



# Meta-analysis of the Microbial Diversity Cultured in Bioreactors Simulating the Gut Microbiome

David Felipe Garcia Mendez<sup>1</sup> · Siobhon Egan<sup>1</sup> · Julien Wist<sup>1,3</sup> · Elaine Holmes<sup>1</sup> · Janeth Sanabria<sup>1,2</sup>

Received: 7 December 2023 / Accepted: 25 March 2024  
© The Author(s) 2024

## Abstract

Understanding the intricate ecological interactions within the gut microbiome and unravelling its impact on human health is a challenging task. Bioreactors are valuable tools that have contributed to our understanding of gut microbial ecology. However, there is a lack of studies describing and comparing the microbial diversity cultivated in these models. This knowledge is crucial for refining current models to reflect the gastrointestinal microbiome accurately. In this study, we analysed the microbial diversity of 1512 samples from 18 studies available in public repositories that employed cultures performed in batches and various bioreactor models to cultivate faecal microbiota. Community structure comparison between samples using t-distributed stochastic neighbour embedding and the Hellinger distance revealed a high variation between projects. The main driver of these differences was the inter-individual variation between the donor faecal inocula. Moreover, there was no overlap in the structure of the microbial communities between studies using the same bioreactor platform. In addition,  $\alpha$ -diversity analysis using Hill numbers showed that highly complex bioreactors did not exhibit higher diversities than simpler designs. However, analyses of five projects in which the samples from the faecal inoculum were also provided revealed an amplicon sequence variants enrichment in bioreactors compared to the inoculum. Finally, a comparative analysis of the taxonomy of the families detected in the projects and the GMRepo database revealed bacterial families exclusively found in the bioreactor models. These findings highlight the potential of bioreactors to enrich low-abundance microorganisms from faecal samples, contributing to uncovering the gut microbial “dark matter”.

**Keywords** In vitro gut · Faecal microbiome · Simulation · Diversity · Hill numbers

## Introduction

In the last 20 years, multiple researchers have contributed to understanding the impact of the gut microbiome on human health, and the growing interest in this field is reflected in the escalating body of publications. What started as a goal to describe the diversity of the microorganisms in the human gut and other body sites has shifted focus to the more challenging task of studying the mechanisms by which host-microbiota interactions occur [1]. Previous work has addressed the limitations surrounding the mechanistic study of the gut ecosystem. One significant limitation is that some of the regions of the intestinal tract remain mostly inaccessible. While invasive methods such as endoscopy and nasoenteric probes have provided valuable information on the microbial activity and composition of the different sections of the gut [2, 3], these procedures are expensive, require highly trained personnel and specialised equipment, and may face practical and ethical limitations. In addition,

---

✉ Janeth Sanabria  
janeth.sanabria@correounivalle.edu.co  
David Felipe Garcia Mendez  
34311041@student.murdoch.edu.au

<sup>1</sup> Australian National Phenome Centre and Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Harry Perkins Building, Perth, WA 6150, Australia

<sup>2</sup> Environmental Microbiology and Biotechnology Laboratory, Engineering School of Environmental & Natural Resources, Engineering Faculty, Universidad del Valle – Sede Meléndez, 76001 Cali, Colombia

<sup>3</sup> Chemistry Department, Universidad del Valle - Sede Meléndez, 76001 Cali, Colombia

this intestinal environment has a very dynamic nature, showing a high intra-individual and inter-individual variation, being influenced by confounding factors such as diet, lifestyle, and medication, among others [4].

Combining high-throughput sequencing with cultivation-independent methods has provided valuable insights into our understanding of the factors that shape gut ecology. It has given access to a large inventory of taxonomic and functional information on the trillions of microorganisms present in the human gut. Notably, it has also raised awareness of the microbial “dark matter”, which includes multiple bacterial and other microbial species still to be cultured and whose biological functions are yet to be understood [5]. Multiple research groups have highlighted the limitations of culture-independent techniques with regard to fully capturing the bacterial diversity in faecal samples. These include the biases introduced by DNA extraction efficiency between taxa, selecting a given set of primers, and variable sequencing depth between samples [6]. In addition, the extrapolation of functions from genomic information remains challenging, with many genes from the microbial dark matter still to be annotated [5, 7]. Culturing and isolating this microbial diversity is still a challenging task, but it is essential not only to aid in the genetic and phenotypic characterisation of these microorganisms but also to facilitate the shift from correlative studies to causative validation of predicted bacterial functions [8]. Promising advances have been made in the cultivation of multiple species previously considered “unculturable” from the gut microbiota, strongly questioning the prevailing belief that much of the gut microbiome cannot be cultured [9, 10]. However, metagenomic approaches suggest that there is still a large microbial diversity that is not included in these culture collections [5, 11].

All these limitations have motivated the development of multiple models for the cultivation and study of the gut microbiota *in vitro* and *in vivo*, including gnotobiotic animals, organoids, cell culture models, microbial cultures, and bioreactors [12, 13]. Microbial cultures performed in batches are a cheap and efficient method (i.e. multiple culture conditions can be tested simultaneously) for studying multiple bacteria-bacteria and bacteria-substrate interactions in short-term experiments. On the other hand, bioreactors are more complex devices whose integrated technology allows them to provide an optimum environment for the desired microbial reactions that better mimic some of the environmental conditions in the human body. This environment can be obtained by precisely controlling the operational parameters of the cultures, such as pH, redox potential, and nutrient feed, among others.

Bioreactors are promising tools that can unlock access to the microbial “dark matter”. The precise control of the

culture conditions allows the recovery and growth of hard-to-culture microorganisms. A pertinent example is the cultivation of environmental anammox bacteria, which exhibit slow growth rates ranging from 2 days to several weeks. Moreover, their growth is inhibited by the accumulation of their product metabolites in the culture, requiring continuous cultivation methods such as up-flow column reactors [14, 15]. The research teams behind the human gut models have demonstrated that they can culture reproducible bacterial communities with a diversity similar to their initial inoculum. Faeces are commonly used as inoculum, but other inocula have been tested, such as ileostomy effluents, microbial isolates from faeces, and synthetic communities [16–18]. Based on the number of publications derived from these models, the most well-known platforms are SHIME [19], TIM-2 [20], SIMGI [21], PolyFermS [22], ARCOL [23], and the three-stage model from Macfarlane et al. [24]. Although bioreactor technology has made significant progress, there are multiple gaps that limit the potential of these devices to uncover new biology and be fully exploited. For instance, there is still a huge gap in our understanding of the impact of operational parameters such as nutrient medium, retention time, flow and agitation, and pH on the ecology of the microbial communities recovered in bioreactors [25]. Moreover, there is a lack of studies that describe and compare the microbial communities cultured in these devices. This knowledge can be useful for refining current models to reflect the composition and behaviour of the microbiome within the human gastrointestinal tract more accurately. Given the strong inter-individual differences and impact of operational parameters on compositional analysis, comparison of multiple systems and conditions is essential to allow appropriate interpretation of data from bioreactor models. However, the analytical cost and required experimental time prohibit high throughput analysis. Co-analysis of archived sequencing data from multiple studies and research groups has the potential to allow a clearer definition of the limits and appropriate use of bioreactor technology.

In this work, we aimed to describe and compare the taxonomic diversity and structure of the microbial communities cultured in these devices by performing a meta-analysis of the 16S rRNA sequences of over 1512 samples available in the NCBI Sequence Read Archive (SRA). These samples were retrieved from projects that used microbial cultures performed in batches and bioreactors seeded with human faeces. We explored the relationships between sample input and operational parameters in regard to their impact on microbial community structure and investigated differences in microbial communities before and after the inoculation of stool samples in three of the bioreactor models.

