Guardians of the gut: harnessing bioinformatics to study the gut microbiome and faecal microbiota transplantation in intestinal disorders

Nooij, S.

Guardians of the gut

Harnessing bioinformatics to study the gut microbiome and faecal microbiota transplantation in intestinal disorders

Sam Nooij

Colophon

Guardians of the gut: Harnessing bioinformatics to study the gut microbiome and faecal microbiota transplantation in intestinal disorders

Sam Nooij
PhD thesis, Leiden University Medical Center, 2025

Illustrations: Sam Nooij
Layout: Wiebke Keck, persoonlijkproefschrift.nl
Printing: Ridderprint
ISBN: 978-94-6522-470-1

Guardians of the gut

Harnessing bioinformatics to study the gut microbiome and faecal microbiota transplantation in intestinal disorders

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op vrijdag 10 oktober 2025
klokke 10.00 uur

door

Sam Nooij
geboren te Amsterdam
in 1990

Promotor

Prof. Dr. A.C.M. Kroes

Copromotores

Dr. E.M. Terveer

Dr. Q.R. Ducarmon

Leden promotiecommissie

Prof. Dr. E.J. Kuijper

Dr. G.F. Zeller

Dr. A.C. Schürch (Universitair Medisch Centrum Utrecht)

Prof. Dr. W. van Schaik (University of Birmingham, Verenigd Koninkrijk)

Printing of this thesis was financially supported by the Leiden University Library (UBL), and the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM).

“Dubium sapientiae initium.”

René Descartes (1596-1650)

Vertaling:
“Twijfel is het begin van wijsheid.”

Table of contents


サム

Chapter 1

General introduction

The human microbiome

The microbiome has enjoyed great interest in recent years. The term 'microbiome' describes a community of microorganisms living in a specific environment. This comprises bacteria, viruses, fungi, and other microscopic eukaryotes, but bacteria are often the primary focus. Microbiomes can exist anywhere, from environments like soil to the bodies of animals. Of all microbiomes, the human microbiome is perhaps the most captivating and best known. The human body harbours roughly as many bacterial cells as human cells[1], and over a hundred times more microbial genes than are encoded in the human genome[2]. Bacteria are spread across multiple body sites that each offer their own environmental conditions to which microbial communities have adapted (Figure 1). For example, researchers have identified different microbiomes on the skin, in the airways or respiratory tract, and the gut or gastrointestinal tract[3,4]. The gastrointestinal tract encompasses a large surface area covered with a vast diversity of microorganisms[5], and has been the subject of extensive study[6]. In fact, among human microbiome studies, the gut microbiome is the best-studied body site.

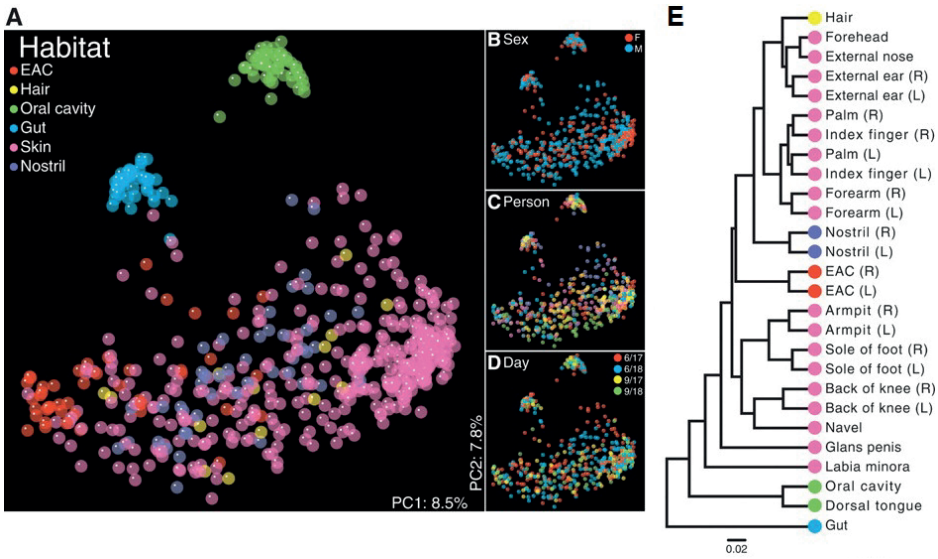


Figure 1. The human microbiome differs per body site. Clustering of microbiome samples based on taxonomic composition reveals that body site or habitat explains the variation most clearly. A) Microbial communities coloured by body site. B-D) the same communities coloured by sex, person or sampling day. E) Hierarchical clustering dendrogram shows that body sites that are closer to one another are more similar in terms of microbial composition. EAC, external auditory canal; L, left; R, right. Figure adapted from reference[4]. Reprinted with permission from AAAS.

Ways of studying the microbiome

The discovery of microorganisms, microscopic organisms invisible to the naked eye, started with the refined microscopes that Antoni van Leeuwenhoek constructed around 1670[7]. It is worth mentioning that Jan Swammerdam and Robert Hooke were also building microscopes at that time, but their magnification was much less than Van Leeuwenhoek's[7]. Up to the present day, microscopes remain an important instrument to study microorganisms and indeed numerous kinds and species, from cyanobacteria to selenomonads, have been identified with microscopes[7,8]. In modern history, the development of DNA sequencing techniques has revolutionised the way we can detect and characterise any type of organism. DNA sequencing was pioneered in the 1960s and -70s by Wu, Gilbert and Maxam[9], after which new techniques were developed and combined, giving rise to the method known as Sanger sequencing (Figure 2)[10,11]. This discovery was deemed so important that Frederick Sanger and Walter Gilbert received a Nobel Prize in Chemistry in 1980. With Sanger sequencing, one could read DNA fragments as a sequence of four nucleotides: A, C, G and T. By collecting multiple reads it was then possible to elucidate increasingly complex DNA molecules. The first completely resolved genome was of bacteriophage lambda (48.5 kilobasepairs (kbp))[12]. The technique was further refined and optimised[13], facilitating increasingly large genome sequencing projects spanning from *Haemophilus influenza* (2 Mbp, 1995)[14], to *Saccharomyces cerevisiae* (12 Mbp, 1996)[15], *Caenorhabditis elegans* (100 Mbp, 1998)[16], and finally, in 2001 the first draft sequence of the human genome was published (*Homo sapiens*, 3.3 Gbp)[17,18]. Concurrently, DNA sequencing methods were further improved and a new generation was created. These so-called next-generation sequencing (NGS) or second generation methods allow for rapid high-throughput sequencing of many DNA molecules, thereby enabling an entirely new field of science. Now, there is a third generation, enabling real-time single-molecule sequencing (Figure 2). Not only have these techniques facilitated the study of complete human genomes (genomics), but it has also enabled researchers to study DNA from multiple organisms in an environment (metagenomics)[19]. This in turn led to the launch of the Human Microbiome Project in 2007[3,20]. The term 'microbiome' is used here to refer to all the genes identified using NGS, although it has been used to mean different things[21]. The microbial species from which the genes derive are commonly referred to as 'microbiota'. NGS methods now yield enormous amounts of data, with a single experiment often generating millions of DNA reads. To process all these data, a new type of biologist was needed: the computational biologist or bioinformatician.

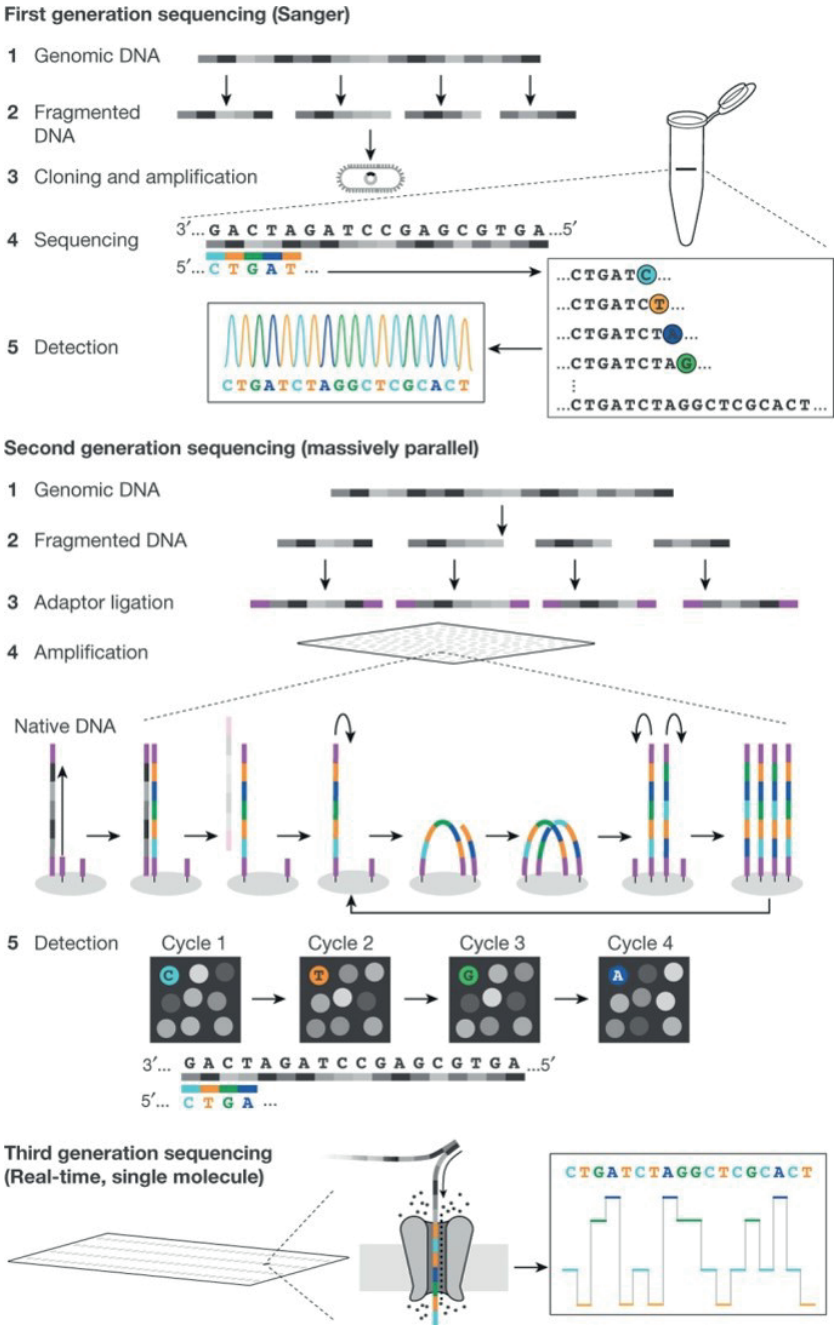


Figure 2. DNA sequencing technologies. Developments of DNA sequencing techniques are listed and illustrated in chronological order from top to bottom. The second generation is generally referred to as next-generation sequencing or NGS. Figure from reference[9].

The computational biologist or bioinformatician

A computational biologist, or bioinformatician, combines skills and knowledge from both biology and computer science[22]. There is some confusion and controversy concerning the terms 'computational biologist' and 'bioinformatician'. However, since there is considerable overlap between the tasks of each I consider them equivalent for the purpose of this text. Bioinformatics as a field has been around since the 1960s, before DNA could be sequenced[23]. Starting out with protein sequence analysis, bioinformatics has expanded to include DNA and many other types of biological data, while benefitting from rapid improvements in computer technology. The advancements in high-throughput biological methods have given rise to biological 'big data'. Due to the high-throughput at which biological data are now generated, the results can no longer be processed and interpreted manually and need increasingly powerful computers and automated workflows. This is the work of a bioinformatician. A bioinformatician needs to understand the biology behind the experiment to formulate the relevant research questions and have a working knowledge of computer science, statistics and/or mathematics to convert those questions into functional computer algorithms, pipelines or tools[24]. The output from computational analyses (often referred to as 'dry lab' work) is typically hypothesis-generating, for instance correlating a specific bacterium or gene to a disease. Consequently, the hypothesis needs to be tested and validated in the so-called wet lab. Therefore, bioinformatics experiments are frequently embedded within larger projects, bringing together collaborators from biomedical backgrounds to computational experts. A good example is the Human Microbiome Project[25], which has run for over a decade and spawned numerous subsidiary studies as well as a large second phase: the Integrative Human Microbiome Project[26]. The Human Microbiome Project was also the largest microbiome project in terms of data generated at the time (Figure 3), underscoring the growing need for data management and computational tools in biology. In a sense, bioinformaticians act as an interpreter in a multidisciplinary team of life science specialists translating between data science and biology. Thus, they must be flexible and able to communicate with people of diverse backgrounds.

Figure 3. Timeline of microbiome projects and size of generated data. Each circle represents a microbiome project and their area and y-coordinate indicate data size in basepairs. The x-coordinate corresponds to the year the study was published and colours separate human (red), animal (black) and environment-associated (green) projects. Figure from reference[25].

Human gut microbiota in health and disease

The advent of second-generation DNA sequencing (NGS) paved the way for a whole new field of microbiology. Researchers were no longer limited to study culturable species, but whole communities were now within reach. Among possible focus areas, the human microbiota has intrigued us for a long time. Early microbiome studies established that microorganisms played key roles in infections, leading to hypotheses as to their involvement in other diseases. The use of metagenomics enabled researchers to compare microbial communities between healthy people and people with various diseases, such as ulcerative colitis and Crohn's disease (UC and CD, together known as inflammatory bowel diseases or IBD), colorectal cancer, obesity, and Parkinson's disease to identify differences in microbiota associated with these conditions[2,27-33]. An important caveat here is that correlation does not imply causation[33]. The presence of certain bacteria in a patient may indicate the cause of disease, or the disease itself could create an environment in which the bacteria can thrive. In this second scenario, the bacteria merely benefit accidentally from conditions brought about by the disease. This was exemplified by a study on type 2 diabetes mellitus in which the researchers found a correlation with the microbiome[34]. Another group later attributed this change to the drug used[35]. Nonetheless, these disease association studies have elucidated the involvement of the microbiota in a number of diseases in which no role for microorganisms was suspected. This has led to the generation of new hypotheses and

further studies into possible mechanisms by which bacteria contribute or exacerbate disease. For instance, discovering the association of *Fusobacterium nucleatum* with colorectal cancer motivated researcher to quickly identify molecular mechanisms by which the bacterium is involved in tumorigenesis[29,36,37].

Sometimes these studies gave conflicting results. Certain bacteria may be disease-associated in one study and associated with good health in another study. These discrepancies may arise from differences in study setup, study population (e.g. geographic location, diet, or other potential confounders[38]), and low statistical power. Another possible explanation can be found in the biology of the bacterium in question. In metagenomics studies, bacteria are often classified to a certain taxonomic level, like genus or species. These taxonomic groups, however, encompass a spectrum of different strains of bacteria. The genomes within a bacterial species are 95% identical[39]. Within the remaining 5% there is ample opportunity for differences that cause fundamental shifts in lifestyle, making one strain pathogenic and the other an innocent symbiont or commensal. It is thought that the exact genetic composition, including the presence or absence of certain genes, and interactions with the environment – in this case its host: the human body – determines the bacterium's functions and the possible implications in disease[2,33]. Conflicting results can be clarified through meta-analyses, which enable more robust statistical evaluation of disease-microbiome associations. This approach is also exemplified in **chapter 7**.

Disturbances to microbiota as marker for disease

For several diseases it is well-established that the microbiota play a fundamental role in disease aetiology. Currently the best example is an infection by the gut pathogen *Clostridioides difficile*, which paradoxically often occurs after antibiotic treatment[40]. The mechanism that underlies this process is that the antibiotics deplete part of the resident microbiota, creating a niche – freeing up space as it were – in which survivors may thrive. *C. difficile* produces spores, which are very resilient and generally unaffected by antibiotics. After antibiotic treatment, these spores germinate and quickly occupy the cleared niche. It then produces toxins which both harm the person in which *C. difficile* resides (causing inflammation), and suppresses other bacteria in the gut[40]. This alteration of the gut microbiota composition is termed 'dysbiosis', describing a deviation from the normal microbial balance and is usually related to disease[41-43]. Even though there is no standard way to describe dysbiosis, several parameters have been used to indicate a deviation in gut microbiota. Consequently, the overall microbiota composition has become an indicator of health. This thesis deals with two microbiota-associated diseases: recurrent *C. difficile* infection (CDI) in **chapters 2 and 3** and IBD in **chapters 4, 5, 6 and 7**.

Inflammatory bowel disease is a chronic condition that is characterised by episodes of gastrointestinal inflammation[44]. IBD encompasses Crohn's disease (CD) and ulcerative colitis (UC). CD typically affects the distal gastrointestinal tract, particularly from the end of the small intestine (ileum) to the colon, but it can involve any part of the intestine in a discontinuous pattern[44]. UC typically affects the distal colon and often includes the rectum, but inflammation can extend proximally to involve larger portions or the entire colon in an uninterrupted pattern[44]. IBD often manifests in early adult life and has a genetic component[45]. It is thought to result from an aberrant inflammatory immune response to commensal gut microbes[45]. Hence, the gut microbiota plays a critical role in the pathogenesis of IBD and it is hypothesised that microbiota modulation could alleviate symptoms in affected patients.

Microbiota therapy: faecal microbiota transplantation (to prevent CDI relapse)

After observing an alteration in gut microbiota, the next step in medical science is to understand how the microbiome can be modulated in such a way that it contributes to disease amelioration or cure. Microbiota-based therapies were already used long before the discovery of microorganisms. Using faecal material to treat diarrhoeal diseases was first described in China in the 4th century, and has supposedly been used throughout Eurasia and North Africa up to the 20th century[46], although the use of camel faeces in North Africa is contested[47]. In recent history, the method has been reintroduced and termed faecal microbiota transplantation (FMT). Initially pioneered in the Netherlands in 2013, this treatment proved highly effective against recurrent *C. difficile* infection and has quickly become the new therapeutic approach[48]. Following this success, FMT has also been attempted with different diseases, with varying results[46,49].

The mechanism underlying FMT is thought to be a restoration of a healthy gut microbiota. This is done by taking the faeces of thoroughly screened healthy volunteer donors, making a suspension from the faeces and transplanting this into the gastrointestinal tract of the patient. This procedure is preceded by antibiotic treatment to eradicate *C. difficile*, but antibiotic pretreatment has not always been used with FMT for other diseases. A new generation of microbiome-modulating therapies are under development, but none have yet reached the stage where they can replace FMT[50-54]. To reach this goal, we need to advance our understanding of human-microbiome interactions and identify specific bacterial strains with therapeutic potential.

Research questions and thesis outline

This thesis describes computational studies of the human gut microbiota with a focus on FMT. The Leiden University Medical Center (LUMC) hosts a dedicated faeces bank, the Netherlands Donor Feces Bank (NDFB), that supplies ready-to-use faecal suspensions for FMT to treating physicians throughout the country. The NDFB is part of the Medical Microbiology and Infection Prevention subdepartment in the Center for Infectious Diseases (LU-CID), creating a multidisciplinary environment to study the microbiota in relation to FMT. Currently, FMT is routinely used to treat multiple recurrent CDI. Besides this, the NDFB has initiated clinical trials for other indications including UC and collaborates with various partners in scientific studies aimed at elucidating the mechanisms underlying FMT. These studies focus primarily on the gut microbiota, employing metagenomic sequencing of faecal samples from healthy donors and recipients before and after FMT. This approach provides a comprehensive view of the microbial composition and its functional potential, enabling exploration of the microbiota and addressing diverse research questions. The main research question of this thesis is: how does FMT affect the recipients' microbiota? The secondary question is: what makes one species variant commensal and the other disease-associated?

Part 1: Concomitant microbiota impacts after faecal microbiota transplantation for recurrent *Clostridioides difficile* infections

Chapter 2 describes a study on putatively procarcinogenic *Escherichia coli* strains within a cohort of recurrent CDI patients and their respective donors. Previous research had established that this *E. coli* strain produces a toxin, colibactin, capable of inducing DNA damage in the host and potentially contributing to the development of colorectal cancer. Furthermore, it was known that the species *E. coli*, and even this specific procarcinogenic strain, is a common inhabitant of the human gut. We screened our metagenomic data for these specific genotoxin genes to evaluate the prevalence and abundance among FMT recipients and their healthy donors. The questions we asked here are: is colibactin-producing *E. coli* present in recurrent CDI patients and healthy faeces donors, and how does FMT affect this *E. coli* variant?

In **chapter 3**, we studied the effect of FMT on antibiotic resistance genes within the gut microbiota. Antibiotic resistance, particularly in hospital settings, is viewed as one of the most significant public health concerns as it jeopardises the ability to treat bacterial infections[55,56]. This not only increases the morbidity and mortality of bacterial infections, but some surgical procedures may also not be feasible anymore. Alternatives to antibiotics are being explored, and FMT has been proposed to control the spread of antibiotic resistance, and possibly eradicate resistant bacteria. In this chapter we

increased the cohort size by including more patient samples compared to **chapter 2**. We also collected long-term follow-up samples after FMT and used both traditional bacterial culture techniques as well as whole-genome and metagenomic sequencing to study the effect of FMT on antibiotic resistance genes. We identified multidrug resistant (MDR) bacteria from cultures. This was compared to metagenomic sequencing data to estimate the relative abundance of these MDR bacteria, while also screening the metagenome for antibiotic resistance genes. In doing so, we provide answers to the questions: 1) does FMT eradicate MDR bacteria? 2) How does FMT affect antibiotic resistance genes in the gut microbiota?

Part 2: Microbiota alterations after faecal microbiota transplantation for ulcerative colitis

In **chapters 4, 5 and 6** we shift our focus to ulcerative colitis (UC, a type of IBD). It has been found that microbial components contribute to this multifactorial disease in addition to disbalances in the immune system and lifestyle factors such as diet. By altering the microbiota it may be possible to induce remission and temporarily control the disease. The NDFB conducted a small randomised clinical trial with two rationally selected donors to assess the colonisation of donor species in patients while also evaluating safety and feasibility of FMT. **Chapter 4** describes the FECBUD clinical trial, and the colonisation of donor-derived bacteria after FMT, known as engraftment, following pretreatment with the anti-inflammatory drug budesonide. The main question here is: does pretreatment with budesonide increase engraftment after FMT? **Chapter 5** extends the FECBUD analysis, applying ecological models to understand bacterial community processes that correlate with clinical remission of IBD symptoms. This chapter addresses the question: what microbial changes in UC patients after FMT correlate with clinical remission or treatment success? **Chapter 6** present a secondary continuation of the analysis in which principles of ecological population dynamics combined with statistical modelling are applied to the FMT metagenomics data. Here, we sought to answer 1) how do dynamics in donor and patient microbial species' abundance correlate with FMT success?, and 2) how can we recognise treatment success early using microbial ecology parameters?

Part 3: Global distribution and genome biology of gut bacterium *Ruminococcus gnavus*

The final part, **chapter 7**, zooms in on one species of bacteria implicated in IBD in several studies: *Ruminococcus gnavus*. This bacterium has been reported to be a common gut bacterium, present in up to 90% of healthy people at low relative abundances. During flares of Crohn's disease (CD), *R. gnavus* has been seen to bloom.

Moreover, several molecular mechanisms have been identified that may play a role in the interaction between *R. gnavus* and the human immune system and contribute to inflammation. We conducted a global meta-analysis to assess the prevalence and abundance of *R. gnavus* in metagenomes, and collected isolates from healthy people and Crohn's disease patients to compare their genetic make-up. Hereby, our aim was to elucidate genomic differences between isolates from Crohn's patients and healthy people, building foundational knowledge of why *R. gnavus* behaves like an innocent commensal in one person and like a pathogen in others.

During this study, we unexpectedly generated a number of whole-genome sequences of *Streptococcus* bacteria. These sequences emerged from presumed pure cultures of *R. gnavus*. Although we cannot be sure about the exact source, we decided to briefly study and summarise these genomes as they might be a valuable resource to future research. These genomes have been deposited in a public repository and are described in **chapter 7.2**.

Chapter 8 concludes the thesis with a general discussion, summarising results from all chapters and placing them in both scientific and societal context. It also proposes ideas for future research directions.

References

1. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology* **14**, e1002533 (2016). https://doi.org/10.1371/journal.pbio.1002533
2. Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nature Medicine* **24**, 392-400 (2018). https://doi.org/10.1038/nm.4517
3. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214 (2012). https://doi.org/10.1038/nature11234
4. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-1697 (2009). https://doi.org/10.1126/science.1177486
5. Marchesi, J. & Shanahan, F. The normal intestinal microbiota. *Current Opinion in Infectious Diseases* **20**, 508-513 (2007). https://doi.org/10.1097/QCO.0b013e3282a56a99
6. Gomaa, E. Z. Human gut microbiota/microbiome in health and diseases: a review. *Antonie Van Leeuwenhoek* **113**, 2019-2040 (2020). https://doi.org/10.1007/s10482-020-01474-7
7. Cocquyt, T. Positioning Van Leeuwenhoek's microscopes in 17th-century microscopic practice. *FEMS Microbiology Letters* **369** (2022). https://doi.org/10.1093/femsle/fnac031
8. Dobell, C. *Antony van Leeuwenhoek and his "little animals"*. (Harcourt, Brace and company, 1932).
9. Shendure, J. *et al.* DNA sequencing at 40: past, present and future. *Nature* **550**, 345-353 (2017). https://doi.org/10.1038/nature24286
10. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463-5467 (1977). https://doi.org/10.1073/pnas.74.12.5463
11. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* **74**, 560-564 (1977). https://doi.org/10.1073/pnas.74.2.560
12. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. Nucleotide sequence of bacteriophage lambda DNA. *Journal of Molecular Biology* **162**, 729-773 (1982). https://doi.org/10.1016/0022-2836(82)90546-0
13. Prober, J. M. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336-341 (1987). https://doi.org/10.1126/science.2443975
14. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496-512 (1995). https://doi.org/10.1126/science.7542800
15. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563-547 (1996). https://doi.org/10.1126/science.274.5287.546
16. Consortium, C. e. S. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282**, 2012-2018 (1998). https://doi.org/10.1126/science.282.5396.2012
17. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001). https://doi.org/10.1038/35057062

18. International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004). https://doi.org/10.1038/nature03001

19. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74 (2004). https://doi.org/10.1126/science.1093857

20. Team, N. I. H. H. M. P. A. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. *Microbiome* **7**, 31 (2019). https://doi.org/10.1186/s40168-019-0620-y

21. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103 (2020). https://doi.org/10.1186/s40168-020-00875-0

22. Berger, B. *et al.* What are the keys to succeeding as a computational biologist in today's research climate? *Cell Systems* **13**, 781-785 (2022). https://doi.org/10.1016/j.cels.2022.09.005

23. Gauthier, J., Vincent, A. T., Charette, S. J. & Derome, N. A brief history of bioinformatics. *Brief Bioinform* **20**, 1981-1996 (2019). https://doi.org/10.1093/bib/bby063

24. Loman, N. & Watson, M. So you want to be a computational biologist? *Nature Biotechnology* **31**, 996-998 (2013). https://doi.org/10.1038/nbt.2740

25. Gevers, D. *et al.* The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS Biology* **10**, e1001377 (2012). https://doi.org/10.1371/journal.pbio.1001377

26. Integrative, H. M. P. R. N. C. The Integrative Human Microbiome Project. *Nature* **569**, 641-648 (2019). https://doi.org/10.1038/s41586-019-1238-8

27. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* **104**, 13780-13785 (2007). https://doi.org/doi:10.1073/pnas.0706625104

28. Gevers, D. *et al.* The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe* **15**, 382-392 (2014). https://doi.org/https://doi.org/10.1016/j.chom.2014.02.005

29. Kostic, Aleksandar D. *et al.* Fusobacterium nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host & Microbe* **14**, 207-215 (2013). https://doi.org/https://doi.org/10.1016/j.chom.2013.07.007

30. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders* **30**, 350-358 (2015). https://doi.org/https://doi.org/10.1002/mds.26069

31. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* **444**, 1022-1023 (2006). https://doi.org/10.1038/4441022a

32. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006). https://doi.org/10.1038/nature05414

33. Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94-103 (2016). https://doi.org/10.1038/nature18850

34. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55-60 (2012). https://doi.org/10.1038/nature11450

35. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**, 262-266 (2015). https://doi.org/10.1038/nature15766

