



OPEN ACCESS

Original research

# Absence of a pancreatic microbiome in intraductal papillary mucinous neoplasm

Marie-Madlen Pust <sup>1,2,3</sup>, Darío Missael Rocha Castellanos <sup>4</sup>, Kara Rzasa <sup>1</sup>,  
Andrea Dame,<sup>1</sup> Gleb Pishchany,<sup>1,5</sup> Charnwit Assawasirisin,<sup>4</sup> Andrew Liss,<sup>4</sup>  
Carlos Fernandez-del Castillo,<sup>4</sup> Ramnik J Xavier<sup>1,3</sup>

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/gutjnl-2023-331012>).

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

<sup>2</sup>Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

<sup>3</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>4</sup>Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

<sup>5</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA

## Correspondence to

Dr Ramnik J Xavier and Dr Marie-Madlen Pust, Broad Institute, Cambridge, MA 02142, USA; [xavier@molbio.mgh.harvard.edu](mailto:xavier@molbio.mgh.harvard.edu), [mpust@mgh.harvard.edu](mailto:mpust@mgh.harvard.edu)

and Dr Carlos Fernandez-del Castillo, Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA; [cfernandez@mgh.harvard.edu](mailto:cfernandez@mgh.harvard.edu)

Received 28 August 2023  
Accepted 6 February 2024  
Published Online First  
1 March 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Pust M-M, Rocha Castellanos DM, Rzasa K, *et al.* *Gut* 2024;**73**:1131–1141.

## ABSTRACT

**Objective** This study aims to validate the existence of a microbiome within intraductal papillary mucinous neoplasm (IPMN) that can be differentiated from the taxonomically diverse DNA background of next-generation sequencing procedures.

**Design** We generated 16S rRNA amplicon sequencing data to analyse 338 cyst fluid samples from 190 patients and 19 negative controls, the latter collected directly from sterile syringes in the operating room. A subset of samples (n=20) and blanks (n=5) were spiked with known concentrations of bacterial cells alien to the human microbiome to infer absolute abundances of microbial traces. All cyst fluid samples were obtained intraoperatively and included IPMNs with various degrees of dysplasia as well as other cystic neoplasms. Follow-up culturing experiments were conducted to assess bacterial growth for microbiologically significant signals.

**Results** Microbiome signatures of cyst fluid samples were inseparable from those of negative controls, with no difference in taxonomic diversity, and microbial community composition. In a patient subgroup that had recently undergone invasive procedures, a bacterial signal was evident. This outlier signal was not characterised by higher taxonomic diversity but by an increased dominance index of a gut-associated microbe, leading to lower taxonomic evenness compared with the background signal.

**Conclusion** The ‘microbiome’ of IPMNs and other pancreatic cystic neoplasms does not deviate from the background signature of negative controls, supporting the concept of a sterile environment. Outlier signals may appear in a small fraction of patients following recent invasive endoscopic procedures. No associations between microbial patterns and clinical or cyst parameters were apparent.

## INTRODUCTION

Intraductal papillary mucinous neoplasms (IPMNs) of the pancreas are precursors of pancreatic cancer and are being diagnosed with increasing frequency. Not unlike other tumours, inflammation is considered a critical culprit of progression in IPMN.<sup>1</sup> Thus, a deeper understanding of this process and searching for ways to arrest it are being actively pursued, for example, in a clinical trial comparing Sulindac (a non-steroidal inflammatory drug) to placebo in this patient population (ClinicalTrials.gov Identifier: NCT04207944). Bacteria can be the

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Inflammation within intraductal papillary mucinous neoplasms (IPMNs) is associated with increased dysplasia. Some studies have suggested that the presence of complex microbial communities within the cyst may be the drivers of the inflammatory response.

## WHAT THIS STUDY ADDS

⇒ Our findings indicate there is no microbiome within IPMNs, and that the presence of bacteria within outlier cyst fluid samples is mainly related to prior invasive intervention.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Other causes of inflammation should be considered in the progression of IPMN. Intervention with antibiotics is unlikely to have any benefit in this disease.

drivers of an inflammatory response, and recent studies have proposed that the pancreatic cyst is colonised by a unique and diverse population of bacteria.<sup>2,3</sup> This notion of a ‘pancreatic cyst microbiome’ contradicts the view, from the presequencing era, that the pancreas is sterile. The presence of a microbiome, which is defined as a niche-specific, stable and interactive micro-ecosystem,<sup>4,5</sup> within the pancreatic cyst would carry profound implications for clinical medicine. If proven correct, it would require a comprehensive review of established concepts on early cyst and tumour development in the pancreas and mechanisms of bacterial translocation in the human body.

However, the available literature on the low-biomass pancreatic cyst microbiome raises concerns. On the one hand, some studies have obtained the cyst fluid samples via endoscopic ultrasound, which risks contamination from the oral, gastric and duodenal microbiota. Most studies did not report the implementation of negative controls sampled, processed, sequenced and analysed alongside the low-biomass patient samples. Microbial DNA contamination is unavoidable, as it is present in all laboratory reagents and kits, often referred to as the ‘kitome’ and is further introduced during sampling and sample processing, including cross-contamination between samples, known as the ‘splashome’.<sup>6–9</sup>

Moreover, the type and degree of DNA contamination fluctuate within and between laboratories, ultimately affecting data analysis, data interpretation, replicability and reproducibility of low-biomass microbiome studies without negative controls.<sup>9</sup> Even though past guidelines have recommended sequencing negative controls, their application has not been standardised in the low-biomass microbiome research field.<sup>7</sup> The lack of standardisation leads to persistent debates regarding the existence of certain low-biomass microbial communities, including prenatal or blood microbiomes, in both health and disease states. Notably, when these niches were reassessed incorporating negative controls, distinct and clinically relevant microbiome signatures were no longer observed.<sup>10–12</sup> Another point of contention emerges from studies on the ‘cyst fluid microbiome’, which have equated the isolation of microbial DNA fragments with the existence of a microbiome, hence suggesting diverse microbial proliferation and stable host colonisation. However, studies using DNA sequencing for bacterial identification, without follow-up verification of bacterial growth and stable metabolic activity, cannot refute the concept of a sterile pancreas and cyst environment; by definition, sterility refers to the absence of living microbes but not to the absence of microbial DNA.<sup>12</sup>

These issues give rise to two key questions: Can we differentiate a unique microbiome signature in pancreatic cyst fluid from the DNA background noise? If so, does this signature indicate the presence of active microbial growth within the pancreatic cyst fluid? To probe these questions, we carried out 16S rRNA amplicon sequencing on 338 cyst fluid samples obtained from 190 patients who underwent surgical resection and pathological evaluation. All samples were obtained with sterile technique during surgery directly from the cystic lesion, avoiding any passage through the GI tract. Most patients (145, 76%) had an IPMN, and the cohort included tumours with low and high-grade dysplasia, as well as 42 patients who had invasive carcinoma within the IPMN. We also included 21 patients with other cystic pancreatic neoplasms such as neuroendocrine cystic tumours, mucinous cystic neoplasms and serous cystadenomas, 16 patients with conventional pancreatic ductal adenocarcinoma and cystic degeneration and 8 patients who either had pseudocysts or other less common pancreatic cystic lesions (table 1). Furthermore, 19 negative controls obtained from the operating room itself using a sterile syringe as the sampling device were implemented. Through reprocessing and resequencing, a subset of samples and negative controls up to three times with the same protocols, we examined the effects of background contamination and cross-contamination on our findings. The existence of bacterial growth for microbiologically relevant signals was verified by follow-up culturing experiments. Overall, our study failed to uncover a pancreatic cyst microbiome that was different from the microbiome signature of the negative controls. This finding supports the hypothesis of a predominantly sterile pancreatic cyst environment.

