *Bacteroidetes* phylum and the class *Clostridia*, but the microbial
 402 profile differed according to the abundance of *Proteobacteria*, being increased in CeD,
 403 particularly *Gammaproteobacteria* from the family *Pasteurallaceae*, and the genus
 404 *Haemophilus* spp. [55].

405 Later, Garcia-Mazcorro et al. (2018) analysed duodenal and stool microbiota from
 406 gluten-related disorders in Mexican patients. Again, they only found differences in
 407 particular taxa abundance but not in the general parameters of diversity. Regarding
 408 duodenal samples, a lower abundance of *Bacteroidetes* and *Fusobacteria* was found in
 409 CeD patients and a higher abundance of *Actinobacillus* (*Gammaproteobacteria*),
 410 *Finegoldia* (*Clostridia*) and TM7 in Non-Celiac gluten sensitivity patients (NCGS), on
 411 the other hand, *Sphingobacterium* (*Bacteroidetes*) was higher in healthy subjects. The
 412 microbial profile of stool samples was different, *Firmicutes* were predominant

413 regardless of disease status, and *Bacteroidetes* were less than 1%. For the first time, this
414 study showed particular microbial changes specific to gluten-related disorders in
415 Mexican people [56].

416 Bodkhe et al. (2019) studied faecal microbiome and duodenal samples from patients
417 with CeD and first-degree relatives. Despite no significant differences in alpha diversity
418 between sample sites were found, a different overall composition of the microbiome
419 was inferred between groups, showing that microbiome composition and diversity were
420 different between duodenum and stool samples for all disease status evidenced by
421 changes in individual taxa abundance [57].

422 More recently, Panelli et al. (2020) studied also duodenum, stool, and saliva from CeD
423 patients. In stool samples, *Bacteroides* were predominant, and no differences in taxa
424 abundance were detected between groups. A decrease in *Firmicutes* and *Actinobacteria*
425 accompanied by an increase in *Proteobacteria* was evident in the duodenum and
426 salivary samples of CeD groups. While at the genus level an expansion of *Neisseria* and
427 a reduction of *Streptococcus* were found in CeD groups compared to controls was also
428 appreciated. Authors conclude that there are different patterns of diversity across groups
429 and body tissues [58].

430 These studies demonstrate that alfa-diversity between samples from patients with CeD,
431 and other groups did not show differences. However, specific changes in taxa
432 abundance were found. Some similarities were evident across studies, for example, the
433 increase of *Proteobacteria* in CeD patients and a decrease in *Firmicutes* and
434 *Actinobacteria*, evidencing the existence of a dysbiosis in CeD with a predominance of
435 gram-negative bacteria.

436 **3.1.3 Influence of GFD on the microbiome of healthy people (gut) and CeD**
437 **and NCGS patients (saliva and gut)**

438 We found five papers studying GFD in different conditions.

439 Bonder et al. (2016) study the influence of a GFD through time by measuring microbial
440 changes in patients with a GFD for four weeks followed by a washout period. The
441 authors did not observe a time-dependent change in alpha diversity or differential
442 abundance in the microbiome composition. There was a substantial difference when
443 comparing individuals regardless of time point, showing that inter-individual effect on
444 the microbiome variation is stronger than the effect of diet. Carbohydrate and starch
445 metabolisms were altered with differences between GFD and a healthy diet in the
446 abundance of pathways related to tryptophan metabolism, butyrate metabolism, fatty
447 acid metabolism, and seleno-compound metabolism. The authors conclude that change
448 in diet did not influence the bacterial diversity within a sample. Inter-individual
449 differences were more influential than the effect of GFD [59].

450 Garcia-Mazcorro et al. (2018) the study was partially discussed in section 1.2. They also
451 investigated the potential microbial signatures associated with GFD by consuming
452 certified gluten-free foods in patients with paired samples. Contrary to Bonder et al.,
453 they found that the groups studied showed group-specific variation over time after
454 consuming the GFD for four weeks, notably NCGS presented an increase in the
455 duodenal *Pseudomonas* on the GFD. In contrast, only half of CeD patients showed
456 increases in *Pseudomonas*, but these increases were pronounced to affect median
457 values. Biomarker analysis of the taxa at the genus level confirmed the results on
458 *Pseudomonas* and showed that other *Proteobacteria* (e.g., *Stenotrophomonas* and
459 *Novosphingobium*) were significantly more abundant on the GFD. However, these
460 differences were only observed in duodenum samples, and there were no differences
461 regarding these bacterial groups in faecal samples [56].

462 Ercolini et al. (2016) studied the influence of a change on GFD over the microbiome of
 463 celiac children following an African- style GFD during two years before the 60 days of
 464 treatment with an Italian-style gluten-free diet. After Italian-style GFD treatment, the
 465 microbiome underwent a time-dependent reduction in alpha diversity compared to
 466 baseline samples on African-style GFD. Many significant differences in microbiome
 467 composition were found. Regardless of changes in abundance, the core microbiome was
 468 conserved in children and is composed of fifteen genera belonging to the phylum
 469 *Firmicutes*. The metagenome prediction confirmed the diet's remarkable effects at a
 470 possible functional level. At baseline, there was a higher abundance of genes related to
 471 carbohydrate metabolism. In contrast, after 60 days of diet change, the saliva was
 472 characterised by a higher abundance of genes related to amino acid metabolism,
 473 vitamins and cofactors. The change in GFD diet style alters the Saharawi children's
 474 salivary microbiome significantly, reducing its alfa diversity, changing their beta
 475 diversity and varying the abundance of microbiome members and functional pathways
 476 [60], suggesting that diet composition is a critical factor in evaluating the effect of GFD
 477 over CeD patients microbiome.

478 D'Argenio. et al. (2016) compared the microbiome of CeD patients with unrestricted
 479 diet, under a GFD, and healthy controls. They found that *Actinobacteria* and *Firmicutes*
 480 phyla were less abundant in the CeD on gluten-containing diet than in the other two
 481 groups. *Betaproteobacteria* class was highly represented in active CeD patients' gut
 482 microbiome, although its levels did not directly correlate with disease severity.
 483 *Neisseriales* order, the *Neisseriaceae* family, and the *Neisseria* genus were significantly
 484 more abundant in active CeD patients than in the other study groups. Bacterial richness
 485 did not differ between CeD and control. The Beta diversity of bacterial communities
 486 was statistically significant within the three groups, suggesting that CeD patients'

487 microbiomes either on GFD or unrestricted diet are more similar to each other than
488 either of them to the control microbiomes. Authors conclude that overall microbial
489 community composition is different among CeD and controls, although richness is
490 comparable [61].

491 In a similar study using samples from duodenum, Nistal. et al. (2016) did not found
492 significant differences in alpha or beta diversity among CeD and controls; however,
493 *Asteuralleaceae* and *Streptococcus* members were frequent in CeD patients, mainly on a
494 gluten-containing diet. The authors did not find dysbiosis in the microbiota composition
495 in the duodenum. They suggested that the microbiome's role in the disease must be
496 associated with the function and not merely the composition [62].

497 In conclusion, there is no consensus in the literature regarding the adherence of CeD
498 patients to a GFD. Some authors mention that inter-individual effect on the microbiome
499 is greater than the effect of diet, whereas some studies report partial changes in the
500 diversity and microbial composition after acquiring this type of diet.

501 **3.1.4 Upper gastrointestinal tract microbiome in CeD.**

502 CeD mainly affects the small intestine (duodenum), its diagnosis often comprises a
503 combination of clinical, serological and histopathological findings, and small-bowel
504 biopsy specimens are fundamental for an accurate diagnosis [23]. However, biopsy-
505 sparing diagnostic guidelines have been proposed and validated in a few recent
506 prospective studies, as the obtention of a duodenal biopsy o duodenal content is an
507 invasive procedure, some studies aim to find microbial markers for other parts of the
508 upper gastrointestinal tract, with minor invasive procedures (saliva. and oropharyngeal
509 exudate) in order to identify possible microbial markers as "surrogate markers" of the

510 duodenum. We found three papers studying the upper gastrointestinal tract microbiome
511 in CeD.

512 In a study by Francavilla et al. (2014), the authors revealed that the composition of the
513 main bacterial phyla differed between the salivary microbiota of CeD children and
514 controls. *Lachnospiraceae*, *Gemellaceae* (genus *Gemella*), and *Streptococcus sanguinis*
515 were most abundant in CeD children's saliva. At the same time, the abundance of
516 *Streptococcus thermophiles* was markedly decreased. Other *Firmicutes* (e.g., *Veillonella*
517 *parvula*) associated with oral health were found at the highest levels in controls' saliva.
518 The saliva of CeD children harboured the highest levels of *Bacteroidetes* (e.g.,
519 *Porphyromonas spp.*, *P. endodontalis*, and *Prevotella nanceiensis*) and the lowest levels
520 of *Actinobacteria*. The authors conclude that the salivary microbiome of CeD children
521 is less diverse, with a different community structure of healthy children's microbiome.
522 CeD microbiome has increased commensal opportunistic pathogens and reduced
523 microbial species associated with health [63].