36. Castellarin, M. *et al.* Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Research* **22**, 299-306 (2012). https://doi.org/10.1101/gr.126516.111

37. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Research* **22**, 292-298 (2012). https://doi.org/10.1101/gr.126573.111

38. Vujkovic-Cvijin, I. *et al.* Host variables confound gut microbiota studies of human disease. *Nature* **587**, 448-454 (2020). https://doi.org/10.1038/s41586-020-2881-9

39. Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**, 81-91 (2007). https://doi.org/10.1099/ijs.0.64483-0

40. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. Clostridium difficile infection. *Nature Reviews Disease Primers* **2**, 16020 (2016). https://doi.org/10.1038/nrdp.2016.20

41. Weiss, G. A. & Hennet, T. Mechanisms and consequences of intestinal dysbiosis. *Cellular and Molecular Life Sciences* **74**, 2959-2977 (2017). https://doi.org/10.1007/s00018-017-2509-x

42. Duan, H. *et al.* Antibiotic-induced gut dysbiosis and barrier disruption and the potential protective strategies. *Critical Reviews in Food Science and Nutrition* **62**, 1427-1452 (2022). https://doi.org/10.1080/10408398.2020.1843396

43. Lee, J.-Y., Tsolis, R. M. & Bäumler, A. J. The microbiome and gut homeostasis. *Science* **377**, eabp9960 (2022). https://doi.org/doi:10.1126/science.abp9960

44. Abraham, C. & Cho, J. H. Inflammatory bowel disease. *New England Journal of Medicine* **361**, 2066-2078 (2009). https://doi.org/10.1056/NEJMra0804647

45. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307-317 (2011). https://doi.org/10.1038/nature10209

46. Stripling, J. & Rodriguez, M. Current Evidence in Delivery and Therapeutic Uses of Fecal Microbiota Transplantation in Human Diseases—Clostridium difficile Disease and Beyond. *The American Journal of the Medical Sciences* **356**, 424-432 (2018). https://doi.org/https://doi.org/10.1016/j.amjms.2018.08.010

47. Koopman, N., van Leeuwen, P., Brul, S. & Seppen, J. History of fecal transplantation; camel feces contains limited amounts of Bacillus subtilis spores and likely has no traditional role in the treatment of dysentery. *PLoS ONE* **17**, e0272607 (2022). https://doi.org/10.1371/journal.pone.0272607

48. van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent Clostridium difficile. *New England Journal of Medicine* **368**, 407-415 (2013). https://doi.org/10.1056/NEJMoa1205037

49. Porcari, S. *et al.* Key determinants of success in fecal microbiota transplantation: From microbiome to clinic. *Cell Host & Microbe* **31**, 712-733 (2023). https://doi.org/https://doi.org/10.1016/j.chom.2023.03.020

50. Feuerstadt, P. *et al.* SER-109, an Oral Microbiome Therapy for Recurrent *Clostridioides difficile* Infection. *New England Journal of Medicine* **386**, 220-229 (2022). https://doi.org/doi:10.1056/NEJMoa2106516

51. Sims, M. D. *et al.* Safety and Tolerability of SER-109 as an Investigational Microbiome Therapeutic in Adults With Recurrent Clostridioides difficile Infection: A Phase 3, Open-Label, Single-Arm Trial. *JAMA Network Open* **6**, e2255758-e2255758 (2023). https://doi.org/10.1001/jamanetworkopen.2022.55758

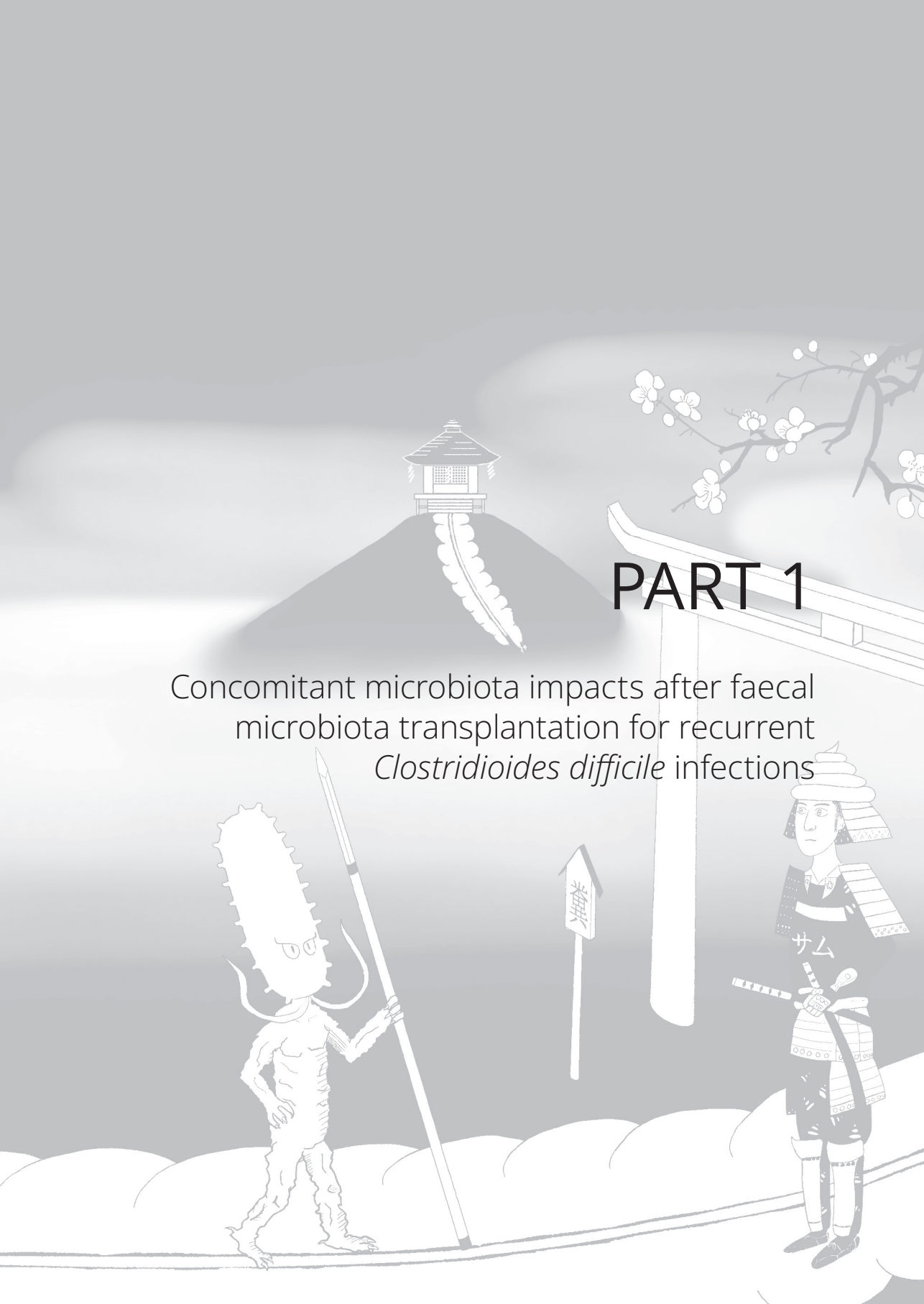
52. Dsouza, M. *et al.* Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host & Microbe* **30**, 583-598. e588 (2022). https://doi.org/10.1016/j.chom.2022.03.016

53. Khanna, S. *et al.* Efficacy and Safety of RBX2660 in PUNCH CD3, a Phase III, Randomized, Double-Blind, Placebo-Controlled Trial with a Bayesian Primary Analysis for the Prevention of Recurrent Clostridioides difficile Infection. *Drugs* **82**, 1527-1538 (2022). https://doi.org/10.1007/s40265-022-01797-x

54. Feuerstadt, P. *et al.* Retrospective Analysis of the Safety and Efficacy of Fecal Microbiota, Live-jslm (REBYOTATM) Administered Under Enforcement Discretion to Patients With Clostridioides difficile Infection. *Open Forum Infectious Diseases* **10** (2023). https://doi.org/10.1093/ofid/ofad171

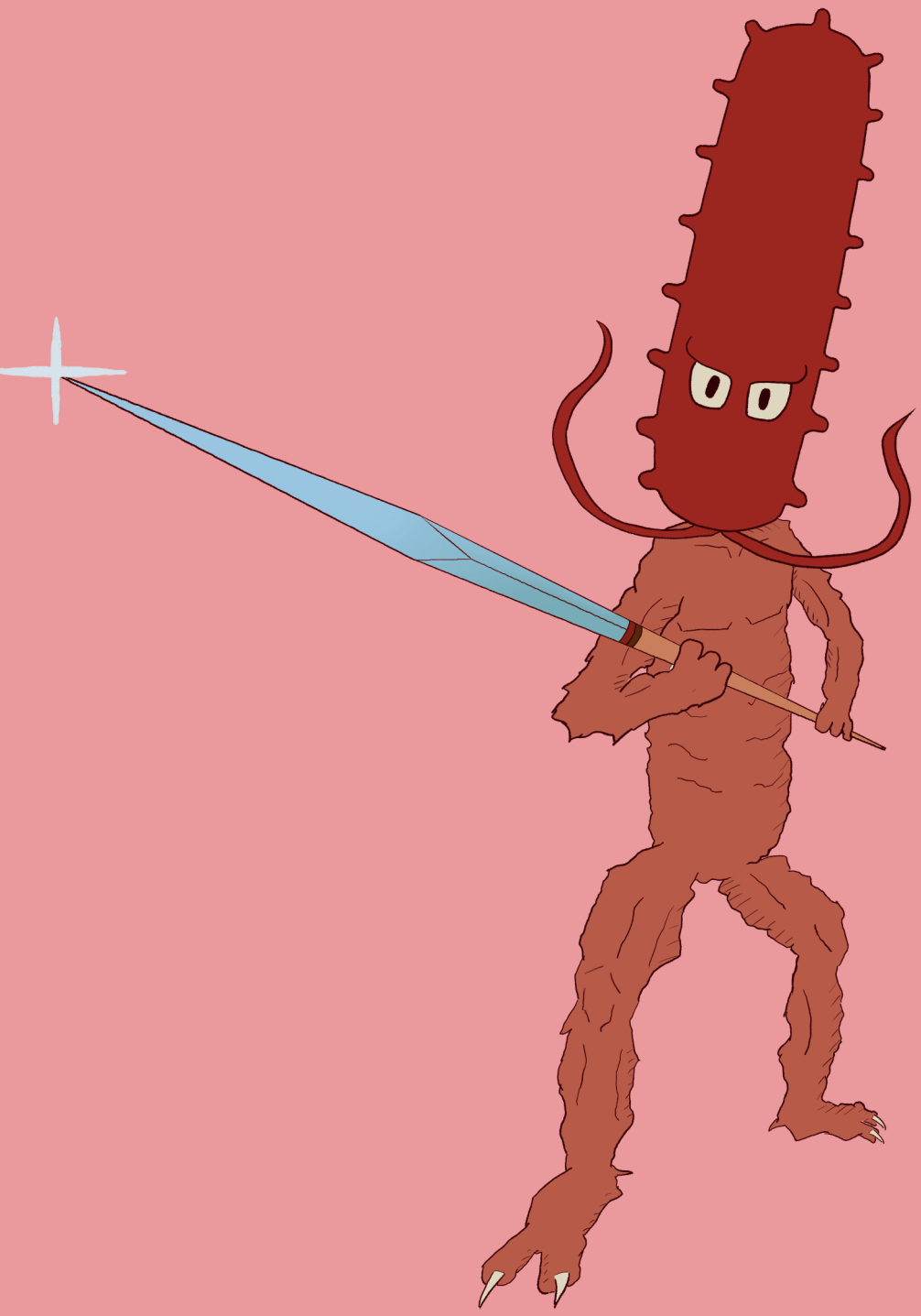
55. Antimicrobial Resistance, C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629-655 (2022). https://doi.org/10.1016/S0140-6736(21)02724-0

56. Tillotson, G. S. & Zinner, S. H. Burden of antimicrobial resistance in an era of decreasing susceptibility. *Expert Review of Anti-infective Therapy* **15**, 663-676 (2017). https://doi.org/10.1080/14787210.2017.1337508



PART 1

Concomitant microbiota impacts after faecal microbiota transplantation for recurrent *Clostridioides difficile* infections



Chapter 2.1

Faecal microbiota transplantation influences procarcinogenic *Escherichia coli* in recipient recurrent *Clostridioides difficile* patients

Sam Nooij[1,2,3], Quinten R. Ducarmon[1,3], Jeroen F.J. Laros[3,4,5,6], Romy D. Zwittink[1,3], Jason M. Norman[7], Wiep Klaas Smits[1,3], Hein W. Verspaget[8,9], Josbert J. Keller[8,10], Elisabeth M. Terveer[1,2,3], Ed J. Kuijper[1,2,3,6] on behalf of the working group of the Netherlands Donor Feces Bank

[1]Experimental Bacteriology, Department of Medical Microbiology, Leiden University Medical Center, Leiden, the Netherlands

[2]Netherlands Donor Feces Bank, Leiden, Netherlands

[3]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, the Netherlands

[4]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands

[5]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

[6]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

[7]Vedanta Biosciences Inc., Cambridge, Massachusetts, United States of America

[8]Department of Gastroenterology, Leiden University Medical Center, Leiden, the Netherlands

[9]Department of Biobanking, Leiden University Medical Center, Leiden, the Netherlands

[10]Department of Gastroenterology, Haaglanden Medical Center, The Hague, the Netherlands

Gastroenterology, Volume 161, Issue 4, 1218 - 1228.e5, October 2021

Abstract

Background and aims

Patients suffering from multiple recurrent *Clostridioides difficile* infections (rCDI) have a disturbed gut microbiota, which can be restored by faecal microbiota transplantation (FMT). Despite extensive screening, healthy faeces donors may carry bacteria in their intestinal tract that could have long-term health effects, such as potentially procarcinogenic *pks*$^{+}$ *Escherichia coli*. Here, we aim to determine whether the *pks* abundance and persistence of *pks*$^{+}$ *E. coli* is influenced by *pks*-status of the donor faeces.

Methods

In a cohort of 49 rCDI patients treated with FMT and matching donor samples – the largest cohort of its kind, to our knowledge – we retrospectively screened faecal metagenomes for *pks*$^{+}$ *E. coli* and compared the presence of *pks* in patients before and after treatment, and to their respective donors.

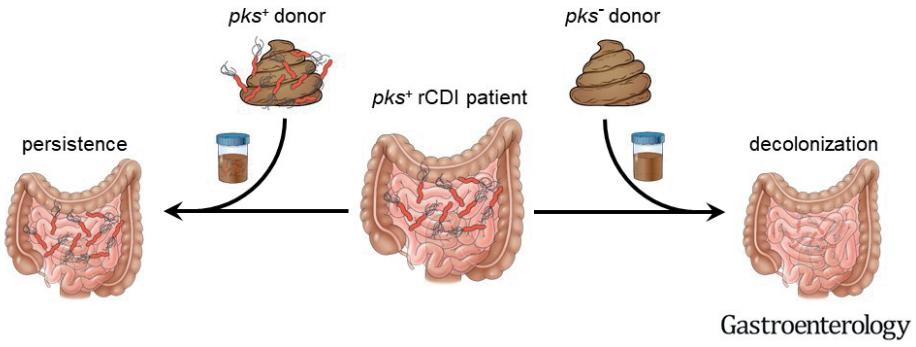
Results

The *pks* island was more prevalent ($P = .026$) and abundant ($P < .001$) in rCDI patients (pre-FMT; 27/49 = 55%, median: 0.46 reads per kilobase per million (RPKM) *pks*) than in healthy donors (3/8 donors (37.5%), 11/38 samples (29%), median: 0.01 RPKM *pks*). The *pks*-status of patients post-FMT depended on the *pks*-status of the donor suspension by which the patient was treated ($P = .046$). Particularly, persistence (8/9 cases) or clearance (13/18) of *pks*$^{+}$ *E. coli* in *pks*$^{+}$ patients was correlated to *pks* in the donor ($P = .004$).

Conclusion

We conclude that FMT contributes to *pks*$^{+}$ *E. coli* persistence or eradication in rCDI patients, but that donor-to-patient transmission of *pks*$^{+}$ *E. coli* is unlikely.

Graphical abstract



Introduction

Multiple recurrent *Clostridioides difficile* infection (rCDI) may occur as a complication after *C. difficile* infection (CDI) and is associated with a disturbance of the colonic microbiota.[1–3] Patients with rCDI can be treated with faecal microbiota transplantation (FMT), which has a cure rate of up to 89%.[1,4] The mechanism of action appears to be restoration of a healthy microbiome.[5] The donor faecal microbiota derives from carefully screened, healthy donors.[4,6,7] Donor screening focuses on the prevention of transfer of pathogens and the potential transmission of microbiota-associated disorders. In recent years there has been increasing attention for the causative role of gut bacteria in the development of colorectal cancer, which may pose a long-term risk.[8]

In 2019, Drewes and colleagues studied transmission and clearance of putative procarcinogenic bacteria, including colibactin-encoding *Escherichia coli*, by FMT in eleven paediatric rCDI patients.[9] Colibactin-producing *E. coli*, also known as *pks*$^+$ *E. coli*, has gained much attention lately, because it can be present in healthy and diseased people, has been used as a probiotic (strain Nissle 1917) and is now suspected to contribute to colorectal carcinogenesis.[10–12] The *pks*$^+$ *E. coli* carries a polyketide synthase (*pks*) gene island of about 54 kb long with 19 genes that encode the machinery to produce the nonribosomal peptide-polyketide hybrid genotoxin colibactin.[13–15] Colibactin can induce double-strand DNA breaks, which cause specific mutational signatures found in colorectal carcinomas.[10,13–15]

Here, we determine the effect of FMT from healthy donors on *pks*$^+$ *E. coli* in rCDI patients by retrospectively screening 49 matching triplets of deep-sequenced faecal metagenomes. Our results show that patients carry higher levels of *pks*$^+$ *E. coli* than donors and that the potential procarcinogen is more likely to persist when *pks*$^+$ donor faeces is used for FMT. When donor material is used with non-detectable levels of *pks*, FMT can contribute to clearance of *pks*$^+$ *E. coli*. Thus, FMT is unlikely to lead to transmission of *pks*$^+$ *E. coli*, but can reduce the levels of this procarcinogenic bacterium in rCDI patients.

Materials and methods

Faecal Microbiota Transplantation treatment

After careful screening of a multidisciplinary FMT expert panel, 49 patients suffering from multiple recurrent CDI were treated with FMT with a faecal suspension provided by the Netherlands Donor Faeces Bank. The complete procedure of donor recruitment, selection and screening, faecal suspension processing, evaluation of FMT requests, FMT treatment and patient follow-up has been described before.[4,16] In short, these patients received a vancomycin treatment for a minimum of four days (median: 12 days) until 24

hours before FMT. Eleven patients had a shorter antibiotics course or an unrecorded duration. Eight patients were treated with fidaxomicin or an unrecorded antibiotic. FMT was administered as a single treatment by a nasoduodenal tube, except in two patients that received a single infusion by colonoscopy. Differences in antibiotic pre-treatment regimen and mode of FMT were not found to influence the study results. The 198 mL faecal suspensions contained 60 gram of donor faeces in saline solution with 10% v/v glycerol and were stored until use at -80°C. All donors are healthy individuals between the ages of 18 and 60 and are extensively screened for disorders associated with microbiota and potential transmissible diseases.

Sample preparation and sequencing

Stool samples of patients were collected for sequencing approximately one day prior to FMT, after antibiotic pre-treatment, and three weeks after FMT. Stool samples of donors used for FMT treatment were also used for sequencing. Stool samples were stored at -80 °C. DNA extraction and sequencing was provided by DNA Genotek Inc. (Ottawa, Ontario, Canada). In short, DNA was extracted using the automated KingFisher™ Flex platform (Thermo Fisher Scientific Inc., Waltham, MA, USA) and MoBio PowerMag® +ClearMag® microbiome RNA/DNA isolation kit (QIAGEN Inc., Hilden, Germany). Sequencing libraries were prepared using Illumina's NexteraXT protocol and libraries were sequenced on the Illumina NextSeq platform to a target depth of 10 million reads per sample (2x150bp paired-end; Illumina, San Diego, CA, USA).

For patients P2 and P5 no matching donor metagenome was available, so instead they were matched to donor metagenomic data from samples taken 19 and 7 days later than those used for FMT, respectively.

Sequence analysis

Raw metagenomic reads were randomly subsampled to 10 million read pairs to facilitate *de novo* assembly using BBtools' reformat.sh (v. 38.79; parameters: 'samplereadstarget=10000000 sampleseed=20'; http://sourceforge.net/projects/bbmap/). The subsampled datasets were then analysed with the assembly-based workflow Jovian version 0.9.6,[17] using the NCBI BLAST nt database from 13 March 2020 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/). In short, Jovian trims raw reads (trimmomatic v. 0.38,[18] parameters 'PE SLIDINGWINDOW:5:30 MINLEN:50'), removes human reads by mapping to the human genome (bowtie2 v. 2.3.4.3,[19] parameters '--local', samtools v. 1.9 and bedtools v. 2.27.1;[20,21] reference genome 'GRCh38.p7' (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.33/)), assembles filtered reads into scaffolds (SPAdes v. 3.11.0,[22] parameters '--meta --only-assembler -k 21,33,55,77') and filters scaffolds to a minimum length of 500 nucleotides, classifies scaffolds taxonomically using a Lowest Common Ancestor approach (BLAST v. 2.9.0 parameters '-evalue 0.05 -max_target_seqs 250, -max_hsps 1',[23] MGKit v. 0.3.4 using quantile threshold .97 and bitscore threshold 100)[24] and quantifies scaffolds by mapping filtered reads back to the

scaffolds (BWA-mem v. 0.7.17,[25] default parameters). From Jovian, we used 1) the quality-trimmed reads from which human reads are removed, 2) assembled scaffolds and 3) taxonomic classifications and quantifications of scaffolds. These reads and scaffolds were screened for *Escherichia coli* and the *pks* island (19 *clb* genes from GenBank accession ID AM229678) using the custom workflow 'Jovian screener' (v. 0.1).[26] This screening extension to Jovian works in three ways. 1) It extracts the species of interest from the taxonomic profile results table (based on the 'tax_name' column). 2) It looks for the (gene) sequences of interest in all scaffolds using BLAST (v. 2.9.0, parameters '-task blastn -perc_identity 75 -qcov_hsp_perc 0 -evalue 1E-20 -culling_limit 1 -max_hsps 1'), which are then filtered to include only matches that cover at least half (50%) of the sequences of interest. And 3) it maps trimmed and filtered reads to the sequences of interest (BWA v. 0.7.17, samtools v. 1.10), after which mapped reads are deduplicated (using Picard MarkDuplicates v. 2.23.3 (https://broadinstitute.github.io/picard/)) and counted (samtools v. 1.10). The resulting tables were analysed and visualised using R version 4.0.2 (Rmarkdown scripts online at https://git.lumc.nl/snooij/pks-in-rcdi-metagenomes-analysis).[27]

We renamed the 'hypothetical protein (clb)' from AM229678 '*clbS*', because it is identical to accession number KX683217 (Escherichia coli c*lbS* gene, complete CDS). Reads per kilobase per million (RPKM) values were calculated by dividing the number of mapped reads by the length of the respective gene (coding sequence), multiplied by 1,000, divided by the number of trimmed and filtered reads, multiplied by one million.

Pilot screening for other procarcinogenic bacteria

Next to *pks* from *E. coli*, we also screened metagenomes for the presence of *Fusobacterium nucleatum* adhesin A (*fadA*; accession ID AVQ22939), *Bacteroides fragilis* toxin (*bft* 1-3; accession IDs AB026624, AB026625 and AB026626), and *Campylobacter jejuni* cytolethal distending toxin (*cdtA, cdtB, cdtC*; acccession ID AB274791). The *fadA* and *cdt* genes were not detected, and *bft* was detected in two patients (P7 pre- and post-FMT, and P42 post-FMT). As these genes are almost completely absent from the current dataset and absent from donors, we did not analyse these further.

Improving the identification of *pks* by combining data from all 19 *clb* genes

To improve the sensitivity of screening for the *pks* island in metagenomics data compared to using a single marker gene only (e.g. *clbB*), we assessed whether the genes always co-occur and may be considered as one unit. In assembled scaffolds, the *clb* genes are mostly assembled on one or a few scaffolds when depth is sufficiently high, suggesting consistent close genomic proximity (Supplementary File 1). For the mapping approach, we compared the RPKM values of *clb* genes in each metagenomics sample with one another and calculated Spearman correlations (Supplementary Figures S1-3). Correlations between *clb* genes are high (r = .54-.99, P < .01), and also higher than

correlations with flanking regions and more distant *E. coli* marker genes. Therefore, we conclude that the *pks* island is present in genomes as complete island and for its detection and quantification we combine all hits to *clb* genes to calculate average *pks* values.

Selecting representative *E. coli* marker genes

To calculate the abundance of *E. coli* in a similar way as the *pks* genes to facilitate the estimation of *pks* island copies per genome, we also mapped the reads to a number of putative single-copy marker genes. The putative marker genes are *dxs*, *rodA* and *uidA* (accession IDs: AF035440, M22857, and S69414, respectively),[28,29] MultiLocus Sequence Typing (MLST) genes *adk, fumC, gyrB icd, mdh, purA* and *recA* from Enterobase,[30] and *adk, gcl, zwf, mdh, metA* and *ppk* by Adiri and colleagues.[31] Based on the depth of coverage, we found that the MLST genes sometimes showed uneven coverage – especially in metagenomes with more non-*E. coli* Enterobacteriaceae. This indicates that these genes may have regions that are homologous to genes from other species, so the MLST genes were excluded. Next, we compared RPKM values for *dxs, rodA* and *uidA* to the relative abundance of *E. coli* and Enterobacteriaceae in all analysed metagenomes. The *uidA* gene has the best correlation with *E. coli* abundance (Supplementary Figure S4A; Spearman correlation $r = .97$, $P < 2.2e\text{-}16$). The *dxs* and *rodA* genes also appear in metagenomes with no *E. coli* (Supplementary Figure S4 B-C) and show better correlations with Enterobacteriaceae (Supplementary Figure S4 E-F, Spearman correlations $r = 0.96$ and $r = 0.97$, respectively). Even though *uidA* is the most specific marker gene, it is not sufficiently present in all metagenomes. Therefore, we calculated the average RPKM of *dxs*, *rodA* and *uidA*, whichever are present, to represent the quantity of *E. coli* genomes in the metagenomes, which was then used to calculate the ratio of the average *pks* island copies per *E. coli* genome.

Comparing *E. coli* assembled scaffolds between subjects

We assessed whether our data allows us to distinguish within-species variants of *E. coli* that could be used to conclusively demonstrate transfer or persistence. By using average nucleotide identity (pyANI version 0.2.10, parameter '-m ANIb')[32] we compared *E. coli* scaffolds from patients with *pks*$^+$ *E. coli* before and after FMT and their donors, and *E. coli* from patients that had no *pks* before FMT but were *pks*$^+$ after FMT and the corresponding donors. Average nucleotide identities (ANI) between *E. coli* scaffolds from different patients were as high as within patients pre- and post-FMT, so they could not be used to support persistence. Furthermore, donor scaffolds were too short to make a reasonable argument for transmission based on ANI. Where possible, *E. coli* assemblies were typed *in silico* using the multi-locus sequence typing (MLST) software 'mlst' (https://www.github.com/tseemann/mlst) and the PubMLST database (downloaded on 14 April 2021)[33]. MLST was successful only for 20 patients and for two patients provided result for both pre-FMT and post-FMT (Supplementary Table S1). These two patients, P21 and P36, had the same sequence type before and after

FMT, suggesting persistence of the same dominant variant. Both these patients lost *pks* after FMT, so these sequence types do not carry *pks*.

Data availability

All raw metagenomics data, from which human-derived reads have been excluded by mapping to a human reference genome (bowtie2 v.2.4.2, using parameter '--very-sensitive-local' to the GRCh38.p7 assembly), have been deposited in the European Nucleotide Archive under accession number PRJEB44737.