## Methods

### Search in the NCBI Sequence Read Archive (SRA)

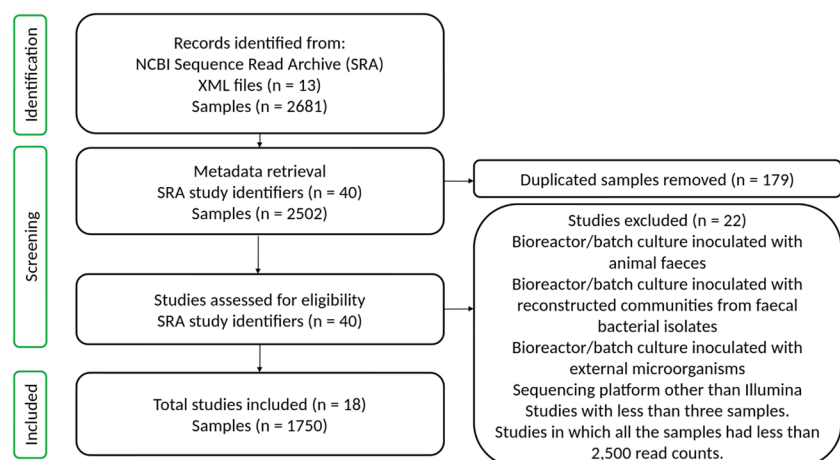
Our search aimed to retrieve 16S rRNA amplicon sequences from samples obtained from studies that used either (i) bioreactors to simulate the gut environment or (ii) microbial cultures performed in batches, both inoculated with human faeces. For this purpose, we performed multiple searches in the SRA (<https://www.ncbi.nlm.nih.gov/sra>) using each one of the following keywords: “continuous flow bioreactor model”, “fecal batch culture”, “gut bioreactor”, “gut microbiota cultured in vitro”, “in vitro digestion”, “PolyfermS”, “SHIME”, “simulated gastrointestinal”, “simulation of the colon”, “simulation of the gut”, and “TIM-2”. After each search, we used the SRA Run selector tool to download the metadata of all the samples into an XML file. This search was performed in September 2022. Then, in-house scripts in R were used to remove duplicates and filter the studies and samples based on selection criteria (Fig. 1) ([https://zenodo.org/doi/10.5281/zenodo.8045625—file:1\\_Metadata\\_retrieve.Rmd](https://zenodo.org/doi/10.5281/zenodo.8045625—file:1_Metadata_retrieve.Rmd)). The accession numbers of the projects were used to retrieve the FTP address of the samples using `ffq` [26], and their sequences were downloaded from the NCBI. Additional metadata, such as the operational conditions of the bioreactor and composition of the culture media, were obtained from the published papers of the projects to support further analyses. During the search, a special focus was given to projects in which the composition of the faecal inoculum could be identified, allowing comparisons with the samples obtained from the bioreactors.

### Informatic Analysis

Data visualisation was performed using R statistical software (v 4.1.2, R Core Team 2022) in RStudio (v2022.12.0,

Posit Team 2022). High-throughput sequence data were analysed using `dada2` (v 1.16) [27]. First, we analysed the samples from the selected projects individually and manually adjusted the input parameters of the `dada2` package to maximise the number of reads that passed the denoising and quality filtering steps, while reducing the number of total errors in the reads. The parameters used to set up the `dada2` algorithms are summarised in Additional file 1: Table S1. Then, sequences were de-replicated, and high-resolution amplicon sequence variants (ASVs) were produced, followed by the removal of chimaeras. Taxonomy was assigned to ASVs using the BLASTN algorithm with NCBI RefSeq 16S rRNA database as the reference (downloaded June 2022—available in Zenodo:[https://zenodo.org/doi/10.5281/zenodo.8045625—file:/supporting\\_metadata/bacteria.16SrRNA.zip](https://zenodo.org/doi/10.5281/zenodo.8045625—file:/supporting_metadata/bacteria.16SrRNA.zip)). Briefly, we filtered and selected the top hit with the highest percentage of identical positions (`pident`) for each ASV. The subject sequence id (`sseqid`) obtained from BLAST was used to search the NCBI unique record identifier (UID) using the package `taxize` (v 0.9.100). Taxonomic lineage was retrieved using the function `classification` from the package `taxize` (code available in Zenodo:<https://zenodo.org/doi/10.5281/zenodo.8045625—file:Taxonomy.Rmd>). Due to the region and length variation of the 16S rRNA sequences obtained from the various studies and limitations in accurate taxonomic assignment to genus and species level for short-read 16S rRNA data, all taxonomy was collapsed to the family level to enable meaningful comparison. Data were analysed using `phyloseq` (v 1.38) and other R packages (code available in Zenodo:<https://zenodo.org/doi/10.5281/zenodo.8045625>).

**Fig. 1** Flow diagram showing the data retrieval process



## Comparison of the Community Structure Between Samples ( $\beta$ -Diversity) and Within Samples ( $\alpha$ -Diversity)

Samples not directly obtained from the bioreactor (e.g. faecal samples), those with zero sequences after bioinformatic analysis, and those with less than 2500 read counts were removed from the data set used for diversity analysis. Beta diversity was assessed using principal component analysis (PCA) and with the non-linear dimension reduction algorithms, t-distributed stochastic neighbour embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP) with packages *Rtsne* (v 0.16) and *umap* (v 0.2.9). Hellinger distances were used for  $\beta$ -diversity analysis to account for large differences in sequencing depth while avoiding rarefying of samples. The limitations of rarefying microbiome data have been addressed previously [28]. The Hellinger distances were calculated by applying the Hellinger transformation to the compositional data matrix and then computing the Euclidean distance among samples. For  $\alpha$ -diversity analyses, a series of Hill numbers were calculated for each sample at diversity orders  $q = 0-3$  using the package *hilldiv* (v 1.5.1). Subsequently, these variables were plotted to obtain a diversity profile of each study. In addition, traditional diversity indices such as Observed, Chao1, Shannon, and inverse Simpson indices were calculated using packages *phyloseq* (v 1.38) and *ggstatsplot* (v 0.11.0). Statistical analysis of  $\alpha$ -diversity measures was performed using non-parametric tests with the holm p-adjustment method, Kruskal–Wallis one-way ANOVA (no. groups > 2), or Mann–Whitney *U* test (no. groups = 2). Pair-wise comparisons for paired combinations of experimental parameters (e.g. single versus pooled donor) were produced for ANOVA analysis using Dunn’s non-parametric all-pairs comparison.

## Analysis of Amplicon Sequence Variants (ASVs) Enrichment in the Bioreactors

We explored the metadata of the studies available in the NCBI to select the ones that allowed assessment of (i) sequences of the initial faecal inoculum; (ii) source donor; (iii) if it was a multistage design, the gut compartment that was simulated; and (iv) if it included a mucosal component, the luminal or mucosal origin of the sample. By mucosal component, we refer to bioreactors in which support media such as plastic carriers covered with mucin were introduced to allow biofilm growth, whereas luminal sample refers to samples taken from the unstructured culture media. Based on these criteria, five projects were further selected to evaluate the enrichment of ASVs in the bioreactors. “ASVs enrichment” in this study is defined as the ASVs not present in the initial faecal sample (inoculum), as determined from the sequencing data analysis that appeared after the

cultivation of the microbial community in the bioreactor. *In-house* scripts in R were used to filter the ASVs absent in the initial inoculum (number of reads = 0) but that were present in the bioreactors (number of reads > 0). For studies with multiple donors, the biodiversity comparison was made between samples from the same donor. On the other hand, if the study included replicates, only ASVs absent in both faecal sample replicates and present in both bioreactor sample replicates were selected. Finally, the taxonomic classification of these ASVs was obtained from the previous identification performed with BLAST (see “Informatic Analysis” for details). A table describing selected samples is available in Additional file 1: Table S5.

## Comparison of the Taxonomy Between Samples and the Data Repository for Human Gut Microbiota (GMrepo)

The overlap of taxonomy at the family level found in the *in vitro* models and that previously documented by direct sequencing of the human faecal microbiome was explored. Firstly, a list of all the species available in the Data Repository for Human Gut Microbiota (GMrepo) [29] was downloaded (February 2024—available in Zenodo:<https://zenodo.org/doi/10.5281/zenodo.8045625>—file:/supporting metadata/GMREPO). Then, the package *taxize* was used to retrieve the taxonomic lineage by searching for their NCBI taxon ID. Finally, the taxa lists obtained from the projects and the samples in GMrepo were compared using the function *setdiff* in R.