524 A similar study conducted by Tian et al. (2017) analysed salivary microbial profiles and
525 protease activities of the microbiome given their potential involvement in gluten
526 digestion and processing in CeD and refractory CeD (R-CeD). Significant differences
527 were found between the CeD and R-CeD groups with respect to *Bacteroidetes* (CeD >
528 R-CeD), *Actinobacteria* (CeD < R-CeD), and *Fusobacteria* (CeD >R-CeD) abundances.
529 The overall microbial diversity was greater in the healthy subjects than in the CeD
530 group, and controls had a higher abundance of TM7 sp., *Treponema sp.*, *Simonsiella*
531 *muelleri*, *Actinomyces sp.*, *Porphyromonas sp* *Alloscardovia omnicolens* in comparison
532 to CeD and R-CeD. The authors studied gluten degradation rates and found that gluten
533 substrate degradation in saliva was relatively low in healthy controls. However, gluten

534 degradation rates were significantly higher in saliva from CeD patients, regardless of
535 whether the tripeptide substrate or the 33-mer substrates were employed [64].

536 Finally, Iaffaldano et al. (2018) found that the oropharynx of CeD patients showed a
537 microbial imbalance similar to that previously described in the CeD duodenal
538 microbiome undergoing an unrestricted diet. They observed an almost coincident
539 oropharyngeal microbiome in controls and GFD subjects. Genes associated with
540 polysaccharide metabolism predominated in control and CeD-GFD microbiomes. In
541 contrast, in CeD patients with an unrestricted diet, the more significant metabolic
542 potential for degradation of amino acids, metabolism of lipid and ketone bodies and
543 microbial antioxidant defence mechanisms was observed. Finally, the authors suggest a
544 continuity of CeD microbial composition from mouth to the duodenum, proposing that
545 the characterisation of the microbiome in the oropharynx could contribute to
546 investigating the role of dysbiosis in CeD pathogenesis [65].

547 Upper intestinal microbiota analysis revealed differences across CeD patients and
548 healthy controls, suggesting the possible use of less invasive samples as a marker. There
549 is a need to perform studies integrating a significant number of samples to determine
550 universal markers related to CeD.

551 **3.1.5 Treatments: dietary interventions, prebiotics and hookworm infections**

552 Due to the implications of the microbiome in CeD onset and progress, dietary
553 modifications aiming to restore microbial eubiosis are of interest. In this sense,
554 probiotics intake and different approaches such as hookworm infections that modify the
555 microbiome and activate the immune response have been proposed as useful tools for
556 modifying gut microbiota in CeD patients. We discuss here two papers evaluating both
557 approaches.

558 Quagliariello et al. (2016) evaluate the probiotic effect of two *Bifidobacterium breve*
 559 strains on the gut microbiota composition of CeD under a GFD.
 560 *Firmicutes/Bacteroidetes* (F/B) ratio was calculated for each group of subjects; this ratio
 561 expresses the relationship between the two dominant phyla found in the gut microbiota
 562 and associated with several pathological conditions [66]. CeD subjects had F/B ratio
 563 values lower than the control group. However, the administration of the probiotic strains
 564 for three months showed to increase this ratio. The authors found that *Firmicutes* were
 565 significantly lower in CeD subjects not receiving the probiotic formulation than the
 566 control and probiotic groups. Similar results were found for *Actinobacteria*,
 567 underrepresented in the CeD group, and increased after the administration of probiotic
 568 strains. The treatment with the *B. breve* strains not caused changes at the levels of genus
 569 or phylum to which the probiotic belongs, but the intake of the probiotic strains acted as
 570 a "trigger" for the increase of *Firmicutes* and restoring F/B ratio. The authors concluded
 571 that three months of administration of *B. breve* strains could make the intestinal
 572 microbiota of CeD patients more similar to that of healthy individuals, restoring the
 573 abundance of some microbial communities that characterise the typical physiological
 574 condition [67].

575 In another study, Giacomini et al. (2016) investigated alterations in duodenal microbiota
 576 of CeD subjects before and following experimental infection with *Necator americanus*
 577 and following the administration of a low and high dose of dietary gluten. Considering
 578 that experimental infections with the human hookworm *N. americanus* could be used to
 579 treat CeD; infection alone increases anti-inflammatory cytokine, hinders colonisation of
 580 *Proteobacteria*, and favours the enrichment of *Clostridiales* in patients on GFD to
 581 improve tolerance to gluten micro-challenges [68]. They revealed that diversity in trial
 582 subjects was higher than in control subjects. Moreover, they observed significant

583 microbial diversity changes throughout the trial: at week 24, a significant increase in
584 diversity compared to baseline samples, whereas at week 36, diversity returned to
585 baseline. These results suggest a correlation between infection status and exposure to
586 escalating gluten doses and the gut microbiota composition. Finally, the authors
587 conclude that low doses but not high doses of gluten in hookworm infected CeD
588 patients previously in long-term GFD have beneficial effects in duodenal microbiome
589 diversity (increase). However, the study does not differentiate between gluten and
590 infection effects, do not measure inflammatory markers or mention symptoms and do
591 not go in-depth on analysing individual taxa abundance changes [69].

592 **3.2 Integrative data analysis of publicly available CeD microbiome data sets.**

593 The limited number of cases present in studies exploring the effects of microbiota over
594 CeD onset, progression, symptoms development and diagnosis, impulses the need for
595 comparison among them. However, the use of different sequencing methodologies and
596 DNA extraction procedures difficult the comparison among results; moreover, the lack
597 of standardisation methodologies for downstream analyses of sequencing introduces
598 statistical biases and challenges for reproducibility and cross-study comparisons [34].

599 We combine 16S RNA sequencing gene datasets for the first time and perform an
600 analysis following the same pipeline. We compared sequences generated from different
601 regions of the 16S rRNA gene by using a reference mapping protocol for ASV
602 assignment, in which sequences from different regions of the 16S rRNA gene will map
603 to the same full-length reference sequence from SILVA SSU v.138 database [43] if they
604 are from the same species. Achieving an integrative analysis including a high number of
605 samples from different body sites and extensive metadata to find microbial biomarkers
606 characteristic of CeD taking into account the type of diet.

We found nine out of the nineteen selected studies meeting inclusion criteria for the merged data analysis (Table 1). Table 2 shows the number of samples for each study, the clinical classification and tissue of origin (stool, duodenum, pharynx or saliva) for the data included in the analysis. Finally, we included 435 total samples, comprising 190 patients with active or treated CeD and 245 controls.

3.2.1 Diversity and microbial composition.

Similar to previously discussed studies in section 2, we did not find differences among healthy controls and CeD patients regarding alpha diversity indexes. However, when we consider GFD (for duodenum and stool samples), we found an increment in the diversity in patients as in healthy controls undergoing a GFD. Alpha diversity of the microbiome was estimated using the Chao1, Shannon and Simpson indices (Figure 2).

For the beta-diversity analysis, we study differences among groups in each tissue sampled by PCoA using weighted UniFrac distance, but we did not find a clear separation when analysing biological variables (age, type of diet or clinical condition), indicating that there are no differences in beta diversity between the microbial communities studied and that the differences in the microbiome are not due to global differences in the abundance of the phylogenetic groups present, but probably due to the abundance of specific taxa.

Our results are in the same line as others that assure that differences in the microbiota across different body sites are enough for grouping samples [72]. Permutational multivariate analysis of variance (PERMANOVA) for each variable by ADONIS function revealed that Study Accession, in the case of the duodenum, saliva, and stool, was a factor influencing the grouping of samples. This may be explained by the experimental protocols used in each study, including differences in the sequencing

platform, the region of 16S rRNA targeted, and the DNA extraction technique used, suggesting that some particular protocols may induce some biases, however, the principal coordinates explain a low percentage of the variance between the samples, indicating that other non-technical factors are mainly responsible for the variance.

3.2.2 Differential analysis of the microbiota, correlation and biomarker finding.

Celiac disease (CeD) mainly affects the small intestine (duodenum); according to the European Society for the Study of Coeliac Disease (ESsCD) guidelines, the diagnosis of CeD relies on a combination of clinical, serological and histopathological findings and small-bowel biopsy specimens are fundamental for an accurate diagnosis [23]. In this regard, we study duodenal microbiota intending to find microbial markers associated with this disease. On the other hand, biopsy-sparing diagnostic guidelines have been proposed and validated in a few recent prospective studies, as the obtention of a duodenal biopsy or duodenal content is an invasive procedure, we aim to find microbial markers for other parts of the gastrointestinal tract, with less invasive procedures (saliva, stool, oropharyngeal exudate) in order to identify possible microbial markers as "surrogate markers" of the duodenum.