Results

Study characteristics

We screened the faecal metagenomes of 49 rCDI patients, pre- and post-FMT, and 38 corresponding metagenomes from donor faeces, of eight donors of the Netherlands Donor Feces Bank, for *pks*$^+$ *E. coli* via read mapping against a *pks* island reference (GenBank accession ID AM229678) and *E. coli* single-copy marker genes (*dxs*, AF035440; *rodA*, M22857; *uidA*, S69414). Eleven donor suspensions were used to treat two patients each, thus yielding 49 sample triplets. The patients were between 27 and 92 years old (median: 73.5), including 31 females and 18 males. Only patients were included that had been treated with a single FMT and for which pre- and post-FMT samples were available. The donors were between 24 and 46 years old (median: 33); four females and four males. Samples were collected between April 2016 and March 2018. For patients P2 and P5 no donor metagenomic data were available from the date of FMT, but from 19 and 7 days later, respectively. For other patients, corresponding donor data were available from the same sample as used for FMT. All patients were cured from rCDI, except for three patients that had a recurrence within three weeks after FMT (P12, P25 and P26).

The *pks* island is absent in most donors or detected in very low quantities

To quantify *pks* levels in metagenomes, we averaged normalised RPKM values of 19 *clb* genes (colibactin-encoding, AM229678). Subjects with zero reads mapped to *clb* genes in their metagenome are considered *pks*-negative (*pks*$^-$) and subjects with one or more reads mapped to one or more *clb* genes are *pks*$^+$. Five out of eight healthy donor metagenomes were always negative for the *pks* island (Figure 1). One donor had multiple samples with *pks*$^+$ metagenomes (D5, see below). Another donor (D2) provided only one sample, which was *pks*$^+$, and a third donor (D6) had one *pks*-positive sample out of nine used for transplantation. Levels of *pks* among positive donors range from 0.0013 to 0.0498 RPKM (median: 0.010). In total, 11 out of 38 donor samples were *pks*$^+$ (29%).

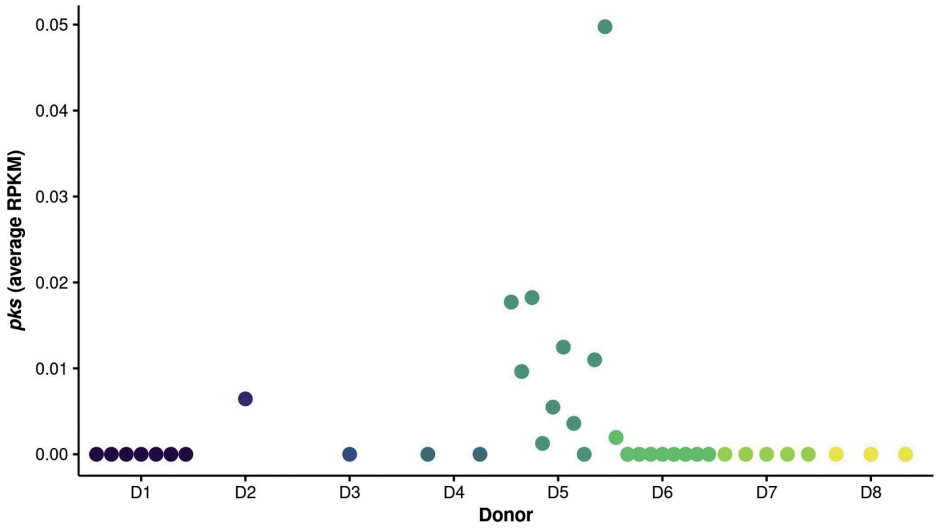


Figure 1. Levels of *pks* in donor metagenomes. Reads from donor metagenomes were mapped to the *pks* island (AM227896) to quantify the *pks* island per sample. Values are normalised by gene length and library size (reads per kilobase per million; RPKM) and the average RPKM value of the 19 *clb* genes is calculated as average *pks*. Out of eight donors, three had *pks*⁺ samples of which only one had multiple *pks*⁺ samples. Out of 38 donor samples included, 11 were *pks*⁺.

Healthy donors may persistently carry *pks*

To study the temporal dynamics of *pks* colonisation of donors, we evaluated longitudinal samples from one donor that was consistently *pks*$^+$ (D5). Out of 15 metagenomes total from donor D5, of which 10 were used for transplantation, 13 had traces of the *pks* island (Figure 2A). Both the first and last samples were positive, suggesting that this donor was persistently colonised with a *pks*$^+$ bacterium during the donation period of six months. Levels of *pks* range from 0 to .05 RPKM and correlate loosely with the relative abundance of *E. coli* (Spearman r = .43, P = .11) or Enterobacteriaceae (Spearman r = .1, P = .72). *E. coli* was not always detected in the metagenomes of donor D5 (Figure 2B-C). It seems most likely that due to the low abundance of *E. coli*, the sequencing depth in the metagenome was low, yielding only short genomic regions that are shared between *E. coli* and other Enterobacteriaceae resulting in these being classified as Enterobacteriaceae.

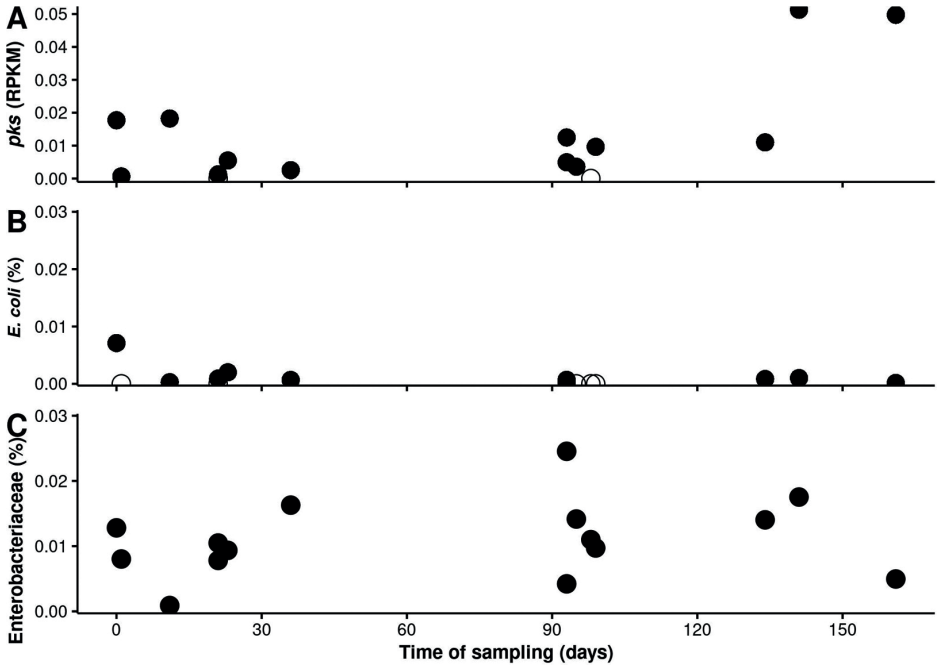


Figure 2. Levels of *pks* levels and relative abundances of *E. coli* and Enterobacteriaceae over time in one healthy donor. To gain more insight in the duration of colonisation by *pks*$^+$ *E. coli* in healthy donors, we quantified average *pks* levels of all available metagenomes of D5 by read mapping and we determined the relative abundances of *E. coli* and Enterobacteriaceae for each metagenome using Jovian. A) Average normalised values of *pks* island (*clb* genes) in donor D5 over approximately six months' time. Samples from days 21 and 93 are duplicates; both are shown. *Pks* levels are generally very low, near the limit of detection. The donor seems to have been persistently colonised by a *pks*$^+$ species for at least half a year, even though two samples in the first and fourth months were *pks*$^-$. B-C) Relative abundances of *E.* coli and Enterobacteriaceae (including *E. coli*) in each metagenomic sample of donor D5. For the duplicate samples only the higher values are shown. There is no clear correlation between either abundance and the level of *pks*. Empty circles denote that no *pks* or *E. coli* was detected at that time point.

The *pks* island is frequent and abundant in pre-FMT rCDI patient metagenomes

Genes from the *pks* island were detected in 27/49 patients before FMT treatment (55%), and 19/49 patients after FMT (39%). Overall, *pks* was more prevalent in pre-FMT rCDI patient metagenomes than in donor metagenomes (χ^2, $P = .026$, odds ratio = 2.246, 95% confidence interval = 1.010–4.589). Patient metagenomes reached *pks* levels up to 59.7 RPKM, corresponding to roughly 0.3% of all analysed reads. The median *pks* levels of *pks*$^+$ patient metagenomes were 0.461 RPKM and 1.348 RPKM, respectively for pre- and post-FMT. The *pks* level in *pks*$^+$ patient metagenomes was significantly higher than in donors (pairwise Wilcoxon rank sum exact test, $P < .001$ for both pre- and post-FMT), but was not different between pre- and post-FMT ($P = .66$; Figure 3A). When including

the *pks*-negative metagenomes in the comparison, the differences are less apparent and *pks* levels of donors only differ from patients before FMT (Supplementary Figure S5).

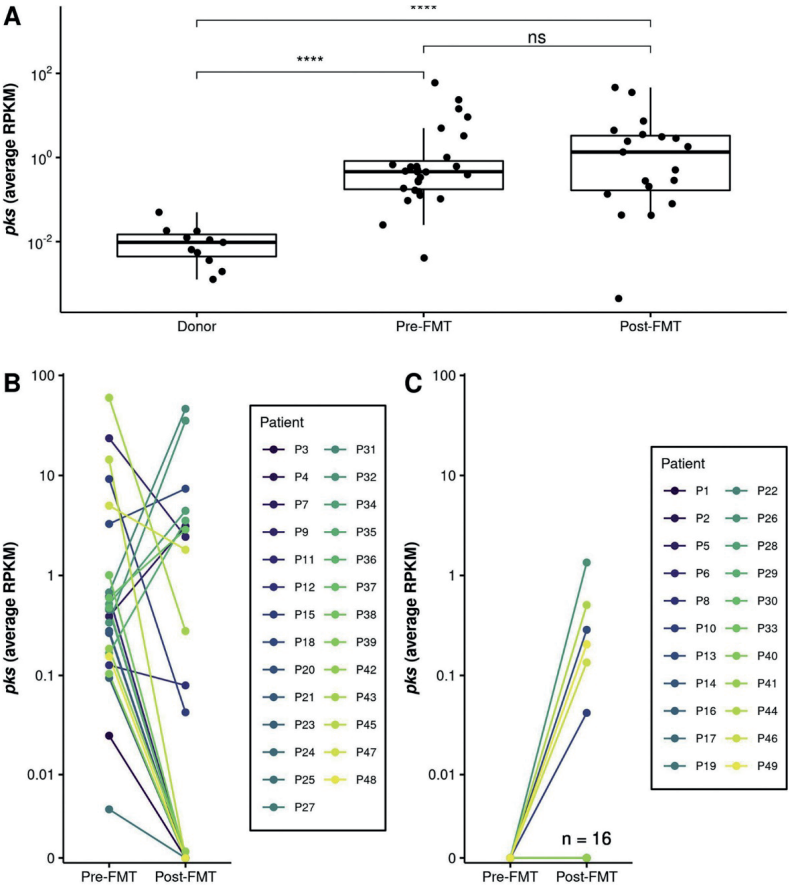


Figure 3. Levels of *pks* levels in *pks*+ faecal donor and rCDI patient metagenomes before and after FMT treatment and changes in *pks* in patients after FMT. To assess the effect of FMT on the *pks* levels in patients, we mapped reads to a *pks* island reference and quantified *pks* by normalising read counts by gene length and library size (RPKM), and averaged values over all 19 *clb* genes. Y-axes use a log10-scale. A) Levels of *pks* in *pks*$^+$ donor samples were lower than those in patients pre- and post-FMT (Wilcoxon rank sum test, $P < .001$). Levels of *pks* were not different between those found in patients pre-FMT and those post-FMT. B) Changes in *pks* levels in patients that were *pks*$^+$ before FMT. Twenty-seven patients were *pks*$^+$ prior to FMT. Twenty of them have lower *pks* levels after FMT, of which 14 lost *pks*. Seven patients had increased levels of *pks* after FMT. C) Changes in FMT levels in patients that were *pks*$^-$ before FMT. Twenty-two patients had no detectable *pks* prior to FMT. Six patients acquired *pks* post-FMT and sixteen remained *pks*$^-$. The number of patients with zero *pks* after FMT is written in the figure to clarify the overlapping dots.

Fecal microbiota transplantation changes *pks* levels in *pks*$^+$ patients with recurrent *Clostridioides difficile* infection

As *pks* levels – when detected – are much higher in rCDI patients than in healthy donors (Figure 3A), we determined whether treatment by FMT leads to a reduction in *pks* levels in the subset of *pks*$^+$ rCDI patients. Overall, there is no significant difference between *pks* levels pre- and post-FMT in patients that had *pks* prior to FMT (paired Wilcoxon signed rank exact test, $P = .2901$). However, when stratified by *pks*-status of the donor we find that the change of *pks* level in the rCDI patient after FMT depends on the *pks*-status of the donor (Fisher's exact test, $P = .0006$).

Twenty-seven patients' metagenomes were *pks*$^+$ prior to FMT treatment, and *pks* levels varied more than thousandfold. Twenty (74%) had decreased levels of *pks* after FMT, of which 14 (52%) were reduced to levels below our detection limit (Figure 3B). Eighteen out of these 20 (90%) were treated with a *pks*$^-$ donor suspension (Table 1, Supplementary Figure S6). In seven cases the level of *pks* increased, of which six were treated with a *pks*$^+$ donor suspension (86%).

Table 1. Summary of *pks*-status for each step in the FMT treatments of the 49 included patients. Forty-nine patients, of which 22 were *pks*$^-$ and 27 *pks*$^+$ prior to FMT, were treated with 35 *pks*$^-$ and 14 *pks*$^+$ donor suspensions. After FMT, 30 patients are *pks*$^-$ and 19 are *pks*$^+$.

Donor suspensions	Patients pre-FMT	Patients post-FMT
35 *pks* -	17 *pks* -	12 *pks* -
		5 *pks* +
	18 *pks* +	13 *pks* -
		5 *pks* +
14 *pks* +	5 *pks* -	4 *pks* -
		1 *pks* +
	9 *pks* +	1 *pks* -
		8 *pks* +

Patients who were *pks*-negative before FMT mostly remained negative

We assessed the effect of FMT on patients that had no detectable levels of *pks* before FMT to evaluate the possibility of transmission from a *pks*$^+$ donor. Overall, as expected with a conversion of *pks*-status in a *pks*-negative population, the *pks* level increased after FMT (paired Wilcoxon signed rank test, $P = .03603$). Twenty-two patients had no detectable *pks* prior to FMT. Sixteen (73%) of them remained negative after FMT and six patients acquired *pks* after FMT (Table 1). Interestingly, from the six patients who became *pks*$^+$ after FMT, five were treated with a *pks*$^-$ donor suspension derived from a total of three different donors (83%; Figure 3C, Supplementary Figure S6). Out of

the 16 patients that remained negative, 12 were treated with *pks*$^-$ donor suspensions (75%). The average *pks* level of the six patient metagenomes that acquired *pks* after FMT is 0.420 RPKM (median: 0.245 RPKM). The *pks* levels in patients who acquired *pks* vary more than tenfold and are generally low compared to *pks* levels in patients with persistent *pks* (median: 2.874 RPKM), although the difference is not significant (Wilcoxon rank sum test, $P = .08742$). Importantly, however, whether a patient in this group of *pks*$^-$ rCDI patients demonstrated increased *pks*-levels after FMT is not significantly correlated with the *pks*-status of the donor (Fisher's exact test, $P = 1$), suggesting that an increase in *pks* is as likely to derive from the patient itself as from the donor.

Persistence or clearance of *pks* in patients after FMT depends on presence of *pks* in donors

In addition to testing the effect of FMT on the quantity of *pks* in patients, we assessed the likelihood of *pks* persistence, clearance and transmission based on the presence of *pks* in patients and their donors. Thirty-five rCDI patients were treated with a *pks*$^-$ donor suspension of which twenty-five were *pks*$^-$ after FMT (71%; Table 1) and four had reduced *pks* (11%; Supplementary Figure S6). Fourteen patients were treated with a *pks*$^+$ donor suspension and nine patients were *pks*$^+$ after FMT (64%). Based on these numbers, we find a direct dependence between the post-FMT *pks*-status and *pks* in the donor suspension (χ^2 with Yates' continuity correction, $P = .046$). The effect of FMT on post-FMT *pks*-status was strongest in patients that were already colonised with a *pks*$^+$ *E. coli* before FMT (Fisher's exact test, $P = .004$). Out of 27 patients that were *pks*$^+$ pre-FMT, 9 were treated with a *pks*$^+$ donor suspension of whom 8 remained *pks*$^+$ post-FMT. Of the 18 patients who were treated with a *pks*-free donor suspension, 5 contained *pks* post-FMT. There was no such effect in patients in which we detected no *pks* pre-FMT (Fisher's exact test, $P = 1$). Although our metagenomic analyses could not demonstrate persistence or transmission using within-species variants (see Methods), taken together, these results indicate that FMT contributes to persistence or clearance rather than donor-to-patient transmission of *pks* in rCDI patients.

Variable within-species *pks* island prevalence suggests persistence of *pks*$^+$ *E. coli* in patients

Using single-copy marker genes *dxs*, *rodA* and *uidA* (encoding D-1-deoxyxylulose 5-phosphate synthase, rod-shape determining protein, and beta-glucuronidase, respectively) to quantify *E. coli* genomes, we calculated the average number of *pks* island copies per *E. coli* genome, or prevalence of *pks* among *E. coli* variant subpopulations. *E. coli* is detected in less than half of the donor metagenomes (14/38; full circles in Figure 4A) and is detected in five out of eleven *pks*$^+$ metagenomes (from donors D2 and D5). *E. coli* markers are present in nearly all patients pre-FMT (48/49; Figure 4B) and *pks* per *E. coli* ratios that are often near 0 indicate that *pks*$^+$ *E. coli* variants are often a minority variant. Post-FMT, still many patients harbour *E. coli* species (46/49;

Figure 4C) and when *pks* is present the *pks* per *E. coli* ratios are often higher. Taken together, the results suggest that rCDI patients treated with vancomycin or fidaxomicin before FMT may have high abundances of different *E. coli* subpopulations or variants. When a patient is colonised with a *pks*$^+$ *E. coli*, the *pks*$^+$ *E. coli* is often a minority of *E. coli* variants. After treatment by FMT using a donor suspension that contains little if any *E. coli*, the number of different *E. coli* variants decreases. And when a *pks*$^+$ variant is present post-FMT, this is often a larger fraction or the only *E. coli* variant, possibly indicating that this variant is more persistent than other *E. coli* variants.

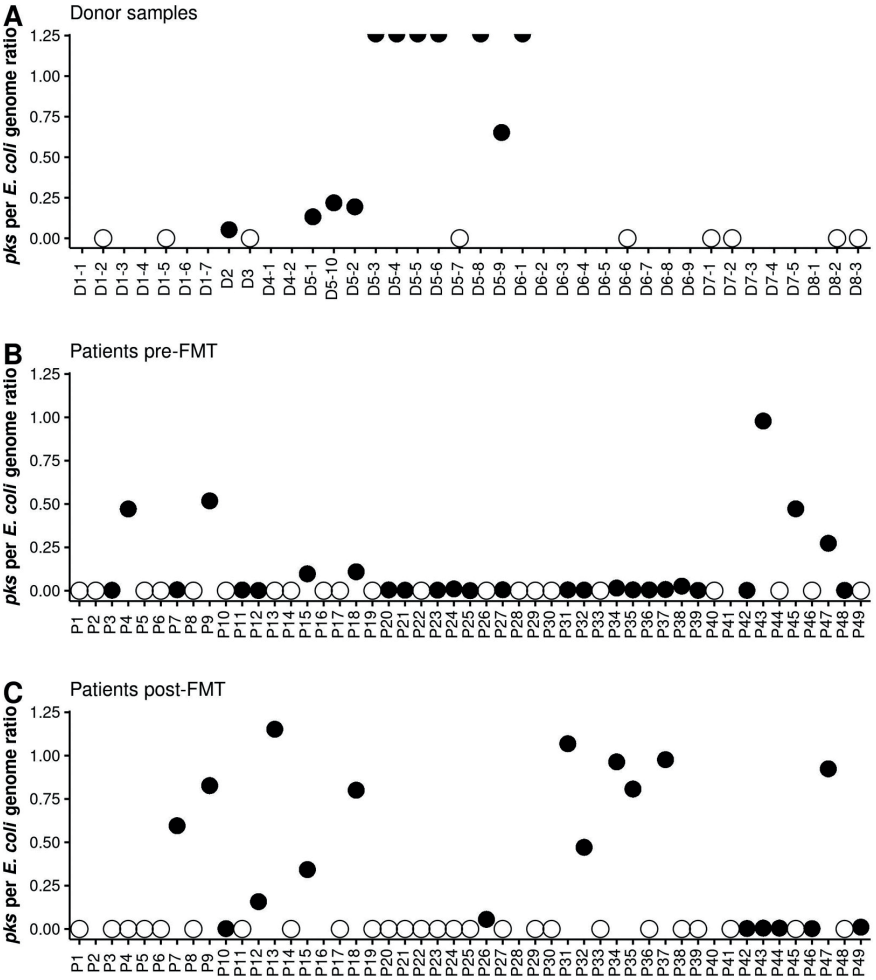


Figure 4. Average number of *pks* island copies per *E. coli* genome, based on three marker genes. Average *pks* island abundances per sample are calculated as the average RPKM value of all 19 *clb* genes. *E. coli* genome abundances are estimated based on the average RPKM value of the single-copy marker genes *dxs*, *rodA* and *uidA*. Empty circles indicate presence of *E. coli* and absence of *pks*, black semicircles at the top of a figure indicate presence of *pks* and absence of *E. coli* and missing circles indicate neither *E. coli* nor *pks* was detected in the metagenome. A) *E. coli* is detected in fourteen donor samples of which five samples from two different donors are *pks*⁺. In these samples, there are between 0.05 and 0.65 *pks* island copies per *E. coli* genome. Six donor metagenomes have *pks* without *E. coli* markers. B) Nearly all patients have *E. coli* in their faecal metagenome pre-FMT (all but P41). *pks* per *E. coli* ratios vary from zero (*E. coli* without the *pks* island) to one (all *E. coli* genomes in the sample have the *pks* island). In most *pks*⁺ cases, it is a minority of *E. coli* genomes that have the *pks* island. C) After FMT, *E. coli* is still detected in most patients and the average number of *pks* copies per *E. coli* genome varies from zero to approximately one. Fewer patients are *pks*⁺ after FMT and *pks* per *E. coli* ratios are often higher: in nine cases the majority of *E. coli* genomes has the *pks* island.

Discussion

In this cohort study of 49 FMT-treated adult rCDI patients and their corresponding healthy faeces donors, we screened faecal metagenomes for the presence of *pks*$^+$ *E. coli* to assess its prevalence and the effect of FMT. To our knowledge, this is the first study to employ metagenomic screening for potentially procarcinogenic bacteria in the largest FMT cohort to date. We find that *pks* was present in a minority of healthy donor samples and, when present, *pks* levels were low. Of eight donors, one donor was likely persistently colonised with a *pks*$^+$ bacterium for six months. In contrast, rCDI patients had high levels of *pks* and high abundances of *E. coli* in their metagenomes. FMT resulted in changes in both status and levels of *pks* detected in the deep-sequenced metagenomes. In particular, we find that treatment of *pks*$^+$ rCDI patients with a *pks*$^-$ donor suspension reduces *pks* levels, in some cases to below the detection limit, whereas treatment with a *pks*$^+$ donor suspension leads to persistence of *pks*$^+$ *E. coli*.

A previous study elucidating the effects of FMT on the presence of intestinal procarcinogenic bacteria in eleven paediatric rCDI patients reported similar findings.[9] Drewes and colleagues found that FMT supported decolonisation depending on the microbiota of both patient and donor. We find that FMT supports decolonisation of *pks*$^+$ bacteria in rCDI patients treated with *pks*$^-$ donor suspensions. Drewes and colleagues reported transmission of putative procarcinogenic bacteria during FMT. Though we cannot exclude the possibility of transmission on the basis of our data, we consider it more likely that *pks*$^+$ *E. coli* variants persist in the patients than that they are transferred from the donor, as *E. coli* is much lower in abundance in donor material than in patients, and engraftment of low-abundant bacteria in the donor is less likely.[34] A study specifically designed to address the question of transmission of procarcinogenic bacteria during FMT, using culturing and whole-genome sequencing, is needed to investigate this aspect further. Other potentially procarcinogenic species were studied previously in the context of FMT, namely enterotoxigenic *Bacteroides fragilis* and *Fusobacterium nucleatum*.[9] We also screened the current set of faecal metagenomes for the presence of these and other putative procarcinogenic bacteria (the *bft* toxin gene from enterotoxigenic *B. fragilis*, the *fadA* adhesin gene from *Fusobacterium nucleatum*, and the *cdtA-C* genes from *Campylobacter jejuni*),[35] but we only detected *bft* in two metagenomes from two patients. As copy numbers of the *bft*, *clbB* and *fadA* genes was reported to be similar in colonised patients,[9] and we readily detect *clb* genes (from the *pks* island), we do not expect this to be a technical limitation of our method, but possibly reflecting differences in the study population. As the other procarcinogenic bacteria appeared to be absent and could not be analysed further, we focused our analyses on *pks*$^+$ *E. coli*.

In a few faecal metagenomes, especially from donors, we detected *pks* in absence of *E. coli*. Possibly, *E. coli* was present below the limit of detection or the *pks* gene island

belonged to a different, pathogenic, species of Enterobacteriaceae that may also carry the *pks* island.[36] Although we detected *Klebsiella*, *Enterobacter*, and *Citrobacter* species using the assembly-based approach in six out of eleven donor samples with *pks*, but that were negative for *E. coli* based on the three single-copy marker genes, pathogenic isolates of these species are unlikely to be present in healthy donors due to thorough screening. Therefore, we consider it more likely that *E. coli* was present below the detection limit of our assay and is in fact the host of the *pks* island.

Earlier studies on procarcinogenic intestinal bacteria applied different technical approaches for detection and quantification. Often, detection of the *pks* island was based on the longest gene (*clbB*; 9.6 kb); either by PCR or by metagenomic screening.[9,11] Metagenomics-based assays require high sequencing depths to reliably detect single genes. We combine data of all 19 *clb* genes, which we show are highly intercorrelated, to improve the sensitivity of our assay and allowing us to detect very low levels of *pks*. Thereby we increased the number of *pks*$^+$ samples by eight compared to detection based on only *clbB*. The *pks* levels (as RPKM values) in our screened faeces donors are slightly lower than previously reported for healthy controls (donors: 0 RPKM median, 0.05 maximum; healthy controls: 0.06 median), while *pks* levels in *pks*$^+$ rCDI patients (pre-FMT: 0.46 median, post-FMT: 1.35 median) are higher than those found in inflammatory bowel disease patients (0.18 RPKM median).[11] This suggests that the screened and approved donors of the National Donor Feces Bank may have lower *pks* levels than average healthy controls. In addition, these rCDI patients seem to have been exposed to *pks* levels higher than those in inflammatory bowel disease patients. By comparing prevalence of *pks* in different groups we come to similar conclusions. The prevalence of *pks* among healthy donors (10 out of 38 samples; 26.3%) is comparable to other studies, which reported prevalence of *pks* among healthy controls between 20 and 30%.[10,11,36] The prevalence of *pks* among rCDI patients is higher (27/49 pre-FMT and 19/49 post-FMT; 55.1% and 38.8%) than in healthy donors. The numbers are between those reported for inflammatory bowel disease (~33%),[11] and those reported among colorectal cancer patients (~60%).[10] However, the reported prevalence is often based on fewer than 50 patients and may therefore not accurately reflect the prevalence in the general population. Regardless, the *pks* levels and prevalence we find are approximately in the same range as described before and suggest that the relatively old rCDI patients might be at an increased risk of developing colorectal cancer.

The age difference of approximately 40 years between our donors and patients is of note, as this may have been a confounding factor in the current study, assuming that age is correlated with *pks*$^+$ *E. coli* colonisation. To our knowledge, however, there has been no evidence indicating that a higher age is associated with a higher incidence of *pks*$^+$ *E. coli* colonisation. Furthermore, it has been suggested that rCDI is correlated to colorectal carcinogenesis.[37] These hypotheses may be tested in a future study by screening a cohort of healthy people of similar age as the rCDI patients and follow

them over time to compare the number of occurrences of colorectal cancer between these two groups.

