nature portfolio

Peer Review File

Comprehensive Cross-Cohort Analysis Reveals Global Gut Microbiome Signatures of Celiac Disease

Corresponding Author: Dr Olivia Ogilvie

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This work represents a very significant attempt to generate meaningful data on the microbiome in celiac disease (CeD). By assembling data sets from a number of sources they, to some extent, address the small population size limitations that bedevil many studies. But limitations remain (as nicely summarized in the opening sections of the discussion. Here are my concerns:
1. The heterogeneity of the populations that provided the data sets. For example, we are told that the prospective stool was derived from children - what were the ages of those who provided the other data sets? Were control samples age-matched?
2. Was the prospective stool group the only longitudinal set? Were these individuals followed through diagnosis to treatment and beyond?
3. A lot of time in the discussion is spent on interpreting the data in relation to the pathophysiology of celiac disease - I find this overly speculative in light of the heterogeneity of the groups who provided samples and the influence of so many confounders some of which are mentioned but when interpreting stool data (the bulk of the samples) factors such as stool consistency and gut transit need to be borne in mind.
3. As is usual in studies like this causality is difficult to establish as the changes seen could be secondary to the disease process (lack of change with treatment does not remove this effect given that many adults are slow to respod or may not respond completely).

Reviewer #2

(Remarks to the Author)
This manuscript presents a large, carefully curated cross-cohort meta-analysis of gut microbiome data in celiac disease (CeD), integrating 16S rRNA and shotgun metagenomic datasets across disease stages, body sites, and geographies.
The compilation of >900 stool and duodenal samples across multiple countries and disease stages is a major asset and clearly advances prior work. Harmonized pipelines, explicit handling of batch effects, and dual approaches (random-effects meta-analysis and pooled models) are appropriate and well executed. The authors are appropriately cautious, emphasizing small effect sizes and the dominance of inter-study heterogeneity.
The finding that microbiome alterations persist on a gluten-free diet and that prediction performance drops under LODO validation is important and realistically framed.
Major Concerns
1. While the manuscript repeatedly notes that disease-associated variance is small (≈1–4%), the Discussion sometimes leans toward strong functional interpretations (e.g., altered SCFA metabolism) that are not directly measured. These claims should be more clearly framed as hypotheses rather than implied mechanisms.
2. Key confounders (age, diet composition beyond GFD status, medications, serology, symptom severity) are inconsistently available and largely unmodeled. This limits causal inference and should be emphasized more strongly as a constraint on conclusions.
3. The lack of a robust, adjusted signal in prospective cohorts undercuts claims about pre-diagnostic microbiome alterations. This point should be stated more bluntly to avoid overinterpretation.
4. High K-fold AUCs, especially in treated cohorts, are potentially misleading. Although the authors acknowledge this, the Results section could more clearly foreground LODO performance as the only clinically meaningful benchmark.
5. Please cite PMID: 36980164
Minor Comments

- The manuscript is long and at times repetitive, particularly in the Results and Discussion. Streamlining would improve clarity.
- Concordance between 16S and shotgun results is often low; a short, explicit discussion on how readers should weigh conflicting signals would help.
- Figures are generally clear, but some volcano plots are dense and hard to interpret without extensive cross-referencing.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
No further comments.

Reviewer #2

(Remarks to the Author)
The authors have adequately addressed the reviewers' comments, and the manuscript has improved in clarity and balance. Key concerns regarding overinterpretation, confounding, and generalizability have been appropriately addressed. The conclusions are now better aligned with the strength and limitations of the data. I have no further major concerns and recommend the manuscript for publication.

Response to reviewers

Title: Comprehensive Cross-Cohort Analysis Reveals Global Gut Microbiome Signatures of Celiac Disease

Manuscript number: **COMMSMED-25-3296-T**

In addition to the specific responses to reviewer comments detailed below, we have made several formatting and editorial changes to ensure compliance with the journal requirements. Firstly, we adapted the abstract to adhere to the structuring guidelines and added a “Plain language summary”. We adapted the final paragraph of the introduction to provide a present-tense summary of what was done and what we found in a short, broad approach. While following the STROBE checklist, we also identified the need to add several sections to the study to ensure strong reporting of observational research. These included a “Variable Definitions” section in the methods to define primary outcomes, predictors, exposures, confounders, and diagnostic criteria and a “Statistics and Reproducibility” section that provides clarity on the methods used and the handling of sample definitions. We also split the “Data and Code Availability” section into separate sections. The “Data Availability" section now clearly states where the source data for figures and the data used in this analysis can be found. We added a separate “Funding” section to clearly specify the funding source and the funders' roles in this study. Finally, we added details to the “Ethics Declarations” section to state the ethical considerations for this study and that it adheres to the Declaration of Helsinki. The following section addresses each reviewer's comment in detail and outlines the corresponding revisions made to the manuscript.