## Data and Code Availability

No new sequencing datasets were generated during the current study. Details of accession numbers for data analysed can be found in Additional file 1: Table S2. Original R scripts and metadata files are available in Zenodo (<https://zenodo.org/doi/10.5281/zenodo.8045625>).

## Results

### Search in the NCBI Sequence Read Archive (SRA)

A total of 18 studies were selected after the search in the SRA, accounting for 1750 samples from all the projects (Additional file 1: Table S2). Subsequently, samples directly from faeces ( $n = 79$ ), those with zero reads after bioinformatic processing ( $n = 63$ ), and those with less than 2500 read counts ( $n = 96$ ) were removed, leaving the final number of samples as 1512. The library size of selected samples was between 2661 and 419,940 reads (Additional file 1: Fig. S1). Moreover, the rarefaction curves obtained from plotting the

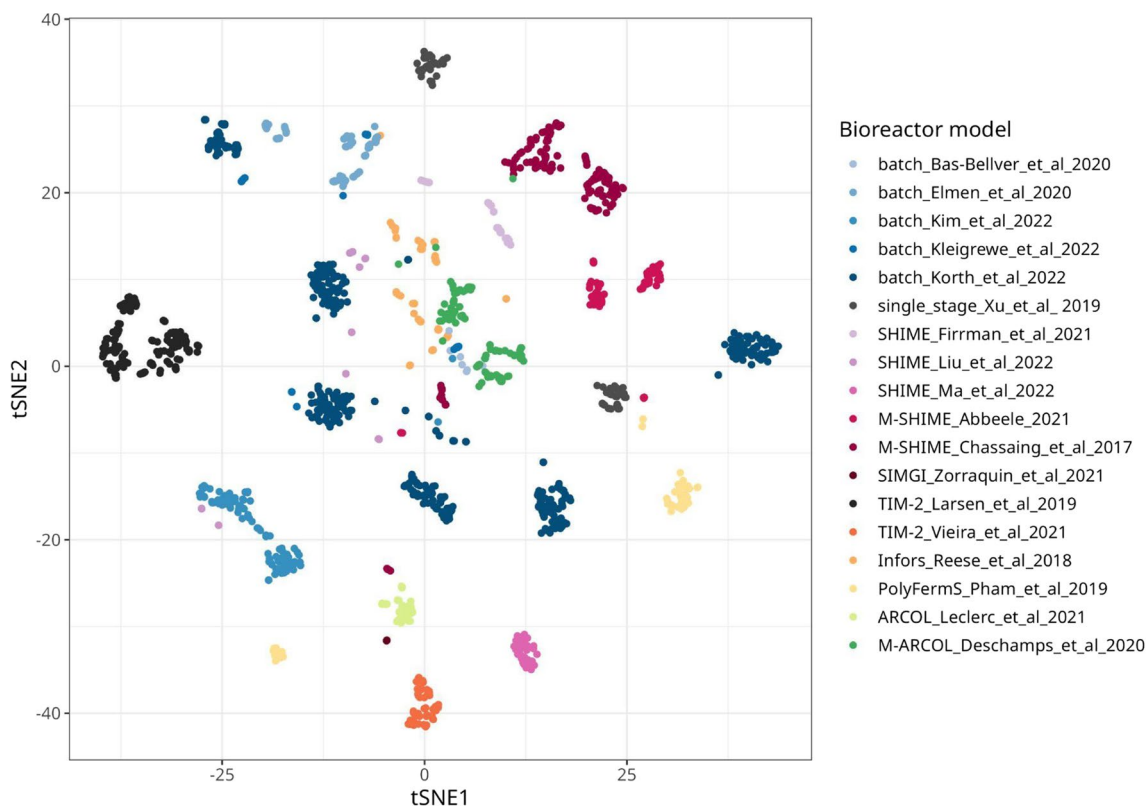
number of observed OTUs against the sequencing depth (number of reads) in the samples showed that most cultures reached a plateau, suggesting that the sequencing depth was adequate (Additional file 1: Fig. S2).

The selected projects included samples from five studies that used cultures performed in batch and seven bioreactors with specific configurations: single-stage model (single\_stage\_Xu\_et\_al\_2019), SHIME, SIMGI, TIM-2, PolyFermS, and ARCOL. These configurations are briefly described in Additional file 1: Table S2. Most of these studies (13/18) targeted the V3-V4 region of the 16S rRNA gene, followed by the V4 region (3/18), and only two studies used each of the V1 or V3 regions, respectively. The metadata on the operational parameters used in these models, including country, inoculum type, pH, and composition of the culture media, were extracted from the papers and are available in Additional file 2.

### Comparison of the Microbial Diversity Between Bioreactor Studies ( $\beta$ -Diversity)

The t-SNE analysis grouped the samples by study, with no overlap between samples from the same bioreactor model (Fig. 2). For instance, the samples from SHIME studies

were not clustered and remained separated in the t-SNE plot. However, microbial communities from different studies did overlap, except for the ones in SHIME\_Ma\_et\_al\_2022, which was notably different from that observed in other studies, as indicated by its remote mapping space (Fig. 2). One of the limitations of using t-SNE for  $\beta$ -diversity analysis is that it preserves the local distances but not the global distances. To overcome this limitation and support our observations, we also performed a PCA over the Hellinger-transformed compositional data. The PCA showed a similar pattern to that observed in the t-SNE (Additional file 1: Fig. S4), indicating that the observations made from the t-SNE are robust. In addition, where studies included more than one donor, the individual donors did not necessarily map closely, indicating the greater impact of the donor over the in vitro model platform. For example, batch\_Korth\_et\_al\_2022 screened individual stool samples from four adult volunteers, showing high dispersion in the t-SNE plot (Fig. 2) and PCA (Additional file 1: Fig. S4) based on inter-individual compositional differences. UMAP analysis confirmed the grouping patterns observed by t-SNE but overestimated sample similarity, as reflected by the tightness of the clusters relating to each study (Fig. S3). Importantly, microbial community differences were not driven by variation in the sequenced



**Fig. 2** Comparison of microbiome community structure from 18 different studies with 1512 samples. Multi-dimensional scaling plot using t-distributed stochastic neighbour embedding (t-SNE) with Hellinger distance measure of compositional data

16S rRNA gene region once taxonomy was collapsed to the family level (Additional file 1: Fig. S4). Based on visual assessment, operational parameters, such as the carbon-to-nitrogen ratio, sample origin, use of vitamins, number of bioreactor stages, gut compartment, and mucin amount, did not exhibit strong patterns across the dataset (Additional file 1: Fig. S5-6).

