

5 Title

Disease signatures in the gut metagenome of a prospective family cohort of inflammatory bowel disease

Authors

10 Malte Rühlemann^{1,2}, Silvio Waschina³, Eike Matthias Wacker¹, Philipp Rausch¹, Ana Schaan¹, Zaira Zafroon¹, Katrin Franzpötter⁴, Gunnar Jacobs⁴, David Ellinghaus¹, Robert Kruse⁵, Daniel Bergemalm⁶, Jonas Halfvarson⁶, Malte Ziemann⁷, Siegfried Görg⁷, Wolfgang Lieb⁴, Stefan Schreiber^{1,8}, Mathilde Poyet^{9,*}, Corinna Bang^{1,*}, Andre Franke^{1,*}, Mathieu Groussin^{1,*}

15 Affiliations

1 - Institute of Clinical Molecular Biology, Kiel University & UKSH, Kiel, Germany

2 - Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School Hannover, Germany

3 - Nutriinformatics Research Group, Institute for Human Nutrition and Food Science, Kiel University, Kiel, Germany

20 4 - Institute of Epidemiology, Kiel University, Kiel, Germany

5 - Department of Clinical Research Laboratory, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

6 - Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

7 - Institute of Transfusion Medicine, UKSH, Lübeck, Germany

25 8 - Department of Internal Medicine I, Kiel University & UKSH, Kiel, Germany

9 - Evolutionary Medicine Department, Institute of Experimental Medicine (IEM), Kiel University & UKSH, Kiel, Germany

Abstract

30 Inflammatory bowel disease (IBD) is associated with dysbiotic microbiomes. However, whether microbiomes of family members of IBD patients harbour microbial disease signatures of IBD is unknown. Here, we generate shotgun metagenomic data of an IBD family cohort and treatment-naïve IBD cases, which we combine with published IBD metagenomes, to perform a meta-analysis of IBD-microbiome associations. Our study reveals microbial shifts that are
35 specific to IBD or IBD subtypes (Crohn's disease and ulcerative colitis). We find that IBD cohorts share signatures of association with disease irrespective of geography, and we report novel species-level identifications of microbial taxa that are universal markers of Crohn's disease. We further demonstrate that the microbiome of healthy family members at high risk for IBD represent transitional states between the microbiome of unrelated healthy controls and IBD
40 cases, and harbor diversity, compositional and ecological features that are also observed among IBD microbiomes. Overall, our study provides valuable insights into the intricate relationship between the gut microbiome and IBD, with implications for better predicting disease onset among individuals with high susceptibility for IBD.

45

Introduction

Inflammatory Bowel Disease (IBD), encompassing conditions such as Crohn's disease (CD) and ulcerative colitis (UC), represents a complex and debilitating group of chronic inflammatory disorders affecting the gastrointestinal tract (Alatab et al., 2020; Baumgart & Sandborn, 2007; Lönnfors et al., 2014) The pathogenesis of IBD involves intricate interactions among genetic predisposition, environmental factors, immune system dysregulation, and the gut microbiome (Baumgart & Sandborn, 2007; Graham & Xavier, 2020; Plichta et al., 2019). Previous microbiome
50

55 studies identified multiple taxonomic, genomic or metabolic signatures of disease compared
to healthy individuals, e.g. with a decrease in *Faecalibacterium* abundance and elevated levels
of Enterobacteriaceae and *Veillonella* (Häsler et al., 2017; IBDMDB Investigators et al., 2019;
Ma et al., 2022; Mallick et al., 2021; Schirmer et al., 2019). Specific bacterial metabolites, such
as the short-chain fatty acid butyrate, have also been shown to be altered in IBD and to be
60 associated with clinical remission following anti-TNF therapy (Aden et al., 2019; Meade et al.,
2023). The transfer of fecal microbiome from IBD patients into germ-free mice elevates pro-
inflammatory T cell responses (e.g., Th17) and exacerbates disease symptoms in a mouse
model of colitis (Britton et al., 2019). While caution is advised in interpreting human
microbiota transfer experiments into mice (Walter et al., 2020), these findings imply that the
65 microbiome may convey IBD disease phenotypes and play a role in IBD development and
progression (Ni et al., 2017).

Despite extensive sequencing studies on the IBD microbiome, reported associations between
IBD and specific microbes vary across studies (Gevers et al., 2014; Halfvarson et al., 2017;
IBDMDB Investigators et al., 2019; Papa et al., 2012). These inconsistencies arise from
70 differences in sample size, clinical heterogeneity, significant inter-individual and inter-
population variations in microbiome compositions, and the construction of cross-sectional
case-control studies that limit our ability to account for temporal variations in the microbiome
within individuals (Gibbons et al., 2018; Schirmer et al., 2018). These inconsistent findings
underscore the need for systematic approaches, such as meta-analyses or the use of validation
75 cohorts, to reveal general patterns of disease-associated microbiome shifts (Abbas-Egbariya
et al., 2022; Abdel-Rahman & Morgan, 2023; Duvallet et al., 2017). Recent clinical and
methodological efforts greatly improved our understanding of the microbial factors that
associate with IBD, incorporating multi-center or meta-analyses of both cross-sectional and
longitudinal microbiome datasets covering various intestinal biospecimens and omics
80 profiling (Ma et al., 2022).

The majority of IBD microbiome studies used 16S rRNA amplicon sequencing, limiting our
ability to identify microbial disease signatures at fine taxonomic resolutions (e.g., species or
strain) due to the restricted information in amplicon data and the incompleteness of 16S-
85 based reference taxonomic databases (Abbas-Egbariya et al., 2022; Ma et al., 2022). For
example, unclassified Lachnospiraceae bacteria are often associated with IBD, but the specific
species involved remain unknown, hindering follow-up experimental investigations into their
functional role in disease (Ma et al., 2022). As of now, the availability of IBD metagenomic
datasets for the community is limited in terms of study number, sample size, and sequencing
90 depth (IBDMDB Investigators et al., 2019; Vich Vila et al., 2018). Compared to 16S,
metagenomic approaches provide the opportunity to build dataset-specific reference genome
collections, profile precise taxonomies, capture functional repertoires (e.g., through gene
family or pathway profiling), and identify bacterial genomic markers of host phenotypes down
to the resolution of structural variants or single nucleotide polymorphisms (Blanco-Míguez et
95 al., 2023; Olm et al., 2021; Rühlemann et al., 2022). As such, accessing large-scale longitudinal
metagenomic datasets from IBD clinical cohorts is crucial to reveal and experimentally validate
the contributions of the gut microbiome to IBD etiology and prognosis.

100 Previous IBD cohorts mainly enrolled patients with active disease, making it challenging to
understand the sequential events leading to dysbiotic microbiomes and to discern whether

changes in the microbiome are attributable to inflammation, drugs, or comorbidities (Raygoza Garay et al., 2023; Wright et al., 2015). First-degree relatives of IBD patients face an increased 5% lifetime risk of developing IBD (Halme, 2006; Russell & Satsangi, 2008), and exhibit higher levels of intestinal inflammation and impaired intestinal barrier function (Keita et al., 2018; Lee et al., 2019). The extent to which microbiomes contribute to this increased susceptibility and display signatures of early disease development is poorly characterized. Interestingly, previous studies demonstrated positivity for various anti-microbial antibodies, such as ASCA (anti-Saccharomyces cerevisiae antibodies) or ALCA (anti-laminaribioside carbohydrate antibodies) in relatives of CD patients (Michielan et al., 2013; Vermeire et al., 2001). In addition, a recent 16S-based analysis of a prospective cohort of CD cases and their first-degree relatives identified taxonomic markers of future onset of CD (Raygoza Garay et al., 2023). This body of evidence points to an early causative role of the gut microbiome in IBD development.

Here, we generated metagenomic data from hundreds of IBD patients and IBD family members with extensive clinical and environmental metadata, and performed a cross-country meta-analysis to investigate associations between microbiome features and disease. We reveal novel microbiome markers of IBD, including features that delineate cases from controls, or that are specific to either CD, UC or IBD, as well as reduced auxotrophic capacity of the IBD microbiome, which suggest lower metabolic cross-feeding interactions in dysbiotic microbiomes.

Main text

125 Cohort description

The Inflammatory Bowel Disease Family Cohort (IBD-FC) cohort is a German nationwide, prospective cohort study, which was established at Kiel University, Germany, in 2013, recruiting individuals affected by IBD, either CD or UC, together with their IBD-unaffected family members, to build a resource for the pro- and retrospective assessment of (incident) IBD cases. After the initial baseline sampling, participants are invited for follow-up appointments and biosample collection every 2-3 years or, in case of a newly diagnosed IBD case, soon after the initial diagnosis.

We sequenced and analyzed fecal metagenomes of 199 individuals diagnosed for IBD ($n_{CD} = 99$; $n_{UC} = 100$) and of 296 in IBD-unaffected family members (family controls, FC) in the IBD-FC cohort. We further generated metagenomes from unrelated healthy population controls (HC; $n=344$). We also generated metagenomes for 50, 53 and 147 of the CD, UC and FC individuals at a second time point, from ~ 2 years after the initial sampling was performed. We additionally generated metagenomes from a cohort of treatment-naïve CD- and UC-affected individuals ($n_{CD}=17$ and $n_{UC}=16$) from Sweden, together with unrelated healthy controls ($n=17$) and individuals with bowel-complaints not related to IBD (symptomatic controls, SC; $n=16$). Further, as an additional cohort with longitudinal data, we included a subset of the HMP2 dataset from the USA for which a second sampling timepoint ~ 180 days after initial sampling was available, resulting in the inclusion of 89 individuals ($n_{CD} = 39$, $n_{UC} = 25$, $n_{HC} = 25$) (IBDMDB Investigators et al., 2019). The overall dataset analyzed in this study includes $n = 155$ and $n = 141$ CD- and UC-affected individuals, respectively, in addition to $n =$