## METHODS

### Study population and sample collection

All cyst fluid samples and negative controls were obtained at Massachusetts General Hospital (Boston, Massachusetts, USA) with sterile technique at the time of surgical resection directly from the cystic lesion, avoiding any passage through the GI tract (figure 1A.1). The samples were transported to a research laboratory where they were aliquoted and stored at  $-80^{\circ}\text{C}$ . Patient samples were obtained under institutional review board approval and patients had signed consent for use of tissue, blood and fluid

**Table 1** Patient characteristics (n=190)

Variable	Response	IPMN N=145* (76%)	Other cysts N=45† (24%)
Cyst location	Body/uncinate	81 (55.9%)	31 (68.8%)
	Body/tail	64 (44.1%)	14 (32.2%)
Grade consensus	Low grade	74 (51.0%)	26 (57.8%)
	High grade	71 (49.0%)	19 (42.2%)
Invasive status	Yes	42 (29.0%)	19 (42.2%)
	No	103 (71.0%)	21 (46.7%)
	N/A	0	5 (11.1%)
Pancreatitis	Yes	17 (11.7%)	4 (8.9%)
	No	128 (88.3%)	41 (91.1%)
Jaundice	Yes	11 (7.6%)	9 (20%)
	No	134 (92.4%)	36 (80%)
Endoscopic ultrasound	Yes	127 (87.6%)	31 (68.9%)
	No	18 (12.4%)	14 (31.1%)
Fine needle aspiration	Yes	117 (80.7%)	24 (53.3%)
	No	28 (19.3%)	21 (46.7%)
Sex	Male	72 (49.7%)	19 (42.2%)
	Female	73 (50.3%)	26 (57.8%)

Data were described as number and percentage.  
 \*Branch-duct IPMN was 65 cases, main-duct was 12 cases and mixed type was 68 cases.  
 †Pancreatic cancer was 16 cases, neuroendocrine tumour was 6 cases, mucinous cystic neoplasm was 8 cases, serous cystadenoma was 7 cases, pancreatic pseudocyst was 4 cases and others for 4 cases.  
 IPMN, intraductal papillary mucinous neoplasm.

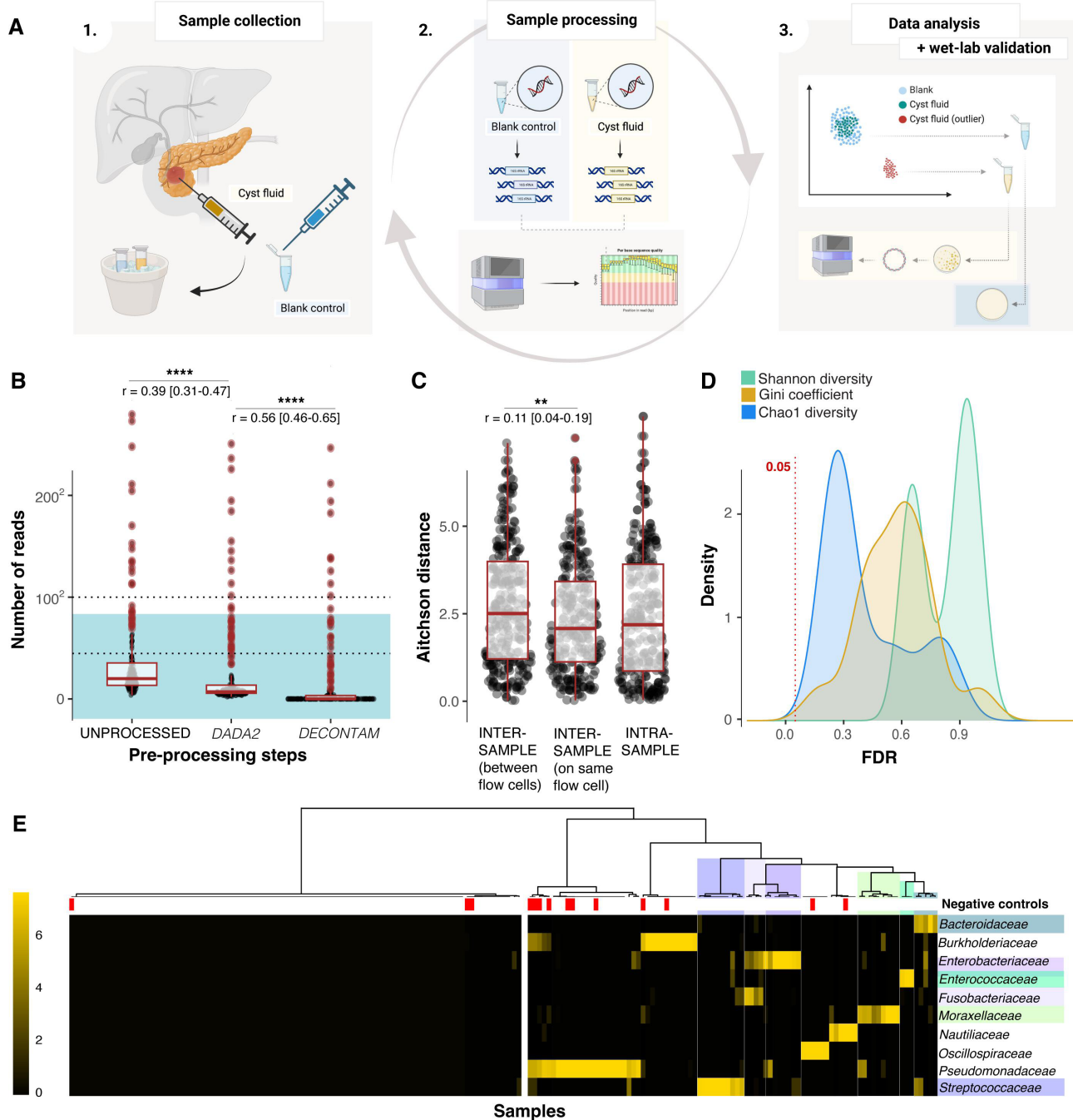
for research purposes. The corresponding diagnosis, cyst and patient characteristics as well as prior endoscopic interventions are summarised in table 1. A detailed, anonymised metadata table is provided in online supplemental table 1.

### ZymoBIOMICS Spike-in Control for low microbial load spike-ins

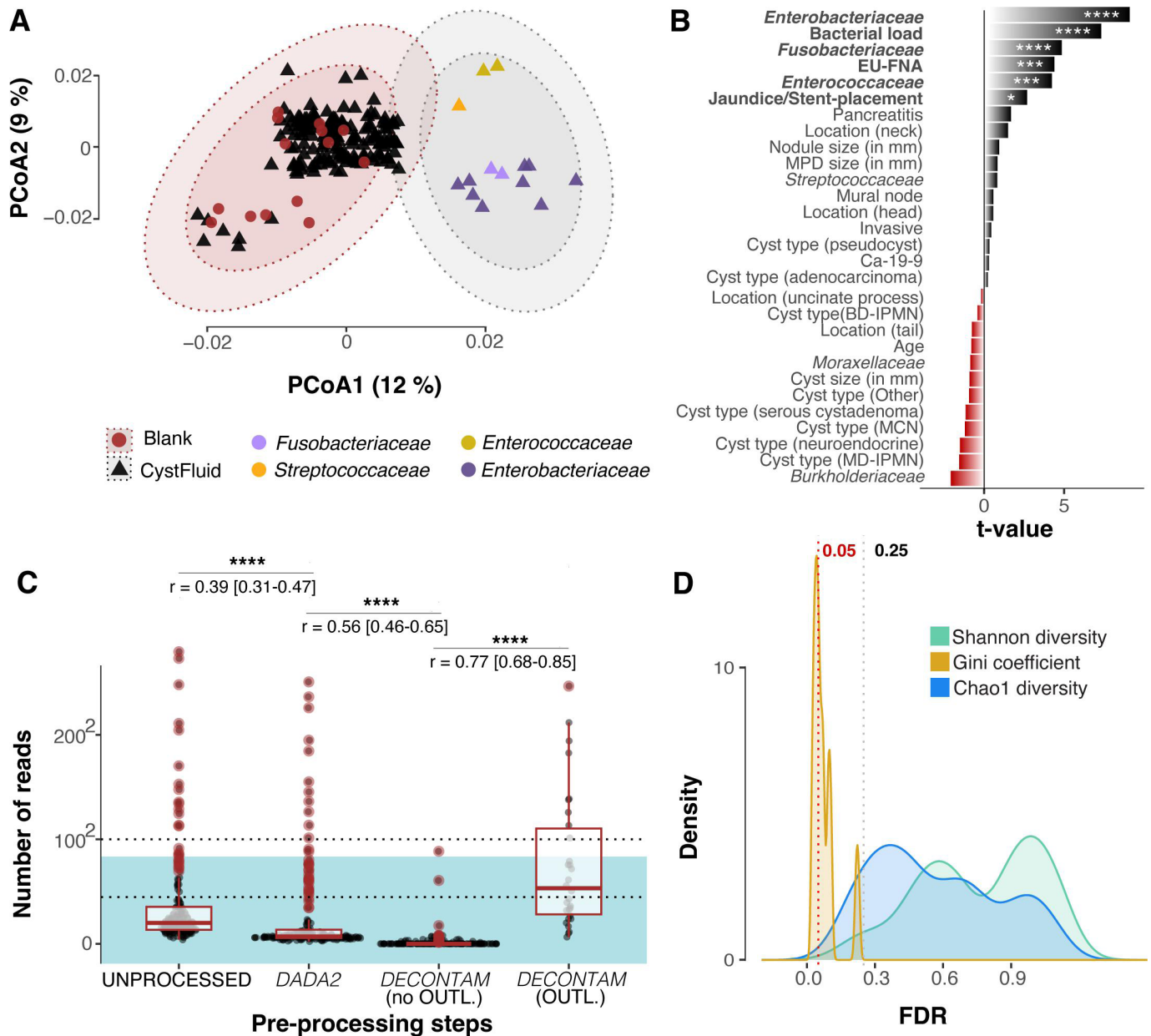
The BIOMICS Spike-in Control II mixture (Low Microbial Load, 20  $\mu\text{L}$ ,  $\sim 0.39$  ng) consisting of three bacterial strains alien to the human microbiome, *Truepera radiovictrix*, *Imtechella halotolerans* and *Allobacillus halotolerans* was added to a subset of cyst fluid samples (n=20) and blank controls (n=5) to ensure the accurate quantification of absolute DNA from microbial traces.