In conclusion, we find that *pks* is prevalent and can be highly abundant in rCDI patients. FMT with *pks*$^-$ donor suspensions generally decreases *pks* levels of patients and *pks* persists in *pks*$^+$ patients when a *pks*$^+$ donor is used. We find no clear evidence of donor-to-patient transmission of *pks*$^+$ *E. coli* and if it were to occur, the impact may be negligible given the age and comorbidity of most patients in this particular cohort. Therefore, we think that current screening protocols for FMT donors are safe and routinely screening for *pks* is not required, at least for the treatment of rCDI. It is currently unknown how long *pks*$^+$ *E. coli* persists in cured rCDI patients and on what time-frame this may contribute to colorectal carcinogenesis. Further research addressing the long-term health effects is needed to evaluate the cancer risk due to *pks*$^+$ *E. coli* in rCDI and other patients, since this might offer opportunities for early intervention in the development of colorectal cancer.

Acknowledgments

The authors wish to thank dr. Szymon M. Kiełbasa and the other members of the interdepartmental next-generation sequencing discussion group at the Leiden University Medical Center for their advice on the analysis and statistics, and the Experimental Bacteriology group for helpful comments during work discussions. Also, the authors thank dr. Bastian V.H. Hornung for assistance with handling the dataset.

Ethics approval

Patients provided informed consent for collection of stool samples and outcome data of FMT for research purposes, which was approved by the Medical Ethics Committee at the Leiden University Medical Center (P15.145).

Contributors

SN, QRD, EMT and EJK conceived and designed the study. EJK, EMT and RDZ supervised the study. HWV, JJK, EMT and EJK were involved in the treatment of patients and collection of samples. SN, QRD and JFJL designed the analysis approach. JMN contributed materials and oversaw logistics. SN performed bioinformatics analyses and wrote the manuscript. SN, QRD, JFJL, RDZ, WKS, HWV, JJK, EMT and EJK discussed results and implications. All authors contributed to and approved the final manuscript.

Competing interests

The Netherlands Donor Feces Bank received an unrestricted grant from Vedanta Biosciences, Inc. (Cambridge, MA, USA) from 2017 to 2020. Metagenomic sequencing was funded by Vedanta Biosciences. The authors received no financial support for the research, authorship and/or publication. EJK has performed research for Cubist, Novartis and Qiagen and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur and Seres Therapeutics. WKS has performed research for Cubist and has received speaker fees from Promega. JMN is an employee of Vedanta Biosciences Inc. The companies had no role in this study or writing of the manuscript. The other authors declare no competing interests.

Abbreviations used in this paper

rCDI (multiple recurrent *C. difficile* infection), FMT (faecal microbiota transplantation), *pks* (polyketide-synthase), RPKM (reads per kilobase per million), ST (sequence type)

Supplementary Material

Supplementary figures and document are available online on the publisher's website:

https://www.gastrojournal.org/article/S0016-5085(21)03120-6/fulltext#app-1

References

1. Smits WK, Lyras D, Lacy DB, et al. Clostridium difficile infection. Nature Reviews Disease Primers 2016;2:1–20.
2. Chang JY, Antonopoulos DA, Kalra A, et al. Decreased Diversity of the Fecal Microbiome in Recurrent Clostridium difficile—Associated Diarrhea. J Infect Dis 2008;197:435–438.
3. Nood E van, Vrieze A, Nieuwdorp M, et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N Engl J Med 2013;368:407–415.
4. Terveer EM, Vendrik KE, Ooijevaar RE, et al. Faecal microbiota transplantation for Clostridioides difficile infection: Four years' experience of the Netherlands Donor Feces Bank. United European Gastroenterology Journal 2020;8:1236–1247.
5. Khoruts A, Sadowsky MJ. Understanding the mechanisms of faecal microbiota transplantation. Nature Reviews Gastroenterology & Hepatology 2016;13:508–516.
6. Cammarota G, Ianiro G, Kelly CR, et al. International consensus conference on stool banking for faecal microbiota transplantation in clinical practice. Gut 2019;68:2111–2121.
7. Keller JJ, Ooijevaar RE, Hvas CL, et al. A standardised model for stool banking for faecal microbiota transplantation: a consensus report from a multidisciplinary UEG working group. United European Gastroenterology Journal 2020:2050640620967898.
8. Scott AJ, Alexander JL, Merrifield CA, et al. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. Gut 2019;68:1624–1632.
9. Drewes JL, Corona A, Sanchez U, et al. Transmission and clearance of potential procarcinogenic bacteria during fecal microbiota transplantation for recurrent *Clostridioides difficile*. JCI Insight 2019;4.
10. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, et al. Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature 2020;580:269–273.
11. Dubinsky V, Dotan I, Gophna U. Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. Trends in Microbiology 2020;0.
12. Olier M, Marcq I, Salvador-Cartier C, et al. Genotoxicity of Escherichia coli Nissle 1917 strain cannot be dissociated from its probiotic activity. Gut Microbes 2012;3:501–509.
13. Nougayrède J-P, Homburg S, Taieb F, et al. Escherichia coli Induces DNA Double-Strand Breaks in Eukaryotic Cells. Science 2006;313:848–851.
14. Li Z-R, Li J, Cai W, et al. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. Nat Chem 2019;11:880–889.
15. Dziubańska-Kusibab PJ, Berger H, Battistini F, et al. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nature Medicine 2020:1–7.
16. Terveer EM, Beurden YH van, Goorhuis A, et al. How to: Establish and run a stool bank. Clin Microbiol Infect 2017;23:924–930.
17. Schmitz D, Zwagemaker F, Nooij S, et al. *Jovian, user-friendly Public Health toolkit*. Zenodo; 2020. Available at: https://doi.org/10.5281/zenodo.3929618.

18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.

19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods 2012;9:357–359.

20. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2079.

21. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–842.

22. Nurk S, Meleshko D, Korobeynikov A, et al. metaSPAdes: a new versatile metagenomic assembler. Genome Res 2017;27:824–834.

23. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.

24. Rubino F, Creevey CJ. MGkit: Metagenomic Framework For The Study Of Microbial Communities. 2014. Available at: https://figshare.com/articles/poster/MGkit_Metagenomic_Framework_For_The_Study_Of_Microbial_Communities/1269288/1 [Accessed November 10, 2020].

25. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio] 2013. Available at: http://arxiv.org/abs/1303.3997 [Accessed November 10, 2020].

26. Nooij S. *jovian-screener*. Zenodo; 2020. Available at: https://doi.org/10.5281/zenodo.4288679.

27. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available at: https://www.R-project.org/.

28. Lee C, Kim J, Shin SG, et al. Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. Journal of Biotechnology 2006;123:273–280.

29. Chern EC, Siefring S, Paar J, et al. Comparison of quantitative PCR assays for Escherichia coli targeting ribosomal RNA and single copy genes. Letters in Applied Microbiology 2011;52:298–306.

30. Wirth T, Falush D, Lan R, et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Molecular Microbiology 2006;60:1136–1151.

31. Adiri RS, Gophna U, Ron EZ. Multilocus sequence typing (MLST) of Escherichia coli O78 strains. FEMS Microbiol Lett 2003;222:199–203.

32. Pritchard L, Glover RH, Humphris S, et al. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 2015;8:12–24.

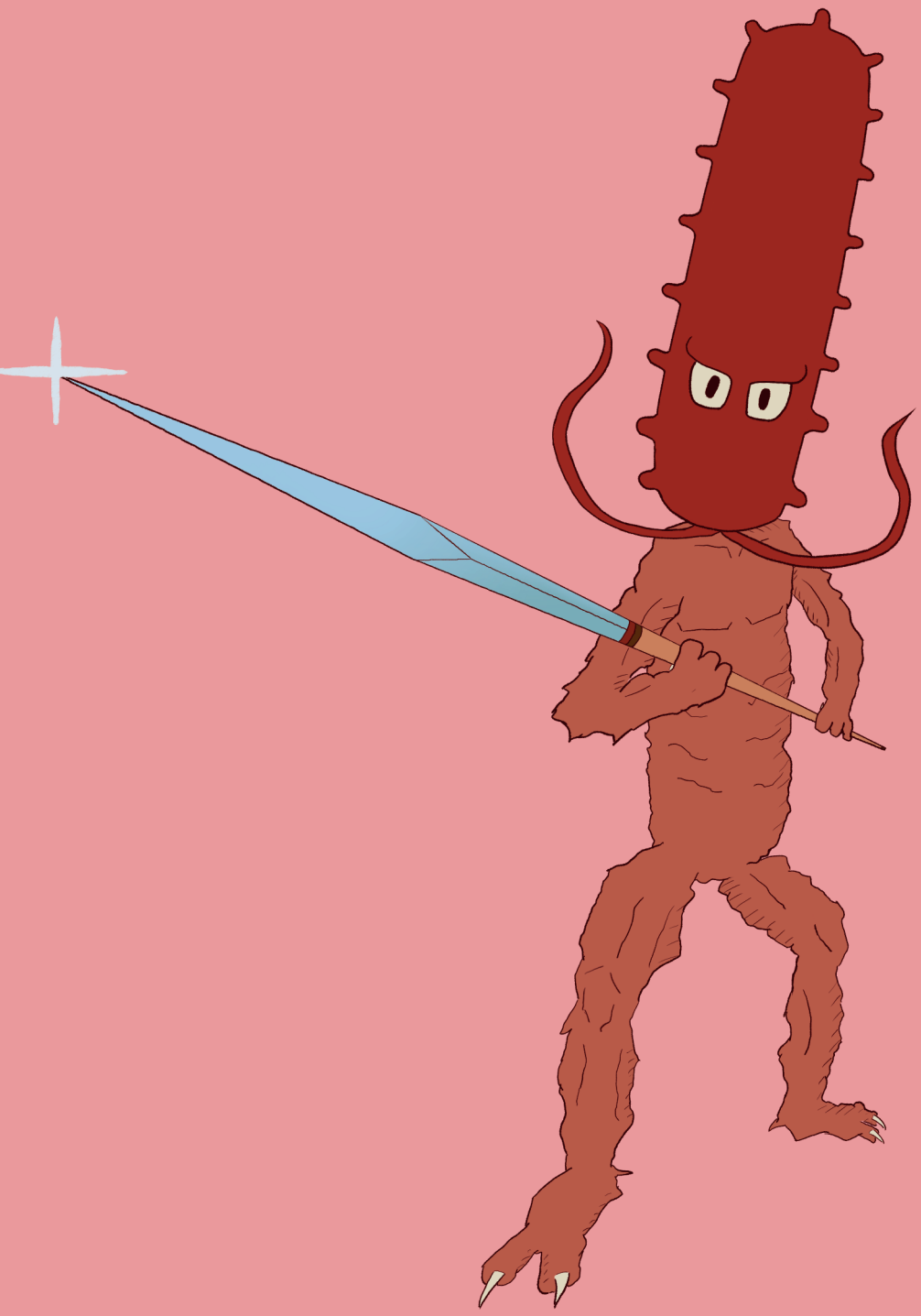
33. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 2010;11:595.

34. Smillie CS, Sauk J, Gevers D, et al. Strain Tracking Reveals the Determinants of Bacterial Engraftment in the Human Gut Following Fecal Microbiota Transplantation. Cell Host & Microbe 2018;23:229-240.e5.

35. He Z, Gharaibeh RZ, Newsome RC, et al. Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. Gut 2019;68:289–300.

36. Putze J, Hennequin C, Nougayrède J-P, et al. Genetic Structure and Distribution of the Colibactin Genomic Island among Members of the Family Enterobacteriaceae. Infection and Immunity 2009;77:4696–4703.

37. Fukugaiti MH, Ignacio A, Fernandes MR, et al. High occurrence of Fusobacterium nucleatum and Clostridium difficile in the intestinal microbiota of colorectal carcinoma patients. Braz J Microbiol 2015;46:1135–1140.



Chapter 2.2

Rational donor selection for fecal microbiota transplantation

Reply

Sam Nooij, Elisabeth M. Terveer, on behalf of the Netherlands Donor Feces Bank

Gastroenterology, Volume 162, Issue 3, 994 – 995, March 2022

Dear editor,

We thank dr. Gianluca Ianiro et al. for the suggestion, in their reaction to our manuscript, to evaluate the donor microbiome for fecal microbiota transplantation (FMT) by carefully assessing donor history and analyzing the microbiome to predict a favorable microbial signature. We agree that studying successful donor-patient pairs is crucial in understanding complex microbiome-related disease pathology and subsequent cure. However, there are several arguments against the existence of a favorable microbial signature for recurrent *Clostridioides difficile* infections (rCDI) (1). For other diseases it is currently not yet reliable as predictor on an individual donor level.

Fecal suspensions from carefully screened donors without specific targeted microbiome screening for the treatment of patients with rCDI always show high (>80%) success rates world-wide. Still, the working mechanism of FMT to prevent recurrent CDI is not completely understood and immunological effects are probably involved in addition to restoring the colonization resistance by reintroducing a healthy microbiome. FMT seems to enhance *C. difficile*-specific cellular and antibody-mediated immunity, as it is associated with increased proportions of toxin B-specific T helper-17 cells as well as IgG and IgA antibodies specific for toxin A and B (2). Of note, the immunological effect that FMT can elicit may be of more relevance for other diseases, but is still difficult to predict. Interestingly, a study using FMT to promote response in immunotherapy-refractory melanoma patients revealed that of the two donors rationally selected based on a preferred microbiota profile, only one was associated with tumor suppression in the patients (3). Nevertheless, donor selection is gaining ground despite the lack of knowledge about a preferred microbial signature to define a super donor on the individual donor level. A large FMT center in China has performed over 60,000 FMTs across more than 5,000 patients for various gastrointestinal diseases (recurrent CDI, inflammatory bowel disease, irritable bowel syndrome) and extraintestinal diseases (autism spectrum disorder and Parkinson's disease) (4). Our Chinese colleagues developed a very stringent donor screening program including 16S rRNA gene amplicon sequencing of stool samples to evaluate bacterial compositions. These data would provide a unique opportunity to evaluate the efficacy of FMT from various donors, (serious) adverse events and long-term follow-up of patients.

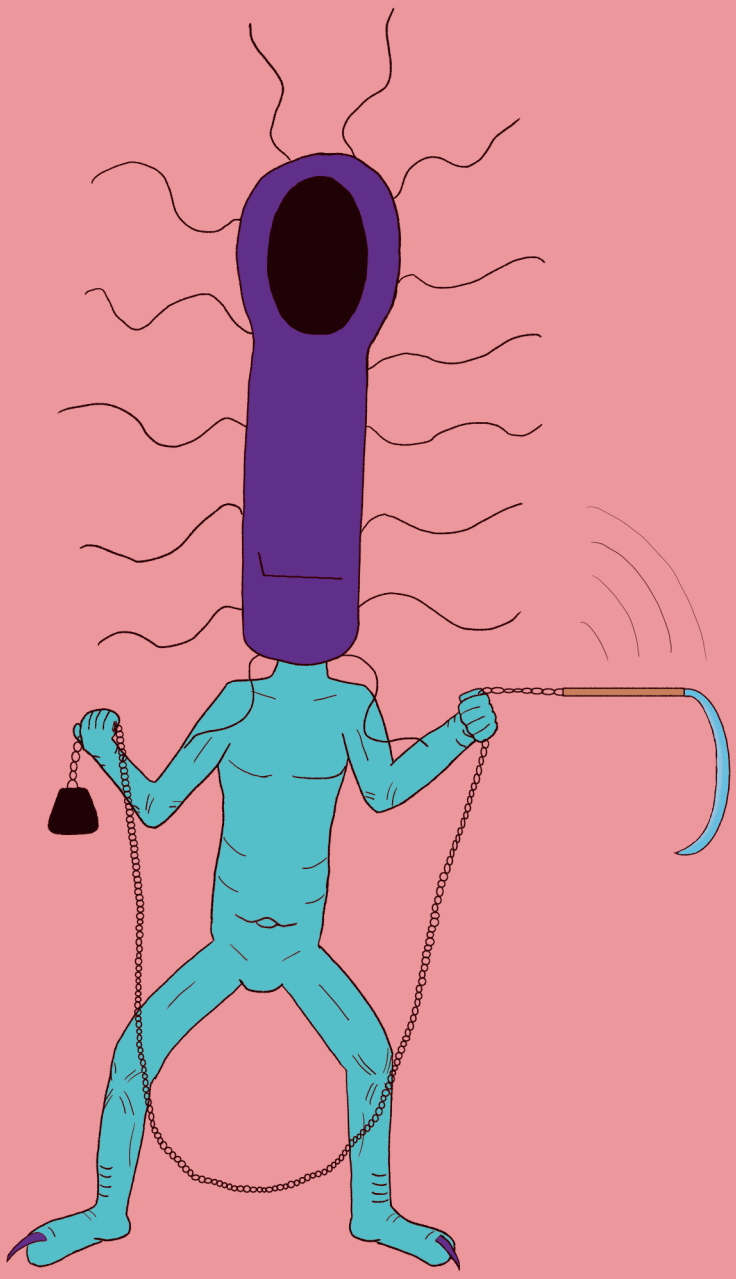
Our Italian colleagues mention that engraftment of donor microbiota is low after a single fecal infusion and disappears gradually over time after FMT. However, Goloshchapov et al. found significant long-term changes of the gut microbiota of healthy people one year after FMT, accompanied with transient changes of systemic immune parameters (5). Similarly, the study of Varun Aggarwala et al. found stable engraftment of 71% of donor microbiota strains in recipients up to 5 years post-FMT (6). Nonetheless, it remains unknown whether long-term bacterial engraftment is important for clinical success, and of clinical significance. Notably, engraftment was not correlated to the clinical success

in studies of five different FMT-treated illnesses (rCDI, ulcerative colitis, Crohn's disease, metabolic syndrome, infection with extended-spectrum betalactamase producing bacteria) (7). In addition, the statement that the identification of specific donor microbiome signatures before FMT can predict outcome, neglects the importance of the recipient microbiota. This was clearly demonstrated by Schmidt and colleagues in a recent meta-analysis of 142 FMTs that showed that recipient factors consistently outweighed donor factors in driving FMT outcomes (7). Besides, it remains difficult to define a 'healthy' microbiome without a thorough understanding of all its constituents, including also Archaea, viruses and fungi and their function in the gut ecosystem.

Long-term adverse events definitively related to FMT have not been reported and seem rare (8), but long-term follow-up with microbiome analyses, immunological parameters and clinical data is needed to recognize persistent engraftment and late adverse events. This is of particular importance for younger patient populations with non-CDI FMT-treated disorders. A recently completed survey in Europe demonstrated that of 31 FMT centers from 17 countries, 42% of all FMTs were administered for experimental indications other than rCDI (9). This illustrates the urgent need for an international registry to collect information from both donors and FMT-treated patients with follow-up of at least 10 years, and storage of stool samples from patients and donors by FMT centers. Such a registry is active in the US (American Gastroenterological Association) and is currently initiated in Europe by the European Working Group. We think this is an essential next step in both rationalized donor selection and establishing the safety of FMT also on the long term and enables action on previously unforeseen potential adverse events.

References

1. Budree S, et al. Do Specific Bacteria Drive Clinical Cure in Fecal Microbiota Transplantation for Clostridium Difficile Infection?: Clinical, Microbial and Metabolomic Characterization of Universal FMT Donors. *Gastroenterology* 2017; 152: S349.
2. Cook L, et al. Fecal Microbiota Transplantation for Recurrent Clostridioides difficile Infection Enhances Adaptive Immunity to C difficile Toxin B. *Gastroenterology*. 2021; 160(6): 2155-2158.e4.
3. Baruch EN, et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. *Science*. 2021; 371(6529): 602-609.
4. Zhang S, et al. Donor Screening for Fecal Microbiota Transplantation in China: Evaluation of 8483 Candidates. *Gastroenterology*. 2021; S0016-5085(21)03726-4.
5. Goloshchapov OV, et al. Long-term impact of fecal transplantation in healthy volunteers. *BMC Microbiology*. 2019; 19(1):312.
6. Aggarwala V, et al. Precise quantification of bacterial strains after fecal microbiota transplantation delineates long-term engraftment and explains outcomes. *Nature Microbiology*. 2021; 6(10):1309-1318.
7. Schmidt SB, et al. Drivers and Determinants of Strain Dynamics Following Faecal Microbiota Transplantation. *biorXiv*. 2021. Doi: c.
8. Ooijevaar RE, et al. Ten-Year Follow-Up of Patients Treated with Fecal Microbiota Transplantation for Recurrent Clostridioides difficile Infection from a Randomized Controlled Trial and Review of the Literature. *Microorganisms*. 2021; 9(3):548.
9. Baunwall SMD, et al. The use of Faecal Microbiota Transplantation (FMT) in Europe: A Europe-wide survey. *The Lancet Regional Health Europe*. 2021; 9:100181.



Chapter 3

Long-term beneficial effect of faecal microbiota transplantation on colonisation of multidrug-resistant bacteria and resistome abundance in patients with recurrent *Clostridioides difficile* infection

Sam Nooij[1,2], Karuna E. W. Vendrik[1,2]*, Romy D. Zwittink[2]*, Quinten R. Ducarmon[2], Josbert J. Keller[1,3], Ed J. Kuijper[2], & Elisabeth M. Terveer[1,2], on behalf of the Netherlands Donor Feces Bank study group

[1] Netherlands Donor Feces Bank, Leiden University Center of Infectious Diseases (LUCID) Medical Microbiology and Infection Prevention, Leiden University Medical Center, Leiden, Netherlands
[2] Center for Microbiome Analyses and Therapeutics, LUCID Research, Leiden University Medical Center, Leiden, Netherlands
[3] Haaglanden Medical Center, Department of Gastroenterology, The Hague, Netherlands
*: current affiliation – Centre for Infectious Disease Control, Netherlands Institute for Public Health and the Environment, Bilthoven, the Netherlands

Genome Medicine ***16***, *37 (28 February 2024) DOI: 10.1186/s13073-024-01306-7*

Abstract

Background

Multidrug-resistant (MDR) bacteria are a growing global threat, especially in healthcare facilities. Faecal microbiota transplantation (FMT) is an effective prevention strategy for recurrences of *Clostridioides difficile* infections and can also be useful for other microbiota-related diseases.

Methods

We study the effect of FMT in patients with multiple recurrent *C. difficile* infections on colonisation with MDR bacteria and antibiotic resistance genes (ARG) on the short (3 weeks) and long term (1-3 years), combining culture methods and faecal metagenomics.

Results

Based on MDR culture (n=87 patients), we notice a decrease of 11.5% in the colonisation rate of MDR bacteria after FMT (20/87 before FMT = 23%, 10/87 three weeks after FMT). Metagenomic sequencing of patient stool samples (n=63) shows a reduction in relative abundances of ARGs in faeces, while the number of different resistance genes in patients remained higher compared to stools of their corresponding healthy donors (n=11). Furthermore, plasmid predictions in metagenomic data indicate that patients harboured increased levels of resistance plasmids, which appear unaffected by FMT. In the long-term (n=22 patients), the recipients' resistomes are still donor-like, suggesting the effect of FMT may last for years.

Conclusions

Taken together, we hypothesise that FMT restores the gut microbiota to a composition that is closer to the composition of healthy donors, and potential pathogens are either lost or decreased to very low abundances. This process, however, does not end in the days following FMT. It may take months for the gut microbiome to re-establish a balanced state. Even though a reservoir of resistance genes remains, a notable part of which on plasmids, FMT decreases the total load of resistance genes.

Background

The discovery of antibiotics altered the natural course of infectious diseases and saved millions of lives. Antibiotics might be the most significant development in modern medicine, but there are important trade-offs to their use. Antibiotic resistant bacteria have emerged that are unaffected by standard therapies, which threatens effective prevention and treatment of infections. Antibiotic resistance is now considered a major threat to public health [1, 2]. Besides, broad spectrum antibiotic therapy disrupts the human microbiota, paradoxically resulting in an increased susceptibility to infections, for example by *Clostridioides difficile* [3-5].

C. difficile can asymptomatically reside in the gut but thrives in an antibiotic-affected microbiota. *C. difficile* causes an infection (CDI) varying from self-limiting and mild diarrhoea to life-threatening pseudomembranous colitis [6]. The disruption of the gut microbiota is essential in maintaining the recurrent nature of CDI, which is supported by the observation that replenishing the gut microbiota by faecal microbiota transplantation (FMT) results in prompt resolution of CDI recurrence (rCDI) [7, 8]. It is thought that FMT restores the gut microbiota diversity after antibiotic treatment, thus preventing outgrowth of *C. difficile* spores [9], and possibly decreasing the risk of other infections as well. FMT has been mentioned in treatment guidelines for rCDI for years [10-12], and rCDI is currently the only disease that is routinely treated with FMT.

A gut microbiota disrupted by antibiotics is also more susceptible to colonisation with multidrug-resistant (MDR) bacteria [13], which in turn increases the risk of infection in critically ill patients [14]. A prominent and problematic group of MDR bacteria are extended-spectrum beta-lactamase-producing (ESBL) Enterobacterales. Most infections with ESBL-producing Enterobacterales have high morbidity and mortality and are preceded by intestinal colonisation [15-17]. Hence, the prevention and eradication of ESBL-producing Enterobacterales from the intestinal tract is of global interest. Spontaneous decolonisation depends on comorbidities and type of species [18, 19], and innovative strategies to promote decolonisation of MDR bacteria are desired. So far, there is no recommended decolonisation method [20]. However, Millan et al., found that FMT in patients with rCDI decreased the number and diversity of antimicrobial resistance genes in their faeces [21]. This observation was followed by various case reports of patients colonised with ESBL-producing Enterobacterales who were successfully treated with FMT [22-31]. One underpowered randomised controlled trial (RCT) has been conducted (n=39 patients) to assess decolonisation of MDR Enterobacterales by treatment with oral non-absorbable antibiotics and FMT [32]. No statistically significant advantage of FMT was found, although colonisation rates were slightly lower in FMT-treated patients compared to untreated control patients. Subsequently, questions were raised about the efficacy of FMT against MDR bacteria and experiments were suggested to further assess this [33]. Interestingly, data from

another RCT using FMT for the decolonisation of MDR bacteria in renal transplant patients indicated that FMT-treated patients had longer time to recurrent infections than patients that did not receive FMT [34]. This underscores the need for longer term sampling in similar FMT studies.

To further explore the effects of FMT in rCDI patients on antibiotic resistance of the gut microbiota, we assess colonisation with MDR bacteria with both culture and faecal metagenomics. We pay special attention to the resistome, defined as the collection of all antibiotic resistance genes (ARG) present. Additionally, we study the long-term effects on the microbiota up to three years after FMT in a subset of patients.

Methods

Study design

In this cohort study, we use stool samples of rCDI patients treated in 34 different healthcare centres across the Netherlands with FMT provided by the Netherlands Donor Feces Bank (NDFB, Leiden, the Netherlands) to assess the presence of MDR bacteria and the resistome. The NDFB uses standardised procedures for the collection, screening, preparation and storage of donor faecal suspensions, and treatment and follow-up of rCDI patients as described previously [35, 36]. In short, patients were first treated with antibiotics against *C. difficile* for at least four days until 24 hours before FMT. The day before FMT, patients received a bowel lavage with macrogol solution [8]. Pre-FMT samples were collected during or shortly after antibiotic treatment and before bowel lavage. Approximately three weeks after FMT a short-term post-FMT sample was requested. Pre- and short-term post-FMT stool samples of rCDI patients and their corresponding donors were collected between May 2016 and March 2021. Additionally, in February 2021 we approached FMT-treated patients of the cohort with informed consent to contact them for later research purposes (n=53) for updated clinical information and requested a long-term follow-up (LTFU) stool sample. Clinical data, including recurrence of CDI after FMT, were recorded for further investigation. Stool samples were stored at -80°C until DNA extraction for metagenomics sequencing or stored in an end concentration of 10% glycerol until MDR culture testing.

Definition of multidrug-resistant bacteria

Definitions and testing methods were used as described previously [37]. MDR bacteria were defined according to the definitions of the Dutch Working Group on Infection Prevention [38]. This includes ESBL-producing Enterobacterales; Enterobacterales and *Acinetobacter* spp. that are resistant to both fluoroquinolones and an aminoglycoside or produce carbapenemases; *Pseudomonas aeruginosa* that produces carbapenemase or is resistant to at least three of the following antibiotic classes or agents: fluoroquinolones, aminoglycosides, ceftazidime or piperacillin, and carbapenems; both penicillin and

vancomycin-resistant *Enterococcus faecium* (VRE); or methicillin-resistant *Staphylococcus aureus* (MRSA).