682 IBD-unaffected individuals, totaling samples from n = 978 individuals from three countries.

150 IBD and subtypes harbor strong, yet distinct microbial signatures

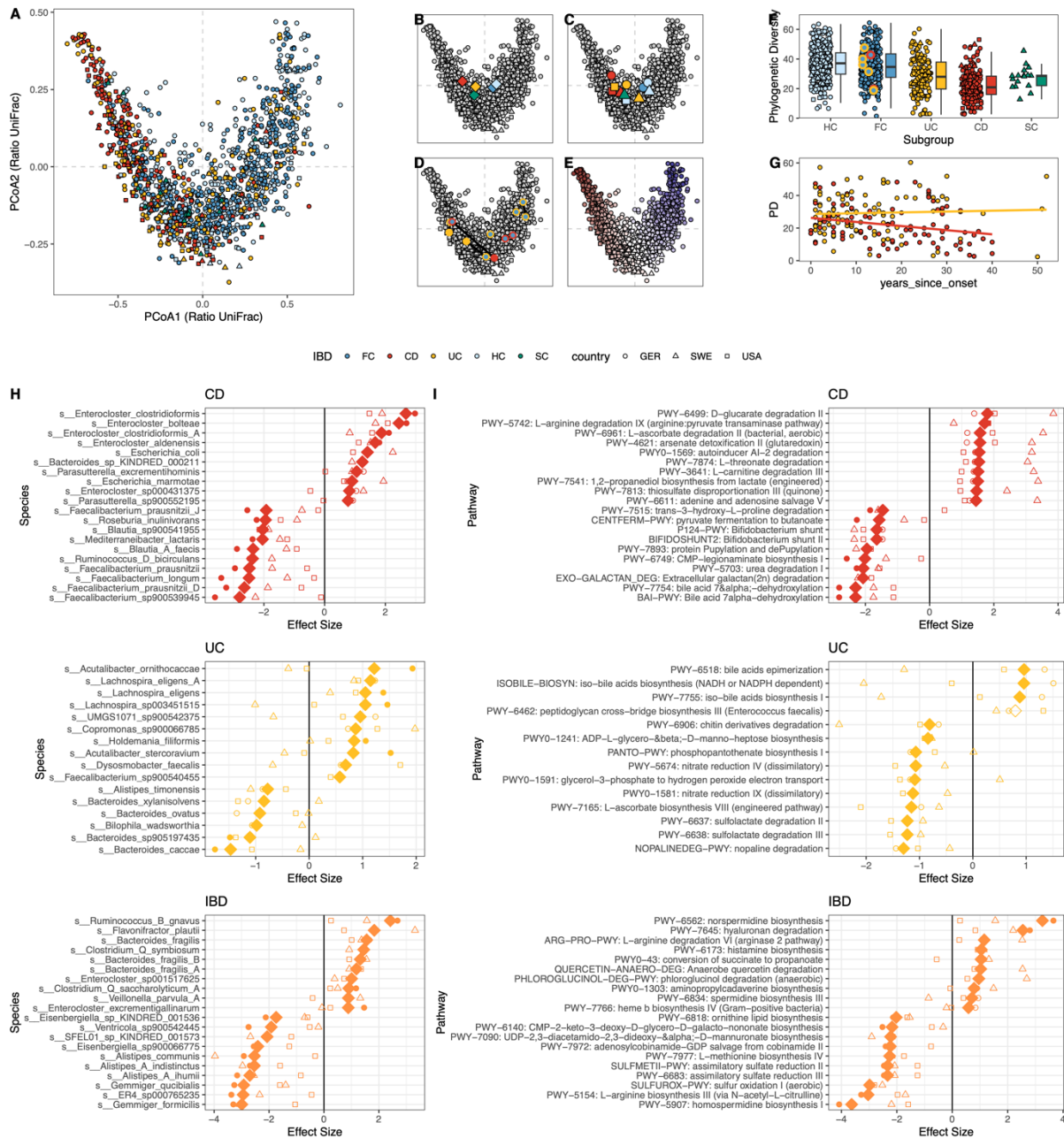


Figure 1: Our meta-analysis of gut metagenomes reveals microbiome changes that are associated with IBD and IBD subtypes. (A) Multivariate analysis (PCoA) plot based on Ratio UniFrac distances. (B) Subgroup centroids; HC = healthy population controls, FC = family controls, CD = Crohn's disease, UC = Ulcerative colitis, SC = symptomatic controls (SWE). (C) Country-wise subgroup centroids. (D) IBD incident cases in the IBD-FC cohort following baseline sampling/ Samples from multiple timepoints are connected by a line. Blue-filled circles are pre-onset samples, yellow/red filled circles are post-diagnosis. (E) Samples are colored by phylogenetic diversity (high/low diversity: blue/red, respectively). (F) Comparison of phylogenetic diversity across subgroups. (G) Phylogenetic diversity decreases in CD over time since disease onset ($P=0.0146$), but not in UC ($P=0.78$); linear regression model. (H) Top 20 bacterial species with the strongest abundance shifts that are specific to either a particular disease subtype (CD, UC) or to general IBD, compared to HC. (I) Top 20 pathways with the strongest abundance

155

160

shifts that are specific to either a particular disease subtype (CD, UC) or to general IBD, compared to HC. The full list of bacterial and pathway hits can be found in Supplementary Table 2 and Supplementary Figure 3.

165

We generated a catalog of metagenome-assembled genomes (MAGs) that we clustered into non-redundant species-level genome bins (SGBs), which were taxonomically and functionally annotated (See Methods). This SGB catalog served as a mapping reference for estimating taxonomic and pathway abundance profiles (See Methods). At the aggregate level, we observed a strong shift in microbial compositions among individuals affected by IBD when compared to both FCs and HCs, with the most pronounced differences observed in CD (Figure 1A and B). At the level of individual countries, this shift was consistently reproduced (Fig. 1C). We also observed country-specific signatures of microbial diversities, with USA HC microbial compositions being close to those of groups affected by disease and intestinal symptoms in Sweden and Germany (UC microbiomes). This pattern is also confirmed by a generally lower phylogenetic diversity in USA HC microbiomes. This heterogeneity between populations confirms the importance of performing cross-geography analyses to reveal universal and replicated association signatures of IBD.

170

175

180

Within the FC group, seven individuals were diagnosed with IBD ($n_{CD} = 2$, $n_{UC} = 5$) following initial recruitment (Figure 1D). While we include these onset cases in our analyses and highlight them in our results (Figure 1D), the low incidence sample size precluded dedicated statistical analysis - future analyses with additional onset cases will be needed in the future to allow longitudinal microbiome comparisons before and after disease manifestation.

185

We further found that shifts in microbiome compositions are linked to a reduction in phylogenetic diversity within samples (Figure 1E). This decrease in diversity is further underscored by between-group comparisons, showing reduced diversity in both IBD subtypes ($P_{CD} = 1.06 \times 10^{-29}$, $P_{UC} = 1.40 \times 10^{-13}$), with no significant difference being observed between HC and FC (Figure 1F). Moreover, substantial differences in phylogenetic diversity were found between HCs from different countries, with the USA HC cohort exhibiting lower phylogenetic diversity compared to their European counterparts (Suppl. Figure S1). This emphasizes the need for proper healthy control cohorts when performing multi-country comparisons of microbiome datasets, which need to account for geographic disparities in the diversity and composition of the baseline healthy microbiome. Interestingly, a decline in phylogenetic diversity over time since disease onset was specifically observed for CD ($P = 0.0146$, linear regression controlling for sex and BMI; Figure 1G), suggesting a potential association between disease progression and the microbial dysbiosis associated with CD, whereas no such correlation was found for UC ($P = 0.78$, controlling for sex and BMI; Figure 1G).

190

195

200

We conducted a differential abundance analysis of microbial taxonomies at the species level (Figure 1H) and of microbial pathways (Figure 1I) across all three included cohorts, using a meta-analysis approach that is based on likelihood-based measures (See Methods). This analysis aimed to determine whether observed changes are generally associated with any form of IBD or specifically driven by one of its subtypes. Out of the 305 species tested for differential abundance, 208 exhibited a decrease in abundance in IBD or one of its subtypes, while 44 exhibited increased abundance ($Q_{FDR,Meta} < 0.05$). Most species were associated with IBD in general ($n_{up} = 17$, $n_{down} = 148$; Figure 1H, Supp. Figure 3). We found elevated levels of *Ruminococcus B gnavus*, *Flavonifractor plautii*, *Bacteroides fragilis* (and its closely related

205

210 species), and *Clostridium Q symbiosum*, and a decreased abundance in multiple *Gemmiger*
and *Alistipes* (A) species (Figure 1H, Supp. Figure 3).

Crohn's Disease (CD) was found to be specifically associated with several species with strong
abundance shifts compared to healthy control. We detected five *Enterocloster* species among
215 the top 10 differentially abundant taxa that have elevated levels in CD, including *E.*
clostridioformis and *E. boltea* (Figure 1H). As in previous reports (Chiodini et al., 2015; IBDMDB
Investigators et al., 2019), we also found elevated levels of Pseudomonadota (previously
named Proteobacteria) taxa, such as *Escherichia coli* and *Parasutterella* species. Several taxa
showed substantial decreases in abundance specifically in CD, such as *Faecalibacterium* and
220 *Blautia* (A) species, which is consistent with findings from previous amplicon-based analyses
(Ma et al., 2022). Signals specific to ulcerative Colitis (UC) were less pronounced, with only six
species showing a decrease in abundance, including four *Bacteroides* species. The taxa most
prominently increased in UC included two *Acutalibacter* and three *Lachnospira* species (Figure
1H).

225 Looking into the differential abundances of 818 functional pathways (Figure 1I), we observed
robust signals associated with CD, with 33 decreased pathways, and 259 increased pathways
($Q_{FDR,Meta} < 0.05$), compared to controls (Supp. Figure 3). The most strongly reduced pathway
is the bile-acid 7 α -dehydroxylation pathway (bai-pathway), responsible for the formation
230 of LCA and DCA from chenodeoxycholate and cholate, respectively—a connection previously
suggested in relation to CD (Larabi et al., 2023). We further identified a reduced abundance in
the central fermentation pathway (CENTFERM-PWY), indicating a diminished production of
butyrate, confirming past reports (Laserna-Mendieta et al., 2018). The elevated abundance of
autoinducer AI-2 degradation suggests a potential shift towards a more selfish behavior in the
235 microbiome, as autoinducers play a crucial role in quorum sensing within bacterial
communities. Interestingly, we found that UC is associated with distinct bile acid metabolism
pathways compared to CD, as the bile acid epimerization and iso-bile acid biosynthesis were
significantly increased in abundance in UC compared to controls. This implies that variations
in microbial bile acid metabolism may contribute significantly to the differences observed
240 between Inflammatory Bowel Disease (IBD) subtypes. Finally, we identified histamine
biosynthesis as being among the most prominent signals universally increased in IBD. While
the role of histamine in IBD and other intestinal diseases is a subject of debate (Jutel et al.,
2006), studies in mice undergoing experimental colitis suggested a potential involvement
through the activation of innate immunity (Wechsler et al., 2018).

245

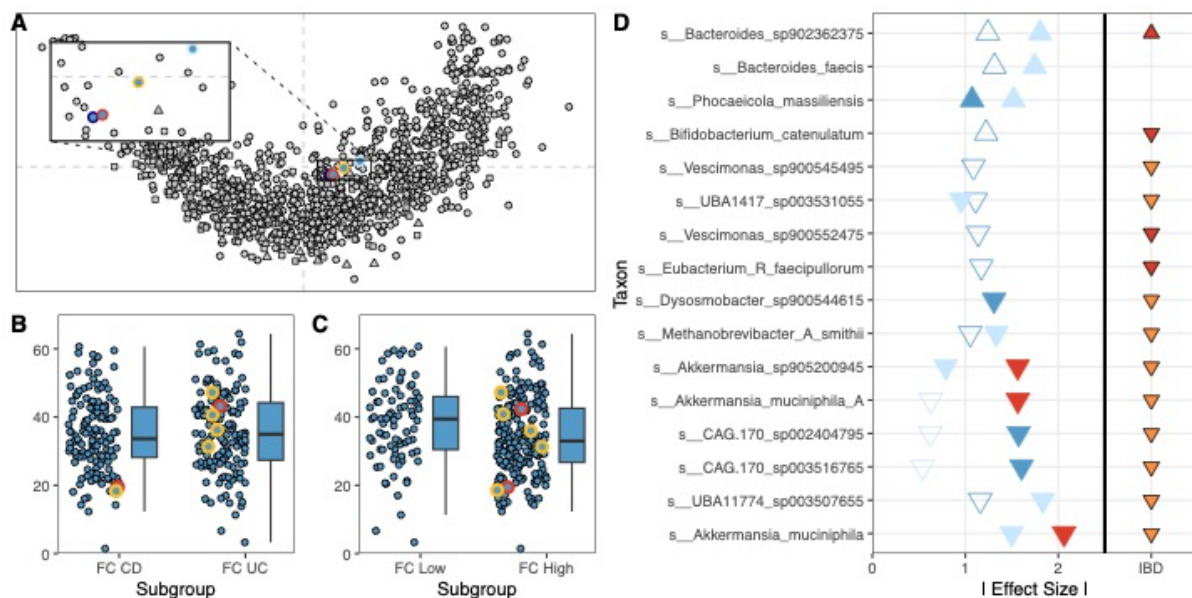
250

255

Healthy siblings and children of IBD patients exhibit microbiome signatures of disease

260 The IBD-FC cohort provides the unique opportunity to explore potential shifts in the microbiota of healthy individuals that have a family history of IBD. To do this, family controls were categorized into groups based on the predominant IBD type in their respective families (FC CD and FC UC). Alternatively, FC were divided based on individual risk, with siblings and children of IBD-affected individuals designated as "high risk" (FC High) and other relatives (parents, partners, higher-degree relatives) as "low risk" (FC Low). All seven IBD incident cases included in the dataset were among the family members classified as FC High. In addition, two of the incident cases were not diagnosed with the IBD subtype predominant in their respective families. Our analysis revealed significant differences in community composition (Figure 2A) and alpha diversity (Figure 2B) between FC High and FC Low categories ($P_{\alpha} = 0.002$, $P_{\beta} = 0.015$). However, no difference in these two diversity metrics was found between FC CD and FC UC groups (Figure 2A & 2B).

270



275 **Figure 2:** Family controls differ based on their risk to develop IBD, but not based on the IBD subtype predominant in their families. (A) Group centroids of healthy individuals from IBD-affected families (family controls = FC) grouped by IBD subtype predominant in the respective family (yellow = UC, red = CD; $p_{\text{PERMANOVA}} > 0.05$), and by disease risk (dark blue = high risk = siblings and children of IBD-affected individuals, light blue = low risk = parents, partners and higher degree relatives of IBD-affected individuals; $p_{\text{PERMANOVA}} < 0.05$). (B) Alpha diversity of FCs grouped by family-predominant IBD subtype ($P > 0.05$) and disease risk ($P < 0.05$). Pre-onset cases healthy at the timepoint of sampling are marked. (C) Differentially abundant species between low risk vs. high risk FCs (dark blue); between unrelated HC and high risk FCs (light blue); and between UC FCs vs CD FC (red). Filled shapes: $q < 0.05$, empty shapes: $0.05 < q < 0.1$. The orientation of arrows reflects the effect direction of the difference in abundance with respect to the second group in the pairwise comparison (e.g. *A. muciniphila* is reduced in CD FCs compared to UC FCs, and reduced in FC High compared to HCs). The IBD column marks associations with IBD (orange) and IBD-subtypes (CD in red) as derived from the meta-analysis (Figure 1).