### DNA extraction and 16S amplicon library preparation

DNA was extracted using Qiagen DNeasy PowerSoil HTP 96 Well Soil DNA Isolation Kit (Cat# 12955-4) with bead-beating on TissueLyzer II at 20 Hz for 10 min per the manufacturer’s recommendations. 16S rRNA gene amplicon libraries targeting the V4 region were constructed with the 515F primer (5′-AATGATACGCGACCACCGAGATCTACA CTATGGTAATTGTGTGC CAGCMGCCGCGGTAA-3′) and unique reverse barcode primers from the Golay primer set.<sup>13</sup> After amplification, the samples were purified via the Agencourt AMPure XP-PCR purification system. DNA concentrations of the purified libraries were measured by Quant-IT (Life Technologies), normalised and pooled. The pooled library was quantified by Qubit (Life Technologies) and analysed on an Agilent 4200 TapeStation with a D1000 ScreenTape. The library was sequenced on an Illumina MiSeq System with MiSeq Reagent Kit v2 (500 cycles, Cat# MS-102-2003) using custom index 5′-ATTAGAWACCCBDGTAGTCC GG CTGACT-GACT-3′ and custom Read 1 (5′-TATGGTAATTGTGTG



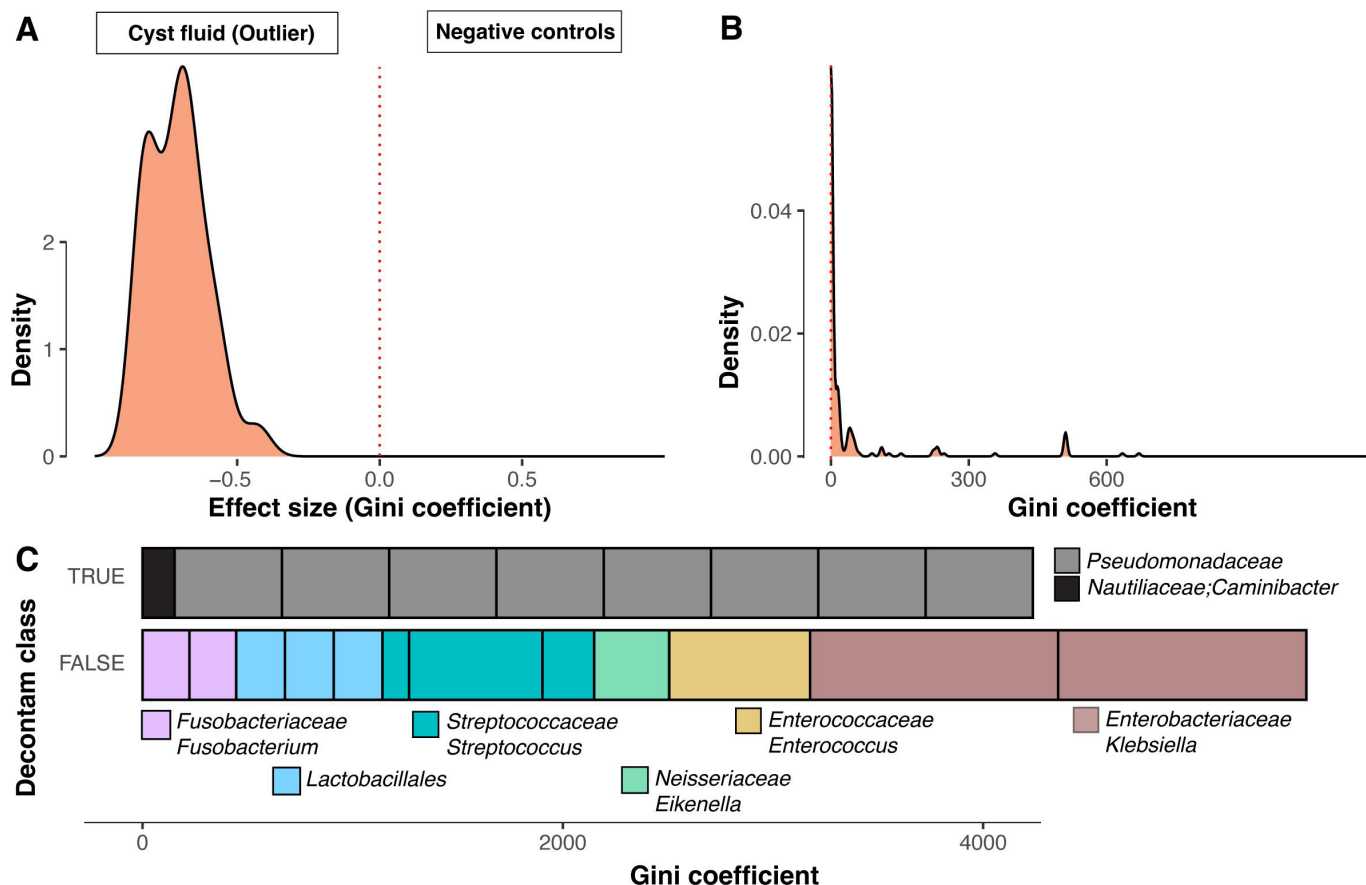
**Figure 1** Overview of experimental design, data quality and microbial signals. (A.1) Sample collection. Cyst fluid (n=190) and 19 controls (n=19) were collected in the operating room and stored at  $-80^{\circ}\text{C}$ . (A.2) Sample processing with DNA extraction and sequencing on the MiSeq Illumina platform. Technical replicates of cyst fluid samples were incorporated. (A.3) Data analysis and validation. Various analytical methods, including hierarchical clustering, and PCoA microbial community analysis were applied. Culturing experiments were conducted validating bacterial viability. The figure was created with BioRender.com (B) Gradual decrease in read quantity throughout preprocessing steps. Boxplots represent changes in central tendency of read counts through preprocessing steps (raw, post-DADA2 and post-DECONTAM). Significant differences were assessed by Mann-Whitney U tests with effect size  $r$  and CI computations. For context, the blue background illustrates the maximum raw read count from negative controls. (C) Extent of cross-contamination. Aitchison distances were compared to assess microbial signal variation between technical replicates (INTRASAMPLE) and samples from different patients on the same flow cell (INTERSAMPLE, same flow cells) or between sequencing runs (INTERSAMPLE, different flow cells). Aitchison distances between technical replicates were comparable with those of intersubject distances from the same flow cell (Wilcoxon test,  $p>0.3$ ;  $r=-0.04$  (-0.1 to 0.03)) and across flow cells (Wilcoxon test,  $p>0.05$ ;  $r=0.07$  (0–0.14)). (D) Comparative analysis of Shannon diversity (green), Chao1 diversity (blue) and the Gini coefficient (orange) between cyst fluid and controls. A resampling approach was deployed, randomly and iteratively selecting five samples from each group for comparison (n=30 iterations). Benjamini-Hochberg correction was applied to Mann-Whitney U p values to account for the multiple testing strategy (FDR). (E) Agglomerative hierarchical clustering based on Euclidean distances of column-scaled ASV counts from cyst fluid samples and controls (x-axis) at the family level (y-axis). The analysis of dominating ASVs on the family level unravelled the presence of controls in most clusters by minimising the total within-cluster variation. Only a subset of samples fell into groups devoid of any controls (coloured cluster background). FDR, false discovery rate; PCoA, principal coordinate analysis.