Reviewer 1

Comment 1: The heterogeneity of the populations that provided the data sets. For example, we are told that the prospective stool was derived from children - what were the ages of those who provided the other data sets? Were control samples age-matched?

Author's response: We agree that heterogeneity across populations and the lack of key confounders, such as age, hindered attempts to separate population-level variance and identify true cross-cohort associations. Age was inconsistently reported in the uploaded metadata across all studies. Regarding prospective studies, only one study reported age (Girdhar et al.), and it used an age bracket (i.e., either 2.5 or 5 years). In each study's publication, an age-matched control is used; however, given the limited availability of this data, we were unable to adjust for these effects. Regarding the ages inferred from the publications, the 16S datasets were similar (Millitich et al. = 1 year; Girdhar et al. = 2.5-5 years), with most samples in the Girdhar et al. cohort at 2.5 years. The shotgun prospective

dataset collected samples at 0 to 6 months, and then at −18, −15, −12, −9, −6, and −3 months before onset. Other metadata variables, such as sex, were also inconsistently reported, hindering adjustment. In response to this comment, we revised the Discussion's conclusion to mention the absence of confounding metadata and how this limitation restricts causal inference. Lines 625- 630 now reads:

Although between-study heterogeneity remained substantial, accounting for cohort-level variation identified reproducible alterations in the gut microbiome consistent with hypothesized alterations in microbial-metabolic function among distinct stage-specific progression "These findings should be interpreted cautiously, as key confounders, including age, sex, diet information beyond GFD status, medications, serology, and symptom severity, were inconsistently available and could not be uniformly modelled, limiting causal inference."

Comment 2: Was the prospective stool group the only longitudinal set? Were these individuals followed through diagnosis to treatment and beyond?

Author's response: In addition to the prospective dataset (Leonard et al.), several datasets included samples taken at the time of diagnosis and following treatment with a gluten-free diet (Nobel et al., Quagliariello et al., Turjeman et al., and Costigan et al.). We agree that the Leonard et al. study design should be more explicitly stated and have added a minor amendment in the Methods Processing 16S rRNA gene and shotgun data section to address this. Lines 183-184 now read:

"The only exception was the longitudinal SG_118_Leonard dataset [26], which was retained as its own phyloseq object because it is the sole prospective shotgun stool dataset"

Comment 3: A lot of time in the discussion is spent on interpreting the data in relation to the pathophysiology of celiac disease - I find this overly speculative in light of the heterogeneity of the groups who provided samples and the influence of so many confounders some of which are mentioned but when interpreting stool data (the bulk of the samples) factors such as stool consistency and gut transit need to be borne in mind.

Author's response: We agree that the original Discussion placed too much emphasis on pathophysiological interpretation, given the heterogeneity of the included cohorts and the limited availability of key confounders. We have therefore revised the Discussion throughout to separate observed microbial associations from mechanistic inference, and we now explicitly note that stool-based microbial profiles may be influenced by factors such as stool consistency, gut transit, diet, and medication exposure in addition to disease status. We have also clarified that the functional implications discussed are hypothesis-generating and not direct evidence of causality. Lines 545-555, 568-572, & 625-630 now read:

Lines 545-555:

"Observed compositional differences should be interpreted cautiously. It is important to note that concordance between 16S rRNA amplicon sequencing and shotgun metagenomic sequencing is often limited, reflecting differences in taxonomic resolution, detection sensitivity, and methodological biases inherent to each approach. As such, conflicting signals between methods should be interpreted with caution, with greater confidence placed in findings that are consistent across both platforms. Moreover, most included datasets were stool-derived, and microbial profiles in stool can be influenced by factors beyond disease status, including stool consistency, gut transit, diet, medications, and other host factors that could not be uniformly modelled. The present findings are best viewed as robust cross-cohort associations, not direct evidence of causal pathophysiological mechanisms in CeD."

Lines 568-572:

This observation is consistent with previous duodenal biopsy–based work reporting impaired fiber fermentation and reduced SCFA-producing capacity in CeD patients. "While these compositional shifts are consistent with impaired SCFA production, functional consequences and causality remain to be established."

Lines 625-630:

"Although between-study heterogeneity remained substantial, accounting for cohort-level variation identified reproducible alterations in the gut microbiome consistent with hypothesized alterations in microbial-metabolic function among distinct stage-specific progression. These findings should be interpreted cautiously, as key confounders, including age, sex, diet information beyond GFD status, medications, serology, and symptom severity, were inconsistently available and could not be uniformly modelled, limiting causal inference."