280

285

We also searched for microbial species that are differentially abundant in High FC vs. Low FC, in UC FC vs. CD FC, and in HC vs. High FC (Figure 2C), and compared these taxonomic hits against those found to have abundance signatures of IBD cases (Figure 1H). We found multiple taxa that are depleted in IBD or CD that are also depleted in either FC CD, such as *Akkermansia muciniphila*, or depleted in FC High compared to either FC Low or HCs, such as *Akkermansia*

290

295 *muciniphila* or *Methanobrevibacter A smithii* (Figure 2C). In addition, the CD-associated species *Bacteroides sp902362375* was found to be increased in FC High individuals compared to HCs. Altogether, these results suggest that FC High and CD FC individuals also exhibit microbiome signatures of IBD at the level of individual taxa or at the level of the community. These early shifts in the microbiome of FC individuals may potentially contribute to predisposing them to disease onset, in combination with other genetic and environmental factors, or serve as early markers for predicting disease risk. Further work will be needed to test and validate this hypothesis.

300

Enterotypes and their dynamics associate with IBD subtypes

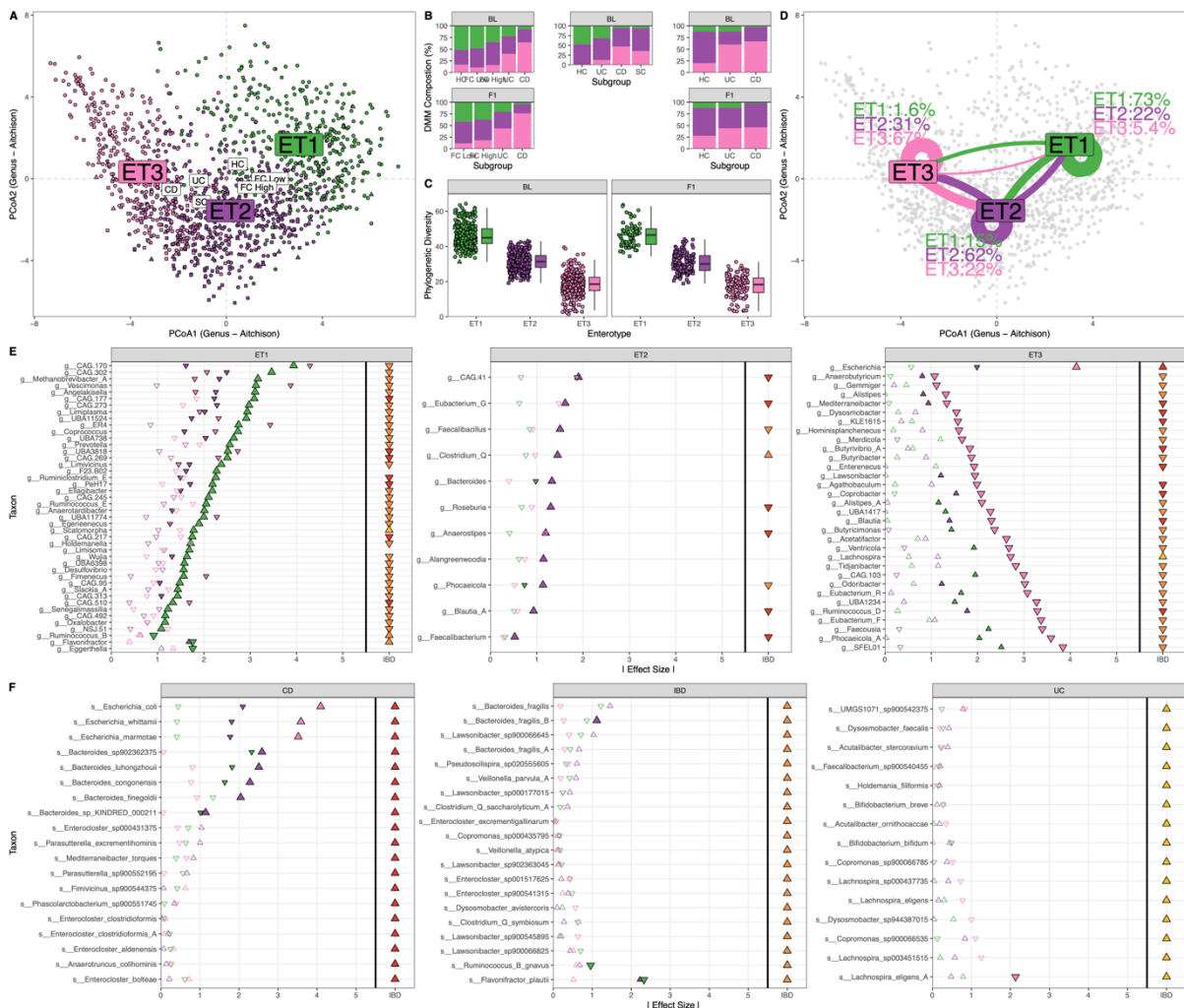
305 Microbiome community configurations, known as enterotypes, have been linked to various chronic inflammatory disorders affecting the gut and other organs (Valles-Colomer et al., 2019; Vieira-Silva et al., 2019). Here, we employed Dirichlet multinomial mixture (DMM) (Holmes et al., 2012) models to assess enterotypes based on genus-level abundances. We identified that a three-type model (ET1-3) explained best the diversity in community compositions in our data, based on Laplace approximation (Figure 3A). Next, we analyzed the distribution of our different disease and family control categories across the three enterotypes. ET1 was found to be the most prevalent among HC and FC Low individuals, particularly in the European cohorts, and poorly represented among CD individuals (Figure 3B). We observed a gradual loss of phylogenetic diversity from ET1 to ET3, ET2 being intermediate (Figure 3C). ET2 is more prevalent than ET1 among UC cases but is also more common in FC High compared to FC Low, again suggesting that FC High microbiomes are exhibiting signatures of IBD dysbiosis. Across all cohorts, ET3 is the most prevalent among CD cases, confirming their overall decreased alpha diversity (Figure 1). Examining community dynamics between timepoints in the German and USA cohorts, we observed that approximately two-thirds of individuals maintain their community configuration. Temporal shifts in enterotype most often involved transitions from or to the intermediate enterotype ET2, rarely occurring between ET1 and ET3 (Figure 3D). The high phylogenetic diversity of ET1 is associated with taxa that are usual markers of high phylogenetic diversity, including genera like *Prevotella* and *Methanobrevibacter A*. Many of ET1-associated taxa are also found to be depleted in IBD and IBD subtypes (Figure 3E). ET2 is associated with a few genera, specifically *Bacteroides*, suggesting that ET1 and ET2 represent the classical *Prevotella* and *Bacteroides* enterotypes, respectively, as found in previous studies (Costea et al., 2018). In contrast, ET3 represents a more 'dysbiotic' community state, characterized by high abundances of *Escherichia coli*, in addition to a low phylogenetic diversity.

310

315

320

325



330

Figure 3: IBD is reflected in microbial community composition and diversity. (A) Distribution of derived enterotypes (ET) in the ordination plot. Colors throughout all panels - ET1: green; ET2: purple; ET3: pink. (B) Distribution of enterotypes within countries, study subgroups and timepoints. (C) Phylogenetic diversity by enterotype and timepoints across all cohorts. (D) Enterotype dynamics between timepoints. (E) Genus-level associations with the three enterotypes. Triangle symbols pointing downward or upward reflect decreased or increased abundances, respectively. Filled shapes represent significant adjusted P-values < 0.05. (F) Species enriched in IBD and IBD subtypes (CD on the left, IBD in the middle, and UC on the right), and their association with enterotypes.

335

340

While enterotypes are traditionally reconstructed based on genus-level profiles, as initially introduced using 16S rRNA gene amplicon data, metagenomic datasets have the potential to significantly enhance the resolution of enterotype analyses. We investigated the link between enterotypes and the 54 species positively associated with IBD (Figure 3F). Interestingly, species found increased in IBD or IBD subtypes are not all associated with specific enterotypes. However, three *Escherichia* species that are markers of CD are significantly enriched in the dysbiotic ET3 enterotype. Five CD-associated *Bacteroides* species were found enriched in the ET2 enterotype, with four of these being less abundant in ET1. None of the *Enterocloster* species that are enriched in CD (Figure 1F and G) are associated to a specific enterotype. As CD cases can be found across all enterotypes, this result suggests that CD-specific *Enterocloster* species have the potential to be used as microbial markers of CD with high sensitivity, independently of other characteristics of dysbiosis, such as the increase in

350

355 *Escherichia* species. We did not find strong associations between UC microbial markers and enterotypes, with the exception of *Lachnospira eligens* A, which is significantly decreased in ET3. For taxa associated to IBD irrespective of disease subtypes, two taxa – *Bacteroides fragilis* B and *Flavonifractor plautii* – are found enriched in ET2. In addition, *Ruminococcus B gnavus* and *Flavonifractor plautii* are found in lower abundance in ET1. This analysis on individual taxa complements observations made at the community level and further suggests that ET2 represents an intermediary, and possibly predisposing, community state for IBD, compared to
360 ET1.

Decreased amino acid auxotrophies in Crohn's disease associated communities

Next, we investigated the association between IBD and species-level or community-level
365 amino acid (AA) auxotrophies. AA auxotrophies in microbial species and communities can be a marker of metabolic cross-feeding and their level in the gut microbiome were found to be associated with long term microbiome stability (Starke et al., 2023). We first measured species-level AA auxotrophies from genome-scale metabolic models that we built from our MAG collection (see Methods). We found strong signals of AA auxotrophy depletion in IBD in
370 general and in CD in particular (Figure 4A). The lack of significant associations between auxotrophy and UC suggests that the significant depletion hits in IBD in general are mainly driven by the depletions found in CD. While there is a global trend for reduced auxotrophy in the microbiome of disease cases, we found that taxa that are auxotroph for glycine (Gly) are increased in abundance in IBD (Figure 4C). We next analyzed the abundance weighted mean
375 number of auxotrophies per metagenomic sample in health and disease as a proxy for community-level measurement of auxotrophy. We found a significant reduction in auxotrophy levels in both UC and CD, with a stronger effect in CD individuals (Figure 4B). AAs with auxotrophy that are the strongest markers of disease are tyrosine (Tyr) auxotrophy (Figure 4D), but also tryptophan (Trp), which was significantly reduced in CD. Tryptophan metabolism
380 was previously shown to be altered in (active) Crohn's disease and was also found to be changed in abundance in the pathway-level analysis (Figure 1I). Community-level auxotrophy estimates are negatively correlated with phylogenetic diversity, independent of cohort subgroups (Figure 4F and 4G). Next, we wondered whether the loss of auxotrophy measured in disease was solely due to the loss of diversity. For this, we tested whether the number of
385 AA auxotrophies in individual species is correlated with their association with IBD and IBD subtypes measured in the differential abundance analysis (Figure 1H). We found significant negative correlations between the number of auxotrophies in species and the effect size of for their association with IBD in general ($P = 0.025$) and with CD ($P = 0.026$) (Figure 4H). These results suggest that the observed reduced auxotrophies in disease are not simply an effect of
390 depleted community diversity but are also driven by the increase in abundance of IBD and CD-associated taxa that have low number of auxotrophies. Overall, these results reveal shifts in the symbiotic relationships between gut bacteria in IBD and suggest that transitions towards more selfish metabolic and ecological behaviors occur in IBD microbial communities. Experimental approaches will be needed to validate this interpretation.