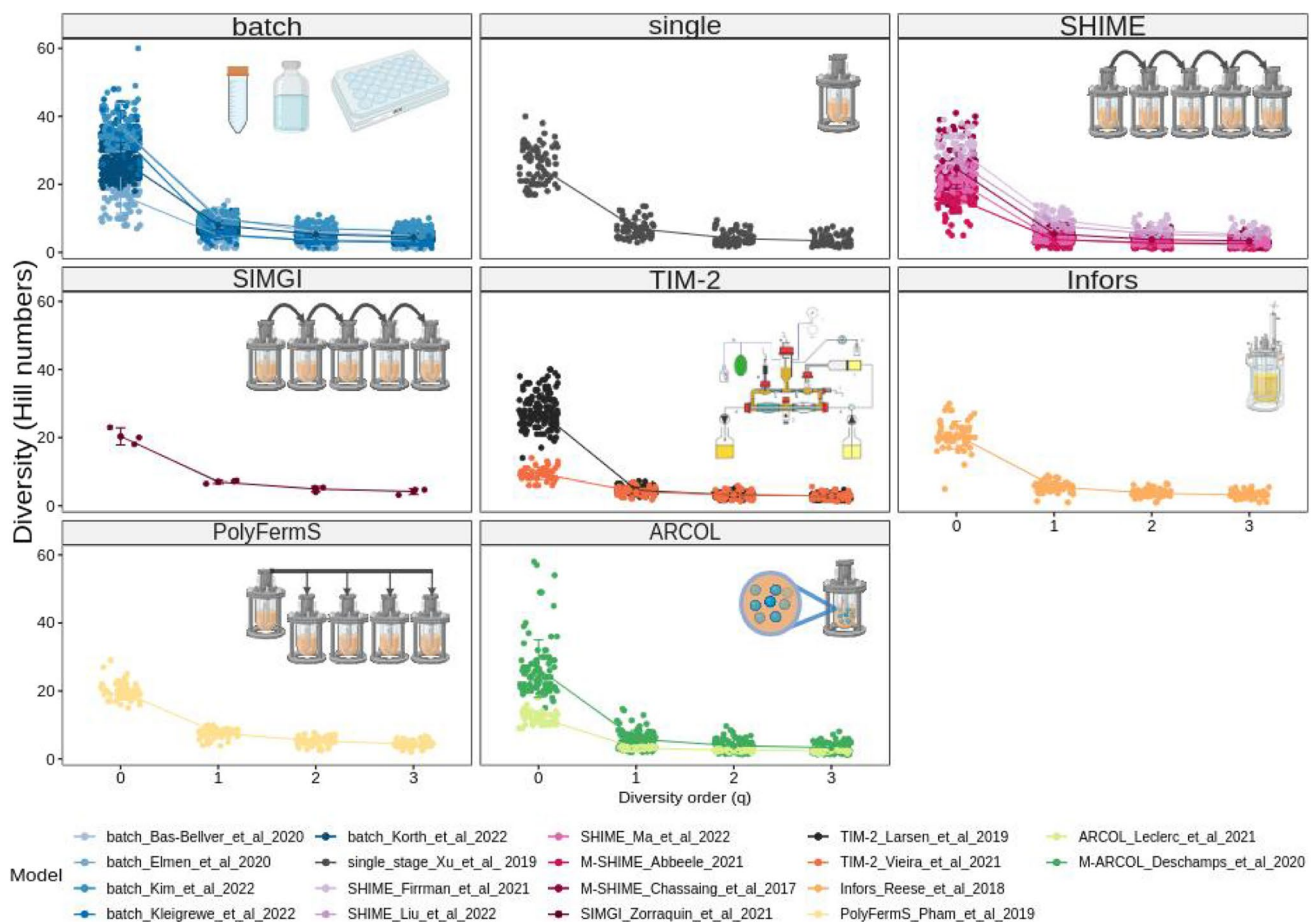
### Mean Species Diversity in a Site at a Local Scale ( $\alpha$ -Diversity)

Based on the  $\alpha$ -diversity profile determined by Hill numbers, the studies that showed the highest diversity of families were batch\_Bas-Bellver\_et\_al\_2020, batch\_Kim\_et\_al\_2022, and batch\_Kleigrewet\_et\_al\_2022. As can be noted, all these projects used cultures performed in batches (Fig. 3, Additional file 1: Fig. S7). In contrast, the studies that showed the

lowest diversity across the profile were ARCOL\_Leclerc\_et\_al\_2021 and TIM-2\_Vieira\_et\_al\_2021. In addition, samples from bioreactors that used formula-fed infants as an inoculum showed significantly higher Shannon diversity indices when compared to other inocula (Additional file 1: Fig. S8c).

### Comparison of the Taxonomy Between Models and the Faecal Microbiome

An evaluation of the effect of the different operational parameters and settings of the bioreactor models showed that, as expected, the most abundant groups at the phylum level in the samples were *Bacillota*, *Pseudomonadota*, *Actinomycetota*, and *Bacteroidota* (Additional file 1: Fig. S9). In particular, the study SHIME\_Ma\_et\_al\_2022 showed enrichment in *Synergistota* (50.9%) and *Fusobacteriota* (29.8%). In the current meta-analysis, at the family



**Fig. 3** Diversity profiles of the 18 selected studies determined by Hill numbers at diversity orders  $q=0-3$ . Diversity profile plots are grouped by bioreactor model type to aid data visualisation and comparison. The sensitivity towards abundant and rare families is modulated using the diversity order “ $q$ ”. The larger the  $q$  value, the higher the importance attributed to abundant families. At diversity order  $q=0$ , the Hill numbers represent the total number of families

in the samples (richness). A value of  $q=1$  weights families by their proportion without disproportionately favouring rare or abundant families. This value yields the exponential of the Shannon index. At  $q=2$ , abundant families are over-weighted, and the number yields the multiplicative inverse of the Simpson index. Vertical bars show the standard deviation. This summarised explanation was adapted from Alberdi and Gilbert et al. 2019 [30]

level, the top 5 families that accounted for up to 50% of the abundance in the samples were *Bacteroidaceae*, *Lachnospiraceae*, *Oscillospiraceae*, *Bifidobacteriaceae*, and *Enterobacteriaceae* (Fig. 4).

To focus on the less abundant taxa, we compared the taxonomy of the enriched families in the projects selected for the ASVs enrichment analysis (Additional file 1: Table S6): a family was considered to be enriched based on two criteria:

(i) when it was not detected in the faecal inoculum; in other words, there were no ASVs associated to this family in the inoculum and (ii) when we detected at least one read from an ASV belonging to this family in the samples obtained from the bioreactors. The most common enriched families detected across all the samples were *Enterobacteriaceae* (11/36), *Desulfohalobiaceae* (10/36), and *Synergistaceae* (9/36).

