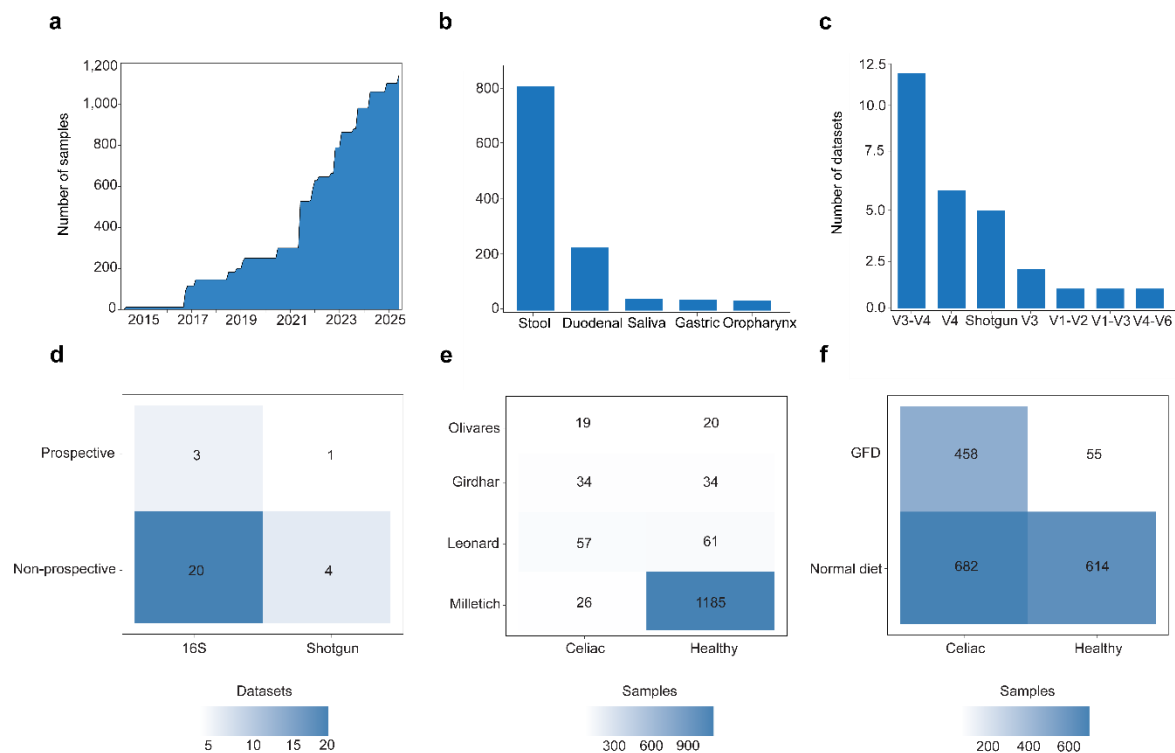
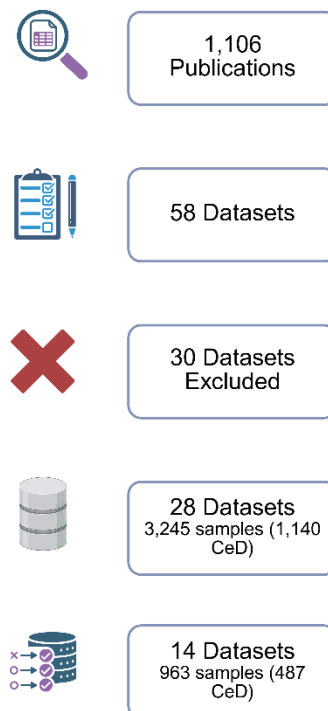


Supplementary Information



Supplementary Fig. 1 | Overview of the Celiac Microbiome Repository version 1.0. **a**, Cumulative count of publicly available samples from individuals diagnosed with CeD since 2014, with the total reaching 1,140 by July 2025. **b**, Distribution of samples from individuals with CeD by body site shows stool dominates (n = 820) followed by duodenal biopsies (n = 224), while salivary, gastric and oropharyngeal samples remain under-represented. **c**, The sequencing methods and 16S amplicon regions of samples from individuals with CeD shows that 73.3% of samples capture the V4 region. **d**, Study design overview. Most datasets are non-prospective 16S rRNA studies (22), with far fewer non-prospective shotgun metagenomics (4), prospective 16S (3), and prospective shotgun datasets (1). **e**, Composition of the four prospective datasets. The largest (16S_1211_Milletich) is extremely control-heavy (1,185 healthy vs 26 prospective CeD), whereas the other three datasets are well balanced. **f**, Breakdown of non-prospective samples by diet and diagnosis: 458 CeD and 55 healthy individuals are on a GFD, while 682 CeD and 614 controls are on a normal diet.



Supplementary Fig. 2 | Pipeline for search, retrieval, curation, and filtering of CeD-related gut microbiome sequencing datasets. Online searches (15 July 2025) returned 1,106 peer-reviewed publications. After applying eligibility criteria, 58 datasets were retrieved. Due to a lack of accessible raw data or adequate metadata, 30 datasets were excluded, leaving 28 datasets and 3,245 samples (1,140 CeD) that comprise version 1.0 of the Celiac Microbiome Repository. Restricting the analysis to datasets with either shotgun sequencing or 16S rRNA sequencing of stool and duodenal samples that cover the V4 region and sufficiently balanced case–control representation yielded a final set of 14 datasets and 963 samples (487 CeD) for analysis. Created in BioRender. Prendergast, Peter. J (2026) <https://BioRender.com/pthdspr>