Culture and antimicrobial susceptibility testing of multidrug-resistant bacteria

To identify MDR bacteria and calculate the prevalence among FMT donors and recipients, stool samples were selectively cultured as described previously [37]. Briefly, an inoculating loop was used to scrape 10 µL faeces from frozen faeces aliquots (containing 10% glycerol). The faeces was enriched in 15 mL of tryptic soy broth and incubated for 18h at 35°C prior to plating on ChromID ESBL, ChromID OXA-48 agar, MacConkey tobramycin (8 mg/L) plus ciprofloxacin (0.5 mg/L) agar and VRE agar (bioMérieux, Marcy l'Etoile, France). For MRSA detection a separate brain heart infusion enrichment broth was used which was supplemented with 2.5 sodium chloride and 10 mg/L colistin sulphate and inoculation on MRSA-ID agar plate. All suspected MDR colonies were identified as bacterial species by matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF) Biotyper (Bruker Daltonik; Bremen, Germany). Antibiotic susceptibility was evaluated by VITEK2 (Card N199, bioMérieux) using the European Committee of Antimicrobial Susceptibility Testing (EUCAST) breakpoints version 11.0 [39]. ESBL production was confirmed using the double disk method. Isolates with a meropenem minimum inhibitory concentration > 0.25 mg/L (ETEST, bioMérieux) were investigated for carbapenemase production with a carbapenem inactivation method (CIM) test and an in-house multiplex PCR to detect *KPC*, *VIM*, *NDM*, *OXA-48* and *IMP* genes. VRE were confirmed by an in-house PCR targeting the *vanA* and *vanB* genes, and MRSA with the BD MAX assay targeting the *MREJ, mecA/mecC* and *Nuc* genes (BD, New Jersey, USA). Six known MDR bacteria-positive and seven MDR bacteria-negative defrosted faeces aliquots (also stored in 10% glycerol) of the NDFB donor screening served as positive and negative controls. Samples were called MDR culture positive if at least one MDR bacterium was cultured on selective media.

Whole-genome sequencing of multidrug-resistant isolates

To assess the antibiotic resistance genotype of MDR isolates and persistence after FMT, cultured MDR bacteria were subjected to whole-genome sequencing (WGS; Figure 1). DNA was isolated using the QIAsymphony DSP Virus/Pathogen Midi Kit (Qiagen, Hilden, Germany) and sent to GenomeScan B.V. (Leiden, Netherlands) to sequence on the Illumina NovaSeq6000 platform (Illumina, Inc., San Diego, California, USA) generating 150 bp paired-end reads. (Reads per bacterial isolate: 780k [258k-1.64M] (median [range]).) Samples were sent in two batches, of which the second failed. We decided to continue with the available data, which includes WGS for 24 out of 32 isolates (15 / 20 from pre-FMT stools, 9 / 10 from short-term post-FMT and 0 / 2 long-term follow-up). The raw sequencing reads were cleared of human-derived reads by mapping to the GRCh38 genome [40] using bowtie2 (version 2.4.2, option '--very-sensitive-local')

[41] and samtools (version 1.11) [42] before adapter and low-complexity read removal and quality-trimming using fastp (version 0.20.1, parameters '--cut_right --cut_window_size 4 --cut_mean_quality 20 -l 50 --detect_adapter_for_pe -y') [43]. High-quality reads were assembled using SPAdes (version 3.15.2, option '--isolate') [44]. All scaffolds were screened for antibiotic resistance genes using ABRicate (version 0.8.13, https://github.com/tseemann/abricate) with both the CARD (from 25 March 2021) [45] and ResFinder (from 25 March 2021) [46] databases, only retaining hits of full-length genes (100% coverage) with at least 97% identity. These cut-offs were used to keep the method consistent with and comparable to the resistome analyses (see below). Furthermore, assembled genomes were taxonomically classified using GTDB-Tk (version 2.1.0) [47]. These classifications were used to verify or further specify classifications made by MALDI-TOF Biotyper as described above and are used as species identification for sequenced isolates. Sequence data have been deposited in the European Nucleotide Archive (ENA) under project number PRJEB64622 (https://www.ebi.ac.uk/ena/browser/view/PRJEB64622) [48].

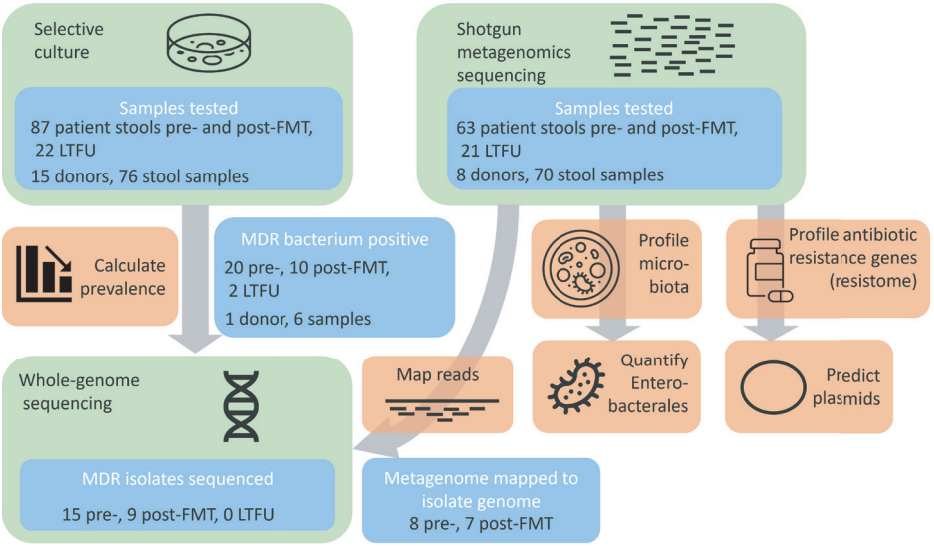


Figure 1. Schematic representation of the study setup. Data sources are shown in blue, data generating (wet lab) techniques in green and major analysis (dry lab) methods in orange boxes. Multiple recurrent *Clostridioides difficile* infected patients were treated with faecal microbiota transplantation in 34 different centres across the Netherlands and samples were requested for research. Only patients are included in the analyses if we received both a pre- and post-FMT sample.

Shotgun metagenomic sequencing

In total, 63 sets of donor-patient FMT triads were sequenced using shotgun metagenomics. Samples collected before 2021 were prepared for sequencing as previously described [49]. This resulted in metagenomes of 49 patients pre- and short-term post-FMT and 56

donor samples of 8 donors that have been deposited in the ENA under project number PRJEB44737 (https://www.ebi.ac.uk/ena/browser/view/PRJEB44737) [50]. An additional 21 sets (tetrads) of patient pre-, short-term post-FMT and now including long-term post-FMT samples, of which 7 were sequenced earlier, as well as 14 donor samples from 8 donors were sequenced at GenomeScan B.V. (Leiden, Netherlands) using the Illumina NovaSeq6000 platform generating a median of 42.6M 150bp paired-end reads per sample. Raw reads, excluding human-derived reads (see below), have been deposited in the ENA under project number PRJEB64621 (https://www.ebi.ac.uk/ena/browser/view/PRJEB64621) [51]. DNA was extracted from 100 mg of unprocessed patient and donor faeces using the Quick-DNA Fecal/Soil Microbe Miniprep Kit (ZymoResearch, Irvine, California, USA), with bead beating step on a Precellys 24 tissue homogeniser (Bertin Technologies, Montigny-le-Bretonneux, France) at 5.5 m/s for three times 1 min with short intervals, as described previously [52]. Libraries were constructed using the NEBNext Ultra II FS DNA kit and NEBNext Ultra II Ligation kit (New England Biolabs, Ipswich, Massachusetts, USA), producing DNA fragments of approximately 500-700 bp. Besides, control samples were included to verify successful DNA isolation and sequencing. These include blank (water) controls, and ZymoBiomics Community Standard (ZymoResearch). Negative controls returned no sequencing reads, while positive controls contained reads of all species present in the communities.

Metagenomic pre-processing

Human-derived reads were removed from raw metagenomic reads by mapping reads to the human reference genome (GRCh38, NCBI accession ID GCF_000001405.26) using bowtie2 (version 2.4.2, option '--very-sensitive-local') and samtools (version 1.11). Remaining non-human reads were then processed by fastp (version 0.20.1) to trim low-quality 3'-ends (parameters: '--cut_right --cut_window_size 4 --cut_mean_quality 20'), remove low-complexity sequences (parameter: '-y'), remove remaining adapter sequences (parameter: '--detect_adapter_for_pe') and remove reads shorter than 50 bases (parameter: '-l 50'). The resulting high-quality metagenomic reads were used in read-based taxonomic profiling and assembly-based ARG profiling.

Quantification of multidrug-resistant isolates in metagenomes

To identify and quantify whole-genome sequenced MDR bacteria in metagenomes, we mapped metagenomic reads derived from the same stool sample to the respective assembled genome using BWA-MEM (version 0.7.17) [53]. Mapped reads were counted and coverage was quantified using samtools coverage (version 1.10). Coverage was calculated as both depth of coverage (mean number of times each position is covered, normalised by the total number of metagenomic reads) and breadth of coverage (percentage of genome covered by at least one read). In all cases, presence of the MDR strain was confirmed by coverage of scaffolds containing ARGs related to the MDR phenotype. Furthermore, presence of antibiotic resistance genes detected in the whole-genome sequence data of each cultured isolate was manually compared

against the resistome data derived from the same stool sample to assess sensitivity of culture and metagenomics.

Taxonomic profiling

Taxonomic microbiota profiles were determined using MetaPhlAn (version 4.0.3) [54], which maps reads to its custom marker database. Resulting taxonomic profiles quantified as percentages of the total microbiota were imported as R phyloseq object to facilitate visualisation and statistical comparisons [55]. To quantify Enterobacterales, we extracted the order of Enterobacterales from the MetaPhlAn output and labelled all other taxa 'other'. Presence of Enterobacterales was defined as a relative abundance > 0%. Species richness and evenness were calculated using the R package 'microbiome', while Shannon diversity was calculated with the 'vegan' package.

Resistome analysis

ARGs were detected using an assembly-based approach. Quality-trimmed reads were assembled into scaffolds using metaSPAdes (version 3.15.4, default parameters) [56]. Next, resistance genes were identified with ABRicate (version 0.8.13) using both the CARD (from 25 March 2021) and ResFinder (from 25 March 2021) databases, only retaining hits of full-length genes (100% coverage) with at least 97% identity. These criteria were selected based on visual inspection of the BLAST hits to balance high specificity and adequate sensitivity. As a control, we repeated the analyses using a coverage cut-off of 50% to include partial genes, which yielded equivalent results. ARGs were annotated with their respective target antibiotic and antibiotic class using the respective databases' annotation files. Scaffolds were quantified by mapping the metagenomic reads back to the scaffolds using BWA-MEM (version 0.7.17) and samtools (version 1.10). Quantifications were normalised to reads per kilobase per million (RPKM) by dividing the number of reads mapped to each contig by the length of the contig and the number of high-quality reads used for the assembly, multiplied by 1,000 * 1,000,000. To annotate scaffolds with additional information, scaffolds were taxonomically classified using the Genome Taxonomy Database Toolkit (GTDB-Tk; version 2.1.0) and the Contig Annotation Tool (CAT, version 5.2.3, parameters: '-r 10 -f 0.5', [57] – which uses Prodigal version 2.6.3 [58]; DIAMOND version 2.0.6 [59]; and the NCBI BLAST nr database from 7 January 2021, https://ftp.ncbi.nlm.nih.gov/blast/db/), using CAT as primary annotation and filling in gaps in classification using the result of GTDB-Tk. The genomic origin of scaffolds with ARGs (chromosome or plasmid) was predicted using viralVerify (version 1.1, option '-p', https://github.com/ablab/viralVerify) and we used only predictions that viralVerify reported as certain. All the scaffold annotation data was loaded into R (version 4.0.2; https://www.R-project.org/) for further analyses. Resistome richness was calculated by counting the number of different genes per sample, total abundance was calculated as the sum of all resistance genes' abundance values (as RPKM) per sample, resistome Shannon diversity was calculated using the 'vegan' R package and Simpson evenness with the 'microbiome' package.

Statistical analyses

The colonisation rate of MDR bacteria among patients was compared between pre- and short-term post-FMT and short-term and the long-term post-FMT using McNemar's chi-square test for paired data (non-exact, without continuity correction). Depth of coverage of MDR bacteria in metagenomic data was compared between pre- and post-FMT with a paired t-test on log-transformed coverage values.

The effect of FMT on the colonisation rate of Enterobacterales in patients was tested using McNemar's test without correction. Total abundances were compared using repeated measures ANOVA, followed by pairwise t-tests using Holm's correction method.

For comparing taxonomic compositions of metagenomes and resistomes between donors and patients, we selected one value for each donor as representative. For principal component analyses (PCA), we picked the middle sample for each donor based on donation date (number of samples / 2, rounded up). Aitchison distance was used to calculate distances between microbiota or resistome compositions. Aitchison uses log-transformed values, which is impossible with zero, so we added pseudocounts. In PCA, donors and patients are compared using PERMANOVA and PERMDISP tests, considering the repeated measures in patients by using their ID as strata. Aitchison distances are compared using Wilcoxon rank sum tests. For comparisons of alpha diversity metrics using boxplots, we selected the median value as representative for each donor. Richness, total abundance, Shannon index, and Simpson evenness are compared between donors and patients using t-tests with Holm's correction method. Abundance values were log-transformed. Within patients, all pre- and short-term post-FMT measures are compared using a paired t-test, while within the subgroup of 22 patients of whom we have collected long-term post-FMT samples values are first compared using repeated measures ANOVA. If $p < 0.05$, paired t-tests were used as post-hoc test to determine differences between pre-FMT and long-term post-FMT and between short and long-term post-FMT. Again, Holm's correction method was used.

To evaluate if antibiotic (vancomycin) treatment duration before FMT influenced the resistome, we compared the pre-treatment duration of patients (n=52) with their resistome richness (number of different ARGs), total abundance, Shannon diversity and Simpson evenness using Spearman correlation.All statistical tests were done in R version 4.0.2, using the base, rstatix, vegan, and pairwiseAdonis packages. A p-value below 0.05 was considered significant. Analysis scripts are available at Zenodo (https://zenodo.org/doi/10.5281/zenodo.10276220) [60].

Results

Donor and patient selection and population characteristics

During the sample collection period the NDFB provided faecal suspensions for 208 FMT treatments of 187 rCDI patients. From 87 patients (median age: 73, interquartile range (IQR): 64-81 years, 56 females (64%)) we obtained stool samples from both pre- and short-term post-FMT to test for MDR bacteria by culture (Figure 1). Twenty-two patients (median age: 73, IQR 64-78 years; 14 females (64%)) provided a long-term post-FMT sample that was culture-tested. The median sampling times for patients are: 1 day pre-FMT (IQR 1-3 days), 27 days post-FMT (IQR 20-48 days; short-term), and 801 days post-FMT (IQR 447-1114 days; long-term). Seventy-six donor samples from 15 different donors (median age: 27, IQR: 24-37.5; 9 females (60%)) were screened for MDR bacteria by culture. For shotgun metagenomic deep sequencing, we used 63 pairs of patient stool samples (patient median age: 73 years, interquartile range (IQR) 65-81 years; 40 females (63%)), 21 LTFU (one failed to provide a pre-FMT sample), and 70 donor stool samples from 11 different donors (median age: 31 years, IQR 27-42 years; 6 females (55%); Figure 1). The resistome analysis includes only complete sample triads (donor, pre-FMT and post-FMT), and sample tetrads with long-term post-FMT if both pre- and short-term post-FMT samples were available.

Prevalence of multidrug-resistant bacteria decreases after FMT

We began our study of the effect of FMT on MDR bacteria with selective cultures. Stool sample cultures of 15 donors and 87 patients pre- and post-FMT were assessed for carriership of MDR bacteria (Figure 1). One donor had MDR bacterium positive samples (1/15 = 6.7%), of which none were used for FMT. At least one MDR bacterium was detected in 20/87 (23.0%) of the patients before FMT (Figure 2A, Table 1). Three weeks after FMT, the colonisation rate decreased to 10/87 (11.5%; p = 0.0075), of which 7 MDR bacteria were also detected before FMT. In the long-term, the colonisation remained similar at 2/22 (9.1%; figure 2B; p = 0.16 compared to short-term post-FMT). Both MDR bacteria present in the LTFU were ESBL-producing *E. coli* also detected in the short-term post-FMT samples. Thereby they appear to be long-term persisters. Within the subgroup of patients that provided long-term samples, there was no decrease in colonisation after FMT as we saw with the whole cohort (Figure 2B; pre-FMT 5/22 = 22.7%, post-FMT 4/22 = 18.2%; p = 0.56). We compared data of MDR bacterial colonisation with CDI recurrence for a comprehensive analysis, but found that the numbers were too small to provide statistically meaningful results.

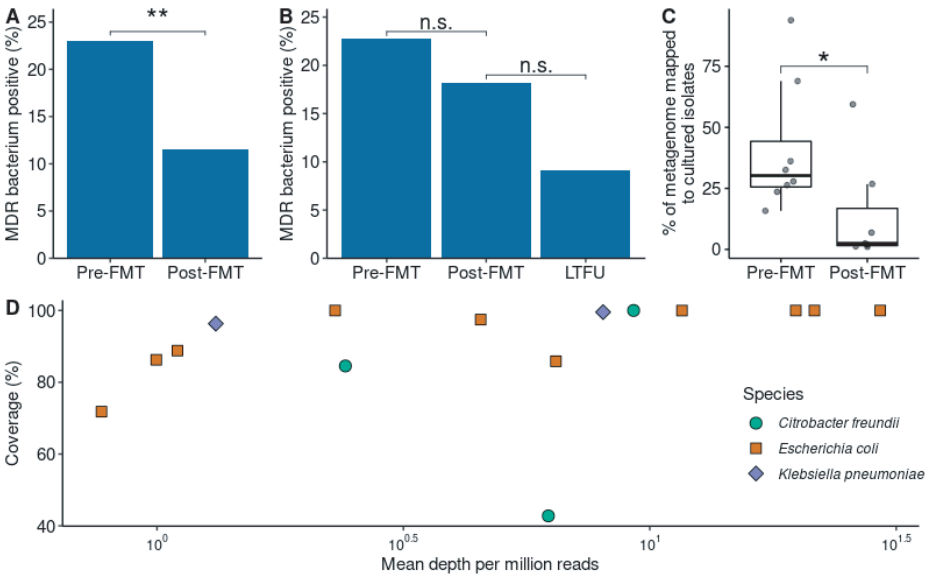


Figure 2. Effect of faecal microbiota transplantation on prevalence and abundance of cultured multidrug resistant bacteria. Stool samples of recurrent *C. difficile* infected (rCDI) patients were selectively cultured to assess the prevalence of multidrug resistant (MDR) bacteria before and after faecal microbiota transplantation (FMT). We called samples MDR positive if at least one MDR bacterium was detected. Cultured isolates were subjected to whole-genome sequencing, and metagenomic sequencing data from the same stool samples were mapped to the assembled genomes to quantify the MDR bacteria in the metagenomes. A) Prevalence of MDR bacteria in 87 rCDI patients. B) Colonisation rates in 22 patients of whom long-term follow-up (~1-3 years after FMT) samples were collected. C) Abundance of MDR bacteria based on metagenome data. D) Breadth of coverage and relative abundance of MDR bacteria in metagenomic sequencing data per species.

Asterisks indicate statistically significant differences, *: $p < 0.05$; **: $p < 0.01$; n.s.: not significant. MDR: multidrug resistant, FMT: faecal microbiota transplantation, LTFU: long-term follow-up.

Table 1. Overview of cultured multidrug-resistant bacteria with genotype and phenotype.

Patient	Sample timepoint	Species	Resistance phenotype	Genotype based on WGS	Detected in metagenome
P22	Post-FMT	*E. coli*	Aminoglycoside, fluoroquinolone, ampC	*APH(3')-Ia, APH(6)-Id, APH(3")-Ib, ANT(2")-Ia, acrD, ampC, QnrB5, emrR*	Yes
P30	Post-FMT	*E. coli*	Aminoglycoside, fluoroquinolone	*acrD, emrR, emrD*	Yes
P31	Pre-FMT	*E. coli*	Aminoglycoside, fluoroquinolone, ceftazidime	*acdD, emrR, emrB, ampC*	Yes
P33	Pre-FMT	*K. pneumoniae**	Aminoglycoside, fluoroquinolone, ESBL	*aadA2, aadA16, AAC(3-IId, TEM-1, SHV-119, CTX-M-14*	Yes
P33	Post-FMT	*K. pneumoniae**	Aminoglycoside, fluoroquinolone, ESBL	*aadA2, aadA16, AAC(3-IId, TEM-1, SHV-119, CTX-M-14*	Yes
P38	Pre-FMT	*E. coli**	Fluoroquinolone, ESBL	*CTX-M-27, ermR, emrB*	Yes
P38	Post-FMT	*E. coli**	Fluoroquinolone, ESBL	*CTX-M-27, ermR, emrB*	Yes
P39	Pre-FMT	*E. coli*	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-15, OXA-1, acrD, AAC(3)-IIe, emrA, emrB, emrR*	Yes
P44	Pre-FMT	*C. freundii**	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-15, OXA-1, AAC(3)-IIe, AAC(6')-Ib-cr, APH(6)-Id, APH(3")-Ib, QnrB6*	Yes
P44	Post-FMT	*C. freundii** + *E. coli*	Aminoglycoside, fluoroquinolone, ESBL + ESBL	*CTX-M-15, OXA-1, AAC(3)-IIe, AAC(6')-Ib-cr, APH(6)-Id, APH(3")-Ib, QnrB17*	Yes
P44	LTFU (3yr)	*E. coli*	ESBL	NA	NA
P51	Pre-FMT	*C. freundii*	ESBL	*CTX-M-9*	Yes
P58	Pre-FMT	*E. coli*	Aminoglycoside, fluoroquinolone	*acrD, emrR*	Yes

Table 1. Overview of cultured multidrug-resistant bacteria with genotype and phenotype. *(continued)*

Patient	Sample timepoint	Species	Resistance phenotype	Genotype based on WGS	Detected in metagenome
P59	Pre-FMT	*E. coli**	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-14, acrD, AAC(3)-IIe, APH(3")-Ib, APH(6)-Id, emrR*	Yes
P59	Post-FMT	*E. coli**	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-14, acrD, AAC(3)-IIe, APH(3")-Ib, APH(6)-Id, emrR*	Yes
P59	LTFU (1yr)	*E. coli**	Aminoglycoside, fluoroquinolone, ESBL	NA	NA
P64	Pre-FMT	*E. coli**	Aminoglycoside, fluoroquinolone	*AAC(6')-Ib-cr, emrA, emrB, emrR*	NA
P64	Post-FMT	*E. coli**	Aminoglycoside, fluoroquinolone	*acrD, APH(3")-Ib, APH(6)-Id, ampC, ampH, emrA, emrB, emrR*	NA
P65	Pre-FMT	*E. hormaechei_A (cloacae)*	Aminoglycoside, fluoroquinolone, ESBL	*ACT-27, CTX-M-15, OXA-1, TEM-1, AAC(3)-IIe, APH(6)-Id, APH(3")-Ib, AAC(6')-Ib-cr, QnrB6*	NA
P66	Pre-FMT	*M. morganii*	ESBL	*DHA-18*	NA
P67	Pre-FMT	*P. mirabilis*	ESBL	*CTX-M-1*	NA
P68	Pre-FMT	*P. mirabilis_B (vulgaris/mirabilis)*	ESBL	(none)**	NA
P69	Pre-FMT	*C. freundii**	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-15, TEM-1, OXA-1, AAC(3)-IIe, APH(3")-Ib, APH(6')-Id, QnrB6*	NA
P69	Post-FMT	*C. freundii**	Aminoglycoside, fluoroquinolone, ESBL	*CTX-M-15, TEM-1, OXA-1, AAC(3)-IIe, APH(3")-Ib, APH(6')-Id, QnrB17*	NA
P70	Pre-FMT	*E. coli**	ESBL	*ampC, ampH, SHV-134*	NA

Table 1. Overview of cultured multidrug-resistant bacteria with genotype and phenotype. *(continued)*

Patient	Sample timepoint	Species	Resistance phenotype	Genotype based on WGS	Detected in metagenome
P70	Post-FMT	*E. coli**	ESBL	*ampC, ampH, SHV-134*	NA
P71	Pre-FMT	*K. pneumoniae*	ESBL	NA	NA
P72	Pre-FMT	*P. hauseri*	ESBL	NA	NA
P73	Pre-FMT	*C. freundii*	ESBL	NA	NA
P74	Pre-FMT	*E. cloacae*	ESBL	NA	NA
P75	Pre-FMT	*E. cloacae**	ESBL	NA	NA
P75	Post-FMT	*E. cloacae**	ESBL	NA	NA

*: same species before and after FMT, persistence is likely based on resistance genotype (Additional file 1: Fig. S1-3) when available. **: no antibiotic resistance genes were detected in the genome sequence data. Species names are listed as in the Genome Taxonomy Database (GTDB), and the alias known by the National Center for Biotechnology Information (NCBI) is given in parentheses when different. When multiple multidrug-resistant bacteria were cultured from the same stool, isolate characteristics are separated by a plus ('+') sign.

FMT: faecal microbiota transplantation, LTFU: long-term follow-up, ESBL: extended-spectrum beta-lactamase, NA: data not available (because the isolate and/or the metagenome were not sequenced), WGS: whole-genome sequencing

Whole-genome sequencing of multidrug-resistant and comparison with metagenomics reveals that MDR bacteria had higher abundances in rCDI patients before FMT than after FMT

Next, we checked the resistance genotype of isolates using WGS and combined the isolate data with faecal metagenomics to quantify the abundance of MDR bacteria in the gut microbiota. Twenty-four cultured isolates of multidrug-resistant bacteria were subjected to WGS. In all but one genome we were able to detect ARGs associated with the resistance phenotype; e.g., ESBL genes in isolates classified as ESBL-producing (Table 1; Additional file 1: Figures S1-3). Furthermore, we mapped metagenomic reads to the assembled isolate genome to compare essay sensitivity and determine relative abundances in the microbiota. As expected in patients pre-treated with antibiotics, we found that MDR bacteria had higher abundances in rCDI patients before FMT than after FMT (Figure 2C; $p = 0.0159$). We detected near-complete genomes of MDR isolates in the metagenomes with various depths (Figure 2D). Only one *Citrobacter freundii* genome was covered less than half (43%) in the metagenome. We then compared resistance genes detected in the WGS data to those detected in metagenomic data to estimate the sensitivity of metagenomic sequencing compared to culturing. These resistance genes of cultured isolates were also found in their respective metagenomes (Table 1). Besides, metagenomic data from patient P44 suggested the presence of an ESBL-producing *E. coli* in the pre-FMT sample, while culture only picked it up in the post-FMT faeces. These data suggest that combining bacterial culture with metagenomic sequencing can be used synergistically and provide more detailed results than either method alone. In summary, we find that both the prevalence and the abundance of MDR bacteria were decreased after FMT.

FMT makes gut microbiota more donor-like and decreases Enterobacterales, also in long-term

To gain deeper understanding in how FMT affected the gut microbiota in this cohort, we profiled faecal metagenomes of donors and recipients using MetaPhlAn4. Donors had microbiota dominated by Bacillota (formerly Firmicutes), Bacteroidota (formerly Bacteroidetes) and Actinomycetota (formerly Actinobacteria; Additional file 1: Figure S4A). Enterobacterales were present in 26/70 donor stools (37%) at ~0.01% abundance (Figure 3A-B). In rCDI patients, that underwent anti-CDI treatment prior to FMT (53 x vancomycin, 6 x fidaxomicin, 1 x metronidazole 1 x metronidazole+vancomycin, 2 unknown), Actinomycetota and Bacteroidota were much less present, while Proteobacteria (mostly *Escherichia coli* or *Klebsiella pneumoniae*) were often dominant (>50% abundance in 31/63 patients = 49%; Additional file 1: Figure S4B). Enterobacterales were present in all pre-FMT patient stools (Figure 3A). Shortly after FMT, the prevalence of Enterobacterales dropped to 58/63 (92%; $p = 0.0253$) and the abundance decreased as well (Figure 3B; adjusted $p < 0.0001$). In the longer-term after FMT, the prevalence of Enterobacterales did not change (18/21 = 86%; $p = 0.655$ compared to 3 weeks post-FMT), but the abundance decreased further (adjusted $p = 0.025$), and was no longer distinct from the donors' ($p = 0.09$).