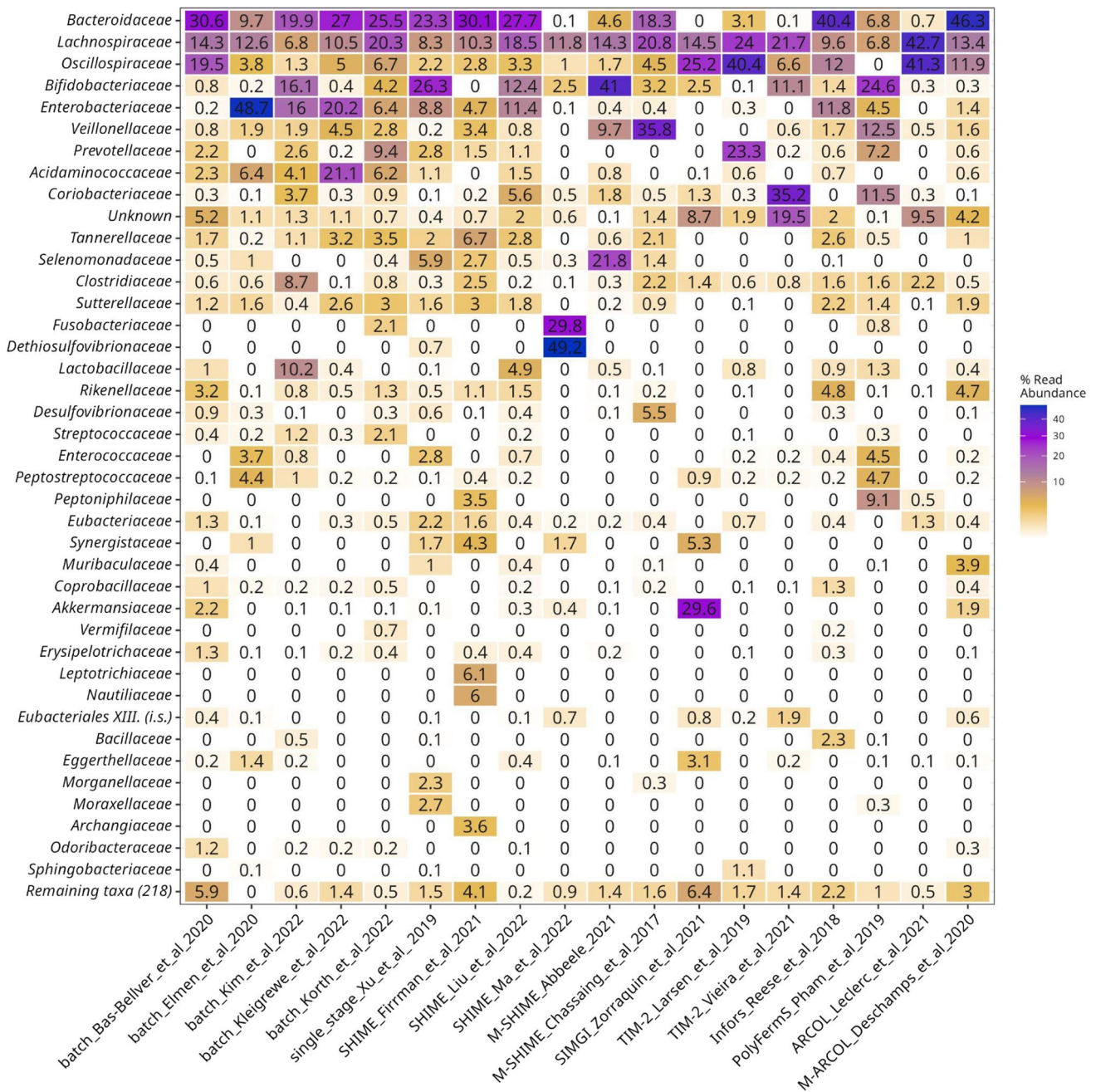


Fig. 4 Heatmap of top 40 bacteria families identified in samples (remaining taxa grouped together in bottom row). Compositional data are represented as the total proportion of sequences from all samples in the project (n=1512)

Subsequently, we explored whether there was any difference between the list of families observed in the bioreactors and the list of faecal taxa obtained from GMRepo. We found 36 families in the bioreactors and batch cultures that have not been detected in the curated projects of GMRepo; the most prevalent ones were the families *Gracilibacteraceae* (13/18), *Muribaculaceae* (11/18), *Defluviitaleaceae* (9/18), and *Vallitaleaceae* (8/18). However, these families were mostly observed in low relative abundances (<0.0001%) and occasionally in higher proportions (>0.1%).

At the phylum level, most of these enriched families belonged to *Pseudomonadota* (25.8%), *Bacillota* (25.8%), *Cyanobacteriota* (11.1%), and *Chloroflexota* (12.9%) (Fig. 5A). In addition, the phylum *Nitrospinota* was only detected in the bioreactors. On the other hand, 310 families found in the faecal samples of the GMRepo database were not detected in any of the studies included in the current meta-analysis. Similarly, these families that are distinct only to the GMRepo database are not highly abundant in faecal samples. Most of them belonged to the phyla *Pseudomonadota* (24.2%), *Actinomycetota* (12.6%), and *Cyanobacteriota* (9.4%) (Fig. 5B).

### Analysis of ASVs Enrichment in the Bioreactors

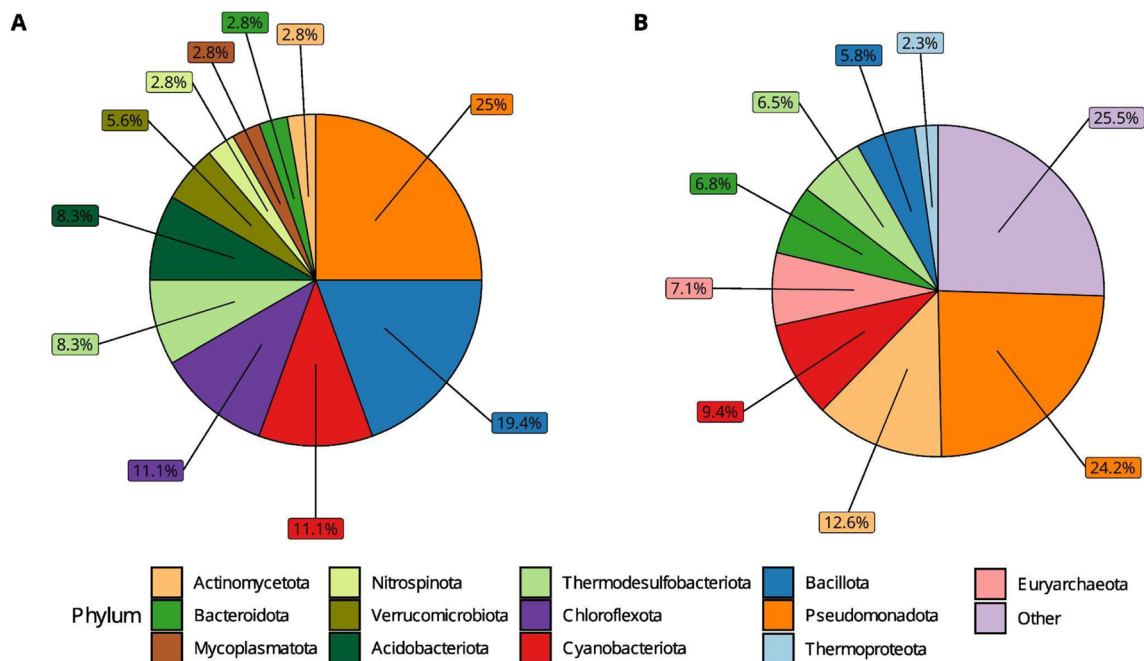
In our analysis, we found that despite the dominance of a few families in the communities developed in the bioreactors, new ASVs appeared in all the selected models when

compared to the inoculum (Fig. 6). Moreover, the results show that modification of the bioreactor design such as adding plastic carriers for promoting biofilm growth increased the number of ASVs detected. For instance, the study of M-ARCOL\_Deschamps\_et\_al\_2020 showed 499 ASVs in the faeces of one of the donors. After inoculation in the bioreactor, 227 and 238 new ASVs were detected in the luminal and mucosal compartments, respectively. A similar trend was observed in the study of SHIME\_Firman\_2021.

## Discussion

### Inter-individual Variation Is the Main Driver of the Difference Between Projects

Our results suggest that inter-individual variation in faecal inoculum diversity is the main driver of the difference in family-level composition between studies. The variability among individuals seems to be greater than the selective pressure exerted by the bioreactor model. However, further research is required to validate this claim. An interesting approach would be to culture faecal samples from the same donors in all these bioreactors to determine how the microbial communities are shaped by the culture conditions across different platforms. Furthermore, these findings show how pooling samples to increase the diversity of the inocula comes at the cost of losing information about the community



**Fig. 5** Composition at the phylum level of unique families observed in selected studies and GMRepo. (A) Phyla of the families only present in the selected studies. (B) Phyla of the families only present in

GMRepo. The percentages were obtained by dividing the number of families within each phylum by the total number of families in each category, which are 36 and 310, respectively



structure of individual samples. Similarly, Isenring et al. [31] suggested that the pooling of samples completely removes inter-individual differences, leading to an artificial community with unpredictable competition and artificial balance among taxa. Moreover, each species that establishes itself in a microbiome plays an important role. Although we cannot observe this importance as a direct metabolic consequence, it may impact both the ecological structure of the community and the downstream microbial metabolites generated by the community as a coherent entity. This information is lost in a pooled sample since the combined microbiota represents multiple independent ecologies. Researchers must carefully consider this effect when selecting the inoculum used in their in vitro models.