To establish microbial biomarkers, first, we conduct a biomarker finding analysis using the LEfSe tool, followed by a differential abundance analysis using DESeq2 to identify ASVs that were differentially expressed according to studied groups. Finally, a correlation analysis looking for an association between CeD and microbial composition was performed. Figure 3 summarise the main findings obtained in each tissue analysed.

3.2.2.1 Microbial changes associated with duodenal microbiota in CeD.

654 We found that bacteria of the phylum *Proteobacteria* were characteristic of CeD
655 patients duodenum, with different genus present according to the type of diet.

656 **(i) Microbial changes of duodenal microbiota from untreated CeD patients**

657 **undergoing an unrestricted diet.** For duodenal samples of patients undergoing an
658 unrestricted diet, we found 23 ASV with an LDA score greater than 3 (Figure 4a). Nine
659 ASV were associated with CeD patients mainly from *Proteobacteria* phylum,
660 particularly bacteria from the *Burkholderia-Paraburkholderia-Caballeronia* clade,
661 *Alphaproteobacteria*, and *Enterobacteria*, and *Actinobacteria* from the family
662 *Corynebacteriaceae*. On the other hand, healthy controls were enriched in *Firmicutes*
663 phylum class *Negativicutes*, and phylum *Epsilonbacteraeota*.

664 After differential relative abundance analysis, we found 93 ASVs with significant
665 changes in abundance (FDR <0.01) (Figure 5a). Among them, 61 were decreased in
666 CeD, and 11 were increased in CeD. Significant ASVs belong to phyla: *Firmicutes*,
667 *Bacteroidetes*, *Proteobacteria*, *Epsilonbacteraeota*, *Actinobacteria*, *Spirochaetes*,
668 *Fusobacteria*, *Synergistetes*.

669 Finally, we found a negative association between CeD and the genus *Actinomyces*,
670 whereas the top five positive associated genus were *Coprococcus* 3,
671 *Hydrocarboniphaga*, *Ruminococcaceae* UCG_010, *Cutibacterium*, and *Deinococcus*.

672 **(ii) Microbial changes of duodenal microbiota from CeD patients undergoing a**

673 **GFD.** We select healthy controls and patients undergoing a GFD to study the microbial
674 composition between the two groups. Biomarker finding analysis revealed 65 ASVs
675 with a LDA score greater than 3 (Figure 6a). Eight ASV were associated with CeD
676 patients, mainly from *Proteobacteria* phylum belonging to *Burkholderia-*
677 *Paraburkholderia-Caballeronia* clade, and *Alphaproteobacteria* *Afipia*, and order

678 *Rhizobiales*. On the other hand, 57 ASV were related with healthy controls following a
679 GFD, particularly from phylum *Firmicutes*, *Fusobacterium*, *Actinobacteria*, and
680 *Epsilonbacteraeota* comprising the genus *Leptotrichia*, *Fusobacterium*, *Rothia*, and
681 *Campylobacter*, and other *Proteobacteria* (*Neisseriaceae*, *Pseudomonales* and
682 *Haemophilus*).

683 After performing the differential abundance analysis, we found 49 ASVs with
684 significant changes in abundance (FDR <0.01) (Figure 5b). Nine were increased in
685 CeD, whereas the other 40 ASV were decreased. Significant ASVs belong to phyla:
686 *Actinobacteria*, *Proteobacteria*, *Firmicutes*, *Epsilonbacteraeota*, *Bacteroidetes*,
687 *Fusobacteria*, and *Spirochaetes*.

688 Finally, the top five genus negative associated with CeD duodenum of patients
689 undergoing GFD were *Haemophilus*, *Neisseria*, *Alloprevotella*, *Fusobacterium* and
690 *Delftia*. In contrast, the top five positive associated genus were mainly from
691 *Proteobacteria*, *Alfaproteobacteriales*: *Falsirhodobacter*, *Asinibacterium*, *Azonexus*,
692 and *Blastomonas*.

693 Higher rates of *Proteobacteria* are linked to dysbiosis and have been suggested to
694 increase low-grade inflammation, affecting tight-junction and leading to leaky gut [70].
695 This phylum was overrepresented in duodenal samples from CeD patients particularly
696 the *Burkholderia-Paraburkholderia-Caballeronia* clade. It is worth mentioning that
697 *Enterobacteriaceae* was overrepresented in the duodenum of untreated patients, but
698 were not observed in patients undergoing a GFD, in fact some genus previously
699 associated with active CeD i.e *Haemophilus*, *Neisseria*, and bacteria from the class
700 *Pseudomonadales* was also diminished in patients undergoing GFD. These changes
701 suggest beneficial variations in the microbial composition of the duodenum after

adherence to GFD. However, these changes were not able to fully restore the microbiota, evidenced by the decrease of SCFA producers such as *Ruminococcaceae*, and *Prevotella*, and by the overrepresentation of *Betaproteobacteriales*.

3.2.2.3 Microbial changes associated with stool microbiota in CeD.

Stool samples are more representative of the colonic microbiota; however, they could also indicate changes related to the disease. Like in duodenum samples, we found an increase in bacteria of the phylum *Proteobacteria*, but different genus was enriched according to the type of diet. Specific changes found were as follows:

(i) Microbial changes of stool microbiota from untreated CeD patients undergoing an unrestricted diet. After biomarker analysis, we found 60 ASVs with an LDA score greater than 3 (Figure 6a). Of them, 17 were associated with CeD, bacteria from phylum *Verrucomicrobia* and *Firmicutes* were characteristics of this group, mainly the genus *Akkermansia*, *Anaerostipes*, *Faecalibacteria* and *Dorea*, on the other hand in healthy controls undergoing an unrestricted diet ASV from phylum *Proteobacteria* (*Betaproteobacteriales*) and *Firmicutes* mainly order *Clostridiales* were overrepresented.

Differential abundance analysis revealed 74 ASVs belonging to phyla: *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria* with significant changes in abundance (FDR <0.01) (Figure 7a). Among them, 56 were decreased in CeD and seven were increased in CeD. Finally, top five genus positively associated with CeD were: *Achromobacter*, *Flavisolibacter*, *Geodermatophilus*, *Candidatus Rubidus*, and *Tepidimonas*.

(ii) Microbial changes of stool microbiota from CeD patients undergoing a GFD.

Biomarker finding analysis revealed the presence of 48 ASVs with LDA > 3 (Figure

6b). Twenty-two were associated with CeD patients, mainly from phylum *Verrucomicrobia*, *Bacteroides*, *Firmicutes* and *Proteobacteria*, particularly genus *Akkermansia*, *Bacteroides*, *Romboutsia*, and *Pseudomonas*. On the other hand, 26 ASV were related with controls undergoing GFD, mainly bacteria from phylum *Firmicutes* and *Actinobacteria* from genus *Lactobacillus*, *Streptococcus*, *Ruminococcus* and *Bifidobacterium*. Differential abundance analysis revealed changes in 77 ASVs (FDR <0.01) (Figure 7b). Among them, 41 were decreased in CeD and 19 were increased in CeD. Significant ASVs belong to phyla: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Verrucomicrobia*, *Actinobacteria*, *Lentisphaerae*.

Finally, correlation analysis showed that 15 ASV with a significant correlation with CeD. 5 have a negative correlation with CeD (*Ruminococcus* 1, *Ruminococcaceae* UCG_014, *Ruminiclostridium* 6, *Coprococcus* 2, *Enterococcus*) and ten ASV have a positive correlation, and the top five genus positive correlated were *Intestinibacter*, *Akkermansia*, *UBA1819*, *Flavonifractor*, and *Terrisporobacter*.

It is worth mentioning that *Betaproteobacteriales* from family *Burkholderiaceae* were overrepresented in the duodenum of CeD patients. However, we found that these bacteria reduce its abundance in stool and diminished notably after treatment with GFD. Also, we found an enrichment in *Akkermansia*, a bacteria associated with anti-inflammatory properties and metabolic health [71]. Despite the beneficial changes observed, stool microbiota of patients undergoing GFD presented a decrease in *Lactobacillus*, and *Bifidobacterium*, SCFA producers bacteria, which also digests intact gluten proteins reducing cytotoxicity and inflammatory responses [15, 72], and the presence of bacteria associated to CeD such as *Bacteroides*, and *Pseudomonas* both genus with species containing metalloproteases able to partially digest gluten,

750 increasing the immunogenic effect of gliadins inducing more severe inflammation [73,
751 74].

752 We found some similitudes in stool microbiota compared to the duodenum, first an
753 increase in *Proteobacteria*, the presence of bacteria associated with CeD (*Bacteroides*,
754 *Neisseria*, or *Pseudomonas*) and a decrease in beneficial SCFA-producers bacteria.
755 These findings suggest that the lower gastrointestinal tract microbiota from CeD
756 patients is enriched in *Proteobacteria* with potential implications in activation of
757 inflammation and immune response and a decrease in beneficial SCFA producers with
758 anti-inflammatory potential, suggesting that gut microbiota favours the inflammatory
759 state of the disease.