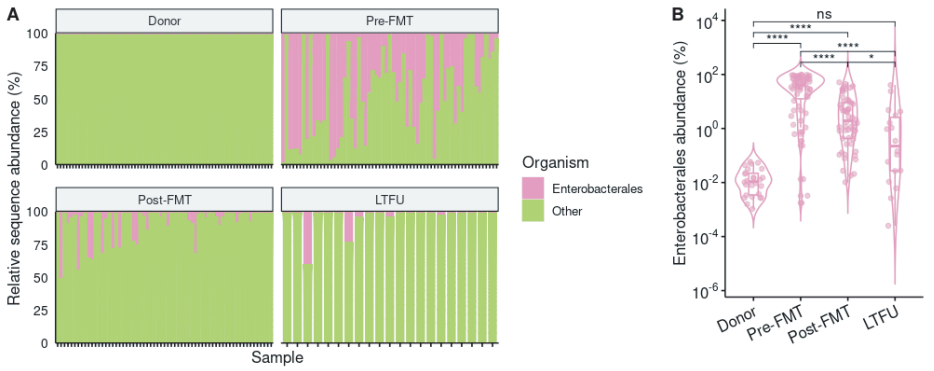


Figure 3. Prevalence and abundance of Enterobacterales in faecal donors and faecal microbiota transplantation recipients. A) Relative abundances of Enterobacterales in metagenomes as determined by MetaPhlAn4. B) Total abundances of Enterobacterales in stool donors and rCDI patients treated with FMT sampled one day before (Pre) FMT, three weeks after (Post) FMT and 1-3 years after FMT (long-term follow-up, LTFU).

Statistically significant differences are indicates by asterisks, *: $p < 0.05$; ****: $p < 0.0001$.FMT: faecal microbiota transplantation, LTFU: long-term follow-up.

We further studied the species composition of donor and patient metagenomes using PCA (Figure 4A). Patient species composition was different from the donors' at all timepoints (PERMANOVA: overall $p = 0.001$, pre-FMT $p = 0.003$, post-FMT $p = 0.014$, LTFU $p = 0.014$; PERMDISP: $p < 0.0001$). However, after FMT the patients' microbiota were more similar to their donors' (Figure 4B; $p < 0.0001$). This shift in microbiota appears at least partly attributable to the FMT, and is not solely due to recovery after antibiotic use, as indicated by a comparison between patient post-FMT species compositions to those of the donor that was used against unrelated donor samples (Additional file 1: Figure S5; $p < 0.0001$). In the long-term, species compositions in FMT recipients moved away from the donors' (post-FMT vs LTFU adjusted $p = 0.0008$), but was still more donor-like than before FMT (adjusted $p = 0.002$).

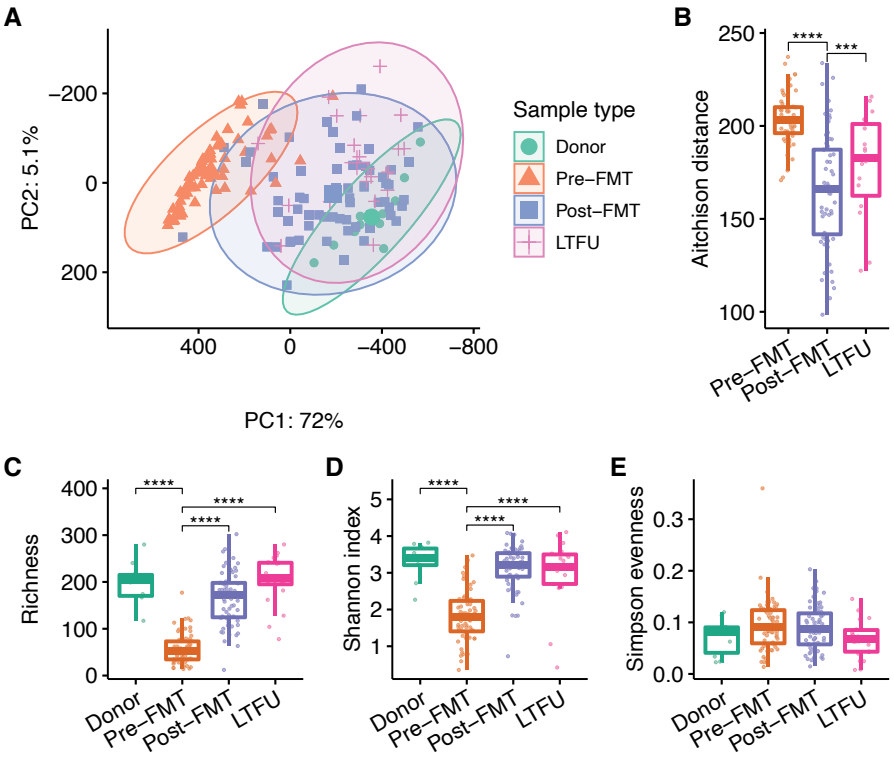


Figure 4. Comparison of gut microbiota composition and diversity. Species composition of metagenomes was determined by MetaPhlAn4. A) Beta diversity expressed as Aitchison distances in a principal component analysis (PCA). Percentages on the X- and Y-axis represent the variance explained by the first two components. B) Aitchison distance from patient species profile to corresponding donor. C-E) Species richness, Shannon index and Simpson evenness compared between donors and recipients, respectively.

Asterisks indicate statistically significant differences, ****: $p < 0.0001$ FMT: faecal microbiota transplantation, LTFU: long-term follow-up.

We then compared the alpha diversity between species profiles of donor and patient metagenomes. Species richness and Shannon diversity were higher in donors than in rCDI patients before FMT (Figure 4C-D; adjusted $p < 0.0001$) and increased dramatically in patients after FMT (adjusted $p < 0.0001$) to levels as seen in donors (adjusted $p > 0.1$). Richness and Shannon index remained high at the long-term. The Simpson evenness, also known as inverse Simpson index or Simpson's dominance, was not different between donors and patients (Figure 4E; adjusted $p > 0.3$). Overall, our data show the expected pattern of lower diversity in rCDI patients, high diversity in FMT donors, and increased diversity in patients after FMT. After FMT, both the prevalence and abundance of Enterobacterales were decreased in patients.

FMT decreases abundance of resistance genes, but not their diversity

Using the same metagenomic sequencing data, we determined the resistome using a custom assembly-based approach. We quantified differences in resistome composition between donors and patients using PCA (Figure 5A). Donors had similar resistomes and often had the same ARGs for aminoglycoside, diaminopyrimidine and tetracycline resistance (Additional file 1: Figure S6), while rCDI patients had a very different resistome (PERMANOVA, $p = 0.003$; PERMDISP, $p < 0.0001$), in which different ARGs for beta-lactam and fluoroquinolone resistance as well as multidrug efflux pumps were prevalent (Additional file 1: Figures S6-7). After FMT, a shift in the patients' resistome toward a donor-like composition is visible (Figure 5B; adjusted $p < 0.0001$), although it remained different from the donors' ($p = 0.003$). At long-term follow-up, the resistome was neither more nor less donor-like than at 3 weeks after FMT (Figure 5B; adjusted $p = 0.123$) and was still statistically different from the donors' ($p = 0.012$). These differences in resistome composition between donors and patients and the shift after FMT are also visible when viewing the resistome as relative abundances per antibiotic class (Additional file 1: Figure S8).

We find that patients before FMT had more different resistance genes (higher resistome richness) in their faecal metagenomes than donors (adjusted $p < 0.0001$; Figure 5C). The duration of vancomycin pre-treatment did not significantly influence the resistome (Additional file 1: Figure S9). After FMT, resistome richness in patients did not change (adjusted $p > 0.1$) and remained higher than in donors (short-term post-FMT: adjusted $p < 0.0001$; long-term: adjusted $p = 0.0002$). The total abundance of resistance genes was also higher in patients pre-FMT than in donors (Figure 5D; adjusted $p < 0.0001$), but in contrast to the resistome richness, abundance decreased in patients shortly after FMT ($p = 0.0003$). In the long term, the abundance lowered further (adjusted $p = 0.02$), although abundances remained higher than in donors (adjusted $p = 0.02$). The Shannon index combines richness and abundance and likewise showed a higher resistome diversity in rCDI patients compared to donors, and a decrease after FMT (Figure 5E). The Simpson evenness shows no statistical difference between donors and patients (Figure 5F; adjusted $p > 0.1$), but indicates a decrease of resistome diversity in patients after FMT ($p = 0.017$). In summary, FMT appears to alter the diversity of the resistome in recipients by lowering relative abundances of ARGs.

We observed different prevalence and abundance patterns of ARGs from different antibiotic classes (Additional file 1: Figures S6-8). To explore this further, we selected classes of which genes were present in both donors and patients and divided them in two groups. One group (beta-lactamase, fluoroquinolone, and multidrug efflux pump) consists of genes that are rare in donors and common and abundant in patients (Additional file 1: Figure S10 A-C and G-I). The abundance of genes in this group decreased shortly after FMT, while the resistome richness decreased only in the long-term. The second group (aminoglycoside, diaminopyrimidine and tetracycline) is

common in donors (Additional file 1: Figure S10 D-F and J-L). Genes from this donor-associated group may have been transferred to the recipients, resulting in greater resistome richness after FMT and their abundance did not decrease after FMT. These results highlight that the effects of FMT on the resistome vary depending on type of antibiotic and the taxa that carry the genes.

Remarkable resistances

We found a number of ESBL genes in our resistome data, and also in donor faeces. Furthermore, we found carbapenamase genes and one colistin resistance gene (mrc-10_1, predicted to be on a plasmid) only in rCDI patients before FMT. Vancomycin genes were detected by metagenomics in 7 out of 63 patients before FMT (11.1%) and 11 / 63 after FMT (17.5%; Additional file 1: Figure S11). Besides, our cultures picked up a vancomycin-resistant *Enterococcus faecalis* from a post-FMT stool, which was penicillin-susceptible and therefore not listed as MDRO. These resistances as well as those predicted to be on plasmids are discussed in more detail in the supplementary results and figures 11-13 (Additional file 2; Additional file 1: Figures S11-13).

Predicted plasmid-mediated antibiotic resistance remains high

We hypothesised that the relative persistence of the resistome after FMT may be linked to plasmids. To test this, we used the plasmid prediction algorithm from viralVerify and assessed which contigs with resistance genes were likely to derive from chromosomes and which from plasmids. Most (4,567 / 6,662 or 68.6%) of the resistance genes were predicted to derive from chromosomes and 400 (6%) likely derived from plasmids. The remaining 1,695 (25.4%) contigs with ARGs could not confidently be classified to either plasmid or chromosome. Unlike chromosomal resistances, which follow the general resistome pattern and decrease after FMT (Figure 6A-C), we find that the resistome richness, abundance, and diversity of ARGs derived from plasmids were higher in rCDI patients than in donors and stayed higher after FMT (adjusted $p \leq 0.01$; Figure 6D-F). This effect persisted up to three years after the FMT, suggesting that FMT may not significantly influence plasmid-mediated antibiotic resistance.

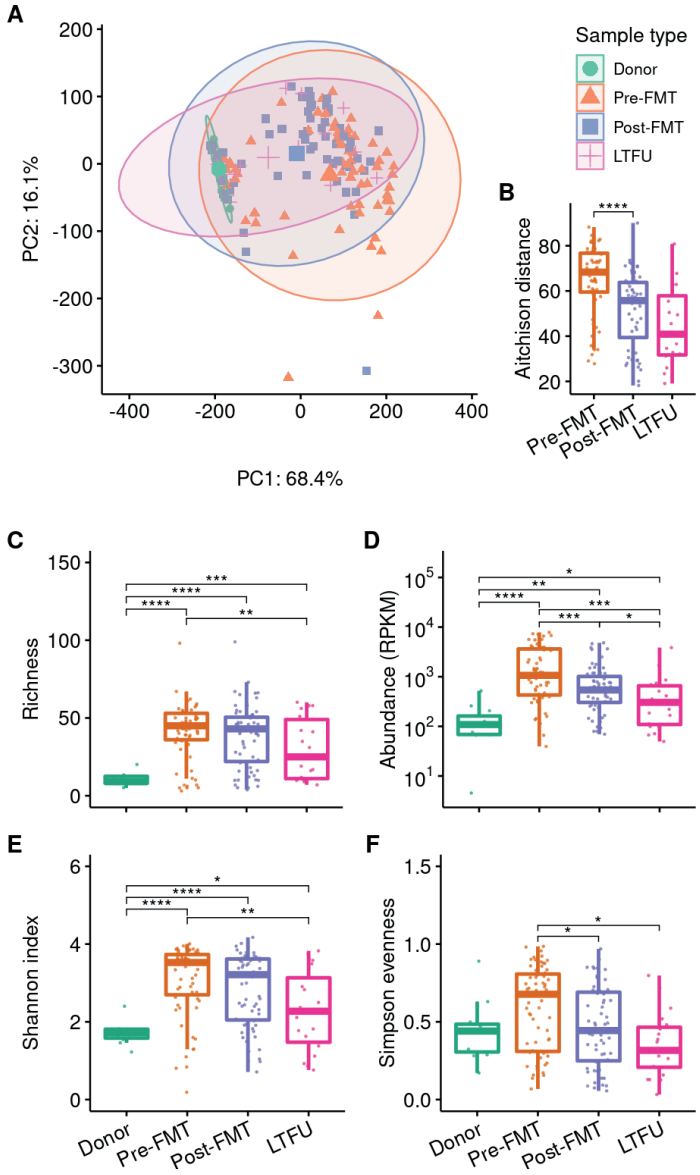


Figure 5. Overview of resistomes of faeces donors and faecal microbiota transplantation recipients. A) Principal component analysis (PCA) of resistomes, based on Aitchison distances. Percentages on the X- and Y-axis represent the variance explained by the first two components. B) Aitchison distance from patient antibiotic resistance gene profiles to corresponding donor. C-F) Antibiotic gene richness, total abundance, Shannon index and Simpson evenness compared between groups and between recipient timepoints, respectively. Asterisks indicate statistically significant differences, *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$. FMT: faecal microbiota transplantation, LTFU: long-term follow-up.

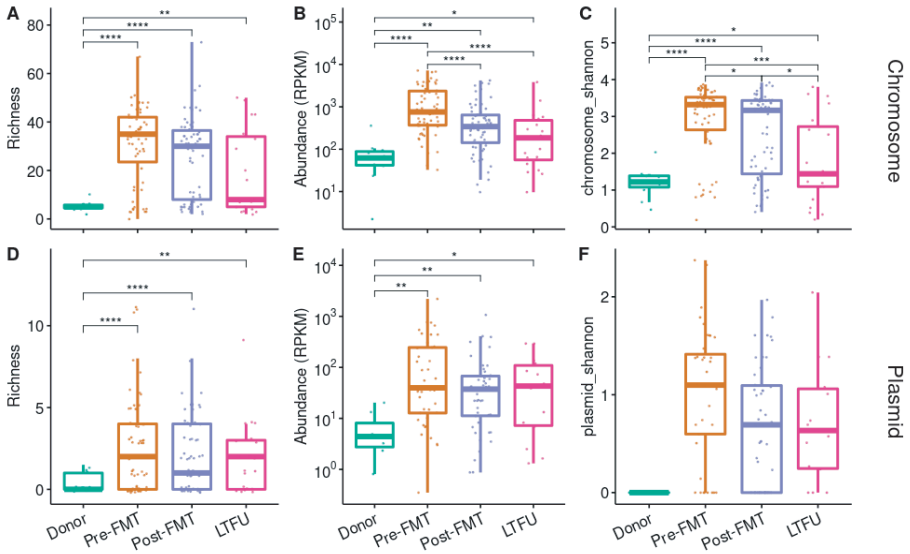


Figure 6. Resistome comparisons for chromosomal resistance genes and plasmid-associated resistance genes. Antibiotic gene-carrying scaffolds' were predicted to derive from chromosomes or plasmids using viralVerify. A) resistome richness, B) total abundance, and C) Shannon index of scaffolds predicted to be chromosomal. D-F) Same parameters for scaffolds predicted to derive from plasmids.

Statistically significant differences are indicates by asterisks, *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$. FMT: faecal microbiota transplantation, LTFU: long-term follow-up.

Discussion

Our current study leverages the strengths of bacterial culture techniques and metagenomic sequencing to provide a comprehensive view of antibiotic resistant bacteria in the intestinal tract in the weeks and years after FMT. We find that FMT decreased prevalence and abundance of MDR bacteria and led to more diverse and donor-like microbiota and ARG compositions. However, the resistome of patients stayed different from the donors' even in the 1–3-year follow-up. This study provides a unique insight into the long-term effects of FMT on the resistome of rCDI patients, although the follow-up is limited in the number of responders and variable time after FMT. Whether there is a correlation of persistence of antibiotic resistance with CDI recurrence could not be assessed given the limited number of relapses in our cohort.

For the interpretation of the resistome of recipients after FMT, we need to consider the characteristics of the recipients' resistome before transplant, and of the donor

resistome. Our patient cohort consists of mostly elderly people with significant comorbidity and disturbance of the gut microbiota by multiple courses of antibiotic treatment for recurrent *C. difficile* infections. This has been described to alter the gut resistome [61]. Contrastingly, the gut of healthy individuals harbours mainly anaerobic commensal bacteria, which frequently carry aminoglycoside and tetracycline resistance genes [62]. Our results indicate that these classes of antibiotic resistance were indeed common in stool donors, and also in rCDI patients. The observed shift in the recipients' resistome composition after FMT is likely a reflection of the introduction of the anaerobic commensal bacteria, resulting in different effects of FMT on various antibiotic classes [63]. Some ARGs and predicted plasmids persisted in the recipients after FMT and, therefore, the resistome composition remained different from the donors'.

While we have no data on the effect of FMT on the prevalence of infections with MDR bacteria post-FMT, in other patient groups it has been found that FMT can decrease the risk of infections [64, 65], or delay the development of MDR infections [34]. Assuming that MDR bacteria are not eradicated after FMT, there may still be a risk of infection when the intestinal environment becomes hospitable for outgrowth of MDR bacteria due to, for example, antibiotics [14-16]. Nonetheless, FMT may reduce the number of infections with MDR bacteria even if the patients' guts are not decolonised with MDR bacteria [66]. The hypothesised mechanism is that the gut microbiota is restored by FMT to a balanced state that is resilient to MDR carrying bacteria [64], for example by nutrient competition [67, 68], restoring short-chain fatty-acid production [69, 70] and production of bacteriocins [71]. This situation may be described as reduced infection susceptibility or infection resistance. Our data give new details on how the taxonomic composition of the microbiota may give shape to reduced MDR infections.

Antibiotic treatment not only affects bacteria, but may also cause fungi to proliferate [72-74]. We recently found using ITS2 sequencing that CDI patients have increased abundances of *Candida* spp., and decreased *Aspergillus* spp. and *Penicillium* spp. compared to controls [75], but the shotgun metagenomic approach in our present study is not adequate to detect fungi with a median relative abundance of 0.003% detected in 5 samples. Thus, to further elucidate the role of the mycobiome, and the fungal resistome, in relation to FMT more targeted experiments are needed.

Our study of the resistome is limited by the annotation of ARGs in publicly available databases and prediction tools for cellular localisation (chromosome or plasmid-based) and do not allow us to definitively link ARGs to specific species. Furthermore, the method we used does not include chromosomal point mutations that confer resistance to antibiotics, which can often be linked to species. Future high-throughput bacterial cultivation efforts will shed light on previously uncharacterised ARGs, what species carry them and on what sort of genetic element [76]. These improved culturomics methods from diverse environments can in turn help alleviate the biases in genome databases,

in which Gram-negative pathogens such as Enterobacterales have been relatively overrepresented. A recent large-scale analysis pointed out that the clinically most relevant ARGs are restricted to particular taxa, most notable Enterobacterales and Bacteroides [77]. Replacement of antibiotic-resistant bacteria may require introduction of susceptible strains [34], which, combined with our data, may suggest a role for FMT donors carrying antibiotic-susceptible Enterobacterales. While FMT may reduce the number or eliminate specific bacteria, horizontal gene transfer may lead to persistence of ARGs when they are transferred to persisting or newly acquired bacteria. Plasmids are potentially mobile genetic elements that facilitate transfer of DNA between bacteria. Plasmid-mediated antibiotic resistance is a growing problem worldwide, and is especially associated with Enterobacterales [78, 79]. We find, however, that most antibiotic resistance genes in the gut microbiota are predicted to be chromosomally encoded, though we cannot exclude the possibility that these ARGs are located on mobile genetic elements. To assess the mobility of resistance genes, techniques are needed that can link ARGs to their host organism, such as Meta-HiC [62], or OIL-PCR [80].

Conclusions

Our study points towards possibilities and limitations of the use of FMT for the eradication of MDR bacteria in the gut. Based on pre- and post-FMT resistome analysis (including a unique LTFU of 1-3 years), we find that FMT induces significant changes in the recipient resistome, that may be associated with a reduction in the abundance of Enterobacterales. However, we also find that specific recipient-ARGs persist. The clinical consequences of this persistence were not included in this study and require further analyses in large cohort of FMT-treated patients. To better assess the possible benefits in MDR eradication, we need larger (randomised controlled) trials and multi-omics studies combined with classical microbiological methods that can link ARGs to bacterial taxa, and to the host's gut ecosystem. Additionally, the use of local, national and international registries for FMT can help collect long-term data to assess infection risks in different patient populations [81, 82]. Besides keeping track of MDR-related outcomes, these registries facilitate evaluation of other long-term microbiota-related risks, such as CDI recurrence or procarcinogenic bacteria [49, 83, 84]. Finally, studies with control patients and more diverse patients are needed to explain the resistome differences and obtain more generalisable results. This will pave the way for evaluating the feasibility of FMT to control antibiotic resistance in infection-susceptible patients.

List of abbreviations

CDI: *Clostridioides difficile* infection
FMT: faecal microbiota transplantation
rCDI: multiple recurrent *C. difficile* infection
MDR: multidrug-resistant
ESBL: extended-spectrum beta-lactamase
RCT: randomised controlled trial
ARG: antibiotic resistance gene
NDFB: Netherlands Donor Feces Bank
LTFU: long-term follow-up
VRE: vancomycin-resistant *Enterococcus*
MRSA: methicillin-resistant *Staphylococcus aureus*
MALDI-TOF: matrix-assisted laser desorption ionization-time of flight
WGS: whole-genome sequencing
RPKM: reads per kilobase per million
ENA: European Nucleotide Archive
PCA: principal component analysis
ANOVA: analysis of variance
IQR: interquartile range
PERMANOVA: permutational analysis of variance
PERMDISP: permutational analyses of multivariate dispersions

Declarations

Ethics approval and consent to participate
Written informed consent was obtained from all patients and donors for use of their faecal samples and follow-up data. Ethical approval was granted for the protocols and practice of the NDFB by the local medical ethics committee at the Leiden University Medical Center (reference P15.145, and long-term follow-up: B21.49). This study conforms to the principles of the Helsinki declaration.

Consent for publication

Not applicable.

Availability of data and materials

Sequencing reads generated for this study are available in the European Nucleotide Archive under project numbers PRJEB64622 (https://www.ebi.ac.uk/ena/browser/view/PRJEB64622) [48], PRJEB44737 (https://www.ebi.ac.uk/ena/browser/view/PRJEB44737) [50], and PRJEB64621 (https://www.ebi.ac.uk/ena/browser/view/PRJEB64621) [51]. Code to reproduce analyses and generate figures are available at Zenodo (https://doi.org/10.5281/zenodo.10276220) [60].
Competing interests
The authors declare that they have no competing interests.

Funding

The Netherlands Donor Feces Bank have received an unrestricted research grant from Vedanta Biosciences.

Authors' contributions

SN, KEWV, RDZ, EJK and EMT conceptualised and designed the study. QRD, EJK, and EMT supervised the study. KEWV, JJK, EJK and EMT supervised treatment of patients. KEWV collected clinical and microbiological data and performed analyses. SN performed genomics, metagenomics and statistical analyses and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We wish to express our gratitude to Eric K.L. Berssenbrugge and Ingrid M.G.J. Sanders from the Experimental Bacteriology group at the LUMC for culturing the MDR bacteria. We thank Prof. Dr. Hein W. Verspaget of the Netherlands Donor Feces Bank for continuous support and supervision. Also, our thanks go to GenomeScan B.V. for providing the DNA sequencing. Finally, we thank the Experimental Bacteriology group and the Center for Microbiome Analyses and Therapeutics of the LUMC for fruitful work discussions, and in particular Dr. Wiep Klaas Smits for feedback on the manuscript.

Supplementary Information

Additional file 1: supplementary figures

This file contains all supplementary figures described in the manuscript. Their titles are listed here.

Fig. S1. Detected antibiotic resistance genes in 5 *Escherichia coli* isolates resistant to aminoglycosides and fluoroquinolones. **Fig. S2.** Detected antibiotic resistance genes in extended-spectrum beta-lactamase producing *Escherichia coli* and *Citrobacter freundii* isolates. **Fig. S3.** Detected antibiotic resistance genes in extended-spectrum beta-lactamase producing *Enterobacter hormaechei_A*, *Klebsiella pneumoniae*, *Morganella morganii* and *Proteus mirabilis* isolates. **Fig. S4.** Taxonomic composition of faecal metagenomes of FMT donors and recipients. **Fig. S5.** Aitchison distance from 63 post-FMT recipients' species composition to 8 used and unrelated donors. **Fig. S6.** Occurrence and abundance of antibiotic resistance genes in faecal metagenomes of FMT donors and recipients. **Fig. S7.** Occurrence and abundance of antibiotic resistance genes of other classes in faecal metagenomes of FMT donors and recipients. **Fig. S8.** Resistome composition as relative abundance of antibiotics classes. **Fig. S9.** Resistome parameters compared with duration of vancomycin pre-treatment in days. **Fig. S10.** Richness and abundance of antibiotic genes of selected classes. **Fig. S11.** Antibiotic resistance genes of high clinical importance. **Fig. S12.** Overview of antibiotic resistance genes predicted to be on plasmids (part 1/2). **Fig. S13.** Overview of antibiotic resistance genes predicted to be on plasmids (part 2/2).

Additional file 2: supplementary results

Additional information regarding the detection of antibiotic resistance genes encoding carbapenemases, ESBL, and colistin and vancomycin resistance.
Supplementary documents are available online on the publisher's website:

https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-024-01306-7#Sec23

References

1. The new EU One Health action plan against antimicrobial resistance [https://ec.europa.eu/health/sites/health/files/antimicrobial_resistance/docs/amr_2017_summary-action-plan.pdf]
2. Antimicrobial Resistance C: Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 2022, 399(10325):629-655.
3. Blaser MJ: Antibiotic use and its consequences for the normal microbiome. *Science* 2016, 352(6285):544-545.
4. Goossens H, Ferech M, Vander Stichele R, Elseviers M, Group EP: Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005, 365(9459):579-587.
5. Karanika S, Karantanos T, Arvanitis M, Grigoras C, Mylonakis E: Fecal Colonization With Extended-spectrum Beta-lactamase-Producing Enterobacteriaceae and Risk Factors Among Healthy Individuals: A Systematic Review and Metaanalysis. *Clin Infect Dis* 2016, 63(3):310-318.
6. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ: Clostridium difficile infection. *Nat Rev Dis Primers* 2016, 2:16020.
7. Quraishi MN, Widlak M, Bhala N, Moore D, Price M, Sharma N, Iqbal TH: Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory Clostridium difficile infection. *Aliment Pharmacol Ther* 2017, 46(5):479-493.
8. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, Visser CE, Kuijper EJ, Bartelsman JF, Tijssen JG *et al*: Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med* 2013, 368(5):407-415.
9. Khoruts A, Sadowsky MJ: Understanding the mechanisms of faecal microbiota transplantation. *Nat Rev Gastroenterol Hepatol* 2016, 13(9):508-516.
10. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, Dubberke ER, Garey KW, Gould CV, Kelly C *et al*: Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* 2018, 66(7):987-994.
11. Ooijevaar RE, van Beurden YH, Terveer EM, Goorhuis A, Bauer MP, Keller JJ, Mulder CJJ, Kuijper EJ: Update of treatment algorithms for Clostridium difficile infection. *Clin Microbiol Infect* 2018, 24(5):452-462.
12. van Prehn J, Reigadas E, Vogelzang EH, Bouza E, Hristea A, Guery B, Krutova M, Noren T, Allerberger F, Coia JE *et al*: European Society of Clinical Microbiology and Infectious Diseases: 2021 update on the treatment guidance document for Clostridioides difficile infection in adults. *Clin Microbiol Infect* 2021, 27 Suppl 2:S1-S21.
13. Isles NS, Mu A, Kwong JC, Howden BP, Stinear TP: Gut microbiome signatures and host colonization with multidrug-resistant bacteria. *Trends in Microbiology* 2022, 30(9):853-865.
14. Dickstein Y, Edelman R, Dror T, Hussein K, Bar-Lavie Y, Paul M: Carbapenem-resistant Enterobacteriaceae colonization and infection in critically ill patients: a retrospective matched cohort comparison with non-carriers. *J Hosp Infect* 2016, 94(1):54-59.