### Complex Bioreactor Designs Did Not Achieve a Higher Microbial Diversity

Hill numbers have been proposed as a general, robust, and flexible statistical framework [30], offering multiple advantages over traditional diversity indices such as richness, Shannon, and Simpson [30, 32, 33]. For instance, the interpretation of the Hill measure and its measurement unit is always the same, and the  $q$  parameter can modulate the sensitivity towards abundant and rare taxa. The steep slope of the profiles in Fig. 3 indicates a high unevenness of the community structure inside the bioreactors, with strong variation in abundance between the represented families. In other words, a few highly abundant families dominate these communities, a common feature of the microbiome of different body sites, such as the vagina and the gut environment. Nevertheless, low-abundance families can still fulfil crucial biological functions, such as those residing in specialised environments like microvilli crypts. If the role of these less common family members is neglected, crucial insights into the community's overall functioning are lost. For instance, methanogenic bacteria in the intestine and anammox in aquatic environments are challenging to cultivate and often go undetected in the direct sequencing of samples. These facts do not remove their impact, as methanogens influence fermentation, and anammox affects the entire nitrogen cycle.

Moreover, the results suggest that the diversity of the original faecal inoculum greatly impacts the diversity recovered in the bioreactor. However, this analysis was limited because the available metadata did not allow us to identify the samples of the faecal inocula in most of the studies.

Consequently, we could not compare the faecal inocula with the bioreactor's communities. Despite this limitation, we did find that microbiome communities originating from individuals had a higher  $\alpha$ -diversity than those where pooled samples were analysed separately (Mann–Whitney  $U$  test on Shannon diversity indices,  $p < 0.005$ . Additional file 1:

Fig. S7, Table S4). These results contrast those obtained by Średnicka et al. (2023), who found higher  $\alpha$ -diversity metrics in pooled samples from healthy donors than in individual stool samples [34].

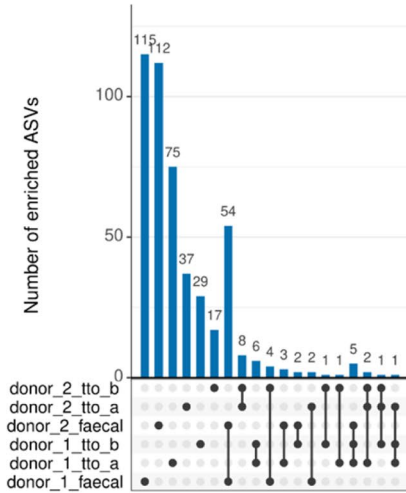
In addition, it is worth highlighting the low total number of reads (less than 10,000) observed in multiple samples from the projects TIM-2\_Vieira\_et\_al\_2021, ARCOL\_Leclerc\_et\_al\_2021, SIMGI\_Zorraquin\_et\_al\_2021, and others (Additional File 1: Fig. S1). This low number of reads is below an appropriate minimum to obtain reliable estimates of diversity statistics from these samples. A low number of reads might be obtained due to errors during sample DNA extraction and sequencing. However, an alternative hypothesis is that the operating conditions in those bioreactors might have impacted the number of reads observed in the samples. For instance, a total low microbial biomass in the system can decrease the DNA yield in the sample during extraction, affecting all the subsequent analytical steps.

### Bioreactors as Tools for Unlocking the Microbial Dark Matter

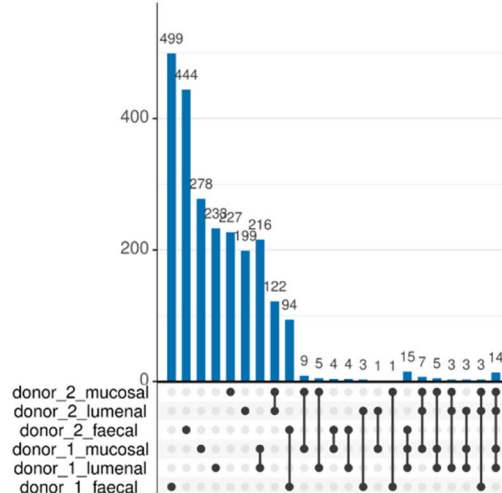
The cultivation of all the microorganisms present in the gut is a challenging task. Although promising advancements have been made in this regard [9, 10], a huge diversity of microorganisms is still to be cultured. This is supported by metagenomic studies that show a huge taxonomic diversity that is not included in culture collections [5, 11]. Moreover, culturomic approaches have demonstrated the distinct benefits of using different culture strategies to analyse the same sample to evidence uncultured species. When multiple plating or enrichment methods are used, a *hidden* diversity appears, extending the number of species that can be detected compared to traditional culture methods. Also, recent research has shown the biases of culture-independent techniques such as 16S rRNA amplicon sequencing in the deep characterisation of the microbiome, which can result in an underestimation of the microbial diversity, especially low-abundant species, which can go undetected [35].

Our findings show that both cultures performed in batch and bioreactors can not only support the growth of the most abundant families in faeces but also uncover a hidden diversity inside the faecal inoculum. Although present in low abundance, we found 36 new families in bioreactors compared to the ones reported in the GMRepo. It has been suggested that modifications promoting biofilm growth or introducing a more complex substrate environment might encourage a re-organisation of the microbial community and encourage the emergence of new functional niches. We observed differences in the detected ASVs between the luminal and mucosal compartments of the bioreactors. For instance, in the study of M-ARCOL\_Deschamps\_et\_al\_2020, adding a mucosal component increased the number

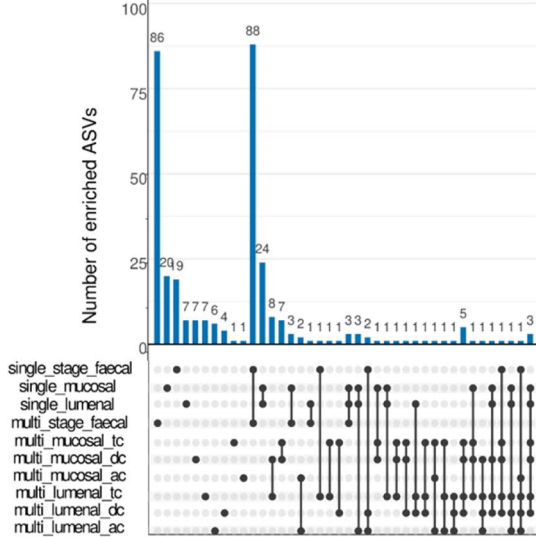
**A. ARCOL Leclerc 2021**



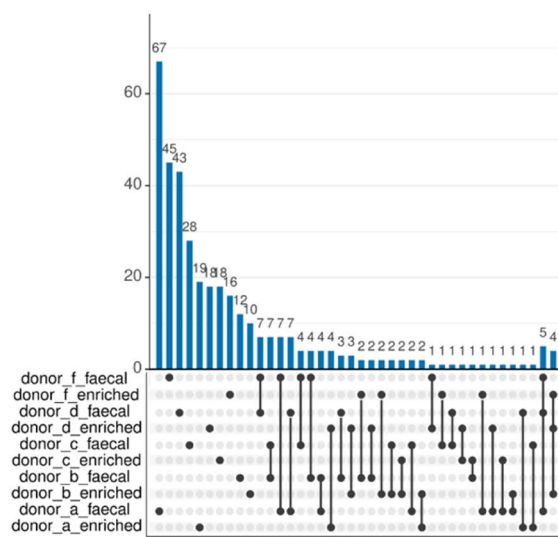
**B. M-ARCOL Deschamps 2020**



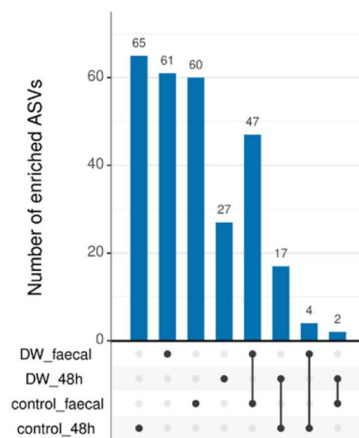
**C. SHIME Firman 2021**



**D. SHIME Liu 2022**



**E. SIMGI Zorraquin 2021**



of newly detected ASVs compared to the faecal sample (Fig. 6B). Nevertheless, the abundance of the ASVs varied greatly, ranging from 500 to 5000. The significance of these

changes needs further exploration, as it is necessary to rule out external contamination as the source of these new ASVs, especially considering their low abundance. Nonetheless,

**Fig. 6** Enrichment analysis of ASVs for selected samples. This graph shows the number of ASVs enriched in five bioreactor models. **A** SHIME\_Firman\_et\_al\_2021. **B** SIMGI\_Zorraquin\_et\_al\_2021. **C** ARCOL\_Leclerc\_et\_al\_2021. **D** M-ARCOL\_Deschamps\_et\_al\_2020. **E** SHIME\_Liu\_et\_al\_2022. The number of ASVs in the faecal samples is plotted as a reference ( $n \text{ reads} > 0$ ). The samples from the original faecal inoculum are labelled as “\_faecal”, while other labels correspond to samples from the bioreactors. The vertical bars represent the number of unique ASVs in each category. The dots and lines represent the comparisons that were performed between samples. Single dots indicate ASVs enriched in a given sample only, while dots connected with a line indicate that ASVs are shared between samples. This plot shows that new ASVs are detected in the bioreactors compared to the faecal inoculum

if a species of interest is enriched in the bioreactor, further enrichment protocols can promote its growth, isolation, and characterisation. This approach has proven successful in cultivating challenging microorganisms such as anammox bacteria at purities of up to 97% [15].