760 **3.2.2.4 Microbial changes associated with pharynx microbiota in CeD.**

761 Anatomically, the pharynx is part of the upper gastrointestinal tract, directly connected
762 with the oesophagus, and is conventionally divided into the nasopharynx, oropharynx,
763 and hypopharynx. Usually, oropharynx exudate, a sample comprising the part of the
764 throat at the back of the mouth behind the oral cavity, are used for clinical diagnosis of
765 microbial infections [75], being a valuable method for studying microbiota from the
766 upper intestinal tract with a less invasive procedure than a biopsy [76, 77].

767 We found two microbial biomarkers characteristic of the pharynx from patients with
768 CeD. *Rothia*, a nitrate-reducing bacteria usually found in the oral cavity of humans [78],
769 and *Peptostreptococcus*, an oral pathogen recently associated with Colorectal Cancer
770 development [79] (Figure 8a). On the other hand, differential abundance analysis
771 showed significant changes in abundance (FDR <0.01) of five ASV. *Veillonella*,
772 *Mogibacterium* (*p_Firmicutes*), *Streptobacillus* (*p_Fusobacteria*) and *Mannheimia*

773 (*p-Proteobacteria*) where increased in CeD whereas *Treponema* 2 (*p_Spirochaetes*)
774 was decreased.

775 These changes revealed a particular microbial composition from the pharynx of CeD
776 patients, mainly associated with proinflammatory bacteria (i.e. *Mogibacterium*) and
777 opportunistic pathogens (i.e. *Peptostreptococcus*, *Streptobacillus*). Although some
778 *Rothia* species have gluten-degrading enzymes [14], the role of this bacteria in CeD
779 pathogenesis remains unexplored. In our analysis, we found it on non-symptomatic
780 patients undergoing GFD. Further studies will be helpful to discriminate its role in CeD
781 and its possible use for monitoring the disease.

782 **3.2.2.5 Microbial changes associated with microbiota from saliva samples in CeD.**

783 The use of saliva samples would be beneficial for diagnosis and monitoring the disease,
784 mainly in children, due to the non-invasiveness of the sample collection procedure.
785 Despite the most significant changes found in the microbial composition on other parts
786 of the gastrointestinal tract, we only found one genus related to CeD in saliva and was
787 identified after differential abundance analysis and biomarker finding, namely genus
788 *Oceanivirga* belonging to the phylum *Fusobacteria*, family *Leptotrichiaceae*.

789 *Leptotrichiaceae* are relatively poor studied gram-negative bacteria, facultative or
790 obligate anaerobes, found as colonisers of mucous membranes in the oral cavity of
791 humans and other animals [80]. Particularly, genus *Oceanivirga* has been isolated in
792 subgingival samples from patients with periodontitis [81]. We found this genus as a
793 marker of CeD patients for the first time; validation as a microbial marker of CeD need
794 to be performed.

795 **3.3 Prediction of the metabolic functions profiles in bacterial communities.**

796 The functional potential of the microbiome in the different tissues was predicted using
797 the PICRUSt tool, followed by a differential abundance analysis performed with
798 Deseq2 and correlation analysis.

799 **3.3.1 Prediction of short-chain fatty acid (SCFA) production genes and** 800 **inflammation-related pathways.**

801 To study the metabolic potential of the microbiome, the analysis was first carried out for
802 all the genes and pathways detected and, since the intestinal microbiome, after its
803 contribution to the digestive process of food, produces short-chain fatty acids that
804 influence the maturation, maintenance, and behaviour of the mucosal immune system,
805 we focus on the genes and pathways involved in the production of these compounds.
806 Figure 3 summarises the main findings observed according to each tissue analysed.

807 First, we search for the 26 genes reported as potentially involved in the biosynthesis
808 pathways of the main SCFAs produced by the human microbiome (butyrate, acetate,
809 and lactate), given their involvement in the modulation of the immune system. Genes
810 included were those identified by Zhao et al. (2019) by gene deletion and
811 overexpression experiments in *E. coli*. These authors identified six genes needed in
812 acetate production (*menI*, *tesA*, *yciA*, *fadM*, *tesB*, *ybgC*), eight genes used in butyrate
813 (*entH*, *tesA*, *ybgC*, *ybhC*, *yicA*, *menI*, *yigI* y *tesB*) production, and two genes required for
814 lactate (*mgsA* y *lldD*) production in addition to the previously known genes *pta-ackA*,
815 *ptb-buk*, *ldhA*, *poxB*, *eutD*, *tdcD*, *dld* y *ykgF* [82]. Table 2 shows the number of genes
816 found according to each tissue and the differentially abundant genes/routes between
817 CeD patients and healthy controls.

818 Differential abundance analysis of metabolic pathways is summarised in Table S3, it
819 revealed an increase in the duodenum of CeD patients, of the degradation of D-

glucarate, L-arabinose, D-galactarate and biogenic amines and a decrease in the degradation of lactose and galactose. Since the cells involved in the immune response require a large amount of energy for their activation, proliferation and recruitment of other cells during inflammation, the availability of carbohydrates is scarce for bacteria to synthesise their components. Therefore, they obtain energy by alternative routes, for example, from acetate through the glyoxalate cycle, or obtaining energy are compounds such as D-glucarate, D-galactarate and biogenic amines [83]. Regarding genes involved in the production of SCFA, we found a decrease in the abundance of genes involved in the production of acetate (*ackA*) and lactate (*ldhA*) and a negative correlation with genes involved in the production of acetate (*pta*), lactate (*lldD*, *ldh*, *dld* and *tesB*) and butyrate (*tesB*). A decrease in the abundance of the fermentation pathway from hexitol to SCFAs involving the *ackA* and *pta* genes was consistently found.

In stool samples of patients with CeD, the synthesis routes of lipopolysaccharide (LPS) components of the membrane of gram-negative bacteria, acetate degradation routes, and vitamin B synthesis (B1 and B9) were found to be increased, which is consistent with our findings in the taxonomic analysis revealing an increase in the abundance of *Proteobacteria*. LPS are key factors in the activation of the immune response through TLR4 signalling, contributing to inflammation and loss of intestinal permeability seen in CeD [1, 70]. Regarding genes implied in SCFA synthesis, correlation analysis revealed that two genes related to acetate and lactate production (*pta* and *dld*) were negatively associated with CeD ($FDR \leq 0.01$; correlation ≥ 0.2).

We found a generalised low presence of SCFA in microbiota of duodenum and stool from patients with CeD. The reduction of these compounds is associated with a reduction in bacterial genus that produces them, such as *Ruminococcus*, *Veillonella*, and *Clostridiales* (as observed in taxonomic analysis) and potentially with a lower fibre

845 intake typical of a gluten-free diet [25]. Low abundance of SCFA has been previously
846 reported in CeD at an early age [84], and may be involved in the predisposition to the
847 development of CeD [85].

848 No statistically significant differences were found in the abundance of any of the 26
849 genes related to SCFA production regarding saliva and pharynx samples. However, in
850 saliva samples of CeD patients, peptidoglycan synthesis was increased. Peptidoglycans
851 are part of the cellular wall of bacteria and, besides LPS, are responsible for microbial-
852 related inflammation mediated by innate immune response [86]. On the other hand, we
853 found an increase in the intermediate degradation routes of aromatic compounds and
854 amino acids in the pharynx samples, both routes associated with energy acquisition
855 from bacteria [83].

856 **3.3.2 Prediction of genes and genes coding for prolyl peptidase enzymes.**

857 Some components of the microbiota can express enzymes different from those produced
858 by humans and promote the digestion of compounds such as the immunogenic peptides
859 of gluten [14, 15]. We analysed four enzymes involved in the degradation of
860 immunogenic gluten peptides that could be involved in reducing CeD symptoms
861 (general N-type aminopeptidase (PepN), X-prolyl dipeptidyl aminopeptidase (PepX),
862 endopeptidase (PepO) and endoprollyl peptidase (PREP)). However, none of the four
863 enzymes involved in the degradation of gliadin peptides were found to be differentially
864 abundant between cases and controls or associated with CeD in any of the tissues.

865 **4. Conclusions and future perspectives.**

866 Our study provides an unprecedented metanalysis of metataxonomic data from CeD
867 patients, including (n=435) samples from CeD patients and healthy controls for

different body sites. One of the most critical limitations currently found in literature is the lack of standardised methodologies for downstream analyses of sequencing approaches, introducing statistical biases and subsequent challenges for reproducibility and cross-study comparisons [34]. Despite some attempts to standardise methods, a gold standard of microbiome research is not established [87]. Herein, we create an extensive dataset processed following the same methodology and leveraging metadata to solve this issue; moreover, we made data publicly available (https://figshare.com/projects/An_lisis_del_microbioma_en_Enfermedad_Cel_aca/82547).