395

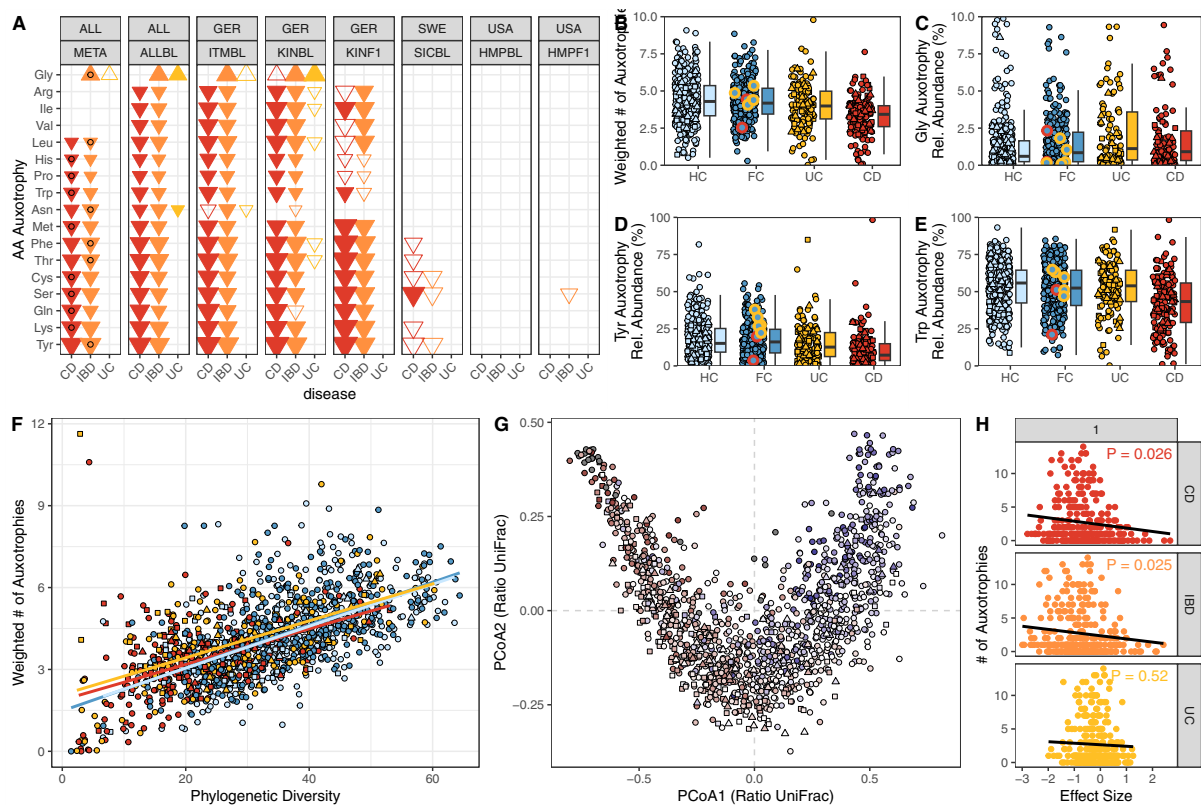


Figure 4: Widespread reduction of amino acid (AA) auxotrophies in IBD and specifically CD. (A) Results of the meta-analysis and individual cohort-level analyses for AA auxotrophies in IBD. Filled shapes represent adjusted P-values < 0.05, empty shapes unadjusted P-values < 0.05. Circles mark the association with the highest likelihood. (B) Abundance weighted levels of AA auxotrophies across study groups. Examples of increased (C: Gly) and reduced (D: Tyr; E: Trp) cumulative abundances of AA-auxotrophic clades in disease subgroups. (F) Correlation of abundance weighted auxotrophy levels and phylogenetic diversity per sample and by study group. (G) Distribution of abundance weighted auxotrophy levels across community compositions. (H) Correlation of species-level number of auxotrophies with effect sizes measured in the differential abundance analysis presented in Figure 1H.

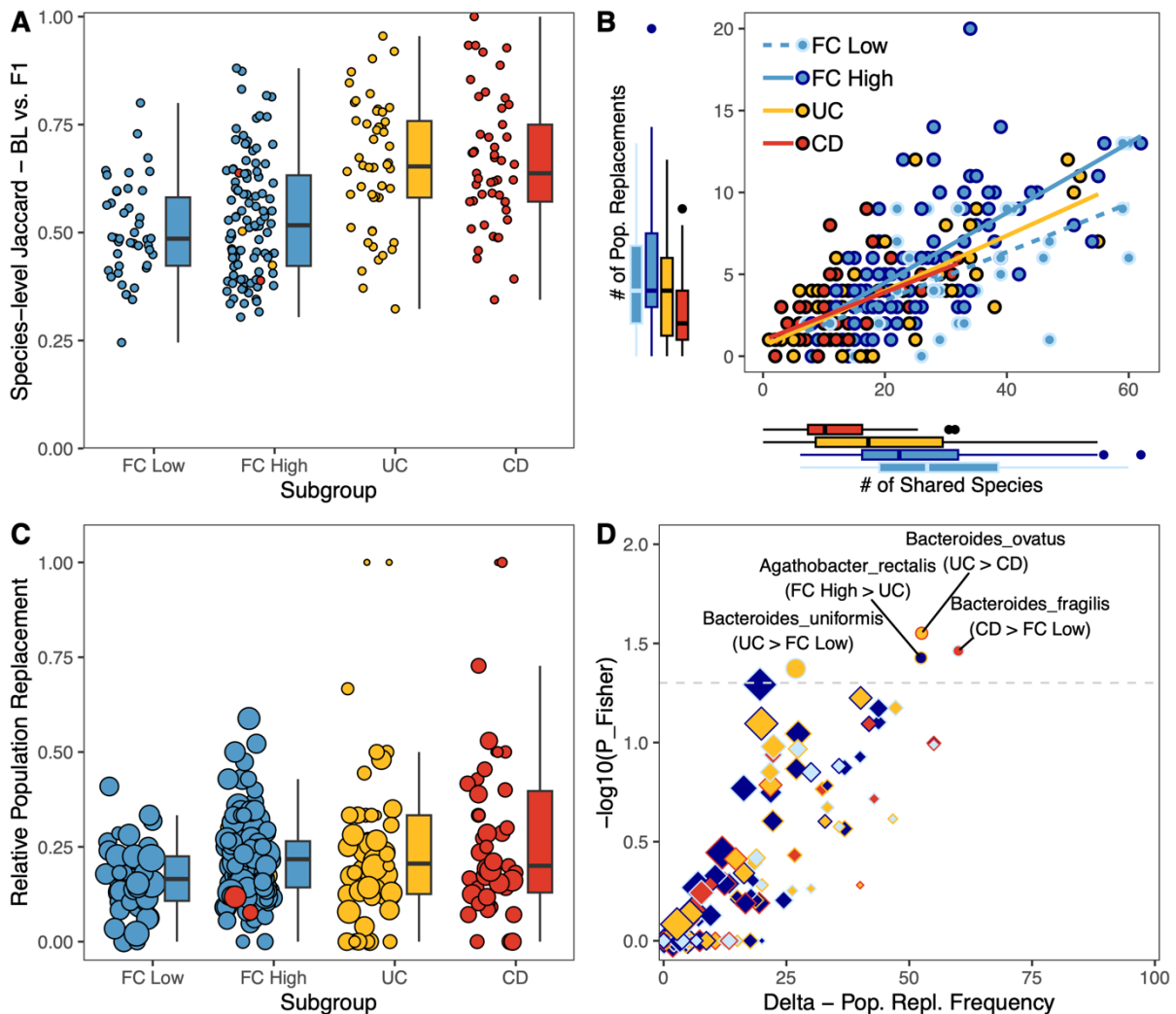
410

Longitudinal microbiome dynamics in IBD are changed at multiple scales

Longitudinal dynamics at the community level have been demonstrated to undergo alterations in IBD (IBDMDB Investigators et al., 2019). In our analysis of presence/absence data in the German cohort over approximately a 2-year span, we identified a significantly larger turnover of species at the species level in both IBD subtypes (Figure 5A). At the strain-level resolution, we observed that occurrences of strain replacements within species ($\text{popANI} > 0.99999$) were closely associated with the total number of shared species between timepoints (Figure 5B). However, when normalized by the total number of shared species between timepoints, we noted an increased relative population replacement in high-risk Family Controls (FCs) and Crohn's Disease (CD) compared to low risk FCs ($P_{\text{Bonferroni}} < 0.05$; Figure 5C). Additionally, relative population replacements in Ulcerative Colitis (UC) compared to low risk FCs and CD compared to high risk FCs were also statistically significant ($P_{\text{unadjusted}} < 0.05$). These findings validate the existence of altered ecological dynamics specifically in Crohn's disease at various taxonomic resolutions. Examining strain-level population replacement rates within

425

individual species, we identified several *Bacteroides* species as well as *Agathobacter rectalis* to be differentially affected ($P_{\text{Fisher}} < 0.05$; Figure 5D).



430

Figure 5: Altered ecological dynamics of individual gut bacteria in IBD subgroups. (A) Species-level jaccard distance between timepoints in the German cohort (B) Correlation of the number of shared species between timepoints and the number of strain-level population replacements derived from the inStrain analysis (C) Relative strain-level population replacement by study subgroup. (D) Pairwise comparison of population replacement frequencies within species and between study subgroups. Fill colors represent the subgroup for which the increased replacement frequencies were found, color of the outline represents the subgroup it was compared to. Shape sizes represent the total number of individuals the species was found shared across both timepoints.

435

440

Discussion

445 We analyzed fecal metagenomic samples from 987 healthy controls and IBD patients from
Germany, Sweden and the USA, including treatment-naïve cases and family controls, using
meta-analysis approaches. We identified consistent cross-cohort community- and species-
level associations at the level of taxa, pathways and auxotrophic metabolisms that are likely
450 universal microbial markers of IBD and IBD subtypes. We also found microbial association
patterns with IBD that reflect the geographic origin of the cohorts. Collectively, our findings
suggest a decline in metabolic interactions in the IBD microbiome, specifically in CD-
associated microbial communities, with an increased tendency toward self-sustained survival
in an inflamed gastrointestinal tract. Whether this loss of altruism precedes dysbiosis or is its
455 eventual outcome remains unclear. The building of the IBD-FC cohort aimed specifically at
identifying early microbiome signatures of IBD. While the number of incident cases in our
cohort remains too low to glean insights into these pre- and peri-onset microbiome dynamics,
the inclusion and analysis of the microbiome of at-risk family controls revealed significant IBD-
like shifts in the microbiome.

460 In our analyses, we combined likelihood-based prioritization of linear regression estimates to
assign taxonomic associations to IBD subtypes, differentiating CD or UC-specific signatures
from responses of microbes to IBD contexts shared between the two diseases, such as
intestinal inflammation. Our analyses identified multiple *Enterocloster* species, including *E.*
clostridioformis and *E. bolteae*, that are specifically enriched in CD and that may represent
465 microbial markers of this IBD subtype. *Enterocloster* taxa have been somewhat overlooked in
discussions about CD, potentially stemming from the limitations in resolution inherent in
amplicon-based studies. In addition, previous investigations reported decreases in
unclassified Lachnospiraceae (Ma et al., 2022), a family to which *Enterocloster* belongs,
possibly contributing to the oversight of *Enterocloster* taxa, which then behaves opposite to
470 the other clades within this family. Furthermore, their relatively recent re-classification from
the (*Lachno*-)*Clostridium* genus into the *Enterocloster* genus may have added to this neglect.
Enterocloster was identified and found associated to IBD in metagenomic data of a large Dutch
cohort (Vich Vila et al., 2018), and to exhibit enriched transcriptional activity in IBD (IBDMDB
Investigators et al., 2019). Yet, a clear association of specific *Enterocloster* species to CD was
475 not investigated. Our high-resolution taxonomic calling approach that is based on dataset-
specific MAGs found clear distinctions between different *Enterocloster* species, identifying
some that are specifically associated with CD, such as *E. clostridioformis* and *E. bolteae*. Other
Enterocloster species, such as *E. excrementigallinarium*, were found to be associated with IBD
in general. The functional implications of CD-specific *Enterocloster* taxa in gut homeostasis or
480 inflammatory processes will require dedicated experimental studies. Whether *Enterocloster*
species, classified as obligate anaerobes, can cope with oxygenic environments in the inflamed
gut also remains to be investigated in future work.

We also found multiple pathways specifically associated with CD. Among those, the elevated
level of the degradation of autoinducer AI-2 pathway is of particular interest. AI-2 plays a
485 crucial role in microbial quorum sensing and cell-to-cell communication (Pereira et al., 2013).
Moreover, the intestinal epithelium secretes AI-2 mimics upon contact with bacteria,

underscoring its involvement in cross-kingdom communication and the interaction between host and microbes (Ismail et al., 2016). Together with the observed decrease in AA auxotrophies in CD, these findings reinforce the idea of an inflammation-associated shift of the microbiome toward a more 'selfish' and less cooperative microbial community.