Although some of the families detected in the in vitro models were not found in GMRepo, a further literature search found documentation of *Gracilibacteraceae*, *Muribaculaceae*, *Defluviitaleaceae*, and *Vallitaleaceae* families identified in faecal microbiota samples [36–39]. Finally, the detection of the family *Vallitaleaceae* is surprising, as members of this family originated from hydrothermal fields and deep oceanic subsurface habitats rich in hydrocarbons [40]. Taken together, these results exhibit the potential of bioreactors to recover low abundant bacteria in faecal inocula and point to vast diversity yet to be cultured in these devices. The detailed list with the names of these families is available in Additional file 1: Table S3.

### Changes in the Microbial Diversity Inside Bioreactors

Regardless of the bioreactor model, once the microbial community is extracted from its natural environment, subsequent ecological changes occur inside this new artificial environment. Moreover, the culture of the faecal microbiota in the bioreactors is marked by a loss of diversity. For instance, in the study M-SHIME\_Chassaing\_et\_al\_2020, there was a decline of 50% in the  $\alpha$ -diversity of the bioreactor after a stabilisation period of 7 days post-inoculation [41]. This loss of diversity might be explained by (1) the loss of species that cannot grow because their nutrient requirements rely on compounds provided by the host, (2) the loss of species whose nutrition relies on co-metabolism and depend on the presence of other species, and (3) the artificial selection of certain bacterial populations that use more effectively the available nutrients.

Alternatively, we propose another complementary explanation to account for the loss of diversity besides the death of

multiple species [42]. We hypothesise that the selective pressure imposed by the conditions in the bioreactor can cause a decrease in the abundance (concentration of cells) of some species, which pushes them below the limit of detection of the sequencing method. In previous work, Gutierrez et al. [43] inoculated a soil sample in packed bioreactors fed with citrate as the sole carbon source and tracked the changes in the microbial community structure using 16S rRNA gene sequencing. Interestingly, some bacterial groups observed in the soil were not detected in the bioreactor. However, the bioreactor allowed the enrichment of new bacterial species undetected in the gene sequences of the soil sample. Then, they used the bioreactor’s effluent to fertilise sterile soil and found some bacterial species detected in the original inoculum, which apparently had disappeared in the bioreactor.

We suggest that the environmental conditions in the bioreactors, especially the absence of a human host, create new niche spaces that drive the replacement and proportional changes in the microbial community composition. These changes allow the growth of microorganisms whose activity is obscured by the presence of other, more dominant species in the human gastrointestinal tract. This feature highlights the potential of using bioreactors to enrich these microorganisms, allowing the role of these low-abundant species to be investigated.

Further research is required to understand the effect of the operational parameters of the bioreactor on diversity and to exploit the full potential of bioreactors to contribute new knowledge on gut microbial communities and to increase the number of species that can be cultured in vitro. We propose using bioreactors in tandem with culturomic approaches as another potential key to unlocking access to the microbial dark matter. More descriptive metadata of the environmental conditions of the communities growing in bioreactors is required to allow data from different studies to be co-modelled, which will accelerate our understanding of the value and limitations of bioreactors as a research tool. Bioreactors have proven to be powerful tools to enrich and even isolate species that are difficult to cultivate; therefore, they are emerging as a resource for studying microbial diversity and the ecological relationships of human microbiomes, especially those that inhabit the intestine. However, we must consider that each configuration and operation allows the enrichment of a group of microorganisms, overlapping or hiding others. Therefore, while the simulations of the intestine do not include all the environmental variables that the intestinal microorganisms experience, we must be cautious with drawing conclusions between the results obtained using in vitro models and a condition in human health.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00248-024-02369-0>.

**Acknowledgements** We thank Australian Research Council Laureate Fellowship funding for supporting EH, DG, and JS (FL200100220). We thank the Department of Jobs, Tourism, Science and Innovation, Government of Western Australian Premier's Fellowship for funding EH.

**Author Contribution** The study was conceived by David Felipe Garcia Mendez and Janeth Sanabria. Data collection, curation, and analysis were performed by David Felipe Garcia Mendez and Siobhon Egan. All authors contributed to the data interpretation and discussion of the results. The first draft of the manuscript was written by David Felipe Garcia Mendez, Siobhon Egan, and Janeth Sanabria. The final version and review of the contents were performed by Julien Wist and Elaine Holmes.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions We thank the Australian Research Council Laureate Fellowship funding for supporting EH, DG, and JS (FL200100220). We thank the Department of Jobs, Tourism, Science and Innovation, Government of Western Australian Premier's Fellowship for funding EH. EH is also supported by the UK Dementia Research Institute at Imperial College, which receives its funding from UK Dementia Research Institute Ltd., funded by the UK Medical Research Council (MRC), Alzheimer's Society, and Alzheimer's Research UK.

**Data Availability** No new sequencing datasets were generated during the current study. Details of accession numbers for data analysed can be found in Additional file 1: Table S2. Original R scripts and metadata files are available in Zenodo (<https://zenodo.org/doi/10.5281/zenodo.8045625>).