We were able to find microbial markers related to CeD at different body sites, including SCFA production and the prevalence of inflammatory pathways (Figure 3). Our results showed coordinated changes throughout the gastrointestinal tract, with specific changes according to each body site. In the upper gastrointestinal tract, we found an enrichment in *Actinobacteria* from genus *Rothia* in pharynx, and *Cutibacterium* in duodenum, and a marked decrease in *Alloprevotella* spp. in both sites. Whereas the lower gastrointestinal tract presents more changes characterised by an increase in *Proteobacteria* and a decrease in *Actinobacteria*, *Campylobacter* and SCFA producers such as *Ruminococaceae*, and *Clostridiales*. Moreover, we define some differences in gut microbiota (from duodenum and stool samples) of untreated CeD patients following an unrestricted diet vs CeD patients following GFD. Although we could not find restoration in microbial dysbiosis, we found that patients following a GFD have an overall higher microbial diversity and an abundance of certain bacteria usually related to health benefits such as *Akkermasia* [71], and a decrease in the presence on *Enterobacteriaceae*. This finding is interesting because differences in patients following GFD may be observed. Recent studies demonstrate that long-term diet is the primary

893 driver of gut microbiota composition [65], so it is expected that following a GFD
 894 characterised by changes in protein, fat, and carbohydrate consumption [24] will have
 895 an impact on the microbial communities composition. In fact, adherence to a GFD leads
 896 to the healing of the duodenal mucosa and the resolution of symptoms and signs of
 897 malabsorption CeD [23]. However, up to 30% of patients with CeD may show signs,
 898 symptoms or persistent enteropathy after one year on a GFD, having the so-called non-
 899 responsive CeD [26]. Non-responsive CeD has been poorly studied, and still, there is no
 900 clear cause-effect association to explain this entity; microbiota may play an essential
 901 role in its pathogenesis.

902 Previous studies have shown that *Proteobacteria* was enriched in CeD [54–58], we
 903 recognise the same finding, but we also were able to discern between differences in the
 904 abundance of particular *Proteobacteria* families according to the type of diet of the
 905 patients. For untreated CeD patients, *Burkholderiaceae*, *Xantobacteriaceae*, and
 906 *Enterobacteriaceae* were enriched, whereas in patients under a GFD we found a
 907 decrease in some *Enterobacteriaceae* (duodenum), and *Betaproteobacteriales* (stool).
 908 *Proteobacteria* may be acting as enhancers of the immune response in CeD due to the
 909 activation of the innate immune system via TLR receptor mediated by LPS [2, 8, 20].

910 Future prospective studies will provide the "solution" to who was before the chicken or
 911 the egg? i.e., Dysbiosis led the disease or the disease produce dysbiosis?, Current
 912 studies have a cross-sectional design and perform descriptive association at a snapshot
 913 of time, CeD is the perfect scenario for studying implications of microbiota over the
 914 pathophysiology of a disease, the trigger of the pathology (*i.e.*: gluten) is traceable, and
 915 the genetic environment predisposing to the disease is also knowledgeable.

916 Finally, there is a need to generate larger datasets and properly apply machine learning
 917 (ML) that would help to generate helpful and more universally applicable results [88].
 918 In this sense, future studies applying ML would be beneficial to corroborate the
 919 microbial and functional markers found in this study, and know the role in the
 920 prevalence of symptoms associated with CeD. Moreover, evaluating the usefulness of
 921 stool, oropharyngeal exudates, and saliva as surrogate markers of the microbial state of
 922 the duodenum for the diagnosis and management of CeD patients would be necessary.
 923 Finally, the potential efficacy of gut microbiota modulators such as probiotics and
 924 prebiotics as adjuvant treatment for CeD need to be further assessed.

925

926 5. References

- 927 1. Levy M, Blacher E, Elinav E. Microbiome, metabolites and host immunity. *Curr*
928 *Opin Microbiol.* 2017;35:8–15. doi: 10.1016/j.mib.2016.10.003.
- 929 2. Valitutti F, Cucchiara S, Fasano A. Celiac Disease and the Microbiome. *Nutrients.*
930 2019;11. doi: 10.3390/nu11102403.
- 931 3. Dieli-Crimi R, Cénit MC, Núñez C. The genetics of celiac disease: A comprehensive
932 review of clinical implications. *J. Autoimmun.* 2015;64:26–41. doi:
933 10.1016/j.jaut.2015.07.003.
- 934 4. Sollid LM, Lundin KEA. Celiac disease. In: *The Autoimmune Diseases.* Elsevier;
935 2019. p. 849–69. doi: 10.1016/B978-0-12-812102-3.00045-2.
- 936 5. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell.*
937 2014;157:121–41. doi:10.1016/j.cell.2014.03.011.
- 938 6. Cénit MC, Codoñer-Franch P, Sanz Y. Gut Microbiota and Risk of Developing
939 Celiac Disease. *J. Clin. Gastroenterol.* 2016;50:S148–52. doi:
940 10.1097/MCG.0000000000000688.
- 941 7. Krishnareddy S. The Microbiome in Celiac Disease. *Gastroenterol. Clin. North Am.*
942 2019;48:115–26. doi: 10.1016/j.gtc.2018.09.008.
- 943 8. Chander AM, Yadav H, Jain S, Bhadada SK, Dhawan DK. Cross-Talk Between
944 Gluten, Intestinal Microbiota and Intestinal Mucosa in Celiac Disease: Recent
945 Advances and Basis of Autoimmunity. *Front. Microbiol.* 2018;9:2597. doi:
946 10.3389/fmicb.2018.02597.
- 947 9. Jabri B, Sollid LM. T Cells in Celiac Disease. *J Immunol.* 2017;198:3005–14. doi:
948 10.4049/jimmunol.1601693.
- 949 10. Naka T, Nishimoto N, Kishimoto T. The paradigm of IL-6: from basic science to
950 medicine. *Arthritis Res.* 2002;4:S233–42. doi: 10.1186/ar565.
- 951 11. Romaldini C, Barbieri D, Okay TS, Raiz RJr, Cançado ELR. Serum soluble
952 interleukin-2 receptor, interleukin-6, and tumor necrosis factor- α levels in children with
953 celiac disease: response to treatment. *J. Pediatr Gastroenterol Nutr.* 2002;35:513–7.
- 954 12. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive
955 promoter-centered chromatin interactions provide a topological basis for transcription
956 regulation. *Cell.* 2012;148:84–98. doi: 10.1016/j.cell.2011.12.014.
- 957 13. Tye-Din JA, Stewart JA, Dromey JA, Beissbarth T, David †, Van Heel A, et al.
958 Comprehensive, Quantitative Mapping of T Cell Epitopes in Gluten in Celiac Disease.
959 *Sci Transl Med.* 2010 Jul 21;2(41):41ra51. doi: 10.1126/scitranslmed.3001012.
- 960 14. Kõiv V, Tenson T. Gluten-degrading bacteria: availability and applications. *Appl.*
961 *Microbiol. Biotechnol.* 2021;105:3045–59. doi: 10.1007/s00253-021-11263-5.
- 962 15. Herrán AR, Pérez-Andrés J, Caminero A, Nistal E, Vivas S, Ruiz de Morales JM, et
963 al. Gluten-degrading bacteria are present in the human small intestine of healthy

volunteers and celiac patients. *Res. Microbiol.* 2017;168:673–84. doi: 10.1016/j.resmic.2017.04.008.

16. Molberg Ø, Mcadam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, et al. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat. Med.* 1998;4:713–7. doi: 10.1038/nm0698-713.

17. Koehler P, Wieser H, Konitzer K. Celiac Disease—A Complex Disorder. In: *Celiac Disease and Gluten*. Elsevier; 2014. p. 1–96. doi: 10.1016/b978-0-12-420220-7.00001-8.

18. de Palma G, Capilla A, Nadal I, Nova E, Pozo T, Varea V, et al. Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. *Curr. Issues Mol. Biol.* 2010;12:1–10.

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

20. Chibbar R, Dieleman LA. The Gut Microbiota in Celiac Disease and probiotics. *Nutrients.* 2019;11. doi: 10.3390/nu11102375.

21. Sanz Y. Microbiome and Gluten. *Ann. Nutr. Metab.* 2015;67 Suppl 2:28–41. doi: 10.1159/000440991.

22. Sjöberg V, Sandström O, Hedberg M, Hammarström S, Hernell O, Hammarström M-L. Intestinal T-cell responses in celiac disease - impact of celiac disease associated bacteria. *PloS one.* 2013;8:e53414. doi: 10.1371/journal.pone.0053414.

23. Al-Toma A, Volta U, Auricchio R, Castillejo G, Sanders DS, Cellier C, et al. European Society for the Study of Coeliac Disease (ESsCD) guideline for coeliac disease and other gluten-related disorders. *United Eur Gastroent.* 2019;7:583–613. doi: 10.1177/2050640619844125.