Comment 4: As is usual in studies like this, causality is difficult to establish as the changes seen could be secondary to the disease process (lack of change with treatment does not remove this effect, given that many adults are slow to respond or may not respond completely).

Author's response: We agree that causality is difficult to establish from this data and that persistence of microbial alterations following GFD initiation does not exclude a secondary disease effect. In response, we have: added an explicit caution paragraph noting that stool-derived compositional data may reflect unmeasured factors such as stool consistency, gut transit, diet, and medications, and should be interpreted as cross-cohort associations over causal evidence; revised mechanistic language throughout the Discussion to distinguish observed associations from inferred pathophysiology; and added a sentence directly following our GFD persistence claim acknowledging that mucosal recovery in CeD can be slow and incomplete, and that true microbial normalization may require longer follow-up than captured in the included datasets Lines 545-555 (above), 579-583, 625-630 (above) now read:

Lines 579-583:

The enrichment of lactate producers and depletion of lactate-utilizing butyrate producers in CeD suggests an alteration in this metabolic link.“This disruption may precede diagnosis and persist after GFD initiation, indicating that microbial normalization may require longer follow-up than captured in the included datasets, as mucosal recovery in CeD is often slow and incomplete.”

Reviewer 2

Comment 1: While the manuscript repeatedly notes that disease-associated variance is small (≈1–4%), the Discussion sometimes leans toward strong functional interpretations (e.g., altered SCFA metabolism) that are not directly measured. These claims should be more clearly framed as hypotheses rather than implied mechanisms.

Author's response: We agree that we should reframe key mechanistic claims as hypotheses. In response to this, we have revised mechanistic language throughout the Discussion to distinguish observed associations from inferred pathophysiology, added caution surrounding unmodelled confounding variables such as stool factors (stool consistency, gut transit, diet, and medications), and incomplete metadata (age, sex, diet composition beyond GFD status, medications, serology, symptom severity). Lines 545-555, 556-572, 593-597, 598-600, 613-618 now read:

Lines 545-555:

"Observed compositional differences should be interpreted cautiously. It is important to note that concordance between 16S rRNA amplicon sequencing and shotgun metagenomic sequencing is often limited, reflecting differences in taxonomic resolution, detection sensitivity, and methodological biases inherent to each approach. As such, conflicting signals between methods should be interpreted with caution, with greater confidence placed in findings that are consistent across both platforms. Moreover, most included datasets were stool-derived, and microbial profiles in stool can be influenced by factors beyond disease status, including stool consistency, gut transit, diet, medications, and other host factors that could not be uniformly modelled. The present findings are best viewed as robust cross-cohort associations, not direct evidence of causal pathophysiological mechanisms in CeD."

Lines 556-572:

The most substantial finding was a consistent alteration in metabolically important taxa across both stool and duodenal samples, with stage-specific patterns. Before disease onset, butyrate-producing species such as *Faecalibacterium* sp., *Agathobacter faecis*, and *Gemmiger qucibialis* [78-80] were depleted, while taxa associated with lactate, acetate, and propionate, including the genus *Lactococcus*, and species *Hominimerdicola aceti*, and *Collinsella* sp. were increased [81-83]. During active disease, the microbial signal became more complex, with further depletion of *Faecalibacterium* and *Gemmiger qucibialis,* and increases in *Prevotella* spp., *Phocaeicola dorei, Bifidobacterium pseudocatenulatum, Lactobacillus,* and *Streptococcus* [84-87]. In individuals on a GFD, depletions persisted for *Agathobacter rectalis*, *Anaerobutyricum hallii*, *Faecalibacterium* sp., and *Gemmiger qucibialis,* alongside reduced acetate-producing *Bifidobacterium* species like *B. adolescentis* and *B. infantis* [88,89]. In the duodenum during active disease, multiple *Prevotella* spp. were depleted while lactate-associated genera like *Lactococcus* increased, suggesting that altered fiber metabolism may

occur at the active site of CeD pathology. This observation is consistent with previous duodenal biopsy–based work reporting impaired fiber fermentation and reduced SCFA-producing capacity in CeD patients [90]. "While these compositional shifts are consistent with impaired SCFA production, functional consequences and causality remain to be established."

Lines 593-597:

The current findings also noted a duodenal-specific loss of commensal oral microbes, such as *Rothia aeria* and *Rothia mucilaginosa,* which possess enzymes capable of hydrolyzing gluten. Their depletion at the site of inflammation could impair local gluten processing, "potentially increasing the concentration of immunogenic peptides and contributing to the local inflammatory response"

Lines 598-600:

Acute increases in potentially opportunistic taxa at the mucosal surface, namely *Helicobacter* and *Campylobacter*, "may" contribute to the inflammatory response in the duodenum, "though further studies are required to clarify their roles" [103,104].