**Figure 2** Principal coordinate analysis (PCoA) of cyst fluid and controls obtained from Bray-Curtis dissimilarity matrices. (A) PCoA representation of microbial community signals in cyst fluid and controls. The main cluster encompassed all negative controls (indicated in red, circle) containing most cyst fluid samples (black, triangle). Every sample within this primary cluster resided either within the 95% or 98% CIs (red ellipses) of the controls. A minor group of outlier cyst fluid samples formed a subcluster (grey ellipse), suggesting a distinct microbiological signal. (B) Investigation into the clinical and ecological factors contributing to the clustering with a linear mixed effect model, incorporating sex as random effect coefficients and the PCoA1 coordinate as response variable. The two-dimensional embedding was primarily influenced by higher bacterial load, dominating presence in the sample of either an *Enterobacteriaceae* sp, *Fusobacteriaceae* sp, *Enterococcaceae* sp and recent invasive procedures (stent-placement, endoscopic ultrasound/fine-needle aspiration: EU-FNA) with FDR below 0.05 (\*), 0.001 (\*\*\*) or 0.0001 (\*\*\*\*). (C) Assessing the gradual decrease in read quantity based on PCoA clusters. The boxplot represents changes in the central tendency of read counts through different preprocessing phases (raw reads, post-DADA2 and post-DECONTAM). The DECONTAM group has been split based on samples falling in the main cluster of the PCoA (DECONTAM, no OUTL.) or in the outlier subcluster (DECONTAM, OUTL.). The significance was quantified using Mann-Whitney U tests and the corresponding effect size  $r$  with CIs. For context, the blue background illustrates the maximum raw read count from unprocessed negative controls. (D) Comparative analysis of Shannon diversity (green), Chao1 diversity (blue) and the Gini coefficient (orange) between cyst fluid outlier samples (PCoA subcluster) and controls. Due to unequal group sizes, a resampling approach was deployed, randomly and iteratively ( $n=30$  iterations) selecting five samples from each group for comparison. Mann-Whitney U tests and effect size  $r$  with CIs were computed with Benjamini-Hochberg correction to account for multiple testing (FDR). FDR, false discover rate; IPMN, intraductal papillary mucinous neoplasm.

YCAGCMGCCGCGGTAA-3') and Read 2 (5'-AGTCAGT-CAGCCGACTACNVGGGTWTCTAAT-3') primers. FASTQ files are publicly available under the European Nucleotide

Archive (ENA) accession number PRJEB64198. Dog stool samples were implemented as positive controls for DNA extraction, library construction and sequencing, representing



**Figure 3** Exploration of the microbial signals observed in the outlier subcluster relative to the controls in the main cluster. (A) Gini coefficient effect size validation. The strong negative effect size identified in the comparison of Gini coefficients suggests a pronounced dominance of particular microbial species in the outlier principal coordinate analysis samples when contrasted with the control samples. (B) Identification of ASVs with elevated Gini coefficients. The y-axis of the bar plots represents *DECONTAM* results, with 'TRUE' referring to a contamination prediction and 'FALSE' referring to a true positive prediction. A dominating pathogen, such as *Klebsiella*, *Enterococcus* or *Streptococcus* was present in outlier samples based on their high Gini coefficients. The outlier sample with dominating *Enterococcus* was found to be infected with *Enterococcus faecium* (> 100 000 CFU/mL), while a representative negative control showed no microbial growth.

a known entity of a unique and diverse microbial community structure (figure 1A.2).

### Bacterial isolation and verification

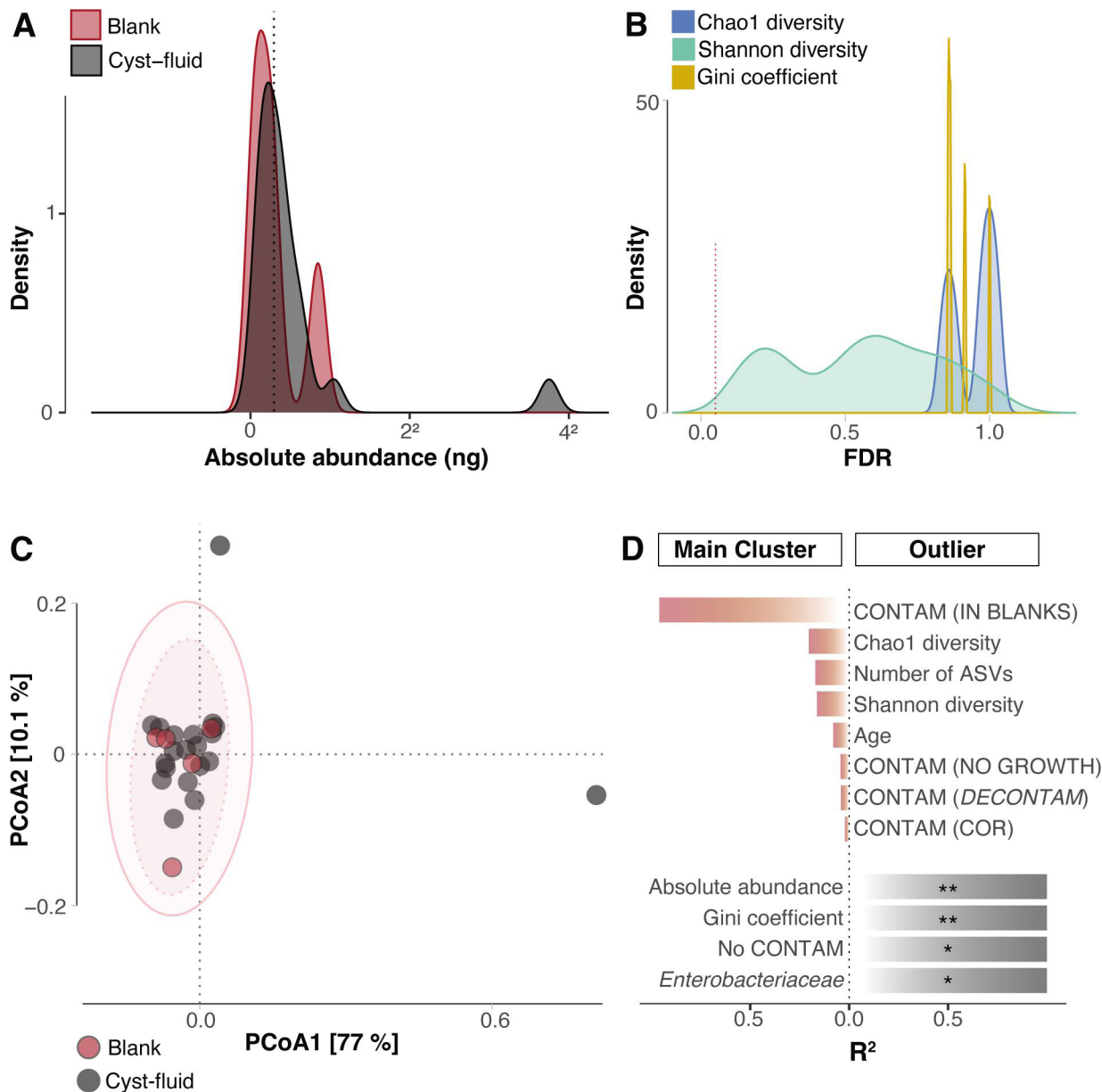
Based on the statistical results of 16S rRNA gene sequencing data, cyst fluid outlier samples were serially diluted and plated on Brucella Blood Agar and Yeast Casitone Fatty Acids with Carbohydrates Agar and incubated under aerobic and anaerobic conditions for 48 hours (figure 1A.3). For two cyst fluid samples, single morphology colonies were observed after 24 hours of growth. Based on the number of colonies, the samples contained  $4.2 \times 10^5$  and  $2.4 \times 10^5$  colony forming units (CFU) per millilitre of cyst fluid, respectively. DNA was isolated from three colonies each.