24. Melini V, Melini F. Gluten-Free Diet: Gaps and Needs for a Healthier Diet. *Nutrients.* 2019;11. doi: 10.3390/nu11010170.

25. Lerner A, O'Bryan T, Matthias T. Navigating the Gluten-Free Boom: The Dark Side of Gluten Free Diet. *Front. Pediatr.* 2019;7:414. doi: 10.3389/fped.2019.00414.

26. Leonard MM, Cureton P, Fasano A. Indications and Use of the Gluten Contamination Elimination Diet for Patients with Non-Responsive Celiac Disease. *Nutrients.* 2017;9. doi: 10.3390/nu9101129.

27. Sánchez E, De Palma G, Capilla A, Nova E, Pozo T, Castillejo G, et al. Influence of environmental and genetic factors linked to celiac disease risk on infant gut colonization by *Bacteroides* species. *Appl Environ Microbiol.* 2011;77:5316–23. doi: 10.1128/AEM.00365-11.

28. Wacklin P, Kaukinen K, Tuovinen E, Collin P, Lindfors K, Partanen J, et al. The Duodenal Microbiota Composition of Adult Celiac Disease Patients Is Associated with

1003 the Clinical Manifestation of the Disease. *Inflamm. Bowel. Dis.* 2013;19:934–41. doi:
1004 10.1097/MIB.0b013e31828029a9.

1005 29. Schippa S, Iebba V, Barbato M, Di Nardo G, Totino V, Checchi MP, et al. A
1006 distinctive “microbial signature” in celiac pediatric patients. *BMC Microbiol.*
1007 2010;10:175. doi: 10.1186/1471-2180-10-175.

1008 30. Nistal E, Caminero A, Herrán AR, Arias L, Vivas S, Ruiz De Morales JM, et al.
1009 Differences of small intestinal bacteria populations in adults and children with/without
1010 celiac disease: Effect of age, gluten diet, and disease. *Inflamm. Bowel. Dis.*
1011 2012;18:649–56. doi: 10.1002/ibd.21830.

1012 31. Hedberg ME, Israelsson A, Moore ERB, Svensson-Stadler L, Wai SN, Pietz G, et
1013 al. *Prevotella jejuni* sp. nov., isolated from the small intestine of a child with coeliac
1014 disease. *Int J Syst Evol Microbiol.* 2013;63 PART 11:4218–23. doi:
1015 10.1099/ijs.0.052647-0.

1016 32. Sánchez E, Donat E, Ribes-Koninckx C, Fernández-Murga ML, Sanz Y. Duodenal-
1017 mucosal bacteria associated with celiac disease in children. *Appl Environ Microbiol.*
1018 2013;79:5472–9. doi: 10.1128/AEM.00869-13.

1019 33. Sellitto M, Bai G, Serena G, Fricke WF, Sturgeon C, Gajer P, et al. Proof of concept
1020 of microbiome-metabolome analysis and delayed gluten exposure on celiac disease
1021 autoimmunity in genetically at-risk infants. *PLoS ONE.* 2012;7. doi:
1022 10.1371/journal.pone.0033387.

1023 34. Moreno-Indias I, Lahti L, Nedyalkova M, Elbere I, Roshchupkin G, Adilovic M, et
1024 al. Statistical and Machine Learning Techniques in Human Microbiome Studies:
1025 Contemporary Challenges and Solutions. *Front. Microbiol.* 2021;12:635781 doi:
1026 10.3389/fmicb.2021.635781.

1027 35. Wooley JC, Ye Y. Metagenomics: Facts and artifacts, and computational
1028 challenges. *J. Comput. Sci. Technol.* 2010;25:71–81. doi: 10.1007/s11390-010-9306-4.

1029 36. Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from
1030 16S rRNA gene sequencing and shotgun metagenomics. *Pac Symp Biocomput.*
1031 2011;165–76. doi: 10.1142/9789814335058_0018.

1032 37. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics,
1033 from sampling to analysis. *Nat Biotechnol.* 2017;35:833–44. doi: 10.1038/nbt.3935.

1034 38. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *Int*
1035 *J Soc Res Methodol.* 2005;8:19–32. doi: 10.1080/1364557032000119616.

1036 39. Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010.

1037 40. Marcel Martin. Cutadapt removes adapter sequences from high-throughput
1038 sequencing reads. 2010. doi:10.14806/ej.17.1.200.

1039 41. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve
1040 genome assemblies. *Bioinformatics.* 2011;27:2957–63 10.1093/bioinformatics/btr507.

1041 42. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.
1042 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*.
1043 2016;13:581–3. doi: 10.1038/nmeth.3869.

1044 43. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, et al. 25 years of
1045 serving the community with ribosomal RNA gene reference databases and tools. *J.*
1046 *Biotechnol.* 2017;261:169–76. doi: 10.1016/j.jbiotec.2017.06.1198.

1047 44. Jari Oksanen, Roeland Kindt, Pierre Legendre, Bob O’Hara, M Henry H Stevens,
1048 Maintainer Jari Oksanen, et al. *Vegan: Community Ecology Package*. 2007.

1049 45. Villanueva RAM, Chen ZJ, Wickham H. *ggplot2: Elegant Graphics for Data*
1050 *Analysis Using the Grammar of Graphics*. 2016;160–7. doi:
1051 10.1080/15366367.2019.1565254.

1052 46. McMurdie PJ, Holmes S. *phyloseq: An R Package for Reproducible Interactive*
1053 *Analysis and Graphics of Microbiome Census Data*. *PLoS ONE*. 2013;8:e61217. doi:
1054 10.1371/journal.pone.0061217.

1055 47. R Core Team. *R: A language and environment for statistical computing*. R
1056 Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
1057 2013.

1058 48. van den Boogaart KG, Tolosana-Delgado R, Matevz B. R Package ‘compositions.’
1059 2021.

1060 49. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion
1061 for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. doi: 10.1186/s13059-014-
1062 0550-8.

1063 50. Douglas GM, Maffei VJ, Zaneveld J, Yurgel SN, Brown JR, Taylor CM, et al.
1064 PICRUSt2: An improved and extensible approach for metagenome inference. *BioRxiv*.
1065 2019 10.1101/672295.

1066 51. Olivares M, Walker AW, Capilla A, Benítez-Páez A, Palau F, Parkhill J, et al. Gut
1067 microbiota trajectory in early life may predict development of celiac disease.
1068 *Microbiome*. 2018;6. doi: 10.1186/s40168-018-0415-6.

1069 52. Rintala A, Riikonen I, Toivonen A, Pietilä S, Munukka E, Pursiheimo JP, et al.
1070 Early fecal microbiota composition in children who later develop celiac disease and
1071 associated autoimmunity. *Scand. J. Gastroenterol.* 2018;53:403–9. doi:
1072 10.1080/00365521.2018.1444788.

1073 53. Leonard MM, Karathia H, Pujolassos M, Troisi J, Valitutti F, Subramanian P, et al.
1074 Multi-omics analysis reveals the influence of genetic and environmental risk factors on
1075 developing gut microbiota in infants at risk of celiac disease. *Microbiome*. 2020;8:130
1076 doi: 10.1186/s40168-020-00906-w.

1077 54. Cheng J, Kalliomaki M, Heilig HG, Palva A, Lahteenoja H, de Vos WM, et al.
1078 Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease.
1079 *BMC Gastroenterol.* 2013;13:113. doi:10.1186/1471-230X-13-113.

1080 55. Pellegrini S, Sordi V, Bolla AM, Saita D, Ferrarese R, Canducci F, et al. Duodenal
1081 Mucosa of Patients With Type 1 Diabetes Shows Distinctive Inflammatory Profile and
1082 Microbiota. *J Clin Endocrinol Metab.* 2017;102:1468–77. doi:10.1210/jc.2016-3222.

1083 56. Garcia-Mazcorro JF, Rivera-Gutierrez X, Cobos-Quevedo ODJ, Grube-Pagola P,
1084 Meixueiro-Daza A, Hernandez-Flores K, et al. First Insights into the Gut Microbiota of
1085 Mexican Patients with Celiac Disease and Non-Celiac Gluten Sensitivity. *Nutrients.*
1086 2018;10. doi: 10.3390/nu10111641.

1087 57. Bodkhe R, Shetty SA, Dhotre DP, Verma AK, Bhatia K, Mishra A, et al.
1088 Comparison of Small Gut and Whole Gut Microbiota of First-Degree Relatives With
1089 Adult Celiac Disease Patients and Controls. *Front. Microbiol.* 2019;10:164. doi:
1090 10.3389/fmicb.2019.00164.

1091 58. Panelli S, Capelli E, Lupo GFD, Schieppati A, Betti E, Sauta E, et al. Comparative
1092 Study of Salivary, Duodenal, and Fecal Microbiota Composition Across Adult Celiac
1093 Disease. *J. Clin. Med.* 2020;9. doi: 10.3390/jcm9041109.