The microbiome of individuals affected by UC exhibits less pronounced differences compared to healthy controls than those observed in CD. Despite the general reduction in diversity associated with IBD in general and widespread taxonomic changes also applicable to UC, subtype-specific alterations are less prominent. Only six species are specifically depleted in UC, including four within the genus *Bacteroides*. This finding starkly contrasts with the positive association of *Bacteroides fragilis* and other *Bacteroides* species with IBD in general (Figure 1I), highlighting the significance of species-level resolution in microbiome studies. This level of granularity enables the investigation of distinct and opposing changes within a single microbial genus associated with the disease. Nonetheless, the UC-specific increases in species belonging to the genera *Acutalibacter*, *Copromonas* (both Oscillospirales), *Lachnospira*, *Dysosmobacter* (both Lachnospiraceae), and *Bifidobacterium* (*B. breve* and *B. bifidum*) could potentially elucidate UC-related processes and contribute to the microbiome-based differentiation of IBD subtypes. The notable rise in *Bifidobacterium breve* is particularly intriguing, presenting a significant departure from previous reports that identified *B. breve*, along with other *Bifidobacterium* species, for its potential to ameliorate DSS-induced colitis in mice (Singh et al., 2020). However, additional studies have documented strain-specific effects for *B. breve*, and human probiotic intervention trials have indicated no discernible impact of oral *B. breve* supplementation on relapse-free survival (Matsuoka et al., 2018). None of the other identified clades have been identified as associated with UC in any metagenome-based analysis, and they are generally insufficiently characterized in the context of the gut microbiome. Their role in UC awaits further clarification.

Pathways exhibiting increased abundances in UC hint at altered bile acid metabolism, specifically linked to shifts in microbial communities. Previous studies have demonstrated modified compositions of primary and secondary bile acids in individuals with IBD, particularly during episodes of acute colitis (Thomas et al., 2022). We found that bile acid epimerization and iso-bile acid biosynthesis pathways are increased in UC. Iso-bile acids are less toxic to bacteria (Devlin & Fischbach, 2015), their stable 7-oxo-intermediates are thought to actively affect host steroid hormone levels (Odermatt & Klusonova, 2015) and are non-active agonists of the farnesoid X receptor (FXR), blocking downstream signaling and production of antimicrobial peptides, such as cathelicidin (Campbell et al., 2012). Thus, bile acid epimerization can manipulate host defense mechanisms in multiple ways and may contribute to pathogenesis in UC (Ridlon et al., 2016).

Regardless of the subtype, the microbiome of IBD cases exhibited a reduced abundance of over 150 microbial species compared to healthy controls. Other species, such as *Ruminococcus B gnavus*, *Flavonifractor plautii*, *Clostridium Q symbiosum*, and *Bacteroides fragilis* including two closely related strains *B. fragilis A* and *B* were found in increased abundances in the samples of IBD-affected individuals. These taxa have previously been identified to be associated with IBD (Hassouneh et al., 2021; Pisani et al., n.d.). *R. gnavus* has been extensively reported to be associated with multiple intestinal and extra-intestinal disorders (Crost et al., 2023) and was shown to actively promote inflammation through TLR4

dependent TNF-alpha secretion in response to the polysaccharide glucorhamnan (Henke et al., 2019). *F. plautii* and *C. symbiosum* were implicated in bloodstream infections (Elsayed & Zhang, 2004; Karpat et al., 2021), and *Bacteroides fragilis* is known to be an opportunistic pathogen (Wexler, 2007). *B. fragilis* virulence is likely expressed in response to oxygen exposure via the oxyR regulon (Rocha et al., 2003; Sund et al., 2008). The increased abundance of ROS in IBD-associated intestinal inflammation is potentially a trigger for the switch of *B. fragilis* from a commensal to pathogenic lifestyle. Yet, the contribution of *B. fragilis* to IBD progression remains to be experimentally demonstrated.

Intriguingly, we found that the hyaluronan degradation pathway (PWY-7645) was among the most increased in IBD. Hyaluronan is known to be deposited in the extracellular matrix (ECM) at inflamed sites, and is actively involved in the recruitment of immune cells (Petrey & de la Motte, 2019). It was also previously shown that IBD-associated bacteria are able to degrade ECM components, including hyaluronan (Porrás et al., 2022). How these dynamics influence IBD onset and progression needs to be investigated. Yet, the dysregulation of hyaluronan deposition and catabolism were shown to be involved in inflammatory processes and coagulation, increasing the risk for microvascular occlusion and thrombosis, which are known comorbidities in IBD (Petrey & de la Motte, 2019). Additionally, our analyses could confirm many associations previously reported in the microbiome of IBD patients, such as the mucus-degrading *Akermansia muciniphila*, *Alistipes spp.*, the *Prevotella copri* complex, *Gemmiger spp.*, and *Methanobrevibacter A smithii* (Vich Vila et al., 2018). The methanoarchaeon *M. smithii* is a known marker of healthy anaerobic gut microbiomes, as demonstrated by several studies focusing on the evolution of mammalian gut microbiomes (Rühlemann et al., 2023). By producing methane as a metabolic product in anoxic conditions through the fermentation of bacterial primary products, such as acetate, methanol, hydrogen and carbon dioxide (Thauer, 1998; Thauer & Shima, 2006), it can engage in syntrophic interactions with a broad range of bacteria (Samuel et al., 2007; Samuel & Gordon, 2006). Several studies demonstrated more efficient polysaccharide fermentation in the human intestine methanogenesis through the presence of *M. smithii*, as it consumes various bacterial fermentation products thereby lowering the overall redox potential (Miller et al., 1984; Samuel et al., 2007; Samuel & Gordon, 2006). Methane production is also linked to low transit time and anti-inflammatory effects in the colon (Boros et al., 2012; Pimentel et al., 2006). Finally, reduced *M. smithii* abundances are inversely correlated with the abundance of sulfate-reducing bacteria (Ghavami et al., 2018). These microorganisms produce toxic hydrogen sulfide, which was also implicated in IBD pathogenesis (Ghavami et al., 2018). Owing to the multiple key metabolic and physiological functions that *M. smithii* perform in the gut ecosystem, it was previously proposed that it may serve as a biomarker for IBD (Ghavami et al., 2018). Our study, which revealed reduced abundances of *M. smithii* in high-risk FCs compared to low-risk FCs or HCs provide additional support to this idea, as it suggests that the loss of *M. smithii* defines a pre-dysbiotic state that is on the path to IBD dysbiosis.

Our study emphasizes the potential to make novel discoveries into the role of the microbiome in IBD by employing meta-analysis approaches on metagenomic datasets derived from both prospective and family cohorts. The future expansion of the IBD-FC cohort is poised to play a

pivotal role in enhancing our comprehension of the intricate interactions among hosts, the environment, and the microbiome in the pathogenesis of IBD.

Methods

580

Cohort description and sampling

The IBD-FC is a German nationwide, prospective cohort study collecting both, questionnaire data and biomaterial from IBD patients and their (affected and unaffected) family members. 585 The recruitment of study participants started in October 2013 and is still ongoing. Follow-up information and new biomaterial samples are collected prospectively at intervals of approximately two years. All participants, including the IBD patients, provided questionnaire data and biomaterial (blood, stool, hair) at baseline and after every 2 years of follow-up. 590 Samples were taken at home by the participants and sent by mail to the Microbiome Laboratory of the Institute of Clinical Molecular Biology for further analysis. In addition, physician-administered questionnaires have been collected from IBD patients and onset cases). Every study participant has given written informed consent on forms that are age-adapted. For participants under the age of 18 years, the informed consent must also be signed by their parents. IBD-FC IBD diagnoses were curated to be grouped into the correct subtype 595 based on questionnaire and physician data of all available timepoints, using the latest diagnosis. The study was conducted in accordance with the Declaration of Helsinki. Ethics committee of the Medical Faculty of Kiel University gave ethical approval for this work (AZ A117/13).

600 Subjects from Northern Germany from a blood donor cohort were recruited by mail and asked to provide stool samples and detailed food, lifestyle and medical history questionnaires. Samples were taken at home and sent by mail within 24 h after sample collection to the Microbiome Laboratory for further analysis. After arrival stool samples were aliquoted and stored at -80°C until further processing. The study was conducted in accordance with the 605 Declaration of Helsinki. Ethics committee of the Medical Faculty of Kiel University gave ethical approval for this work (AZ A103/14). Metadata were assessed for the presence of chronic diseases and samples were included in case of missing questionnaire data, and if individuals were affected by IBD, irritable bowel syndrome, primary sclerosing cholangitis, primary biliary cirrhosis, or any type of diabetes. In total 112 out of 456 samples were excluded from the ITM 610 dataset.

The Swedish Inception Cohort in IBD (SIC IBD) was previously described and includes treatment-naïve patients, between 20–77 years of age (Bergemalm et al., 2021).

615 Quality-filtered metagenomic data from the human microbiome project (HMP) were downloaded for a subset of individuals for which a baseline sample and a sample at 6 months (180 day +/- 30 days) with $> 3 \times 10^6$ reads were available based on the information provided by the curatedMetagenomicData package for R (Pasolli et al., 2017), study_name == "HMP_2019_ibdmdb". 620

Data and material of the IBD-FC cohort are available upon the submission of a research proposal to <https://portal.popgen.de/>.

Data of the ITM cohort will be uploaded to ENA upon acceptance of this manuscript at a peer-reviewed journal.

625

Metagenome sequencing and data processing

Sample preparation and sequencing: Total fecal DNA samples were extracted from 200mg stool aliquots starting with bead beating in a SpeedMill PLUS (Analytik Jena AG) for 45 s at 50 Hz mixed with 1.1 mL ASL buffer in 0.70mm Garnet Bead tubes (Qiagen). Subsequently, samples were heated to 95°C (5 min) followed by centrifugation. 200 µl of the resulting supernatant were used for automated DNA extraction with the QIAcube system (Qiagen) using the QIAmp DNA Stool Mini Kit (Qiagen). DNA quality was assessed by Qubit and Genomic DNA ScreenTape (Agilent). Illumina Nextera DNA Library Preparation Kit was used to construct shotgun metagenomic libraries, and subsequently sequenced with 2 × 150 bp reads on a NovaSeq 6000 system (Illumina).

635

Quality control and binning: All metagenomic data were processed using the TOFU-MAaPO pipeline for metagenomic data (<https://github.com/ikmb/TOFU-MAaPO>). Briefly, sequencing data were quality controlled to remove low quality reads, sequencing adaptors and PhiX spike-in reads using BBTools (Bushnell, 2014). Human host-derived reads were removed using Bowtie2 (Langmead & Salzberg, 2012) and the hg19 version of the human reference genome. HMP data were already quality filtered and the above steps were skipped for these samples. Assembly of metagenomic contigs was performed with Megahit (D. Li et al., 2015) using k=27....127 for each samples individually, retaining contigs > 2kbp. Reads were mapped to the resulting contigs using minimap2, resulting SAM files were converted to ordered BAM files with samtools and per-contig coverages were calculated using the jgi_summarize_bam_contig_depths (H. Li et al., 2009). MaxBin2 (Wu et al., 2016), Metabat2 (Kang et al., 2019), and CONCOCT (Aneberg et al., 2014) were used for binning. Additionally, samples were allocated to groups of 37-54 based on their sample IDs for binning with VAMB (Nissen et al., 2021). Within each group, contigs were concatenated and reads were mapped to this catalog using minimap2, following the same procedure as described above to create VAMB input files. GTDBtk (Chaumeil et al., 2022) bac120 and ar53 marker genes were identified from the prodigal (Hyatt et al., 2010) output of the contigs in --meta mode, which were annotated using HMMer (Eddy, 2011) and used for bin scoring, refinement and estimation of completeness/contamination with the MAGScoT tool (Rühlemann et al., 2022), resulting MAGs with a score > 0.5 and contamination < 10% were retained for downstream analyses.

640

645

650

655

Species-level genome bin (SGB) catalog: We applied a multi-step clustering to the collection of MAGs to minimized SGB catalog inflation due to low-quality MAGs. Within groups, MAGs were dereplicated at 97% identity using dRep (Olm et al., 2017), for each cluster the highest scoring MAG was chosen as representative. Pairwise mash distances of good quality (scores >= 0.7) cluster representatives across all groups were calculated and used for primary single-linkage clustering at 90% similarity. Within primary clusters, Average nucleotide identities (ANI) were calculated using fastANI (Jain et al., 2018) and average linkage clustering was performed resulting in a set of high-quality SGBs selecting the highest scoring MAGs as SGB representatives. Medium quality MAGs (MQ MAGs; 0.5 < score < 0.7) were compared against SGB representatives using fastANI and MQ MAGs with >= 95% ANI to an SGB were assigned

660

665

670 to its cluster. MQ MAGs without matching SGBs (all ANI < 0.95) were subjected to ANI-based
average-linkage clustering as described above and added to the SGB catalog only if they were
recovered from at least two samples across the dataset, removing singleton MQ MAGs from
the dataset. SGB representatives were subjected for taxonomic annotation using GTDBtk
675 (v2.1.0; (Chaumeil et al., 2022)) and GTDB release 214 as reference database. GTDBtk marker
gene alignments of representative SGVs were used to reconstruct a phylogenetic tree using
IQTREE (Minh et al., 2020; Nguyen et al., 2015). Representative SGB genomes were
concatenated and indexed as reference for salmon (Patro et al., 2017).