The full-length 16S rRNA gene was amplified with 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-GGTACCTTGTACGACTT-3') primers and Sanger sequenced with the 27F primer<sup>15</sup>.

### Statistical analysis

The statistical analysis was executed in R (V.4.2.1) on an Ubuntu 20.04.5 LTS system. The *DADA2* package (V.1.24.0) was used to model, correct and filter Illumina-sequenced amplicon errors.<sup>16</sup> Functions such as *filterAndTrim*, *derepFastq*

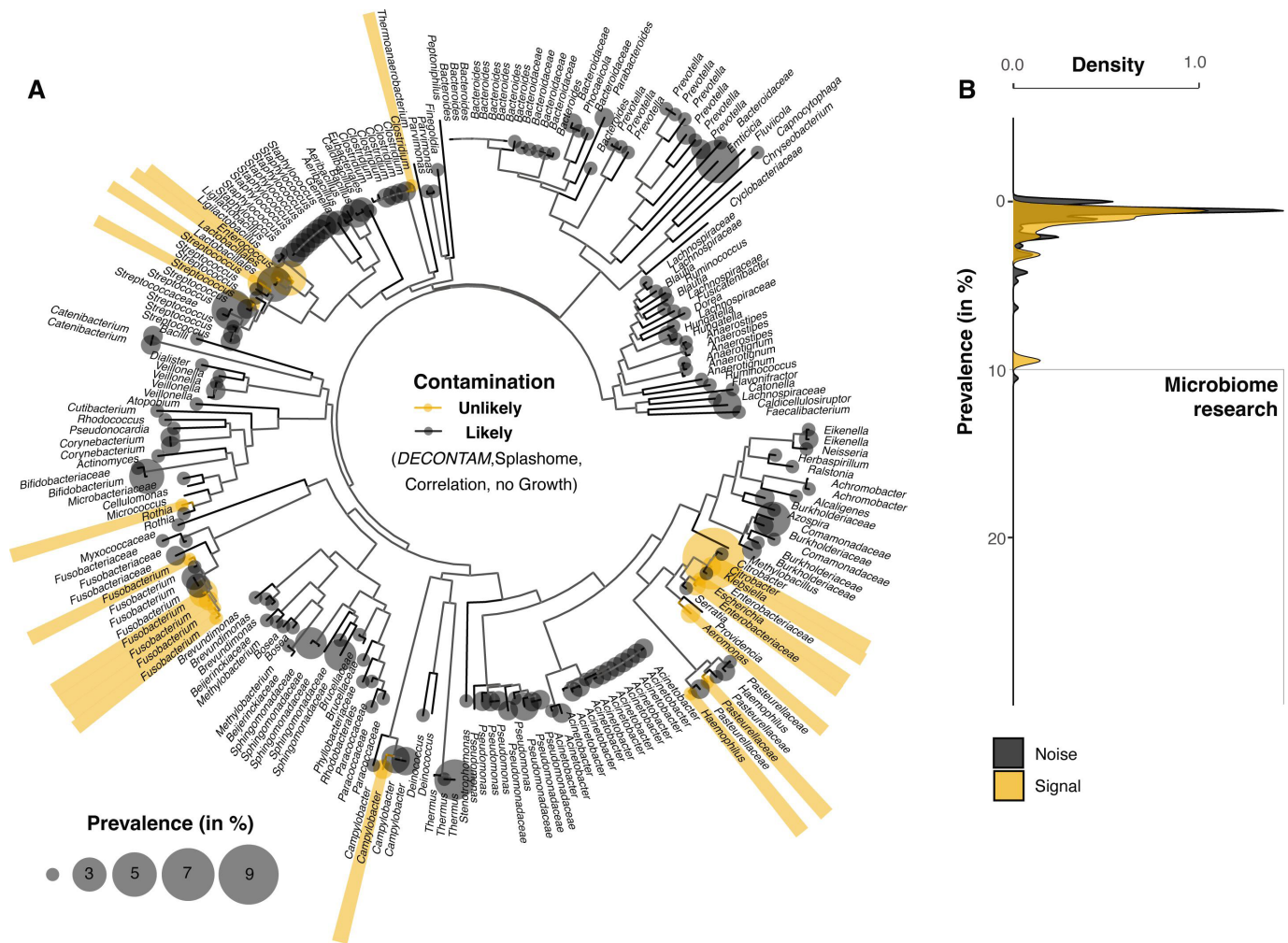
and *removeBimeraDenovo* were employed for adapter trimming, duplicate removal and chimera detection, respectively. Furthermore, both the *learnErrors* and core sample inference *dada* function, executed in multithreaded modes, employed an expectation-maximisation algorithm, computing error rate profiles across filtered reads, thereby enhancing the accuracy of inferred biological variants that were too prevalent to be attributed to amplicon errors. *mergePairs* denoised forward and reverse read pairs separately. A sample-by-sequence observation matrix was created using *makeSequenceTable*, and amplicon sequence variants (ASVs) were annotated with three different reference databases, namely the Genome Taxonomy Database for bacterial and archaeal 16S amplicon taxonomic analysis (GTDB),<sup>17</sup> the SILVA ribosomal RNA gene database<sup>18</sup> and the nucleotide collection database from the National Center for Biotechnology Information (NCBI)<sup>19</sup>. Lastly, the *isContaminant* algorithm provided in R's *DECONTAM* package (V.1.16.0) was applied, flagging all ASVs that were more prevalent in negative controls than in biological samples (prevalence=0.5).<sup>20</sup> For a subset of cyst fluid samples (n=20) and blank controls (n=5) with spiked-in bacterial cells of known concentration, the absolute abundance (in ng) was inferred for remaining microbial DNA reads per sample, so that *DECONTAM*'s frequency and prevalence filter could be used. Next to *DECONTAM*, ASV contamination was



**Figure 4** Evaluation of trace microbial communities in cyst fluid with bacterial spike-ins. (A) Density plot of absolute abundance estimates in controls (red) and cyst fluid (black). Overall, no significant group difference was unravelled (Wilcoxon test,  $p > 0.25$ ). The dotted black line depicts the median absolute abundance value of 0.09 ng. (B) Alpha-diversity analysis of potential microbial communities based on absolute microbial abundance. Due to unequal group sizes, a resampling approach was deployed, randomly and iteratively ( $n = 100$  iterations) selecting five samples from each group for comparison. Mann-Whitney U tests and the corresponding effect sizes  $r$  with CIs were computed using the Benjamini-Hochberg correction to account for multiple testing. Neither Chao1 diversity (blue), Shannon diversity (green) nor Gini coefficients (orange) were found to differ significantly between controls and cyst fluid. (C) Assessment of microbial community structure between controls and cyst fluid. Microbial community compositions were analysed with PCoA based on Bray-Curtis dissimilarity indices derived from absolute microbial abundance matrices. The PCoA plot demonstrates that except for two outliers, all cyst fluid samples cluster within the 80% and 95% CIs of the blank controls, as depicted by red ellipses with dotted and solid lines, respectively, suggesting similar microbial community structures. (D) Identifying the variables contributing to the two-dimensional feature space using a regression model. While both outlier samples showed significantly higher bacterial load, an elevated dominance of distinct ASVs (Gini coefficient), annotated as *Enterobacteriaceae*, and a higher number of ASVs passing contamination evaluation (No CONTAM), outlier samples also showed a tendency of lower Shannon diversity and ASV counts. ASV contamination was identified through prediction by *DECONTAM*, their presence in negative controls (IN BLANKS), the absence of bacterial growth in culture (NO GROWTH), not consistently found across technical replicates (Splashome) or their positive correlation with previously predicted ASV contaminants (COR). FDR below 0.05 (\*) or 0.01 (\*\*) is shown. FDR, false discovery rate; PCoA, principal coordinate analysis.