1094 59. Bonder MJ, Tigchelaar EF, Cai X, Trynka G, Cenit MC, Hrdlickova B, et al. The
1095 influence of a short-term gluten-free diet on the human gut microbiome. *Genome Med.*
1096 2016;8:45. doi: 10.1186/s13073-016-0295-y.

1097 60. Ercolini D, Francavilla R, Vannini L, Filippis F de, Capriati T, Cagno R di, et al.
1098 From an imbalance to a new imbalance: Italian-style gluten-free diet alters the salivary
1099 microbiota and metabolome of African celiac children. *Sci. Rep.* 2015;5:18571.
1100 doi:10.1038/srep18571.

1101 61. D’Argenio V, Casaburi G, Precone V, Pagliuca C, Colicchio R, Samataro D, et al.
1102 Metagenomics Reveals Dysbiosis and a Potentially Pathogenic *N. flavescens* Strain in
1103 Duodenum of Adult Celiac Patients. *Am J Gastroenterol.* 2016;111:879–90. doi:
1104 10.1038/ajg.2016.95.

1105 62. Nistal E, Caminero A, Herrán AR, Pérez-Andres J, Vivas S, Ruiz de Morales JM, et
1106 al. Study of duodenal bacterial communities by 16S rRNA gene analysis in adults with
1107 active celiac disease vs non-celiac disease controls. *J Appl Microbiol.* 2016;120:1691–
1108 700. doi: 10.1111/jam.13111.

1109 63. Francavilla R, Ercolini D, Piccolo M, Vannini L, Siragusa S, Filippis F de, et al.
1110 Salivary microbiota and metabolome associated with celiac disease. *Appl Environ*
1111 *Microbiol.* 2014;80:3416–25. doi:10.1128/AEM.00362-14.

1112 64. Tian N, Faller L, Leffler DA, Kelly CP, Hansen J, Bosch JA, et al. Salivary Gluten
1113 Degradation and Oral Microbial Profiles in Healthy Individuals and Celiac Disease
1114 Patients. *Appl Environ Microbiol.* 2017;83. doi: 10.1128/AEM.03330-16.

1115 65. Iaffaldano L, Granata I, Pagliuca C, Esposito MV, Casaburi G, Salerno G, et al.
1116 Oropharyngeal microbiome evaluation highlights *Neisseria* abundance in active celiac
1117 patients. *Sci. Rep.* 2018;8:11047. doi:10.1038/s41598-018-29443-1.

1118 66. Magne F, Gotteland M, Gauthier L, Zazueta A, Pessoa S, Navarrete P, et al. The
1119 *Firmicutes/Bacteroidetes* Ratio: A Relevant Marker of Gut Dysbiosis in Obese
1120 Patients? *Nutrients.* 2020;12. doi: 10.3390/nu12051474.

1121 67. Quagliariello A, Aloisio I, Cionci NB, Luiselli D, D'Auria G, Martinez-Priego L, et
1122 al. Effect of *Bifidobacterium breve* on the Intestinal Microbiota of Coeliac Children on
1123 a Gluten Free Diet: A Pilot Study. *Nutrients*. 2016;8. doi:10.3390/nu8100660.

1124 68. Giacomini P, Zakrzewski M, Croese J, Su X, Sotillo J, McCann L, et al.
1125 Experimental hookworm infection and escalating gluten challenges are associated with
1126 increased microbial richness in celiac subjects. *Sci. Rep.* 2015;5:13797. doi:
1127 10.1038/srep13797.

1128 69. Giacomini P, Zakrzewski M, Jenkins TP, Su X, Al-Hallaf R, Croese J, et al. Changes
1129 in duodenal tissue-associated microbiota following hookworm infection and
1130 consecutive gluten challenges in humans with coeliac disease. *Sci. Rep.* 2016;6:36797.
1131 doi:10.1038/srep36797.

1132 70. Salguero M v, Al-Obaide MAI, Singh R, Siepmann T, Vasylyeva TL. Dysbiosis of
1133 Gram-negative gut microbiota and the associated serum lipopolysaccharide exacerbates
1134 inflammation in type 2 diabetic patients with chronic kidney disease. *Exp. Ther. Med.*
1135 2019;18:3461–9. doi: 10.3892/etm.2019.7943.

1136 71. Derrien M, Belzer C, de Vos WM. *Akkermansia muciniphila* and its role in
1137 regulating host functions. *Microbial pathogenesis*. 2017;106:171–81. doi:
1138 10.1016/j.micpath.2016.02.005.

1139 72. de Almeida NEC, Esteves FG, dos Santos-Pinto JRA, Peres de Paula C, da Cunha
1140 AF, Malavazi I, et al. Digestion of Intact Gluten Proteins by *Bifidobacterium* Species:
1141 Reduction of Cytotoxicity and Proinflammatory Responses *J. Agric. Food Chem.*
1142 2020;68:4485–92. doi: 10.1021/acs.jafc.0c01421.

1143 73. Caminero A, McCarville JL, Galipeau HJ, Deraison C, Bernier SP, Constante M, et
1144 al. Duodenal bacterial proteolytic activity determines sensitivity to dietary antigen
1145 through protease-activated receptor-2. *Nat. Comm.* 2019;10:1198. doi: 10.1038/s41467-
1146 019-09037-9.

1147 74. Sánchez E, Laparra JM, Sanz Y. Discerning the role of *Bacteroides fragilis* in celiac
1148 disease pathogenesis. *Appl Environ Microbiol.* 2012;78:6507–15. doi: 10.1128/AEM.00563-
1149 12.

1150 75. Lieberman D, Shleyfer E, Castel H, Terry A, Harman-Boehm I, Delgado J, et al.
1151 Nasopharyngeal versus oropharyngeal sampling for isolation of potential respiratory
1152 pathogens in adults. *J. Clin. Microbiol.* 2006;44:525–8. doi: 10.1128/JCM.44.2.525-
1153 528.2006.

1154 76. Lemon KP, Klepac-Ceraj V, Schiffer HK, Brodie EL, Lynch S v, Kolter R.
1155 Comparative analyses of the bacterial microbiota of the human nostril and oropharynx.
1156 *mBio*. 2010;1. doi:10.1128/mBio.00129-10.

1157 77. Boutin S, Depner M, Stahl M, Graeber SY, Dittrich SA, Legatzki A, et al.
1158 Comparison of Oropharyngeal Microbiota from Children with Asthma and Cystic
1159 Fibrosis. *Mediators Inflamm.* 2017;2017:5047403. doi:10.1155/2017/5047403.

1160 78. Rosier BT, Moya-Gonzalvez EM, Corell-Escuin P, Mira A. Isolation and
1161 Characterization of Nitrate-Reducing Bacteria as Potential Probiotics for Oral and
1162 Systemic Health. *Front. Microbiol.* 2020;11:555465. doi: 10.3389/fmicb.2020.555465.

1163 79. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, Letellier E. Microbiome in
1164 Colorectal Cancer: How to Get from Meta-omics to Mechanism? *Trends Microbiol.*
1165 2020;28:401–23. doi: 10.1016/j.tim.2020.01.001.

1166 80. Eisenberg T, Glaeser SP, Blom J, Kämpfer P. *Leptotrichiaceae*. John Wiley & Sons,
1167 Inc; 2018. doi: 10.1002/9781118960608.fbm00155.pub2.

1168 81. Hagenfeld D, Koch R, Jünemann S, Prior K, Harks I, Eickholz P, et al. Do we treat
1169 our patients or rather periodontal microbes with adjunctive antibiotics in periodontal
1170 therapy? A 16S rDNA microbial community analysis. *PloS one*. 2018;13:e0195534.
1171 doi: 10.1371/journal.pone.0195534.

1172 82. Zhao C, Dong H, Zhang Y, Li Y. Discovery of potential genes contributing to the
1173 biosynthesis of short-chain fatty acids and lactate in gut microbiota from systematic
1174 investigation in *E. coli*. *npj Biofilms Microbiomes*. 2019;5:1–8. doi: 10.1038/s41522-
1175 019-0092-7.

1176 83. Chauhan P, Saha B. Metabolic regulation of infection and inflammation. *Cytokine*.
1177 2018;112:1–11. doi: 10.1016/j.cyto.2018.11.016.

1178 84. Cagno R di, Angelis M de, Pasquale I de, Ndagijimana M, Vernocchi P, Ricciuti P,
1179 et al. Duodenal and faecal microbiota of celiac children: molecular, phenotype and
1180 metabolome characterization. *BMC Microbiol.* 2011;11:219. doi:10.1186/1471-2180-
1181 11-219.

1182 85. Verdu EF, Galipeau HJ, Jabri B. Novel players in coeliac disease pathogenesis: role
1183 of the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* 2015;12:497–506.
1184 doi:10.1038/nrgastro.2015.90.