Abundance estimation: Clean metagenomic reads were mapped against the SGB catalog
680 using salmon quant in --meta mode and using the --validateMappings option (Patro et al.,
2017). Resulting contig-level mappings were used to estimate SGB-level abundances. For this,
a contig was considered to be present in a sample if it had at least 10% coverage, calculated
as the number of mapping reads x 2 x 150bp / effective contig length. An SGB was considered
685 present if at least 20% of the cumulative effective contig lengths qualified as present and at
least 1000 reads and 250 counts per million (CPM) were mapped to the SGB. Using GTDBtk
annotations of SGBs, domain- to species-level abundances were calculated as cumulative CPM
abundances of the respective level and clade. CPMs are genome-size and sequencing-depth
adjusted with an upper bound of 10⁶ per sample, thus representing relative abundances. At
690 each taxonomic level, CPMs were used to calculate centered-log ratio (CLR) abundances to
account for compositionality of microbiome data, applying the clr() function of the
compositions package for R and adding a pseudocount of 1 (van den Boogaart & Tolosana-
Delgado, 2008). Resulting negative CLR values were set to 0.

Gene/protein catalog: Prodigal (Hyatt et al., 2010) outputs of predicted amino acid sequences
695 of coding regions for all samples were concatenated and clustered at 95% identity using
mmseqs v2 in “easy-linclust” mode with options “--cov-mode 1 -c 0.8 --kmer-per-seq 80”.
Resulting representative sequences were renamed to carry standardized IDs and subjected to
annotation with emapper.py (Cantalapiedra et al., 2021) using the eggNOG 5.0 (release
220425; (Huerta-Cepas et al., 2019)) database as reference. Further, the 95% identity catalog
700 was used to create catalogs at coarser-level resolution of 80% and 50% identity. EggNOG
annotations of 95% identity clusters were collapsed to represent broader-level protein
clusters. The 50% identity catalog was used to create a DIAMOND reference (Buchfink et al.,
2015) to which metagenomic reads were mapped and gene/protein-level functional CPM
abundances were estimated.

705 **Metabolic modelling and prediction of amino acid auxotrophies:** Amino acid auxotrophies
were predicted using computational metabolic modelling based on genome-scale metabolic
networks as described previously (Starke et al., 2023). In brief, genome-scale metabolic
networks were reconstructed for species-representative MAGs using *gapseq* version 1.2
710 (Zimmermann et al., 2021). Amino acid auxotrophies were predicted using flux balance
analysis, where the objective function was defined as the flux through the biomass formation
reaction. An organism was defined as auxotrophic for a specific amino acid if the
corresponding metabolic model could form biomass in the original medium but not in the
medium without the amino acid of interest. Flux balance analysis was performed in R (v4.3.1),
715 the R package *sybil* v2.2.0 (Gelius-Dietrich et al., 2013), and the IBM ILOG CPLEX optimiser as
linear programming solver.

Quantifying the abundance of predicted metabolic pathways: Quantitative compositional pathway abundances per metagenome sample were estimated by combining TPM values for species-representative MAGs and the MAGs' pathway presence/absence predictions from the metabolic modelling approach (see above). Specifically, the abundance q of pathway i in sample j was calculated as

$$q_{ij} = \sum_{j \in M} a_{ij} t_j.$$

M is the set of all MAGs, a_{ij} is the prediction of pathway i in organism j (0 if absent; 1 if present), and t_j is the TPM of MAG j .

Downstream data processing and statistical analysis

All further processing and statistical analyses were carried out in R 4.2.1 (R Core Team, 2022) and utilizing the tidyverse collection of packages (Wickham et al., 2019). Alpha diversity was assessed as Faith's phylogenetic diversity using SGB-level abundances and the reconstructed phylogenetic tree in the `pd()` function of the `picante` package (Kembel et al., 2010). Differences in alpha diversity between groups were assessed using linear regression. Ordinations are based on ratio UniFrac distances of SGB-level abundances and the reconstructed phylogenetic tree (Lozupone et al., 2011; Wong et al., 2016). Differences in beta diversity were assessed using permutational multivariate analysis of variance using distances matrices implemented in the `adonis2()` function of the `vegan` package (Oksanen et al., 2022).

Differential abundance analysis

Stratified by country, abundance differences of taxonomic groups in healthy controls and IBD-affected individuals were assessed using CLR-transformed abundances and linear models with age, sex and BMI as covariates, except for the SIC-IBD cohort where BMI data were not available. Tested were all taxa with overall prevalence of $\geq 30\%$ or within-group prevalence $\geq 20\%$. Group comparisons performed were (1) all IBD vs HC, (2) CD vs. all non-CD (HC + UC), and (3) UC vs. all non-UC (HC + CD). Similarly, assessing differences in family controls and IBD-affected individuals, linear mixed models were used with 'family ID' as random effects variable in addition to the fixed effects named above. For the cross-country comparison of HC and IBD cases, linear mixed models were used with 'country' as random effects variable and age + sex as fixed effect variables. This model was used to sort taxa into being enriched/depleted with either IBD in general or one of its subtypes based on the log-likelihood of the model. Inverse-variance-weighted meta-analysis was used on the analyses of HC and IBD-affected individuals to obtain final association statistics. P-values from the meta-analysis and individual linear models were FDR- or Bonferroni-corrected for multiple testing, respectively. Pathway abundance-based analyses were performed as described above for taxonomic groups, however abundances were log-transformed as these data are not compositional and CLR-transformation would not be appropriate.

Enterotype inference and analysis

The number of enterotypes was assessed based on HCs from all countries using genus-level abundances and Dirichlet multinomial mixtures implemented in the `dmn()` function of the `DirichletMultinomial` package for R (Holmes et al., 2012). Optimal number of clusters between 1 and 8 were assessed using the Laplace estimator for goodness-of-fit. The respective model

765 with $k=3$ clusters was used to classify non-HC samples into one of three enterotypes. Differential abundances of taxonomic groups between enterotypes were assessed using linear models, including age + sex + BMI as covariates and comparing samples from one enterotype against the other two. Assignments of enterotype-specific increase or decrease of abundances were based on a Bonferroni-corrected P-values < 0.05 and highest log-likelihood.

770 **Auxotrophy analysis**

Differential abundance analysis of amino acid auxotrophic microbes is described above. Mean per-sample auxotrophies were estimated from SGB-level auxotrophy counts weighted by their abundances. Auxotrophy abundance-based analyses were performed as described above for taxonomic groups, however abundances were log-transformed as these data are not compositional and CLR-transformation would not be appropriate.

Analysis of strain-level dynamics

780 MAGs of samples for which two timepoints were subjected to dereplication with dRep to obtain a set of non-redundant SGBs per sample, the SGB representative sequence was chosen based on its completeness and contamination and concatenated in a per-sample SGB (sampSGB) catalog. This catalog was used as bowtie2 reference, reads from both timepoints were mapped against the reference individually and the resulting SAM file was used as input in inStrain (Olm et al., 2021) in 'profile' mode with an additional file listing contig-to-sampSGB mappings. The resulting profiles were then compared in the inStrain 'compare' mode. An
785 sampSGB was considered present at a timepoint if it had a higher than 3x coverage. A strain-level population replacement was present in the case of population average nucleotide identity (popANI) < 0.99999 between timepoints. Pairwise species-level enrichments of population replacements between groups were assessed using Fisher's exact test based on the counts of species-sharing and detected population replacements. Species were included
790 in the comparison for a subgroup if they were found shared between timepoints in at least 5 samples in the respective groups.

References

- 795 Abbas-Egbariya, H., Haberman, Y., Braun, T., Hadar, R., Denson, L., Gal-Mor, O., & Amir, A. (2022). Meta-analysis defines predominant shared microbial responses in various diseases and a specific inflammatory bowel disease signal. *Genome Biology*, *23*(1), 61. <https://doi.org/10.1186/s13059-022-02637-7>
- 800 Abdel-Rahman, L. I. H., & Morgan, X. C. (2023). Searching for a Consensus Among Inflammatory Bowel Disease Studies: A Systematic Meta-Analysis. *Inflammatory Bowel Diseases*, *29*(1), 125–139. <https://doi.org/10.1093/ibd/izac194>
- 805 Aden, K., Rehman, A., Waschina, S., Pan, W.-H., Walker, A., Lucio, M., Nunez, A. M., Bharti, R., Zimmerman, J., Bethge, J., Schulte, B., Schulte, D., Franke, A., Nikolaus, S., Schroeder, J. O., Vandeputte, D., Raes, J., Szymczak, S., Waetzig, G. H., ... Rosenstiel, P. (2019). Metabolic Functions of Gut Microbes Associate With Efficacy of Tumor Necrosis Factor Antagonists in Patients With Inflammatory Bowel Diseases. *Gastroenterology*, *157*(5), 1279–1292.e11. <https://doi.org/10.1053/j.gastro.2019.07.025>
- 810 Alatab, S., Sepanlou, S. G., Ikuta, K., Vahedi, H., Bisignano, C., Safiri, S., Sadeghi, A., Nixon, M. R., Abdoli, A., Abolhassani, H., Alipour, V., Almadi, M. A. H., Almasi-Hashiani, A., Anushiravani, A., Arabloo, J., Atique, S., Awasthi, A., Badawi, A., Baig, A. A. A., ... Naghavi, M. (2020). The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology & Hepatology*, *5*(1), 17–30. [https://doi.org/10.1016/S2468-1253\(19\)30333-4](https://doi.org/10.1016/S2468-1253(19)30333-4)
- 815 Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), Article 11. <https://doi.org/10.1038/nmeth.3103>
- 820 Baumgart, D. C., & Carding, S. R. (2007). Inflammatory bowel disease: Cause and immunobiology. *The Lancet*, *369*(9573), 1627–1640. [https://doi.org/10.1016/S0140-6736\(07\)60750-8](https://doi.org/10.1016/S0140-6736(07)60750-8)
- Baumgart, D. C., & Sandborn, W. J. (2007). Inflammatory bowel disease: Clinical aspects and established and evolving therapies. *The Lancet*, *369*(9573), 1641–1657. [https://doi.org/10.1016/S0140-6736\(07\)60751-X](https://doi.org/10.1016/S0140-6736(07)60751-X)
- 825 Bergemalm, D., Andersson, E., Hultdin, J., Eriksson, C., Rush, S. T., Kalla, R., Adams, A. T., Keita, Å. V., D'Amato, M., Gomollon, F., Jahnsen, J., Ricanek, P., Satsangi, J., Repsilber, D., Karling, P., Halfvarson, J., Arnott, I. D., Bayes, M., Bonfiglio, F., ... Dominique. (2021). Systemic Inflammation in Preclinical Ulcerative Colitis. *Gastroenterology*, *161*(5), 1526–1539.e9. <https://doi.org/10.1053/j.gastro.2021.07.026>
- 830 Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., ... Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, *41*(11), 1633–1644. <https://doi.org/10.1038/s41587-023-01688-w>
- 835 Boros, M., Ghyczy, M., Érces, D., Varga, G., Tőkés, T., Kupai, K., Torday, C., & Kaszaki, J. (2012). The anti-inflammatory effects of methane. *Critical Care Medicine*, *40*(4), 1269–1278. <https://doi.org/10.1097/CCM.0b013e31823dae05>
- 840 Britton, G. J., Contijoch, E. J., Mogno, I., Vennaro, O. H., Llewellyn, S. R., Ng, R., Li, Z., Mortha, A., Merad, M., Das, A., Gevers, D., McGovern, D. P. B., Singh, N., Braun, J., Jacobs, J.