15. Carlet J: The gut is the epicentre of antibiotic resistance. *Antimicrob Resist Infect Control* 2012, 1(1):39.

16. Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM *et al*: Antimicrobial-Resistant Klebsiella pneumoniae Carriage and Infection in Specialized Geriatric Care Wards Linked to Acquisition in the Referring Hospital. *Clin Infect Dis* 2018, 67(2):161-170.

17. Tillotson GS, Zinner SH: Burden of antimicrobial resistance in an era of decreasing susceptibility. *Expert Rev Anti Infect Ther* 2017, 15(7):663-676.

18. Weterings V, van den Bijllaardt W, Bootsma M, Hendriks Y, Kilsdonk L, Mulders A, Kluytmans J: Duration of rectal colonization with extended-spectrum beta-lactamase-producing Escherichia coli: results of an open, dynamic cohort study in Dutch nursing home residents (2013-2019). *Antimicrob Resist Infect Control* 2022, 11(1):98.

19. van Weerlee C, van der Vorm ER, Nolles L, Meeuws-van den Ende S, van der Bij AK: Duration of carriage of multidrug resistant Enterobacterales in discharged hospital and general practice patients and factors associated with clearance. *Infect Prev Pract* 2020, 2(3):100066.

20. Tacconelli E, Mazzaferri F, de Smet AM, Bragantini D, Eggimann P, Huttner BD, Kuijper EJ, Lucet JC, Mutters NT, Sanguinetti M *et al*: ESCMID-EUCIC clinical guidelines on decolonization of multidrug-resistant Gram-negative bacteria carriers. *Clin Microbiol Infect* 2019, 25(7):807-817.

21. Millan B, Park H, Hotte N, Mathieu O, Burguiere P, Tompkins TA, Kao D, Madsen KL: Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent Clostridium difficile Infection. *Clin Infect Dis* 2016, 62(12):1479-1486.

22. Bilinski J, Grzesiowski P, Sorensen N, Madry K, Muszynski J, Robak K, Wroblewska M, Dzieciatkowski T, Dulny G, Dwilewicz-Trojaczek J *et al*: Fecal Microbiota Transplantation in Patients With Blood Disorders Inhibits Gut Colonization With Antibiotic-Resistant Bacteria: Results of a Prospective, Single-Center Study. *Clin Infect Dis* 2017, 65(3):364-370.

23. Crum-Cianflone NF, Sullivan E, Ballon-Landa G: Fecal microbiota transplantation and successful resolution of multidrug-resistant-organism colonization. *J Clin Microbiol* 2015, 53(6):1986-1989.

24. Davido B, Batista R, Michelon H, Lepainteur M, Bouchand F, Lepeule R, Salomon J, Vittecoq D, Duran C, Escaut L *et al*: Is faecal microbiota transplantation an option to eradicate highly drug-resistant enteric bacteria carriage? *J Hosp Infect* 2017, 95(4):433-437.

25. Dinh A, Fessi H, Duran C, Batista R, Michelon H, Bouchand F, Lepeule R, Vittecoq D, Escaut L, Sobhani I *et al*: Clearance of carbapenem-resistant Enterobacteriaceae vs vancomycin-resistant enterococci carriage after faecal microbiota transplant: a prospective comparative study. *J Hosp Infect* 2018, 99(4):481-486.

26. Huttner BD, Galperine T, Kapel N, Harbarth S: 'A five-day course of oral antibiotics followed by faecal transplantation to eradicate carriage of multidrug-resistant Enterobacteriaceae' - Author's reply. *Clin Microbiol Infect* 2019, 25(7):914-915.

27. Lagier JC, Million M, Fournier PE, Brouqui P, Raoult D: Faecal microbiota transplantation for stool decolonization of OXA-48 carbapenemase-producing Klebsiella pneumoniae. *J Hosp Infect* 2015, 90(2):173-174.

28. Manges AR, Steiner TS, Wright AJ: Fecal microbiota transplantation for the intestinal decolonization of extensively antimicrobial-resistant opportunistic pathogens: a review. *Infect Dis (Lond)* 2016, 48(8):587-592.

29. Singh R, Nieuwdorp M, ten Berge IJ, Bemelman FJ, Geerlings SE: The potential beneficial role of faecal microbiota transplantation in diseases other than Clostridium difficile infection. *Clin Microbiol Infect* 2014, 20(11):1119-1125.

30. Singh R, de Groot PF, Geerlings SE, Hodiamont CJ, Belzer C, Berge I, de Vos WM, Bemelman FJ, Nieuwdorp M: Fecal microbiota transplantation against intestinal colonization by extended spectrum beta-lactamase producing Enterobacteriaceae: a proof of principle study. *BMC Res Notes* 2018, 11(1):190.

31. Stalenhoef JE, Terveer EM, Knetsch CW, Van't Hof PJ, Vlasveld IN, Keller JJ, Visser LG, Kuijper EJ: Fecal Microbiota Transfer for Multidrug-Resistant Gram-Negatives: A Clinical Success Combined With Microbiological Failure. *Open Forum Infect Dis* 2017, 4(2):ofx047.

32. Huttner BD, de Lastours V, Wassenberg M, Maharshak N, Mauris A, Galperine T, Zanichelli V, Kapel N, Bellanger A, Olearo F *et al*: A 5-day course of oral antibiotics followed by faecal transplantation to eradicate carriage of multidrug-resistant Enterobacteriaceae: a randomized clinical trial. *Clin Microbiol Infect* 2019, 25(7):830-838.

33. Kuijper EJ, Vendrik KEW, Vehreschild M: Manipulation of the microbiota to eradicate multidrug-resistant Enterobacteriaceae from the human intestinal tract. *Clin Microbiol Infect* 2019, 25(7):786-789.

34. Woodworth MH, Conrad RE, Haldopoulos M, Pouch SM, Babiker A, Mehta AK, Sitchenko KL, Wang CH, Strudwick A, Ingersoll JM *et al*: Fecal microbiota transplantation promotes reduction of antimicrobial resistance by strain replacement. *Sci Transl Med* 2023, 15(720):eabo2750.

35. Terveer EM, van Beurden YH, Goorhuis A, Seegers J, Bauer MP, van Nood E, Dijkgraaf MGW, Mulder CJJ, Vandenbroucke-Grauls C, Verspaget HW *et al*: How to: Establish and run a stool bank. *Clin Microbiol Infect* 2017, 23(12):924-930.

36. Terveer EM, Vendrik KE, Ooijevaar RE, Lingen EV, Boeije-Koppenol E, Nood EV, Goorhuis A, Bauer MP, van Beurden YH, Dijkgraaf MG *et al*: Faecal microbiota transplantation for Clostridioides difficile infection: Four years' experience of the Netherlands Donor Feces Bank. *United European Gastroenterol J* 2020, 8(10):1236-1247.

37. Vendrik KEW, Terveer EM, Kuijper EJ, Nooij S, Boeije-Koppenol E, Sanders I, van Lingen E, Verspaget HW, Berssenbrugge EKL, Keller JJ *et al*: Periodic screening of donor faeces with a quarantine period to prevent transmission of multidrug-resistant organisms during faecal microbiota transplantation: a retrospective cohort study. *Lancet Infect Dis* 2021, 21(5):711-721.

38. Bijzonder resistente micro-organismen (BRMO) (in Dutch) [https://www.rivm.nl/sites/default/files/2018-11/130424%20BRMO.pdf]

39. Breakpoint tables for interpretation of MICs and zone diameters. Version 11.0, 2021. [http://www.eucast.org]

40. Genome assembly GRCh38 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.26/]

41. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, 9(4):357-359.

42. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, 25(16):2078-2079.

43. Chen S, Zhou Y, Chen Y, Gu J: fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018, 34(17):i884-i890.

44. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A: Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics* 2020, 70(1):e102.

45. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN *et al*: CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2017, 45(D1):D566-D573.

46. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV: Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 2012, 67(11):2640-2644.

47. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH: GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022, 38(23):5315-5316.

48. Netherlands Donor Feces Bank: Multidrug-resistant organisms cultured from pre- and post-faecal microbiota transplantation patient stools *European Nucleotide Archive* 2023. https://www.ebi.ac.uk/ena/browser/view/PRJEB64622

49. Nooij S, Ducarmon QR, Laros JFJ, Zwittink RD, Norman JM, Smits WK, Verspaget HW, Keller JJ, Terveer EM, Kuijper EJ *et al*: Fecal Microbiota Transplantation Influences Procarcinogenic Escherichia coli in Recipient Recurrent Clostridioides difficile Patients. *Gastroenterology* 2021, 161(4):1218-1228 e1215.

50. Netherlands Donor Feces Bank: Metagenomic shotgun sequencing of healthy stool bank donors and multiple recurrent Clostridioides infected recipients from the Netherlands Donor Feces Bank. *European Nucleotide Archive* 2022. https://www.ebi.ac.uk/ena/browser/view/PRJEB44737

51. Netherlands Donor Feces Bank: Faecal shotgun metagenomes of healthy stool donors from the Netherlands Donor Feces Bank and multiple recurrent Clostridioides difficile infection patients with long-term follow-up *European Nucleotide Archive* 2023. https://www.ebi.ac.uk/ena/browser/view/PRJEB64621

52. Ducarmon QR, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD: Toward Standards in Clinical Microbiota Studies: Comparison of Three DNA Extraction Methods and Two Bioinformatic Pipelines. *mSystems* 2020, 5(1).

53. Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In: *arXiv: 13033997 [q-bioGN].* 2013.

54. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas AM *et al*: Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol* 2023.

55. McMurdie PJ, Holmes S: phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013, 8(4):e61217.

56. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA: metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017, 27(5):824-834.

57. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE: Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol* 2019, 20(1):217.

58. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010, 11:119.

59. Buchfink B, Reuter K, Drost HG: Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021, 18(4):366-368.

60. Nooij S: Resistome and MDRO analyses of NDFB FMT cohort. *Zenodo* 2023. https://zenodo.org/records/10276220

61. Fredriksen S, de Warle S, van Baarlen P, Boekhorst J, Wells JM: Resistome expansion in disease-associated human gut microbiomes. *Microbiome* 2023, 11(1):166.

62. McCallum GE, Rossiter AE, Quraishi MN, Iqbal TH, Kuehne SA, Schaik Wv: Noise reduction strategies in metagenomic chromosome confirmation capture to link antibiotic resistance genes to microbial hosts. *bioRxiv* 2022:2022.2011.2005.514866.

63. Langdon A, Schwartz DJ, Bulow C, Sun X, Hink T, Reske KA, Jones C, Burnham CD, Dubberke ER, Dantas G *et al*: Microbiota restoration reduces antibiotic-resistant bacteria gut colonization in patients with recurrent Clostridioides difficile infection from the open-label PUNCH CD study. *Genome Med* 2021, 13(1):28.

64. Ghani R, Mullish BH, Davies FJ, Marchesi JR: How to adapt an intestinal microbiota transplantation programme to reduce the risk of invasive multidrug-resistant infection. *Clin Microbiol Infect* 2022, 28(4):502-512.

65. Tariq R, Pardi DS, Tosh PK, Walker RC, Razonable RR, Khanna S: Fecal Microbiota Transplantation for Recurrent Clostridium difficile Infection Reduces Recurrent Urinary Tract Infection Frequency. *Clin Infect Dis* 2017, 65(10):1745-1747.

66. Bilsen MP, Lambregts MMC, van Prehn J, Kuijper EJ: Faecal microbiota replacement to eradicate antimicrobial resistant bacteria in the intestinal tract - a systematic review. *Curr Opin Gastroenterol* 2022, 38(1):15-25.

67. Momose Y, Hirayama K, Itoh K: Competition for proline between indigenous Escherichia coli and E. coli O157:H7 in gnotobiotic mice associated with infant intestinal microbiota and its contribution to the colonization resistance against E. coli O157:H7. *Antonie Van Leeuwenhoek* 2008, 94(2):165-171.

68. Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ, Contreras H, Libby SJ, Fang FC, Raffatellu M: Probiotic bacteria reduce salmonella typhimurium intestinal colonization by competing for iron. *Cell Host Microbe* 2013, 14(1):26-37.

69. Litvak Y, Byndloss MX, Bäumler AJ: Colonocyte metabolism shapes the gut microbiota. *Science* 2018, 362(6418):eaat9076.

70. Roe AJ, O'Byrne C, McLaggan D, Booth IR: Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. *Microbiology (Reading)* 2002, 148(Pt 7):2215-2222.

71. Cotter PD, Ross RP, Hill C: Bacteriocins - a viable alternative to antibiotics? *Nat Rev Microbiol* 2013, 11(2):95-105.

72. Mason KL, Erb Downward JR, Falkowski NR, Young VB, Kao JY, Huffnagle GB: Interplay between the gastric bacterial microbiota and Candida albicans during postantibiotic recolonization and gastritis. *Infect Immun* 2012, 80(1):150-158.

73. Seelbinder B, Chen J, Brunke S, Vazquez-Uribe R, Santhaman R, Meyer AC, de Oliveira Lino FS, Chan KF, Loos D, Imamovic L *et al*: Antibiotics create a shift from mutualism to competition in human gut communities with a longer-lasting impact on fungi than bacteria. *Microbiome* 2020, 8(1):133.

74. Tan CT, Xu X, Qiao Y, Wang Y: A peptidoglycan storm caused by beta-lactam antibiotic's action on host microbiota drives Candida albicans infection. *Nat Commun* 2021, 12(1):2560.

75. Jannie GEH, Monique JTC, Elisabeth MT, Wiep Klaas S, Ed JK, Romy DZ: Fungal and bacterial gut microbiota differ between Clostridioides difficile colonization and infection. *Microbiome Research Reports* 2023, 3(1):8.

76. Lin X, Hu T, Chen J, Liang H, Zhou J, Wu Z, Ye C, Jin X, Xu X, Zhang W *et al*: The genomic landscape of reference genomes of cultivated human gut bacteria. *Nat Commun* 2023, 14(1):1663.

77. Diebold PJ, Rhee MW, Shi Q, Trung NV, Umrani F, Ahmed S, Kulkarni V, Deshpande P, Alexander M, Thi Hoa N *et al*: Clinically relevant antibiotic resistance genes are linked to a limited set of taxa within gut microbiome worldwide. *Nat Commun* 2023, 14(1):7366.

78. Dolejska M, Papagiannitsis CC: Plasmid-mediated resistance is going wild. *Plasmid* 2018, 99:99-111.

79. Crits-Christoph A, Hallowell HA, Koutouvalis K, Suez J: Good microbes, bad genes? The dissemination of antimicrobial resistance in the human microbiome. *Gut Microbes* 2022, 14(1):2055944.

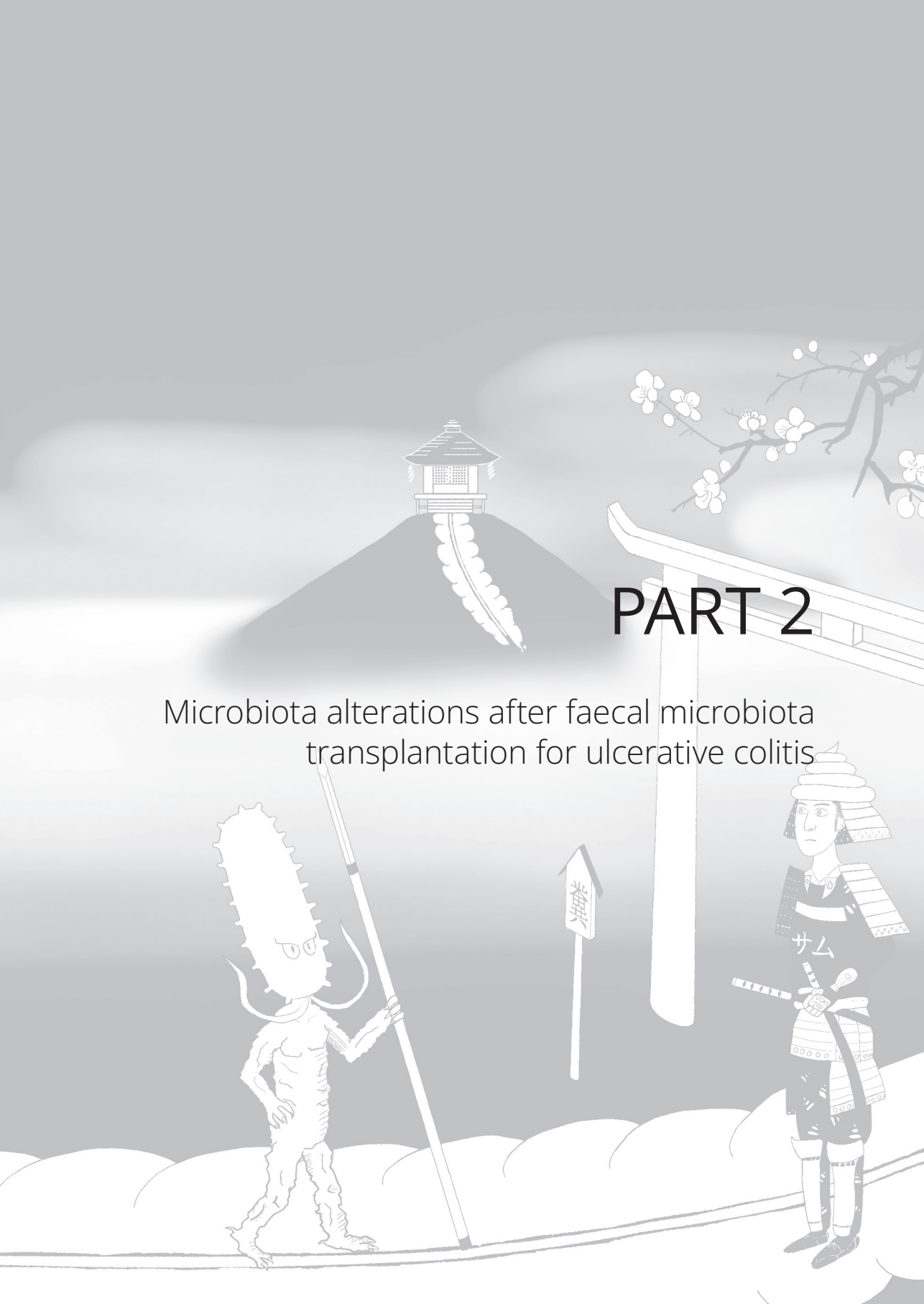
80. Diebold PJ, New FN, Hovan M, Satlin MJ, Brito IL: Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR. *Elife* 2021, 10.

81. Kelly CR, Yen EF, Grinspan AM, Kahn SA, Atreja A, Lewis JD, Moore TA, Rubin DT, Kim AM, Serra S *et al*: Fecal Microbiota Transplantation Is Highly Effective in Real-World Practice: Initial Results From the FMT National Registry. *Gastroenterology* 2021, 160(1):183-192 e183.

82. Lodberg Hvas C. KJ, Dahl Baunwall S. M., Edwards L. A., Ianiro G., Kupcinskas J., Link A., Mullish B. H., Satokari R., Terveer E., Vehreshild M. J.G.: European academic faecal microbiota transplantation (EURFMT) network: improving the safety and quality of microbiome therapies in Europe. *Microb Health Dis* 2023, 5:e954.

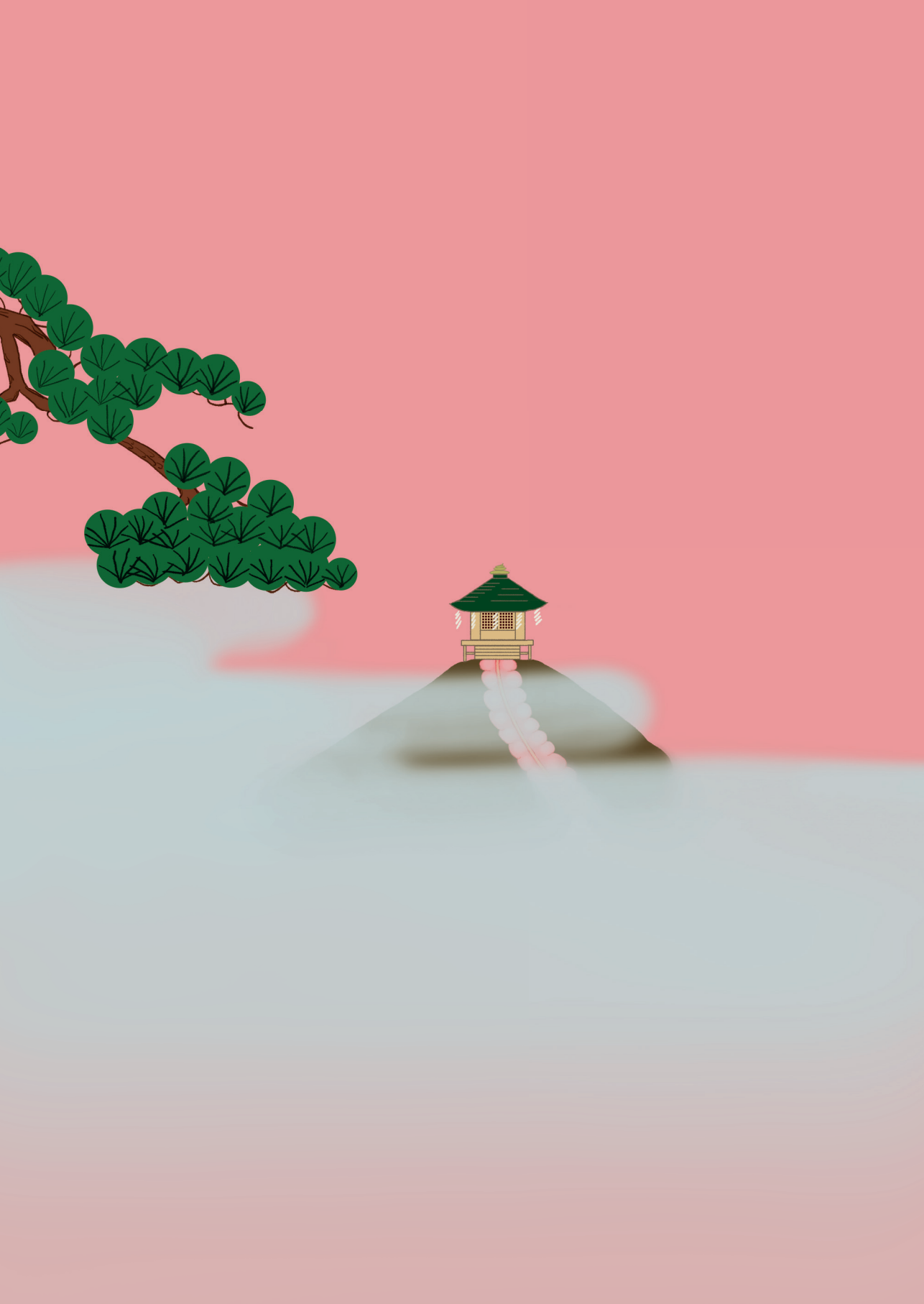
83. Khoruts A: Can FMT Cause or Prevent CRC? Maybe, But There Is More to Consider. *Gastroenterology* 2021, 161(4):1103-1105.

84. Drewes JL, Chen J, Markham NO, Knippel RJ, Domingue JC, Tam AJ, Chan JL, Kim L, McMann M, Stevens C *et al*: Human Colon Cancer-Derived Clostridioides difficile Strains Drive Colonic Tumorigenesis in Mice. *Cancer Discov* 2022, 12(8):1873-1885.



PART 2

Microbiota alterations after faecal microbiota transplantation for ulcerative colitis



Chapter 4

Faecal Microbiota Transplantation Engraftment After Budesonide or Placebo in Patients With Active Ulcerative Colitis Using Pre-selected Donors: A Randomized Pilot Study

E. van Lingen[1], S. Nooij[2], E.M. Terveer[2], Emily Crossette[5], Amanda L. Prince[5], Shakti K Bhattarai[6], Andrea Watson[5], Gianluca Galazzo[5,] Rajita Menon[5], Rose L. Szabady[5,7], Vanni Bucci[6], Jason M. Norman[5,] S. van der Marel[3], C.J. van der Woude[4], H.W. Verspaget[1], A.E. van der Meulen – de Jong[1*], J.J. Keller[1,3*]

[1]Department of Gastroenterology and Hepatology, Leiden University Medical Center
[2]Department of Medical Microbiology, Leiden University Medical Center
[3]Department of Gastroenterology and Hepatology, Haaglanden Medisch Centrum
[4]Department of Gastroenterology and Hepatology, Erasmus Medical Center
[5]Vedanta Biosciences, Inc. Cambridge, MA 02139 USA
[6]University of Massachusetts Medical School, Department of Microbiology and Physiological Systems, Worcester, MA 01605 USA
[7]Current: Ferring Pharmaceuticals, San Diego, CA 92121 USA
*these authors contributed equally

Journal of Crohn's and Colitis, 4 April 2024, jjae043, DOI: 10.1093/ecco-jcc/jjae043

Abstract

Background

Faecal microbiota transplantation [FMT] shows some efficacy in treating patients with ulcerative colitis [UC], although variability has been observed among donors and treatment regimens. We investigated the effect of FMT using rationally selected donors after pretreatment with budesonide or placebo in active UC.

Methods

Patients ≥18 years old with mild to moderate active UC were randomly assigned to 3 weeks of budesonide [9 mg] or placebo followed by 4-weekly infusions of a donor faeces suspension. Two donors were selected based on microbiota composition, regulatory T cell induction and short-chain fatty acid production in mice. The primary endpoint was engraftment of donor microbiota after FMT. In addition, clinical efficacy was assessed.

Results

In total, 24 patients were enrolled. Pretreatment with budesonide did not increase donor microbiota engraftment [$p = 0.56$] nor clinical response, and engraftment was not associated with clinical response. At week 14, 10/24 [42%] patients achieved [partial] remission. Remarkably, patients treated with FMT suspensions from one donor were associated with clinical response [80% of responders, $p < 0.05$] but had lower overall engraftment of donor microbiota. Furthermore, differences in the taxonomic composition of the donors and the engraftment of certain taxa were associated with clinical response.

Conclusion

In this small study, pretreatment with budesonide did not significantly influence engraftment or clinical response after FMT. However, clinical response appeared to be donor-dependent. Response to FMT may be related to transfer of specific strains instead of overall engraftment, demonstrating the need to characterize mechanisms of actions of strains that maximize therapeutic benefit in UC.

Introduction

The inflammatory bowel disease [IBD] ulcerative colitis [UC] is a chronic disorder of the colon. The exaggerated immune response appears to be mediated by an interaction between genetic and environmental factors and the gut microbiota.[1] Current treatment modalities are based on local immune suppression with 5-aminosalicylates [5-ASAs], temporary prednisolone/budesonide, thiopurines, or targeted therapy with small molecules and/or biologics.[2] Despite maintenance medication, most patients suffer from periods of flares and ongoing inflammation with a significant impact on quality of life, and the 20-year cumulative colectomy rate after UC diagnosis is 14%.[3] In addition, a subset of patients experience significant side effects from currently available medication. These observations underlie the need for new treatment modalities.