Lines 625-630:

"Although between-study heterogeneity remained substantial, accounting for cohort-level variation identified reproducible alterations in the gut microbiome consistent with hypothesized alterations in microbial-metabolic function among distinct stage-specific progression. These findings should be interpreted cautiously, as key confounders, including age, sex, diet information beyond GFD status, medications, serology, and symptom severity, were inconsistently available and could not be uniformly modelled, limiting causal inference."

Comment 2: Key confounders (age, diet composition beyond GFD status, medications, serology, symptom severity) are inconsistently available and largely unmodeled. This limits causal inference and should be emphasized more strongly as a constraint on conclusions.

Author's response: We agree and as per above, we have added in dialog to the Discussion's conclusion to mention the absence of confounding metadata and how this limitation restricts causal inference Lines 625-630 (above).

Comment 3: The lack of a robust, adjusted signal in prospective cohorts undercuts claims about pre-diagnostic microbiome alterations. This point should be stated more bluntly to avoid overinterpretation.

Author's response: We agree this point was understated. The adjusted analysis did not produce a significant signal in the prospective cohort (P = 0.27), and we have revised the Discussion to state this more plainly: the present data do not provide evidence for a broad pre-diagnostic microbiome shift after accounting for cohort-level variation. Lines 530-534 now read:

By statistically conditioning on these effects, a reproducible, disease-associated signature was isolated in the large stool cohorts of both active CeD (partial R2 = 1.2%, P = 0.001) and treated CeD (partial R2 = 1.1%, P = 0.001). "No adjusted effect was seen in the prospective stool group (P = 0.27), indicating that a broad pre-diagnostic microbiome shift was not resolved, although subtler or taxon-specific shifts cannot be excluded."

Comment 4: High K-fold AUCs, especially in treated cohorts, are potentially misleading. Although the authors acknowledge this, the Results section could more clearly foreground LODO performance as the only clinically meaningful benchmark.

Author's response: We agree that this distinction should be made clearer in the Results section. Therefore, we have added a sentence acknowledging this important aspect of the interpretation of the machine learning performances. We also adjusted the section header to emphasise generalisation as a key concept. Lines 505, 506-511 now read:

Line 505:

"Cross-Dataset Prediction Reveals Partial Generalizability of Celiac Classification"

Lines 506-511:

To assess microbiome-based prediction of CeD, machine learning models were evaluated on 16S datasets using three cross-validation methods: (1) K-fold, where all datasets were pooled, models were trained on 90% of samples and tested on the rest; (2) LODO, where models were trained on all but one dataset and tested on the unseen dataset; and (3) X-set, where models were trained on a single dataset and tested on another. "Of these, LODO is the most clinically relevant because it evaluates generalization to entirely unseen cohorts."

Comment 5: Please cite PMID: 36980164

Author's response: We have cited this review in the Introduction section in response to this comment. Line 90-93 now reads:

The gut microbiome, a complex ecosystem dominated by microorganisms co-existing with their host and influencing digestion, metabolism, and immune modulation, is increasingly implicated in CeD [18,19].

Comment 6: The manuscript is long and at times repetitive, particularly in the Results and Discussion. Streamlining would improve clarity.

Author's response: We agree and have shortened the discussion (1,809 to 1,439 words), along with subtle, succinct word changes throughout the text, reducing the main text (introduction, results, discussion) from 4950 to 4600 words (within the journal's 5000-word guidelines).

Comment 7: Concordance between 16S and shotgun results is often low; a short, explicit discussion on how readers should weigh conflicting signals would help

Author's response: We have added a couple of sentences that discuss why the concordance was low and the importance of seeing taxa that occur across both with greater confidence. Lines 545-551 now reads:

Observed compositional differences should be interpreted cautiously. "It is important to note that concordance between 16S rRNA amplicon sequencing and shotgun metagenomic sequencing is often limited, reflecting differences in taxonomic resolution, detection sensitivity, and methodological biases inherent to each approach. As such, conflicting signals between methods should be interpreted with caution, with greater confidence placed in findings that are consistent across both platforms."

Comment 8: Figures are generally clear, but some volcano plots are dense and hard to interpret without extensive cross-referencing.

Author's response: We agree that when presenting data across multiple groups and levels of analysis, plots can become difficult to interpret. Due to the number of disease groups and the fact that the differential abundance results are from multiple phylogenetic levels, the plots are complex. We have tried to limit the number of annotated taxa to only those with high relative

abundance and effect. We have also made each panel use as much space as possible while conforming to journal guidelines. As such, no change was made to the volcano plot figure.