1185 86. Amoureux M-C, Rajapakse N, Stipkovits L, Szathmary S. Peptidoglycan and
1186 bacterial DNA induce inflammation and coagulation markers in synergy. *Mediators
1187 Inflamm.* 2005;2005:118–20. doi: 10.1155/MI.2005.118.

1188 87. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best
1189 practices for analysing microbiomes. *Nat. Rev. Microbiol.* 2018;16:410–22. doi:
1190 10.1038/s41579-018-0029-9.

1191 88. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P,
1192 Trajkovic V, Aasmets O, et al. Applications of Machine Learning in Human
1193 Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease
1194 Prediction and Treatment. *Front. Microbiol.* 2021;12:634511. doi:
1195 10.3389/fmicb.2021.634511.

1196

1197

1198 **6. Figures, tables additional files**

1199 **Figure 1.** Schematic representation of the sequential process of interaction between
1200 gluten, the microbiome and the immune response in the intestine, which leads to the
1201 destruction of enterocytes in the pathogenesis of CeD suggested by the evidence to date.

1202

1203 I = intestinal lumen; II = epithelium; III = lamina propria; APC = antigen presenting cell; B = cell B; P = plasma cell, o =
1204 opportunistic or pathogenic commensal bacteria; c = beneficial commensal bacteria, T = CD4 + T cell. A. tTG-gluten peptide
1205 complex: deaminated (right brown), unmodified (left blue), Diet degradation peptides: gliadin peptides (right blue), other dietary
1206 peptides (various colors left). C. Tight junctions between epithelial cells: claudin (green), occludin (pink). D. Cytokines: pro-
1207 inflammatory (blue), and anti-inflammatory (green). E. HLA DQ2/8 heterodimer (blue), immunogenic gliadin peptide (blue circle),
1208 TCR (pink).

1209

1210

1211 **Figure 2.** Shannon (A), Simpson (B) and Chao1 (C) diversity indices for (1) the pooled samples in cases and controls for each sampled tissue according to
 1212 diet. The p-value was calculated using the Wilcoxon test in A and B and the Welch's t-test in C. The limits of the rectangle indicate the 25th and 75th
 1213 percentiles and the horizontal bars indicate the median, in C, the median equals average. Vertical bars indicate upper and lower distribution limits, and dots
 1214 represent mild outliers. See in methods the codes of significance of the p-value.

1215

1216 **Figure 3:** Summary of microbial and functional changes observed after data integration.

1217

1218 SCFA= short-chain fatty acids. CeD=Celiac disease patients.

1219

1220

Figure 4: Linear discriminant analysis (LDA) integrated with effect size (LEfSe).

Cladogram representing the differentially abundant taxonomic groups in the microbiota from the duodenum of CeD following an unrestricted diet (a.1) or a GFD (b.1). Graph bar representing the microbial biomarkers found in the microbiota from the duodenum of CeD following an unrestricted diet (a.2) or CeD patients following a GFD (b.2) (LDA score >3 , $P < 0.001$).

Control: Control group, represented in green. Case: CeD patients, represented in red.

Figure 7: Differential abundance analysis performed on samples from the stool of CeD patients following an unrestricted diet (a) and CeD patients under a GFD (b) compared with healthy controls.

1248 **Figure 8:** Linear discriminant analysis (LDA) integrated with effect size (LEfSe). a) Cladogram representing the differentially abundant
1249 taxonomic groups in the microbiota from the pharynx of CeD patients. b) Graph bar representing the microbial biomarkers found in the
1250 microbiota from the pharynx of CeD patients (LDA score >3 , $P < 0.001$).

1251

1252

1253 Tables

1254 **Table 1.** The number of samples for each study, clinical classification and tissue of origin for the data included in the analysis. The cases (n =
1255 190) include patients with celiac disease, on a diet with or without gluten, the controls (n = 245) include non-celiac patients with gluten
1256 intolerance, gluten-free diet or without any enteropathy. The total samples add up to 435.

References	Accession N° ENA ¹	Sampled tissue	Sequencing technology	16S region
Bodkhe, et al. (2019)	PRJNA385740	duodenum and stool	Illumina Miseq	V4
Garcia-Mazcorro et al., (2018)	PRJNA401920	duodenum and stool	Illumina Miseq	V4
Laffaldano et al., (2018)	PRJNA371697	pharynx	Illumina Miseq	V4-V6
Olivares et al., (2018)	PRJEB23313	stool	Illumina Miseq	V1-V2
Tian, et al., (2017)	PRJNA321349	saliva	Illumina Miseq	V3-V4
Bonder et al., (2016)	PRJEB13219	stool	454	V3-V4
Giacomin et al., (2016)	PRJNA316208	duodenum	454	V1-V3
Quagliariello et al., (2016)	PRJEB14943	stool	Illumina Miseq	V3-V4
Francavilla et al., (2014)	PRJNA231837	saliva	454	V1-V3

¹. Accession numbers in the "European Nucleotide Archive" (ENA) database

Table 2. Statistics of Input data to analyse (After filtering samples and ASVs).

Tissue	Feeding Habit	ASVs	Samples	Cases	Controls
duodenum	GFD	475	30	12	18
duodenum	Unrestricted	475	89	40	49
stool	GFD	374	89	52	37
stool	Unrestricted	374	107	24	83
saliva	Control Unrestricted and Case GFD	120	78	40	38
pharynx	Control Unrestricted and Case GFD	71	42	22	20

ASV= Amplicon Sequence Variant. GFD= Gluten free diet. Cases=Samples from patients with Celiac disease.

1270 **Table 3.** Statistics of significant results obtained on Picrust2 predictions analysis.

1271

Tissue	Feeding Habit	Gene Differential abundance	Gene Correlation	Pathway Differential abundance	Pathway Correlation
duodenum	GFD	412	0	13	0
duodenum	Unrestricted	404	0	8	0
stool	GFD	31	69	0	1
stool	Unrestricted	261	0	2	0

1272 GFD= Gluten-free diet. FC>=3, $p < 0.01$

1273

1274 **Availability of data and material**

1275 The datasets generated during and/or analysed during the current study are available
1276 available in FigShare.

1277 https://figshare.com/projects/An_lisis_del_microbioma_en_Enfermedad_Cel_aca/82547

1278 The computer code was developed with the R and "bash shell" programming
1279 environments for GNU / Linux, The workflow and the specific criteria for each step in
1280 the analysis for each dataset are available on GitHub

1281 [https://github.com/juearcilaga/Assesing-microbiome-profiles-of-celiac-disease-](https://github.com/juearcilaga/Assesing-microbiome-profiles-of-celiac-disease-patients.git)
1282 [patients.git](https://github.com/juearcilaga/Assesing-microbiome-profiles-of-celiac-disease-patients.git)

1283 **List of abbreviations**

1284 **CeD:** Celiac disease

1285 **GFD:** gluten-free diet

1286 **TG2:** transglutaminase type 2

1287 **APC:** antigen presenter molecules

1288 **ENA:** the European Nucleotide Archive

1289 **ASV:** exact amplicon sequence variants

1290 **PCoA:** principal coordinate analysis

1291 **CLR:** Centered log ratio Normalization

1292 **LEfSe:** LDA Effect Size analysis

1293 **TSS:** Total Sum Scaling normalization

1294 **FDR:** False Discovery Rate

1295 **NCGS:** non-celiac gluten sensitivity patients

1296 **GI:** gastro intestinal

1297 **SCFA:** short-chain fatty acids

1298 **FDR:** First Degree Relative.

1299 **HC:** Healthy Control.

1300 **Declarations**

1301 **Ethics approval and consent to participate**

1302 "Not applicable".

1303 **Consent for publication**

1304 "Not applicable".

1305 **Competing interests**

1306 "The authors declare that they have no competing interests".

1307 **Funding**

1308 This study has been funded by a Research Grant 2020 of the European Society of
1309 Clinical Microbiology and Infectious Diseases (ESCMID) to L.J.M.-Z. Laura Marcos-
1310 Zambrano has a Juan de la Cierva Grant (IJC2019-042188-I) from the Spanish State
1311 Research Agency of the Spanish Ministerio de Ciencia e Innovación y Ministerio de
1312 Universidades.

1313 **Authors' contributions**

1314 Conceptualisation, L.J.M.-Z. and E.C.d.S.P.; methodology, J.A. and L.J.M.-Z.;
1315 writing—original draft preparation, J.A. and L.J.M.-Z; writing—critical review and

1316 editing, V. L-K, A.R.d.M.; funding acquisition, L.J.M.-Z. and E.C.d.S.P. All authors
1317 have read and agreed to the published version of the manuscript.

1318 **Corresponding author**

1319 Correspondence to Enrique Carrillo de Santa Pau and Laura Judith Marcos-Zambrano

1320 **Acknowledgements**

1321 Figure 1 and Figure 2 were designed with Biorender.com.

1322