identified through their detection in negative controls, the absence of corresponding bacterial growth in culture, not consistently found across technical replicates (Splashome) or positive correlation with previously predicted ASV contaminants. We

performed non-parametric Mann-Whitney U tests for pairwise group comparisons and computed the effect size  $r$ , which is calculated by dividing the test statistic by the square root of the sample size. We also obtained CIs and if necessary, applied the



**Figure 5** Sequence similarity and ASV prevalence across negative controls and cyst fluid samples by merging the datasets of both experiments with and without absolute abundance information. (A) Phylogenetic tree of ASVs across experiments. The tree was generated for all ASVs identified in controls and cyst fluid using neighbour-joining tree estimation. Node size indicates the prevalence of ASVs across cyst fluid samples from both experiments, with each node representing a distinct ASV. While orange-coloured nodes represent potential true positive ASVs, with bacterial viability confirmed in culture for *Enterococcus* and *Citrobacter*, black nodes depict predicted contamination. ASV contamination was identified through prediction by *DECONTAM*, their presence in negative controls, the absence of bacterial growth in culture, not consistently found across technical replicates (Splashome) or their positive correlation with previously predicted ASV contaminants. (B) Prevalence of ASVs in our cohort. Both predicted true (signal, orange) and false positive (noise, black) ASVs occurred at low prevalence (<10%), indicating the absence of a cohort-, disease-specific microbiome signal.

Benjamini-Hochberg correction to account for multiple testing. In instances of significantly unbalanced group sizes, for example, blank controls versus cyst fluid samples or outliers, we adopted a resampling approach. This involved randomly and iteratively drawing subsets of equal size from both groups and conducting repeated statistical comparisons. Microbial community composition was compared with a hierarchical agglomerative clustering strategy of cyst fluid and controls, based on scaled Euclidean distances of family-level ASV profiles. To investigate the influence of clinical and ecological factors on sample clustering in principal coordinate analysis (PCoA) of Bray-Curtis dissimilarities, we either employed a linear mixed-effects model, incorporating patient's sex as a random effect coefficient with the first dimension as the response variable, or a regression model to integrate the variables into the ordination process. Information about the R environment, package versions and scripts are publicly available (dataset).<sup>21</sup>

## RESULTS

### Reads derived from cyst fluid samples exhibit comparable quality and quantity to controls

16S rRNA amplicon sequencing of the unspiked cyst fluid samples produced an average raw read count of 4607, accompanied by a substantial SD of 11 445 (figure 1B). The high deviation suggests that a small number of samples were outliers with higher read recovery rates, compared with the generally low recovery rates for most of the samples. Meanwhile, the positive standards processed alongside yielded an almost six times higher average raw read count, at 26 911. Following data cleaning procedures, including trimming, filtering and error correction (*DADA2*), the mean read count per cyst fluid sample dropped to 3359 (SD of 9558) (figure 1B). This represented a significant reduction in read count, evidenced by a moderate effect size of 0.39 (0.31–0.47). Subsequent application of the *DECONTAM* software resulted in a steeper decrease in read count, down to a mean

of 2093 (SD of 7457) (figure 1B). This decrease occurred after discarding ASVs that were more prevalent in negative controls than cyst fluid samples. Throughout the preprocessing stages, an increasing percentage of cyst fluid samples, from 90% to 92%, and ultimately more than 94%, demonstrated read counts similar to those of negative controls after filtering with *DADA2* and *DECONTAM* (figure 1B), respectively. Prior studies suggested a minimum amount of 2000 short reads as necessary to accurately characterise a microbial community profile.<sup>13</sup> Therefore, only a limited number of unspiked outlier cyst fluid samples (~10%) met or exceeded this lower read limit post-*DECONTAM*, revealing the absence of an interpretable microbiome signal in approximately 90% of cyst fluid samples when subtracting potential background noise with *DECONTAM*. Given that prior research studies did not highlight prevalent or high-frequency *DECONTAM*-labelled contaminants on account of missing negative controls, we decided to perform the follow-up analysis on count matrices before running *DECONTAM*, but we assessed negative controls alongside cyst fluid samples.

Moreover, a subset of unspiked samples underwent reprocessing and resequencing up to three times (technical replicates), adhering to the same wet-lab protocols. The primary objective was to evaluate the effect of cross-contamination on microbial signatures. For this purpose, Aitchison distances were calculated pre-*DECONTAM* in two scenarios: within technical replicates of the same cyst fluid sample (intrasubject) and between samples from different patients (intersubject). These intersubject samples were either processed together on the same flow cell or on different flow cells on separate days. Intersubject cyst fluid, processed simultaneously on the same flow cell, had more similar ASV signatures as indicated by significantly lower Euclidean distances when compared with those processed on different days and run on separate flow cells (figure 1C). Importantly, ASV signatures from technical replicates were not more similar than intersubject profiles from the same flow cell (Wilcoxon test,  $p > 0.3$ ,  $r = -0.04$  (-0.1 to 0.03), figure 1C) or different flow cells (Wilcoxon test,  $p > 0.05$ ;  $r = 0.07$  (0–0.14), figure 1C). Thus, no reproducible, individual microbial community signal was recovered from technical replicates, demonstrating remarkable low signal-to-noise ratios in our ultra-low biomass cyst fluid samples.

### Alpha- and beta-diversity indices are comparable between cyst fluid samples and controls

To address the statistical challenges posed by the disparity in group sizes between unspiked negative controls ( $n=14$ ) and cyst fluid samples ( $n=318$ ), we employed an iterative analytical approach pre-*DECONTAM* for comparing alpha-diversity metrics, including Chao1 diversity, Shannon diversity and the Gini coefficient, a measure of dominance. This involved repeatedly selecting random combinations of samples ( $n=5$ ) and blanks ( $n=5$ ) from the larger dataset and performing the statistical comparison on subgroups of equal size (figure 1D). The false discovery rate density peaks for all three ecological parameters were observed to be far greater than 0.25, indicating no statistically significant difference in alpha diversity between controls and cyst fluid (figure 1D).

Next, we incorporated a hierarchical agglomerative clustering technique to unravel differentiating microbiome signals between unspiked cyst fluid samples and controls. This approach, based on Euclidean distances derived from column-scaled, family-level ASV profiles pre-*DECONTAM*, underlined the overarching similarity between controls and

biological samples (figure 1E). Notably, only a small subset of cyst fluid samples, highlighted with purple and green backgrounds in the dendrogram, formed distinct clusters. These clusters were marked by a dominance of scaled ASV counts from certain bacterial families, namely either Streptococcaceae, Moraxellaceae, Fusobacteriaceae, Enterococcaceae, Enterobacteriaceae or Bacteroidaceae. Furthermore, a PCoA was conducted using the highest-level microbial ASV annotations prior to running *DECONTAM* for generating Bray-Curtis dissimilarity matrices of unspiked cyst fluid samples and controls. The main cluster encompassed all negative controls (illustrated in red, figure 2A). Most cyst fluid samples (depicted in black) were within the 95% CI of the controls, indicating the absence of a distinct microbiome signature when compared with the DNA background. However, a separate subcluster emerged representing an outlier group with a unique microbial signal (figure 2A). To investigate the ecological parameters and clinical features influencing the distinct clustering of this subgroup, a linear mixed-effect model was used with the first PCoA coordinate as the response variable (figure 2B). This analysis revealed that the distinct microbial signal in cyst fluid samples of the outlier cluster was associated with recent endoscopic diagnostic procedures including fine needle aspiration or was linked to jaundice cases, and hence endoscopic retrograde cholangiopancreatography (ERCP) with stent placement (figure 2B).