- P., Clemente, J. C., Grinspan, A., Sands, B. E., Colombel, J.-F., ... Faith, J. J. (2019). Microbiotas from Humans with Inflammatory Bowel Disease Alter the Balance of Gut Th17 and ROR γ t+ Regulatory T Cells and Exacerbate Colitis in Mice. *Immunity*, *50*(1), 212–224.e4. <https://doi.org/10.1016/j.immuni.2018.12.015>
- 845 Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), Article 1. <https://doi.org/10.1038/nmeth.3176>
- Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner*. <https://www.osti.gov/biblio/1241166>
- Campbell, Y., Fantacone, M. L., & Gombart, A. F. (2012). Regulation of antimicrobial peptide gene expression by nutrients and by-products of microbial metabolism. *European Journal of Nutrition*, *51*(8), 899–907. <https://doi.org/10.1007/s00394-012-0415-4>
- 850 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- 855 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2022). GTDB-Tk v2: Memory friendly classification with the genome taxonomy database. *Bioinformatics*, *38*(23), 5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>
- Chiodini, R. J., Dowd, S. E., Chamberlin, W. M., Galandiuk, S., Davis, B., & Glassing, A. (2015). Microbial Population Differentials between Mucosal and Submucosal Intestinal Tissues in Advanced Crohn’s Disease of the Ileum. *PLOS ONE*, *10*(7), e0134382. <https://doi.org/10.1371/journal.pone.0134382>
- 860 Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S. D., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O’Toole, P. W., Quince, C., Relman, D. A., ... Bork, P. (2018). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, *3*(1), 8–16. <https://doi.org/10.1038/s41564-017-0072-8>
- 865 Crost, E. H., Coletto, E., Bell, A., & Juge, N. (2023). Ruminococcus gnavus: Friend or foe for human health. *FEMS Microbiology Reviews*, *47*(2), fuad014. <https://doi.org/10.1093/femsre/fuad014>
- 870 Devlin, A. S., & Fischbach, M. A. (2015). A biosynthetic pathway for a prominent class of microbiota-derived bile acids. *Nature Chemical Biology*, *11*(9), 685–690. <https://doi.org/10.1038/nchembio.1864>
- 875 Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., & Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nature Communications*, *8*(1), 1784. <https://doi.org/10.1038/s41467-017-01973-8>
- Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLOS Computational Biology*, *7*(10), e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- 880 Elsayed, S., & Zhang, K. (2004). Bacteremia Caused by Clostridium symbiosum. *Journal of Clinical Microbiology*, *42*(9), 4390–4392. <https://doi.org/10.1128/JCM.42.9.4390-4392.2004>
- Gelius-Dietrich, G., Desouki, A. A., Fritzemeier, C. J., & Lercher, M. J. (2013). Sybil – Efficient constraint-based modelling in R. *BMC Systems Biology*, *7*(1), 125. <https://doi.org/10.1186/1752-0509-7-125>
- 885 Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C.,

- 890 González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., ... Xavier, R. J. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe*, 15(3), 382–392. <https://doi.org/10.1016/j.chom.2014.02.005>
- Ghavami, S. B., Rostami, E., Sephay, A. A., Shahrokh, S., Balaii, H., Aghdaei, H. A., & Zali, M. R. (2018). Alterations of the human gut *Methanobrevibacter smithii* as a biomarker for inflammatory bowel diseases. *Microbial Pathogenesis*, 117, 285–289. <https://doi.org/10.1016/j.micpath.2018.01.029>
- 895 Gibbons, S. M., Duvall, C., & Alm, E. J. (2018). Correcting for batch effects in case-control microbiome studies. *PLOS Computational Biology*, 14(4), e1006102. <https://doi.org/10.1371/journal.pcbi.1006102>
- Graham, D. B., & Xavier, R. J. (2020). Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature*, 578(7796), 527–539. <https://doi.org/10.1038/s41586-020-2025-2>
- 900 Halfvarson, J., Brislawn, C. J., Lamendella, R., Vázquez-Baeza, Y., Walters, W. A., Bramer, L. M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., McClure, E. E., Dunkleberger, M. F., Knight, R., & Jansson, J. K. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nature Microbiology*, 2(5), 17004. <https://doi.org/10.1038/nmicrobiol.2017.4>
- 905 Halme, L. (2006). Family and twin studies in inflammatory bowel disease. *World Journal of Gastroenterology*, 12(23), 3668. <https://doi.org/10.3748/wjg.v12.i23.3668>
- Häsler, R., Sheibani-Tezerji, R., Sinha, A., Barann, M., Rehman, A., Esser, D., Aden, K., Knecht, C., Brandt, B., Nikolaus, S., Schäuble, S., Kaleta, C., Franke, A., Fretter, C., Müller, W., Hütt, M.-T., Krawczak, M., Schreiber, S., & Rosenstiel, P. (2017). Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease. *Gut*, 66(12), 2087–2097. <https://doi.org/10.1136/gutjnl-2016-311651>
- 910 Hassouneh, S. A.-D., Loftus, M., & Yooseph, S. (2021). Linking Inflammatory Bowel Disease Symptoms to Changes in the Gut Microbiome Structure and Function. *Frontiers in Microbiology*, 12, 673632. <https://doi.org/10.3389/fmicb.2021.673632>
- 915 Henke, M. T., Kenny, D. J., Cassilly, C. D., Vlamakis, H., Xavier, R. J., & Clardy, J. (2019). *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences*, 116(26), 12672–12677. <https://doi.org/10.1073/pnas.1904099116>
- 920 Holmes, I., Harris, K., & Quince, C. (2012). Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. *PLOS ONE*, 7(2), e30126. <https://doi.org/10.1371/journal.pone.0030126>
- 925 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(Database issue), D309–D314. <https://doi.org/10.1093/nar/gky1085>
- 930 Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>
- IBDMDB Investigators, Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J.,

- 935 Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., ... Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, *569*(7758), 655–662. <https://doi.org/10.1038/s41586-019-1237-9>
- 940 Ismail, A. S., Valastyan, J. S., & Bassler, B. L. (2016). A Host-Produced Autoinducer-2 Mimic Activates Bacterial Quorum Sensing. *Cell Host & Microbe*, *19*(4), 470–480. <https://doi.org/10.1016/j.chom.2016.02.020>
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), Article 1. <https://doi.org/10.1038/s41467-018-07641-9>
- 945 Jutel, M., Blaser, K., & Akdis, C. A. (2006). The role of histamine in regulation of immune responses. *Chemical Immunology and Allergy*, *91*, 174–187. <https://doi.org/10.1159/000090280>
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. <https://doi.org/10.7717/peerj.7359>
- 950 Karpat, I., Karolyi, M., Pawelka, E., Seitz, T., Thaller, F., & Wenisch, C. (2021). Flavonifractor plautii bloodstream infection in an asplenic patient with infectious colitis. *Wiener Klinische Wochenschrift*, *133*(13–14), 724–726. <https://doi.org/10.1007/s00508-021-01877-0>
- 955 Keita, Å. V., Lindqvist, C. M., Öst, Å., Magana, C. D. L., Schoultz, I., & Halfvarson, J. (2018). Gut Barrier Dysfunction—A Primary Defect in Twins with Crohn’s Disease Predominantly Caused by Genetic Predisposition. *Journal of Crohn’s and Colitis*, *12*(10), 1200–1209. <https://doi.org/10.1093/ecco-jcc/jjy045>
- 960 Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. <https://doi.org/10.1093/bioinformatics/btq166>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- 965 Larabi, A. B., Masson, H. L. P., & Bäumlér, A. J. (2023). Bile acids as modulators of gut microbiota composition and function. *Gut Microbes*, *15*(1), 2172671. <https://doi.org/10.1080/19490976.2023.2172671>
- 970 Laserna-Mendieta, E. J., Clooney, A. G., Carretero-Gomez, J. F., Moran, C., Sheehan, D., Nolan, J. A., Hill, C., Gahan, C. G. M., Joyce, S. A., Shanahan, F., & Claesson, M. J. (2018). Determinants of Reduced Genetic Capacity for Butyrate Synthesis by the Gut Microbiome in Crohn’s Disease and Ulcerative Colitis. *Journal of Crohn’s and Colitis*, *12*(2), 204–216. <https://doi.org/10.1093/ecco-jcc/jjx137>
- 975 Lee, Y. W., Lee, K.-M., Lee, J. M., Chung, Y. Y., Kim, D. B., Kim, Y. J., Chung, W. C., & Paik, C.-N. (2019). The usefulness of fecal calprotectin in assessing inflammatory bowel disease activity. *The Korean Journal of Internal Medicine*, *34*(1), 72–80. <https://doi.org/10.3904/kjim.2016.324>
- 980 Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- 985 Lönnfors, S., Vermeire, S., Greco, M., Hommes, D., Bell, C., & Avedano, L. (2014). IBD and health-related quality of life—Discovering the true impact. *Journal of Crohn's and Colitis*, *8*(10), 1281–1286. <https://doi.org/10.1016/j.crohns.2014.03.005>
- 990 Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: An effective distance metric for microbial community comparison. *The ISME Journal*, *5*(2), Article 2. <https://doi.org/10.1038/ismej.2010.133>
- 995 Ma, S., Shungin, D., Mallick, H., Schirmer, M., Nguyen, L. H., Kolde, R., Franzosa, E., Vlamakis, H., Xavier, R., & Huttenhower, C. (2022). Population structure discovery in meta-analyzed microbial communities and inflammatory bowel disease using MMUPHin. *Genome Biology*, *23*(1), 208. <https://doi.org/10.1186/s13059-022-02753-4>
- 1000 Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, *17*(11), e1009442. <https://doi.org/10.1371/journal.pcbi.1009442>
- 1005 Matsuoka, K., Uemura, Y., Kanai, T., Kunisaki, R., Suzuki, Y., Yokoyama, K., Yoshimura, N., & Hibi, T. (2018). Efficacy of Bifidobacterium breve Fermented Milk in Maintaining Remission of Ulcerative Colitis. *Digestive Diseases and Sciences*, *63*(7), 1910–1919. <https://doi.org/10.1007/s10620-018-4946-2>
- Meade, S., Liu Chen Kiow, J., Massaro, C., Kaur, G., Squirell, E., Bressler, B., & Lunken, G. (2023). Gut microbiome-associated predictors as biomarkers of response to advanced therapies in inflammatory bowel disease: A systematic review. *Gut Microbes*, *15*(2), 2287073. <https://doi.org/10.1080/19490976.2023.2287073>
- 1010 Michielan, A., Basso, D., Martinato, M., Pathak, S., Banerjee, A., Oliva, L., Plebani, M., Sturniolo, G. C., & D'Incà, R. (2013). Increased antibody response to microbial antigens in patients with Crohn's disease and their unaffected first-degree relatives. *Digestive and Liver Disease*, *45*(11), 894–898. <https://doi.org/10.1016/j.dld.2013.01.017>
- 1015 Miller, T. L., Weaver, G. A., & Wolin, M. J. (1984). Methanogens and anaerobes in a colon segment isolated from the normal fecal stream. *Applied and Environmental Microbiology*, *48*(2), 449–450. <https://doi.org/10.1128/aem.48.2.449-450.1984>
- 1020 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- 1025 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ni, J., Wu, G. D., Albenberg, L., & Tomov, V. T. (2017). Gut microbiota and IBD: Causation or correlation? *Nature Reviews Gastroenterology & Hepatology*, *14*(10), 573–584. <https://doi.org/10.1038/nrgastro.2017.88>

- 1030 Nissen, J. N., Johansen, J., Allesøe, R. L., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Jensen, L. J., Nielsen, H. B., Petersen, T. N., Winther, O., & Rasmussen, S. (2021). Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology*, *39*(5), Article 5. <https://doi.org/10.1038/s41587-020-00777-4>
- 1035 Odermatt, A., & Klusonova, P. (2015). 11 β -Hydroxysteroid dehydrogenase 1: Regeneration of active glucocorticoids is only part of the story. *The Journal of Steroid Biochemistry and Molecular Biology*, *151*, 85–92. <https://doi.org/10.1016/j.jsbmb.2014.08.011>
- 1040 Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., ... Weedon, J. (2022). *vegan: Community Ecology Package*. <https://github.com/vegandevs/vegan>
- 1045 Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, *11*(12), Article 12. <https://doi.org/10.1038/ismej.2017.126>
- 1050 Olm, M. R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B. A., Morowitz, M. J., & Banfield, J. F. (2021). InStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology*, *39*(6), 727–736. <https://doi.org/10.1038/s41587-020-00797-0>
- 1055 Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S. P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., Schauer, D. B., Ward, D. V., Korzenik, J. R., Xavier, R. J., Bousvaros, A., & Alm, E. J. (2012). Non-Invasive Mapping of the Gastrointestinal Microbiota Identifies Children with Inflammatory Bowel Disease. *PLoS ONE*, *7*(6), e39242. <https://doi.org/10.1371/journal.pone.0039242>
- 1060 Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., Beghini, F., Malik, F., Ramos, M., Dowd, J. B., Huttenhower, C., Morgan, M., Segata, N., & Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nature Methods*, *14*(11), Article 11. <https://doi.org/10.1038/nmeth.4468>
- 1065 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon: Fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods*, *14*(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- 1070 Pereira, C. S., Thompson, J. A., & Xavier, K. B. (2013). AI-2-mediated signalling in bacteria. *FEMS Microbiology Reviews*, *37*(2), 156–181. <https://doi.org/10.1111/j.1574-6976.2012.00345.x>
- 1075 Petrey, A. C., & de la Motte, C. A. (2019). Hyaluronan in inflammatory bowel disease: Cross-linking inflammation and coagulation. *Matrix Biology*, *78–79*, 314–323. <https://doi.org/10.1016/j.matbio.2018.03.011>
- Pimentel, M., Lin, H. C., Enayati, P., van den Burg, B., Lee, H.-R., Chen, J. H., Park, S., Kong, Y., & Conklin, J. (2006). Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *290*(6), G1089–G1095. <https://doi.org/10.1152/ajpgi.00574.2004>
- Pisani, A., Rausch, P., Bang, C., Ellul, S., Tabone, T., Marantidis Cordina, C., Zahra, G., Franke, A., & Ellul, P. (n.d.). Dysbiosis in the Gut Microbiota in Patients with Inflammatory Bowel Disease during Remission. *Microbiology Spectrum*, *10*(3), e00616-22. <https://doi.org/10.1128/spectrum.00616-22>