## Declarations

**Ethics Approval** Not applicable.

**Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

**Competing Interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- de Vos WM, Tilg H, Hul MV, Cani PD (2022) Gut microbiome and health: mechanistic insights. *Gut* 71:1020–1032. <https://doi.org/10.1136/gutjnl-2021-326789>
- Byrne CS, Blunt D, Burn J et al (2019) A study protocol for a randomised crossover study evaluating the effect of diets differing in carbohydrate quality on ileal content and appetite regulation in healthy humans. *F1000Research* 8:258. <https://doi.org/10.12688/f1000research.17870.2>
- Vuik F, Dicksved J, Lam S et al (2019) Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. *United Eur Gastroenterol J* 7:897–907. <https://doi.org/10.1177/2050640619852255>
- Hasan N, Yang H (2019) Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* 7:e7502. <https://doi.org/10.7717/peerj.7502>
- Thomas AM, Segata N (2019) Multiple levels of the unknown in microbiome research. *BMC Biol* 17:48. <https://doi.org/10.1186/s12915-019-0667-z>
- Knight R, Vrbanac A, Taylor BC et al (2018) Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422. <https://doi.org/10.1038/s41579-018-0029-9>
- Jiao J-Y, Liu L, Hua Z-S et al (2021) Microbial dark matter coming to light: challenges and opportunities. *Natl Sci Rev* 8:nwaa280. <https://doi.org/10.1093/nsr/nwaa280>
- Neville BA, Forster SC, Lawley TD (2018) Commensal Koch's postulates: establishing causation in human microbiota research. *Curr Opin Microbiol* 42:47–52. <https://doi.org/10.1016/j.mib.2017.10.001>
- Forster SC, Kumar N, Anonye BO et al (2019) A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat Biotechnol* 37:186–192. <https://doi.org/10.1038/s41587-018-0009-7>
- Browne HP, Forster SC, Anonye BO et al (2016) Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 533:543–546. <https://doi.org/10.1038/nature17645>
- Almeida A, Mitchell AL, Boland M et al (2019) A new genomic blueprint of the human gut microbiota. *Nature* 568:499–504. <https://doi.org/10.1038/s41586-019-0965-1>
- Costa J, Ahluwalia A (2019) Advances and current challenges in intestinal in vitro model engineering: a digest. *Front Bioeng Biotechnol* 7. <https://doi.org/10.3389/fbioe.2019.00144>
- Renwick S, Ganobis CM, Elder RA et al (2021) Culturing human gut microbiomes in the laboratory. *Annu Rev Microbiol*. <https://doi.org/10.1146/annurev-micro-031021-084116>
- Kuenen JG (2008) Anammox bacteria: from discovery to application. *Nat Rev Microbiol* 6:320–326. <https://doi.org/10.1038/nrmicro1857>
- Wu G, Zhang T, Gu M et al (2020) Review of characteristics of anammox bacteria and strategies for anammox start-up for sustainable wastewater resource management. *Water Sci Technol* 82:1742–1757. <https://doi.org/10.2166/wst.2020.443>
- Mabwi HA, Kim E, Song D-G et al (2021) Synthetic gut microbiome: advances and challenges. *Comput Struct Biotechnol J* 19:363–371. <https://doi.org/10.1016/j.csbj.2020.12.029>
- Krause JL, Schaepe SS, Fritz-Wallace K et al (2020) Following the community development of SIHUMIX – a new intestinal in vitro model for bioreactor use. *Gut Microbes* 11:1116–1129. <https://doi.org/10.1080/19490976.2019.1702431>
- Stolaki M, Minekus M, Venema K et al (2019) Microbial communities in a dynamic in vitro model for the human ileum resemble the human ileal microbiota. *FEMS Microbiol Ecol* 95. <https://doi.org/10.1093/femsec/fiz096>
- Molly K, Woestyne MV, Smet ID, Verstraete W (1994) Validation of the simulator of the human intestinal microbial ecosystem (SHIME) reactor using microorganism-associated activities. *Microb Ecol Health Dis* 7:191–200. <https://doi.org/10.3109/08910609409141354>
- Venema K (2015) The TNO In Vitro Model of the Colon (TIM-2). In: Verhoeckx K, Cotter P, López-Expósito I, et al (eds) The impact of food bioactives on health: in vitro and ex vivo models. Springer, Cham (CH)
- Barroso E, Cueva C, Peláez C et al (2015) Development of human colonic microbiota in the computer-controlled dynamic SIMulator

- of the GastroIntestinal tract SIMGI. *LWT - Food Sci Technol* 61:283–289. <https://doi.org/10.1016/j.lwt.2014.12.014>
22. Berner AZ, Fuentes S, Dostal A et al (2013) Novel polyfermentor intestinal model (PolyFermS) for controlled ecological studies: validation and effect of pH. *PLoS One* 8:2–11. <https://doi.org/10.1371/journal.pone.0077772>
  23. Cordonnier C, Thévenot J, Etienne-Mesmin L et al (2015) Dynamic in vitro models of the human gastrointestinal tract as relevant tools to assess the survival of probiotic strains and their interactions with gut microbiota. *Microorganisms* 3:725–745. <https://doi.org/10.3390/microorganisms3040725>
  24. Macfarlane GT, Macfarlane S, Gibson GR (1998) Validation of a three-stage compound continuous culture system for investigating the effect of retention time on the ecology and metabolism of bacteria in the human colon. *Microb Ecol* 35:180–187. <https://doi.org/10.1007/s002489900072>
  25. García Mendez DF, Sanabria J, Wist J, Holmes E (2023) Effect of operational parameters on the cultivation of the gut microbiome in continuous bioreactors inoculated with feces: a systematic review. *J Agric Food Chem*. <https://doi.org/10.1021/acs.jafc.2c08146>
  26. Gálvez-Merchán Á, Min KH (Joseph), Pachter L, Boeshaghi AS (2022) Metadata retrieval from sequence databases with ffq. 2022.05.18.492548
  27. Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>
  28. McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLOS Comput Biol* 10:e1003531. <https://doi.org/10.1371/journal.pcbi.1003531>
  29. Dai D, Zhu J, Sun C et al (2022) GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res* 50:D777–D784. <https://doi.org/10.1093/nar/gkab1019>
  30. Alberdi A, Gilbert MTP (2019) A guide to the application of Hill numbers to DNA-based diversity analyses. *Mol Ecol Resour* 19:804–817. <https://doi.org/10.1111/1755-0998.13014>
  31. Isenring J, Bircher L, Geirnaert A, Lacroix C (2023) In vitro human gut microbiota fermentation models: opportunities, challenges, and pitfalls. *Microbiome Res Rep* 2:2. <https://doi.org/10.20517/mrr.2022.15>
  32. Ma Z, (Sam), (2018) Chapter 8 - Measuring microbiome diversity and similarity with Hill numbers. In: Nagarajan M (ed) *Metagenomics*. Academic Press, pp 157–178
  33. Loreau M (2010) *The Challenges of Biodiversity Science. Excellence in Ecology*. Book 17. International Ecology Institute, Oldendorf/Luhe
  34. Średnicka P, Roszko MŁ, Popowski D et al (2023) Effect of in vitro cultivation on human gut microbiota composition using 16S rDNA amplicon sequencing and metabolomics approach. *Sci Rep* 13:3026. <https://doi.org/10.1038/s41598-023-29637-2>
  35. Lau JT, Whelan FJ, Herath I et al (2016) Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med* 8:72. <https://doi.org/10.1186/s13073-016-0327-7>
  36. Richarte V, Sánchez-Mora C, Corrales M et al (2021) Gut microbiota signature in treatment-naïve attention-deficit/hyperactivity disorder. *Transl Psychiatry* 11:1–7. <https://doi.org/10.1038/s41398-021-01504-6>
  37. Lagkouvardos I, Lesker TR, Hitch TCA et al (2019) Sequence and cultivation study of Muribaculaceae reveals novel species, host preference, and functional potential of this yet undescribed family. *Microbiome* 7:28. <https://doi.org/10.1186/s40168-019-0637-2>
  38. Smith BJ, Miller RA, Schmidt TM (2021) Muribaculaceae genomes assembled from metagenomes suggest genetic drivers of differential response to acarbose treatment in mice. *mSphere* 6:e00851-21. <https://doi.org/10.1128/msphere.00851-21>
  39. Shi M, Wei Y, Hu W et al (2018) The subgingival microbiome of periodontal pockets with different probing depths in chronic and aggressive periodontitis: a pilot study. *Front Cell Infect Microbiol* 8:124. <https://doi.org/10.3389/fcimb.2018.00124>
  40. Quéménéur M, Erauso G, Frouin E et al (2019) Hydrostatic pressure helps to cultivate an original anaerobic bacterium from the Atlantis Massif seafloor (IODP Expedition 357): *Petrocella atlantisensis* gen. nov. sp. nov. *Front Microbiol* 10. <https://doi.org/10.3389/fmicb.2019.01497>
  41. Chassaing B, Wiebe TVD, Bodt JD et al (2017) Dietary emulsifiers directly alter human microbiota composition and gene expression ex vivo potentiating intestinal inflammation. *Gut* 66:1414–1427. <https://doi.org/10.1136/gutjnl-2016-313099>
  42. Bellali S, Lagier J-C, Million M et al (2021) Running after ghosts: are dead bacteria the dark matter of the human gut microbiota? *Gut Microbes* 13:1–12. <https://doi.org/10.1080/19490976.2021.1897208>
  43. Gutiérrez CF, Rodríguez-Romero N, Egan S et al (2022) Exploiting the potential of bioreactors for creating spatial organization in the soil microbiome: a strategy for increasing sustainable agricultural practices. *Microorganisms* 10:1464. <https://doi.org/10.3390/microorganisms10071464>