### Dominance of individual bacteria in a subset of patients

We then separated the data from the previous sequencing depth analysis (figure 1B) based on the two PCoA cluster assignments. It was observed that samples in the subcluster exhibited significantly higher read counts post-*DECONTAM* compared with most cyst fluid samples in the primary cluster (figure 2C). The microbial signals in the outlier samples, although not exhibiting different Shannon or Chao1 diversity relative to the controls, demonstrated a significantly elevated Gini coefficient (figures 2D and 3A), suggesting the lack of a distinct or diverse microbial community even in the outliers, but hinting the presence of a dominating taxon, or potential pathogen in this group, respectively. Indeed, while most ASVs exhibited a low Gini coefficient, suggesting a relatively equitable distribution of ASV abundance across samples and controls (figure 3B), the most pronounced Gini coefficient was associated with either *Klebsiella* sp, *Enterococcus* sp and *Streptococcus* sp, and to a lesser extent with *Lactobacillales* sp, *Fusobacterium* and *Eikenella* sp (figure 3C). One representative cyst fluid sample from the subcluster with *Enterococcus* dominance and a control sample from the primary cluster were selected for further investigation of *Enterococcus* growth and viability. No bacterial growth was detected in the control sample, while the outlier exhibited growth of *Enterococcus faecium*—as confirmed by the full length 16S rRNA gene sequencing—with more than 100 000 CFU/mL, suggesting bacterial infection. Unfortunately, we were unable to culture additional pathogens due to insufficient leftover samples. The limited microbial biomass in cyst fluid required the use of all available material for repeated sequencing, to ensure reproducible results and assess cross-contamination effects.

### Spike-in controls do not reveal traces of microbial communities

Most cyst fluid samples and controls exhibited microbial DNA quantities below or at the detection threshold of the Qubit



dsDNA HS quantification assay. Also, the Illumina sequencing platform requires a minimum amount of input DNA for a successful run, thus, our initial methodology did not rule out the presence of an ultra-low biomass microbial community, undetectable by current techniques. To address this, we introduced a low biomass mock community of known concentration, consisting of *Truepera radiovictrix*, *Imtechella halotolerans* and *Allobacillus halotolerans*, into a subset of previously unseen cyst fluid samples (n=20) and negative controls (n=5). This allowed for normalisation of sequenced microbial 16S rRNA genes based on the number of reads obtained per mock bacterium, facilitating the inference of absolute microbial abundance estimates. In this context, no statistical difference in absolute abundance between controls and cyst fluid samples was observed with only two outliers exhibiting higher absolute abundances of non-spike-in microbial content compared with controls (figure 4A). Moreover, a subsampling approach implemented to account for the unbalanced group sizes and employed to statistically analyse the alpha diversity of a potential microbiome indicated no significant differences in Chao1 diversity, Gini coefficient or Shannon diversity between controls and cyst fluid (figure 4B). Once more, except for two outlier samples, cyst fluid clustered within the CIs of the controls during PCoA of Bray-Curtis dissimilarity matrices (figure 4C), indicating the absence of a unique microbiome in our cohort. Notably, the two outliers were neither characterised by higher Shannon or Chao1 diversity estimates or more ASVs. They rather showed significantly higher bacterial load and increased dominance (Gini coefficient) of ASVs annotated as Enterobacteriaceae, as well as a higher number of ASVs passing the contamination testing (figure 4D).

However, the spike-in approach helped to identify additional, ultra-low abundant ASVs (<0.09 ng) in a subset of samples, including *Acinetobacter baumannii*, *Chitinophagales sp*, *Fusobacteriaceae sp*, *Parabacteroides goldsteinii*, *Phocaeicola sp* and *Prevotella intermedia* (online supplemental figure 1), but none of them demonstrated growth in culture (figure 5A), potentially indicating a past endoscopic procedure-based contamination with bacteria, which were unable to sustain themselves in the pancreatic niche, likely perished, leaving behind residual traces detectable in a subset of the samples. Only one sample exhibited growth of *Citrobacter koseri*, taxonomically confirmed with full-length 16S rRNA gene sequencing, with more than 200 000 CFU/mL, suggesting bacterial infection. Overall, ASV prevalence was low in our cohort (<10%, figure 5B, online supplemental file 4). Thus, in typical microbiome bioinformatics pipelines that retain only those hits with a cohort prevalence exceeding 10%, all our ASVs would have been discarded as irrelevant in discovering disease-specific microbiome signatures.<sup>22 23</sup>

## DISCUSSION

In this study, we analysed 16S rRNA amplicon sequencing data of 338 cyst fluid samples obtained from 145 patients with IPMN harbouring different degrees of dysplasia, from low-grade to invasive carcinoma, 45 patients with other cystic lesions of the pancreas, including benign cystic neoplasm, and 16 patients who had non-IPMN pancreatic ductal adenocarcinoma with cystic degeneration. The objective was to explore the potential role of a cyst fluid microbiome in the development and progression of pathophysiological processes in the pancreas. Our study was initiated in response to emerging studies suggesting the pancreas and cyst fluid, traditionally considered sterile, may harbour a distinct microbiome. These studies have equated the detection of diverse microbial DNA in pancreatic cyst fluid with the stable

colonisation of a unique microbial community in the pancreatic niche, including taxa associated with the upper or lower airways (*Neisseria* spp, *Moraxella* spp, *Haemophilus* spp, *Corynebacterium* spp and *Staphylococcus* spp) and the skin (*Cutibacterium acnes*), along with laboratory contaminants. However, the likelihood of such a diverse collection of viable species migrating from the oral and skin regions to the pancreas is ecologically questionable. To our knowledge, available studies neither presented nor analysed data from negative controls of the sampling environment, nor did they perform repeated processing and sequencing of the same sample. Such measures are crucial in low-biomass microbiome studies to rule out the possibility of contamination.<sup>6-9 12</sup>

To address these concerns, we collected cyst fluid via sterile surgical methods directly from the cystic lesions, bypassing the GI tract to minimise the risk of contamination. We implemented and analysed negative controls sourced from the same surgical environment and equipment. Plus, a subset of samples was reprocessed and resequenced up to three times with the same wet-lab procedure to account for kit-, environmental- and cross-contamination. By doing so, we failed to uncover a unique pancreatic cyst microbiome in any of the fluid samples deviating from the DNA background noise; a finding supporting the hypothesis of a predominantly sterile pancreatic environment. While our positive standards designed to emulate a diverse microbial community reached the expected sequencing depth and quality, the majority of cyst fluid samples yielded only minor amounts of microbial DNA. Furthermore, most of the DNA data were attributable to contamination arising from the sampling and processing procedures, including the kitome and splashome, thus creating a remarkably low signal-to-noise ratio. After discarding sequences that were flagged as potential contaminants by the DECONTAM software—identified based on prevalence or frequency in the negative controls—we found the microbial DNA quantity by definition<sup>13</sup> insufficient to indicate the presence of a complex microbial community within IPMNs or other pancreatic cystic lesions. Given that prior research studies in the field did not assess potential contamination on account of missing negative controls, we proceeded the analyses without removing such DECONTAM-predicted contaminants. Instead, we incorporated the negative controls into the analytical workflow for a continuous evaluation of the background noise. Even then, we found no support for the existence of a unique pancreatic cyst microbiome in terms of microbial community compositions and alpha diversity. No notable correlation between cyst attributes, disease progression and microbial markers was observed. As a final validation, we augmented a small group of cyst fluid samples (n=20) and control samples (n=5) with bacterial cells of low biomass, which are not typically part of the human microbiome. This approach was undertaken to estimate absolute microbial abundance of DNA traces and to investigate the existence of microbial communities present in quantities that fall below the detection capabilities of the Qubit dsDNA HS quantification assay or are otherwise insufficient for effective sequencing. However, both alpha diversity estimates and microbial community composition continued to be comparable between samples and controls.