- Plichta, D. R., Graham, D. B., Subramanian, S., & Xavier, R. J. (2019). Therapeutic Opportunities in Inflammatory Bowel Disease: Mechanistic Dissection of Host-Microbiome Relationships. *Cell*, *178*(5), 1041–1056. <https://doi.org/10.1016/j.cell.2019.07.045>
- 1080 Porras, A. M., Zhou, H., Shi, Q., Xiao, X., JRI Live Cell Bank, Longman, R., & Brito, I. L. (2022). Inflammatory Bowel Disease-Associated Gut Commensals Degrade Components of the Extracellular Matrix. *MBio*, *13*(6), e02201-22. <https://doi.org/10.1128/mbio.02201-22>
- 1085 R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- 1090 Raygoza Garay, J. A., Turpin, W., Lee, S.-H., Smith, M. I., Goethel, A., Griffiths, A. M., Moayyedi, P., Espin-Garcia, O., Abreu, M., Aumais, G. L., Bernstein, C. N., Biron, I. A., Cino, M., Deslandres, C., Dotan, I., El-Matary, W., Feagan, B., Guttman, D. S., Huynh, H., ... Panacionne, R. (2023). Gut Microbiome Composition Is Associated With Future Onset of Crohn's Disease in Healthy First-Degree Relatives. *Gastroenterology*, *165*(3), 670–681. <https://doi.org/10.1053/j.gastro.2023.05.032>
- Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J., & Hylemon, P. B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes*, *7*(1), 22–39. <https://doi.org/10.1080/19490976.2015.1127483>
- 1095 Rocha, E. R., Herren, C. D., Smalley, D. J., & Smith, C. J. (2003). The complex oxidative stress response of *Bacteroides fragilis*: The role of OxyR in control of gene expression. *Anaerobe*, *9*(4), 165–173. [https://doi.org/10.1016/S1075-9964\(03\)00118-5](https://doi.org/10.1016/S1075-9964(03)00118-5)
- 1100 Rühlemann, M. C., Bang, C., Gogarten, J. F., Hermes, B. M., Groussin, M., Waschina, S., Poyet, M., Ulrich, M., Akoua-Koffi, C., Deschner, T., Muyembe-Tamfum, J. J., Robbins, M. M., Surbeck, M., Wittig, R. M., Zuberbühler, K., Baines, J. F., Leendertz, F. H., & Franke, A. (2023). *Functional host-specific adaptation of the intestinal microbiome in hominids* (p. 2023.03.01.530589). bioRxiv. <https://doi.org/10.1101/2023.03.01.530589>
- 1105 Rühlemann, M. C., Wacker, E. M., Ellinghaus, D., & Franke, A. (2022). MAGScoT - a fast, lightweight, and accurate bin-refinement tool. *Bioinformatics (Oxford, England)*, *btac694*. <https://doi.org/10.1093/bioinformatics/btac694>
- Russell, R. K., & Satsangi, J. (2008). Does IBD run in families?: *Inflammatory Bowel Diseases*, *14*, S20–S21. <https://doi.org/10.1002/ibd.20573>
- 1110 Samuel, B. S., & Gordon, J. I. (2006). A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(26), 10011–10016. <https://doi.org/10.1073/pnas.0602187103>
- 1115 Samuel, B. S., Hansen, E. E., Manchester, J. K., Coutinho, P. M., Henrissat, B., Fulton, R., Latreille, P., Kim, K., Wilson, R. K., & Gordon, J. I. (2007). Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(25), 10643–10648. <https://doi.org/10.1073/pnas.0704189104>
- 1120 Schirmer, M., Denson, L., Vlamakis, H., Franzosa, E. A., Thomas, S., Gotman, N. M., Rufo, P., Baker, S. S., Sauer, C., Markowitz, J., Pfeifferkorn, M., Oliva-Hemker, M., Rosh, J., Otley, A., Boyle, B., Mack, D., Baldassano, R., Keljo, D., LeLeiko, N., ... Xavier, R. J. (2018). Compositional and Temporal Changes in the Gut Microbiome of Pediatric Ulcerative Colitis Patients Are Linked to Disease Course. *Cell Host & Microbe*, *24*(4), 600-610.e4. <https://doi.org/10.1016/j.chom.2018.09.009>

- Schirmer, M., Garner, A., Vlamakis, H., & Xavier, R. J. (2019). Microbial genes and pathways in inflammatory bowel disease. *Nature Reviews Microbiology*, *17*(8), 497–511.
1125 <https://doi.org/10.1038/s41579-019-0213-6>
- Singh, S., Bhatia, R., Khare, P., Sharma, S., Rajarammohan, S., Bishnoi, M., Bhadada, S. K., Sharma, S. S., Kaur, J., & Kondepudi, K. K. (2020). Anti-inflammatory Bifidobacterium strains prevent dextran sodium sulfate induced colitis and associated gut microbial dysbiosis in mice. *Scientific Reports*, *10*(1), Article 1. <https://doi.org/10.1038/s41598-020-75702-5>
1130
- Starke, S., Harris, D. M. M., Zimmermann, J., Schuchardt, S., Oumari, M., Frank, D., Bang, C., Rosenstiel, P., Schreiber, S., Frey, N., Franke, A., Aden, K., & Waschina, S. (2023). Amino acid auxotrophies in human gut bacteria are linked to higher microbiome diversity and long-term stability. *The ISME Journal*, *17*(12), Article 12.
1135 <https://doi.org/10.1038/s41396-023-01537-3>
- Sund, C. J., Rocha, E. R., Tzianabos, A. O., Wells, W. G., Gee, J. M., Reott, M. A., O'Rourke, D. P., & Smith, C. J. (2008). The *Bacteroides fragilis* transcriptome response to oxygen and H₂O₂: The role of OxyR and its effect on survival and virulence. *Molecular Microbiology*, *67*(1), 129–142. <https://doi.org/10.1111/j.1365-2958.2007.06031.x>
- 1140 Thauer, R. K. (1998). Biochemistry of methanogenesis: A tribute to Marjory Stephenson:1998 Marjory Stephenson Prize Lecture. *Microbiology*, *144*(9), 2377–2406. <https://doi.org/10.1099/00221287-144-9-2377>
- Thauer, R. K., & Shima, S. (2006). Biogeochemistry: Methane and microbes. *Nature*, *440*(7086), 878–879. <https://doi.org/10.1038/440878a>
- 1145 Thomas, J. P., Modos, D., Rushbrook, S. M., Powell, N., & Korcsmaros, T. (2022). The Emerging Role of Bile Acids in the Pathogenesis of Inflammatory Bowel Disease. *Frontiers in Immunology*, *13*, 829525. <https://doi.org/10.3389/fimmu.2022.829525>
- Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E. F., Wang, J., Tito, R. Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., & Raes, J. (2019). The neuroactive potential of the human gut microbiota in quality of life and depression. *Nature Microbiology*, *4*(4), 623–632.
1150 <https://doi.org/10.1038/s41564-018-0337-x>
- van den Boogaart, K. G., & Tolosana-Delgado, R. (2008). “compositions”: A unified R package to analyze compositional data. *Computers & Geosciences*, *34*(4), 320–338.
1155 <https://doi.org/10.1016/j.cageo.2006.11.017>
- Vermeire, S., Peeters, M., Vlietinck, R., Joossens, S., Hond, E. D., Bulteel, V., Bossuyt, X., Geypens, B., & Rutgeerts, P. (2001). Anti-*Saccharomyces Cerevisiae* Antibodies (ASCA), Phenotypes of IBD, and Intestinal Permeability: A Study in IBD Families: *Inflammatory Bowel Disease*, *7*(1), 8–15. <https://doi.org/10.1097/00054725-200102000-00002>
1160
- Vich Vila, A., Imhann, F., Collij, V., Jankipersadsing, S. A., Gurry, T., Mujagic, Z., Kurilshikov, A., Bonder, M. J., Jiang, X., Tigchelaar, E. F., Dekens, J., Peters, V., Voskuil, M. D., Visschedijk, M. C., Van Dullemen, H. M., Keszthelyi, D., Swertz, M. A., Franke, L., Alberts, R., ... Weersma, R. K. (2018). Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Science Translational Medicine*, *10*(472), eaap8914.
1165 <https://doi.org/10.1126/scitranslmed.aap8914>
- Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleylen, I., van der Merwe, S., Vermeire, S., & Raes, J. (2019). Quantitative

- 1170 microbiome profiling disentangles inflammation- and bile duct obstruction-
associated microbiota alterations across PSC/IBD diagnoses. *Nature Microbiology*,
4(11), 1826–1831. <https://doi.org/10.1038/s41564-019-0483-9>
- Walter, J., Armet, A. M., Finlay, B. B., & Shanahan, F. (2020). Establishing or Exaggerating
Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated
1175 Rodents. *Cell*, 180(2), 221–232. <https://doi.org/10.1016/j.cell.2019.12.025>
- Wechsler, J. B., Szabo, A., Hsu, C.-L., Krier-Burris, R., Schroeder, H., Wang, M. Y., Carter, R.,
Velez, T., Aguiniga, L. M., Brown, J. B., Miller, M. L., Wershil, B. K., Barrett, T. A., &
Bryce, P. J. (2018). Histamine drives severity of innate inflammation via histamine 4
receptor in murine experimental colitis. *Mucosal Immunology*, 11(3), 861–870.
1180 <https://doi.org/10.1038/mi.2017.121>
- Wexler, H. M. (2007). Bacteroides: The Good, the Bad, and the Nitty-Gritty. *Clinical
Microbiology Reviews*, 20(4), 593–621. <https://doi.org/10.1128/CMR.00008-07>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, G.,
Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M.,
1185 Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019).
Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686.
<https://doi.org/10.21105/joss.01686>
- Wong, R. G., Wu, J. R., & Gloor, G. B. (2016). Expanding the UniFrac Toolbox. *PLoS One*, 11(9),
e0161196. <https://doi.org/10.1371/journal.pone.0161196>
- 1190 Wright, E. K., Kamm, M. A., Teo, S. M., Inouye, M., Wagner, J., & Kirkwood, C. D. (2015).
Recent Advances in Characterizing the Gastrointestinal Microbiome in Crohn's
Disease: A Systematic Review. *Inflammatory Bowel Diseases*, 1.
<https://doi.org/10.1097/MIB.0000000000000382>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning
1195 algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*,
32(4), 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Zimmermann, J., Kaleta, C., & Waschina, S. (2021). gapseq: Informed prediction of bacterial
metabolic pathways and reconstruction of accurate metabolic models. *Genome
Biology*, 22(1), 81. <https://doi.org/10.1186/s13059-021-02295-1>
- 1200

Acknowledgements

1205 We would like to thank all patients and probands involved in the sampling of biomaterials for
this study. We are thankful to the staff of the IKMB microbiome laboratory for processing the
stool samples. This work was supported by the DFG Research Infrastructure NGS_CC (project
407495230) as part of the Next Generation Sequencing Competence Network (project
423957469). NGS analyses were carried out at the Competence Centre for Genomic Analysis
(Kiel).

1210

Funding

1215 This study was financed through the EU Horizon miGut-Health project (Personalised blueprint
of chronic inflammation in health-to-disease transition; Grant agreement ID: 101095470, AF,
MG) and the DFG-sequencing grant "Identification of early fine-scale microbial signatures in
inflammatory bowel disease" (project 433152305, CB, AF, WL). Microbiome sequencing and
data analysis received infrastructure support from the DFG Excellence Cluster 2167 "Precision
Medicine in Chronic Inflammation" (PMI) and the DFG Research Unit 5042 "miTarget". MR
was supported by the Bruhn foundation for the advancement of medical research.