Major limitations in our study design include the sole sampling of cyst fluid so that we could not address the possible presence of tissue-associated microbes. Moreover, we performed partial 16S rRNA amplicon sequencing targeting only the V4 region of the bacterial marker gene, which compromised the taxonomic resolution of our study, thereby precluding the possibility of tracking DNA viruses, the presence of a bacteriophage

community, cross-kingdom or species-and-strain-level analysis and may have led to an underestimation of the samples' genetic diversity. However, most cyst fluid samples exhibited DNA trace concentrations from the outset, falling either below or at the detection threshold set by Qubit quantification, resulting in a low yield of sample-specific DNA sequences. We even attempted to bypass the detection thresholds by using a method where we added small amounts of DNA to initiate sequencing runs for samples that could otherwise not be sequenced due to insufficient microbial DNA material. Our goal was to randomly detect trace signals from a potentially very low-density community in the sequenced output. However, no unique microbial community structure became apparent. Given these considerations, the likelihood that our sequencing approach overlooked a unique, stable and diverse microbial community within cyst fluid appears minimal.

Even though no niche-specific, unique pancreatic microbiome was detected, we did unravel a bacterial signature in a small subset of patients—with and without spike-in strategy—, characterised by an increased number of reads mapping, for example, to either *Enterococcus* sp, *Klebsiella* sp, *Streptococcus* sp or *Fusobacterium* sp, which deviated from the background signal. The discovery of potential bacterial infection within pancreatic pathologies is not unexpected. Previous research studies detected *Enterococcus* spp in pancreatic juice of patients suffering from cancers within the pancreaticobiliary-duodenal region.<sup>24</sup> In this context, higher serum antibody levels against *Enterococcus faecalis* capsular polysaccharide have been measured in patients with chronic pancreatitis and pancreatic cancer.<sup>24</sup> Moreover, in patients who have undergone preoperative ERCP, increased bacterial colonisation of the pancreatic duct was observed with GI tract-associated bacteria, including *Enterococcus* and *Streptococcus* spp.<sup>25</sup> Another study reassessed 1743 ERCP operations and potential postoperative infections and highlighted *E. faecium*, followed by *Escherichia coli* as the most prevalent taxon causing subsequent bacteremia in a subset of patients.<sup>26</sup> Another study failed to isolate viable bacteria from 22 of 29 patients with cystic precursors of pancreatic cancer (~76%). From the remaining minor subset of patients, they recovered *Enterococcus*, *Enterobacter*, *Klebsiella* or *Citrobacter* and linked the occurrence of bacterial growth to age, elevated C-reactive protein (CRP) and a history of invasive endoscopy.<sup>27</sup>

Overall, the finding of a non-existent pancreatic cyst microbiome in our study cohort holds clinical relevance as it emphasises the need to explore alternative pathways to understand the role of inflammation in early cyst and tumour development in most patients. Investigating the alternative pathways should be the primary objective in forthcoming research studies. Additionally, we recommend the incorporation of negative controls in future microbiome-centric studies, particularly those involving low-biomass habitats. The rationale behind this recommendation lies in the fact that varying levels of DNA contamination within sampling and laboratory environments as well as cross-contamination among samples yields distinct yet misleading positive microbiome signals within biological samples.

X Darío Missael Rocha Castellanos @dariochamrd

**Acknowledgements** We are grateful for the patients who participated in this study; their contributions and consent were invaluable in advancing our understanding of the subject. We thank Dr med. Eva Winter (Charité, Berlin) for sharing her medical expertise and recommendations. This work has been supported by the German Research Foundation (DFG, 530694780) in the form of a Walter-Benjamin fellowship granted to M-MP.

**Contributors** CF-dC and RJX conceived and supervised the study. CF-dC, DMRC, GP, KR and M-MP developed the experimental design. CA, AL, CF-dC and DMRC

recruited patients, obtained patient consent and managed the sample collection, transfer and storage. GP, KR and AD performed the wet-lab experiments. M-MP performed the data analysis. M-MP, GP and CF-dC wrote the first draft of the manuscript. RJX is the guarantor for this paper. All authors read and approved the manuscript.

**Funding** The study has been funded by the Center for the Study of Inflammatory Bowel Disease at Massachusetts General Hospital (CSIBD #43351) and the IPMN Global Foundation.

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study follows all relevant ethical regulations. Patient samples were obtained under IRB approval (protocol number is 2003POD1289). Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Marie-Madlen Pust <http://orcid.org/0000-0001-6870-2488>

Darío Missael Rocha Castellanos <http://orcid.org/0000-0002-6227-9161>

Kara Rzasa <http://orcid.org/0000-0002-5753-7517>

#### REFERENCES

- Sadot E, Basturk O, Klimstra DS, et al. Tumor-associated neutrophils and malignant progression in Intraductal papillary mucinous neoplasms: an opportunity for identification of high-risk disease. *Ann Surg* 2015;262:1102–7.
- Li S, Fuhler GM, Bn N, et al. Pancreatic cyst fluid harbors a unique microbiome. *Microbiome* 2017;5:147.
- Gaiser RA, Halimi A, Alkharaan H, et al. Enrichment of oral microbiota in early cystic precursors to invasive pancreatic cancer. *Gut* 2019;68:2186–94.
- Konopka A. What is microbial community ecology. *ISME J* 2009;3:1223–30.
- Berg G, Rybakova D, Fischer D, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome* 2020;8:103.
- Salter SJ, Cox MJ, Turek EM, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;12:87.
- Eisenhofer R, Minich JJ, Marotz C, et al. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends Microbiol* 2019;27:105–17.
- Olomu IN, Pena-Cortes LC, Long RA, et al. Elimination of 'Kitome' and 'Splashome' contamination results in lack of detection of a unique placental microbiome. *BMC Microbiol* 2020;20:157.
- Weyrich LS, Farrer AG, Eisenhofer R, et al. Laboratory contamination over time during low-biomass sample analysis. *Mol Ecol Resour* 2019;19:982–96.
- Kennedy KM, de Goffau MC, Perez-Muñoz ME, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature* 2023;613:639–49.
- de Goffau MC, Lager S, Sovio U, et al. Human placenta has no microbiome but can contain potential pathogens. *Nature* 2019;572:329–34.
- Blaser MJ, Devkota S, McCoy KD, et al. Lessons learned from the prenatal microbiome controversy. *Microbiome* 2021;9:8.
- Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1:4516–22.
- Caporaso JG, Lauber CL, Walters WA, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and Miseq platforms. *ISME J* 2012;6:1621–4.
- Lane DJ. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, eds. *Nucleic Acid Techniques in Bacterial Systematics*. New York: John Wiley and Sons, 1991: 115–75.

- 16 Callahan BJ, McMurdie PJ, Rosen MJ, *et al.* DADA2: high resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–3.
- 17 Parks DH, Chuvochina M, Chaumeil P-A, *et al.* A complete domain-to-species taxonomy for bacteria and archaea. *Nat Biotechnol* 2020;38:1079–86.
- 18 Quast C, Pruesse E, Yilmaz P, *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.
- 19 Sayers EW, Beck J, Bolton EE, *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2021;49:D10–7.
- 20 Davis NM, Proctor DM, Holmes SP, *et al.* Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;6:226.
- 21 Pust M-M, Castellanos DMR, Rzasa K, *et al.* Gitlab repository: IPMN microbiome. GitLab; 2024. Available: <https://gitlab.com/mpust/IPMN-microbiome> [Accessed 05 Feb 2024].
- 22 Nearing JT, Douglas GM, Hayes MG, *et al.* Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun* 2022;13:342.
- 23 Yang L, Chen J. A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. *Microbiome* 2022;10:130.
- 24 Maekawa T, Fukaya R, Takamatsu S, *et al.* Possible involvement of enterococcus infection in the pathogenesis of chronic pancreatitis and cancer. *Biochem Biophys Res Commun* 2018;506:962–9.
- 25 Demir E, Yilmaz BS, Jäger C, *et al.* Bacterial content of the human pancreatic duct: an observational study. *Heliyon* 2022;8:e11771.
- 26 Du M, Suo J, Liu B, *et al.* Post-ERCP infection and its epidemiological and clinical characteristics in a large Chinese tertiary hospital: a 4-year surveillance study. *Antimicrob Resist Infect Control* 2017;6:131.
- 27 Halimi A, Gabarrini G, Sobkowiak MJ, *et al.* Isolation of pancreatic microbiota from cystic precursors of pancreatic cancer with intracellular growth and DNA damaging properties. *Gut Microbes* 2021;13:1983101.