IBD is characterized by a proinflammatory cytokine microenvironment with a reduced ratio of regulatory T cells [Tregs] to proinflammatory T helper 17 cells [Th17].[4] Tregs are a subpopulation of T cells that are able to inhibit T cell proliferation and cytokine production and thereby have a role in regulating or suppressing the immune response. Analysis of the microbiome in UC patients with active disease indicates a less diverse composition compared with healthy individuals.[5] An imbalance of dominant species has also been observed in IBD patients, with a reduction in members of the phyla Firmicutes, Bacteroidetes, or Verrucomicrobia [specifically, *Faecalibacterium prausnitzii*, Clostridium cluster IV and XIVa, and *Akkermansia muciniphila*] and an overgrowth of facultative anaerobes from the family Enterobacteriaceae [*Escherichia coli* or *Klebsiella* species].[6,7] Furthermore, short-chain fatty acids [SCFAs], such as butyrate, are typically reduced in patients with active IBD; these organic acids are described as important metabolites to maintain intestinal homeostasis.[8]

Faecal microbiota transplantation [FMT] is an effective treatment for recurrent *Clostridioides difficile* infections by restoring a healthy gut microbiota composition.[9] Several studies have addressed the effects of FMT for UC, showing a response rate of around 36% for patients with various treatment regimens.[10] Of note, efficacy may be donor-dependent[11] and response appears to be associated with engraftment of the donor microbiota.[12] These previous studies highlight many unanswered questions about treatment of UC patients with FMT: how to identify preferred donors, and determine which pretreatments promote the donor microbiome and response.

In this randomized controlled pilot study, we investigated the effect of pretreatment with budesonide on engraftment of FMT donor microbiota in patients with mild to moderately active UC. In addition, donor microbiota-dependent effects were compared using faecal material preparations from two rationally selected donors.

Methods

Study design

This was a single-centre, double-blind, randomized controlled trial in adult UC patients with active mild to moderate disease. The primary outcome was the effect of pretreatment with budesonide or placebo on the engraftment of the donor microbiota after FMT in UC patients. Secondary outcomes were safety, clinical response at week 10 [4 weeks after the last FMT] based on the partial Mayo score,[13] and at week 14 [8 weeks after the last FMT] based on the full Mayo score [partial Mayo score + endoscopic outcomes] on FMT as an induction therapy with and without budesonide pretreatment. Additional secondary outcomes included the effects of pretreatment, donor selection, and donor microbiota engraftment on clinical response. In the absence of reliable information to estimate the effect size of the treatment on the primary endpoint in this pilot study, a sample size of 12 patients per group was chosen.[14] Due to the COVID-19 pandemic and the following national lock-down measures implemented by the Dutch government, the inclusion of patients was temporarily halted from March 2020 to June 2020.

Patient selection

Patients ≥18 years old with a confirmed endoscopic diagnosis of UC with mild to moderate active disease [a full Mayo score of 4–9] were enrolled. A colonoscopy with a Mayo endoscopic sub score of 1 or 2 performed <4 weeks before study entry was required. Concomitant therapy at a stable dose for at least 12 weeks with aminosalicylates, azathioprine, 6-mercaptopurine, methotrexate, and/or anti-tumour necrosis factor [TNF]-α, was permitted.

Exclusion criteria were: patients not able to give written informed consent, pregnancy, disease limited to the rectum [<15 cm from the anal verge], recent use of antibiotics [<6 weeks] or current need for systemic antibiotics, recent use of oral corticosteroids or budesonide [<8 weeks], previous surgery specifically for UC or recent intra-abdominal surgery [<12 weeks], signs of active infectious gastroenteritis/enterocolitis, cytomegalovirus [CMV] infection, abnormal renal function (estimated glomerular filtration rate [eGFR] > 30 mL/min), pre-existing leucopenia [<2.0 mm^{-3}] or thrombopenia [<90.000 mm^{-3}], liver function test abnormalities [>2 ULN (upper limit of normal)], previous treatment with more than two biologics, treatment with any investigational drug in another trial [<12 weeks of randomization], or any other significant medical illnesses that might interfere with this study.

Preparation and Characterization of Donor Fecal Suspensions

The donor faecal suspensions were provided by the Netherlands Donor Feces Bank [NDFB], which uses standardized procedures for donor screening, collection, preparation, and storage.[15–17] All faecal suspensions consisted of 60 g of faeces from each individual donor, with an end volume of 10% glycerol, and stored in a −80°C freezer

until the morning of the FMT. From a set of 12 eligible fecal donors, a rational selection of the 2 donors used in this study (D07 and D08) was based on microbiota community composition, ability to induce T regulatory cells (Tregs) *in vivo*, and capacity to produce short-chain fatty acids (SCFAs) *in vivo*.

Characterization of donor stool

From 12 eligible donors, 71 stool samples were investigated by metagenomic sequencing using the Nextera XT DNA library prep kit and sequenced on the Illumina NextSeq at DNA Genotek's CLIA certified sequencing lab [DNA Genotek Inc.a]. An average of 43 million reads was obtained for each of the 71 unique donor faecal samples [83 total metagenomes]. Faecal metagenomic reads were assigned taxonomies using the One Codex platform and a custom database containing around 150 000 reference genomes developed by Vedanta Biosciences, Inc. The One Codex platform uses a *k-mer*-based assignment algorithm where metagenomic reads are compared to a database of taxon-specific k-mers and assigned to the lowest common ancestor on the One Codex phylogenetic tree. Donors were clustered by their mean relative abundance of taxonomic families across samples, using hierarchical clustering with Euclidean distance and complete linkage. Raw data for the metagenomic sequences are available upon request.

Metagenomic sequencing demonstrated that all donors had high abundance of bacteria within the families Clostridiaceae, Bifidobacteriaceae, Ruminococcaceae, and Lachnospiraceae [Figure 1A]. When the donor microbiome was examined by hierarchal clustering, four donors [D13, D04, D12, and D01] clustered near each other and had high levels of Prevotellaceae, which prior studies demonstrate is detrimental in animal models of IBD.[18,19] D02 had low levels of Prevotellaceae; however, this donor also had low levels of the propionate-producing family of bacteria Bacteroidaceae.[20] Thus, this donor was excluded from the study. A cluster of four donors [D09, D05, D08, and D07] had low levels of Prevotellaceae with high levels of Bacteroidaceae and additional taxa with potential benefit in IBD. D09 had a variety of unique taxa including additional bacterial families associated with remission in UC, such as Oscillospiraceae and Firmicutes [Peptoniphilaceae]. However, D09 did have low abundances Enterobacteriaceae and Enterococcaceae that are associated with UC.[21] D05 had high levels of bacteria demonstrated to be beneficial in UC, and D07 and D08 had similar taxonomy to D05 but also contained Akkermansiaceae, which has been shown to provide benefit in UC.[22] Together, the metagenomic analysis selected D05, D07, D08, and D09 for preclinical studies to examine Treg induction and SCFA production in germ-free mice.

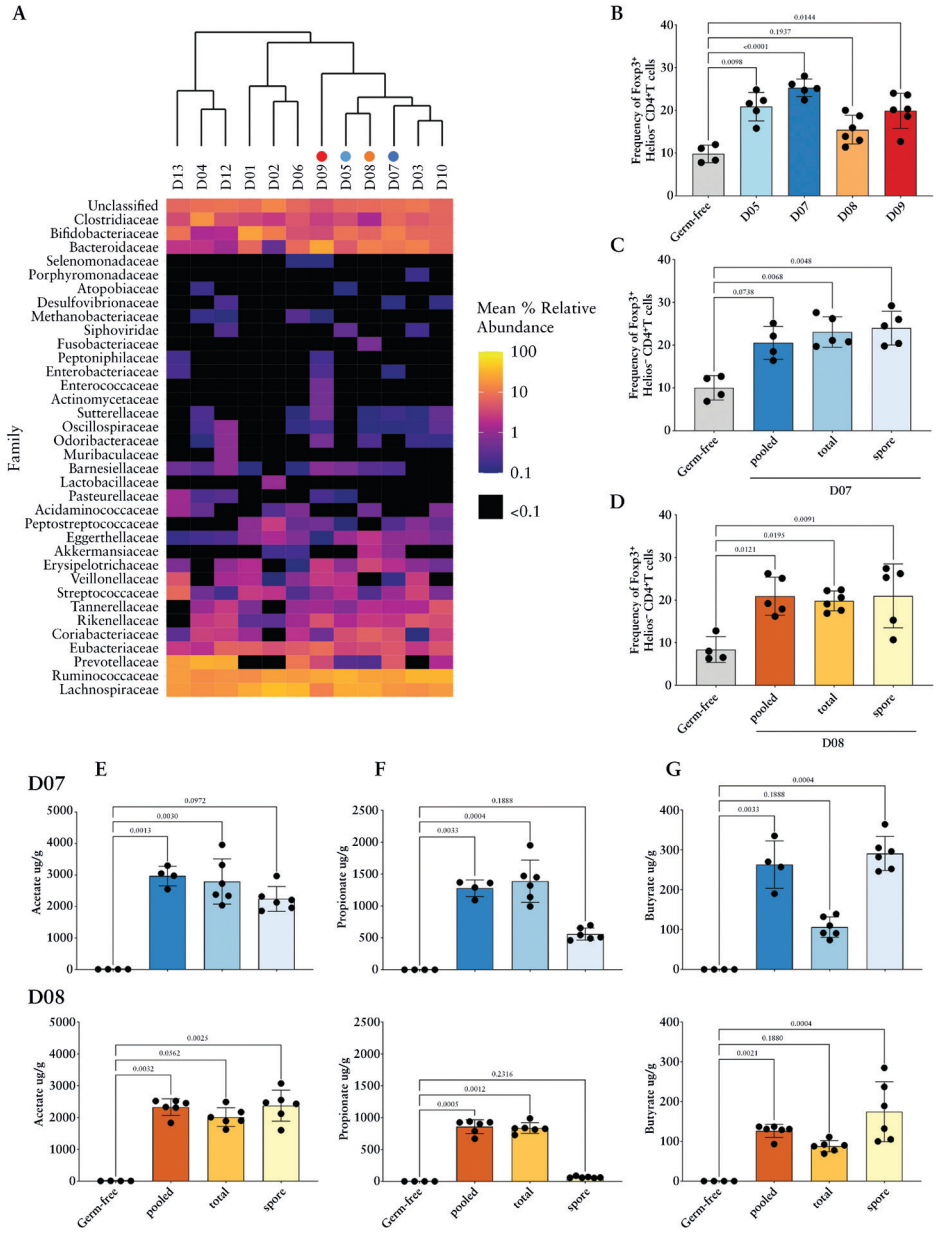


Figure 1. Selection of faecal microbiome transplant donors. [A] Heatmap depicting the mean abundance of taxa at the family level in stool samples from 12 healthy FMT donors. The donors are clustered based on their faecal microbiota composition. The donors selected for further evaluation based on their microbiota composition and donor availability are indicated by the filled in circles and correspond with panel B. [B–D] Donor stool induces T regulatory cells in the colonic lamina propria of mice. Germ-free mice were given either pooled [total + ethanol treated], spore-forming, or total stool fraction libraries [SFL]. Four weeks post-treatment, induced

T regulatory cells in the colonic lamina propria were examined by flow cytometry [gating: lymphocytes → singlets → Live CD45+ → CD4+ CD3+ → Foxp3 vs Helios]. [B] Total SFLs from donors D09, D05, and D07 significantly induced T regulatory cells [Foxp3+ Helios−] when compared to germ-free mice; *n* = 4–6 mice per group. [C, D] Induction of T regulatory cells by pooled, total, and spore-forming SFLs from [C] D07 and [D] D08; *n* = 4–6 mice per group. [E–G] Short-chain fatty acids from caecal contents were measured by mass spectrometry. Acetate [E], propionate [F], and butyrate [G] were examined. Non-spore fractions induced higher propionate [F] while spore fractions induced higher butyrate [G].

Examination of Treg induction in the colonic lamina propria

For preclinical stool selection experiments, stool fraction libraries [SFLs] were generated from selected donors by culturing live bacteria and collecting plate scrapings. Spore-forming SFLs were generated by ethanol treatment of cultured live bacteria from stool. Pooled SFLs were generated by combining the spore-forming fractions with plate scrapings from total SFLs. Germ-free mice were inoculated with SFL material for 28 days prior to examining induction of Tregs in the colonic lamina propria and SCFA production in the caecum.

SFLs of total plate scrapings, ethanol-treated, and pooled [plate scrapings combined with ethanol-treated] samples were given to germ-free mice (according to the Institutional Animal Care and Use Committee [IACUC] at the Massachusetts Host-Microbiome Center Gnotobiotics Core [Boston, MA, USA]). At 28 days post-inoculation with donor SFL, colons from formerly germ-free mice and controls were collected. Mouse colonic tracts were dissected, sectioned longitudinally, cut into 2–4-mm pieces, and washed with phosphate buffered saline [PBS] to remove luminal contents. Cell suspensions from the lamina propria were prepared as described previously [23] with the following modifications:intestinal tissue was cut into small pieces, treated with RPMI with 10% FBS, 1 mM of DTT, and 5 mM EDTA to remove epithelial cells, and then digested with 0.2 mg/mL Collagenase VIII, and 0.075 mg/mL dispase at 37°C with shaking (180 rpm) for two 40 min periods. Cells were counted, stained with a viability dye, using the surface markers CD45, CD3, and CD4, and stained intracellularly for RORγt, Helios, Foxp3, and GATA3. Stained cells were run on a BD FACSCelesta and analysed using Flowjo and GraphPad Prism software.

When SFLs were inoculated into germ-free mice, we saw significant induction of Tregs [Foxp3+ Helios−] in mice receiving faecal material from D05, D07, and D09 [$p < 0.02$, Figure 1B]. D09 containing taxa associated with UC and D07 and D08 containing Akkermansiaceae were further assessed in germ-free experiments of non-spore, spore, and pooled SFL donor material. We found that various stool fractions from D07 and D08 were able to significantly induce Tregs [Foxp3+ Helios−] in gnotobiotic mice [$p < 0.02$, Figure 1C and D] with the exception of the pooled SFL from D07, which was not statistically significant but trended toward Treg induction [$p < 0.08$, Figure 1C].

Short-chain fatty acid analysis

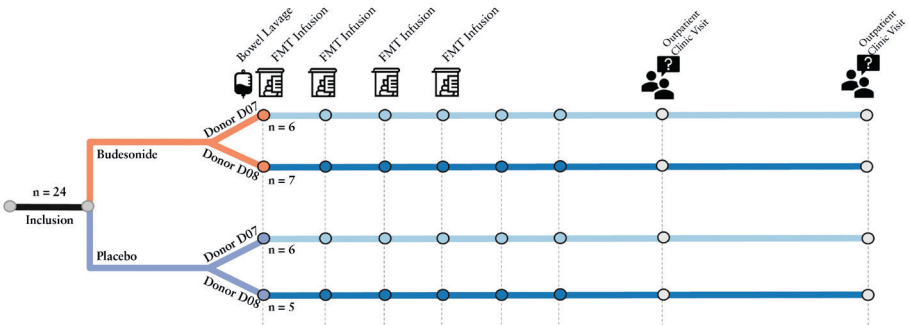
SFLs of total plate scrapings, ethanol-treated, and pooled [plate scrapings combined with ethanol-treated] samples were given to germ-free mice. At 28 days post-inoculation with donor SFL, caecal contents from formerly germ-free mice and controls were collected and mass spectrometry for SCFAs was performed [Metabolon].

Acetate was produced in germ-free mice inoculated by all SFL fractions, the non-spore, spore, and pooled SFL material, from donor microbiomes D07 and D08 [$p < 0.1$, Figure 4E]. Propionate was significantly produced by pooled and non-spore SFL fractions [$p < 0.05$, Figure 1F] as expected due to Bacteroides being the main producers of propionate. Likewise, butyrate was significantly produced by pooled and spore-forming SFL fractions [$p < 0.005$, Figure 1G] as butyrate is mainly produced by spore-forming bacteria, such as Clostridia.

Taken together, the microbial composition, Treg induction, and SCFA production of donor microbiota were used to narrow down two donor samples. Donor microbiomes were narrowed to D05, D07, D08, and D09, based on observed lower relative abundances of Prevotellaceae and abundance of the bacterial families Bacteroidaceae and Akkermansiaceae. Following observed Treg induction and SCFA production in mouse models, donors D08 and D07 were selected as FMT donors in this study.

Treatment schedule

Patients were included and treated at the Department of Gastroenterology at Leiden University Medical Center [LUMC, Leiden, Netherlands]. After inclusion, patients were randomly assigned to a 3-week course of oral budesonide [9 mg] or identical looking placebo once a day [Figure 2]. The randomization procedure was conducted by an independent coordinator of the LUMC hospital pharmacy. A block diagram was used of 8 [4 to 4] patients per block for randomization, which also included stratification for donors [6 to 6]. The pretreatment phase was followed by four weekly infusions [weeks 3–6] of a donor faecal suspension from either donor D07 or D08 produced by the NDFB as described above. The donor faecal samples were infused in the duodenum via either a nasoduodenal tube or directly via the gastroscope, both according to the LUMC protocol. One day before the first FMT, patients underwent bowel lavage with 2 L of a macrogol solution [Kleanprep]. Prior to every FMT, all patients were sober for at least 6 h. During the study, no changes were allowed regarding the patients' medication or diet.



Phase	Pre-treatment	Fecal Microbiota Transplantations				Follow-Up							
Week	0–3	3	4	5	6	7	8	9	10	11	12	13	14
Questionnaire	X	X	X	X	X				X				X
Fecal Sample	X	X* "FMT 1"	X* "FMT 2"	X* "FMT 3"	X* "FMT 4"	X	X		X				X

*Sample Collected before FMT Procedure

Acute Engraftment | Durable Engraftment

Figure 2. Timeline of the FECBUD study. At inclusion in the study, 24 patients were randomized into budesonide and placebo treatment groups. Patients received a bowel lavage, followed by four FMT infusions once weekly from week 3 to 6. The acute time period for assessing engraftment is defined from week 4 to 7, which corresponds to samples collected 1 week after each FMT. The durable time period for assessing engraftment is defined by samples collected 2 weeks or later following the final FMT infusion [week 8 to week 14]. In the table, X indicates when a questionnaire was used or when a sample was collected. FMT = faecal microbiome transplantation, *n* = number.

Data collection

Baseline characteristics included age, sex, weight, disease duration, location of disease, previous medication, and concomitant drug treatment. A full Mayo score, both clinical and endoscopic, was assessed, and all patients underwent physical examination [weight, temperature, and pulse]. The project team [AM, SM, JK] reviewed the endoscopic Mayo score at inclusion. Analysis of blood samples included C-reactive protein, leukocytes, thrombocytes, renal function [eGFR, creatinine], liver function tests and faecal calprotectin [FCP]. To assess quality of life, the EQ-5D-5L and a disease-specific questionnaire were used.[24] During follow-up, disease activity and adverse events were registered in a structured way. At every study visit, stool and blood samples were collected for research purposes. The questionnaires were administered again at weeks 10 and 14. At the end of the study, a sigmoidoscopy was performed to assess the endoscopic Mayo score [Figure 2].

Patient outcomes and clinical response

At week 10 [4 weeks after the last FMT], clinical outcome was based on the partial Mayo score, including stool frequency in 24 h [score 0–3], the presence of blood [score 0–3] and the Physician Global Assessment [score 0–3]. At week 10, remission was defined as no complaints [partial Mayo score of ≤2, with no individual sub score of >2], and partial response was defined as a decrease of at least 3 points in the partial Mayo score; all other situations were considered as no response.

At week 14 [8 weeks after the last FMT], clinical outcome was based on the full Mayo score. All endoscopic pictures were reviewed blinded by two experienced endoscopists [AM and SM]. In case of disagreement, a third reviewer [JK] scored. At week 14, remission was defined as no complaints [partial Mayo score of ≤2, with no individual sub score of >2] and an endoscopic Mayo score 0–1a [erythema without friability]. Partial remission was defined as a decrease of at least 3 points in the partial Mayo score and a decrease of at least 1 point in the endoscopic Mayo score [e.g. also a decrease from 1b to 1a]. No response at week 14 was defined as an occurrence or recurrence of symptoms and an equal or increased endoscopic Mayo score. Patients who did not complete the study because of progressive symptoms/disease were considered treatment failures [no response].

Metagenomic sequencing of faecal samples

Stool samples were collected according to standardized procedures[15,17] at inclusion, before placebo or budesonide treatment, and prior to faecal microbiome transplantations and following treatment on weeks 7, 8, 10, and 14 [Figure 2]. Total DNA was extracted from donor and recipient stool samples using the DNeasy PowerSoil Pro kit [Qiagen Cat. No. 47016] on the QIAcube and the MO BIO MagAttract PowerMicrobiome [+ClearMag] kit [Cat. No. 27500-40EP] automated on the KingFisher Flex [ThermoFisher]. Metagenomic sequencing was performed by Diversigen using the Illumina NovaSeq 6000 platform [Illumina Inc.]. Samples were sequenced using the 1 × 100 bp single end protocol to a median depth of 2.9 million reads. Raw data from metagenome sequencing are available upon request.

Engraftment analysis

Raw metagenomic reads were uploaded to the One Codex analysis platform [One Codex]for taxonomic classification, which assigns taxonomy based on a custom, curated databased as described previously in 'Characterization of donor stool' above. In short, the raw reads are divided into k-mers sets then assigned taxonomy comparing the k-mers to a custom in-house reference database of microbes. Taxonomic profiling with this approach has been used previously to assess the impact of live biological therapeutics on healthy volunteer microbiomes,[25] evaluate host microbiome response to sterile faecal filtrate transplantation in patients with C. difficile infection,[26] and classify gut microbiome composition in patients with bloodstream infections.[27] The taxonomic assignment table was then imported into R to calculate the engraftment score, visualize the results, and perform statistical tests [R v.4.0.2].

After applying a 0.1% relative abundance threshold, the median cumulative abundance per sample was 97%, i.e. only 3% of data were discarded. Next, we determined which species were putatively donor-derived by determining the donor core species. Species were considered 'core' if they were present above the detection limit in at least six samples [out of 13 or 14 for donor D07 and D08, respectively]. This yielded

core microbiomes of 160 and 115 species for donor D07 and D08, respectively. These species were also abundant in all donor samples: their total abundance was always >70%. To determine which of these species engrafted in the recipient patients, these putatively donor-derived species were compared to the bacterial profiles in the recipients before FMT and overlapping species were excluded, i.e. species already present in the patient before FMT were deemed not engrafted. The species that were unique to the donor for each subject were used to quantify engraftment. The resulting median number of putatively donor-derived species per recipient was 82 [range: 62–135]. Engraftment metrics were calculated for the complete duration of the experiment ['overall'], for the phase directly following the FMTs [Weeks 4–7, Figure 2, 'acute'] and for the weeks after the FMTs [Weeks 8–14, Figure 2, 'durable']. Three different engraftment measures were then summed to produce a total engraftment score. [1] The median Species Engraftment Index [SEI], which is the median total relative abundance of the engrafted species. This quantifies the extent of displacement of the recipient microbiome with the donor microbiome. [2] The second metric is the Species Engraftment Fraction [SEF], which is the fraction of bacteria that engrafted divided by all putatively engrafted, donor-derived species [and is therefore a number between 0 and 1]. This metric describes the diversity of donor species, regardless of abundance, that are introduced to the recipient in FMT, which was previously associated with donor efficacy and clinical response.[28–30] [3] Engraftment was assessed by comparing bacterial species compositions using pairwise Aitchison distances. Recipient to donor distances were evaluated as median distance to the corresponding donor microbiomes and normalized by the distance to the donor at the timepoint before the first FMT [after pretreatment and bowel lavage; timepoint 'week 3']. Next, we converted this to similarity scores by taking [1/Aitchison distance] for ease of interpretation [high score = more engraftment]. Microbiota similarity has also been used to study FMT efficacy.[12,31] To combine the three different metrics for engraftment in a total engraftment score, we normalized the SEI, SEF, and median Aitchison distance to donor by multiplying the SEI and Aitchison distance so that each had an approximately equal median, and then summed them. The scripts used to quantify engraftment are available online at https://doi.org/10.5281/zenodo.8158623.

Machine learning modelling to identify microbiota signatures of response

To determine recipient microbiome taxonomic signatures that are strongly predictive of response, random forest classification [RFC] modelling was used as in the study from Haran and colleagues.[32] Fewer microbiome samples were available at later timepoints for training and testing models because treatment failures dropped out of the study over time. This created a trade-off between maximizing the number of patients included in a model or maximizing the breadth of timepoints, which was accounted for by creating four different models. Model 1 predicted response using only baseline microbiome composition of subjects [number of subjects included, n = 24].

Model 2 included baseline and post-FMT microbiome samples which resulted in the exclusion of two subjects [*n* = 22]. Model 3 included baseline, post-FMT, and week 7 and week 8 samples [*n* = 18], and model 4 included all timepoints up to week 14 [*n* = 16]. Each model was run using aggregated microbial abundances at different taxonomic levels [e.g. species, genus, class, order]. To focus on highly predictive features, the RFC models were wrapped within the Boruta feature selection algorithm.[33] The different models were compared in terms of predictive ability by estimating area under the curve [AUC] and F1 score from a leave-one-out cross-validation [LOOCV] scheme. The most predictive model was run on the entire dataset to identify important taxa associated with response via permutated importance calculations.[34] Finally, to derive human interpretable rules on the bacterial abundance of the different taxa that best discriminates between responders and non-responders, we passed the RFC modelling results to the Stable and Interpretable Rule Set [SIRUS] pipeline.[35]

Statistical analysis

Continuous variables are presented as mean with standard deviation [SD] or as median with interquartile range [IQR] depending on the normality of the underlying distribution. Baseline characteristics were compared using an independent sample t-test or Mann–Whitney U-test, and paired variables were compared using a paired sample t-test or Wilcoxon signed-rank test. Categorical variables are presented as a total percentage and compared by using the chi-square test or Fisher's exact test. A p-value of ≤0.05 was considered significant for all tests. All statistical analyses other than those on engraftment scores were performed using SPSS v.25.0.

Ethical consideration

This research project was reviewed and approved by the Medical Ethical Committee in the LUMC, with reference number NL 65976.098.18. Informed consent was obtained from all participants prior to inclusion in the study. The study was conducted in accordance with the principles of the Declaration of Helsinki. The study was registered in the Netherlands Trial Register, with reference number NL9858.

Results

From May 7, 2019 to October 27, 2020, 24 patients were enrolled and treated. The study timeline is summarized in Figure 2. The median age of enrolled patients was 42 years [33.0–57.5] and 50% of patients were male. The median UC disease duration was 13.5 years [5.5–20.3]. The baseline median full Mayo score was 7 [5–9] with a median faecal calprotectin level of 944 µg/g [369–1719]. Although more patients in the budesonide pretreatment group had used anti-TNF therapy in the past [eight vs two; $p < 0.05$] this was not observed with respect to concomitant treatment in the study [Table 1].

Table 1. Patient characteristics.

	Total group [*n* = 24]	Pre-treatment			Donor		
		Placebo [*n* = 11]	Budesonide [*n* = 13]	*p*-value	D07 [*n* = 12]	D08 [*n* = 12]	*p*-value
Age [years], median [IQR]	42 [33–57.5]	52 [37–73]	34 [32–49.5]	0.077	42.5 [33.3–57.5]	40.5 [31.3–61.0]	0.583
Male sex, *n* [%]	12 [50]	5 [45.5]	7 [53.8]	0.682	5 [41.7]	7 [58.3]	0.414
Disease duration [years], median [IQR]	13.5 [5.5–20.3]	15 [4–22]	13 [6–19.5]	0.663	11.0 [3.2–15.0]	15.0 [7.3–21.8]	0.214
Clinical Mayo score, median [IQR]	5.0 [4.0–7.0]	5.0 [4.0–7.0]	5.0 [4.0–7.0]	0.878	6.5 [4.0–7.0]	5.0 [4.0–7.0]	0.284
Endoscopic Mayo score, *n* [%]				1.000			1.000
Mayo 1A	0	0	0		0	0	
Mayo 1B	2 [8.3]	1 [9.1]	1 [7.7]		1 [8.3]	1 [8.3]	
Mayo 2	22 [91.7]	10 [90.9]	12 [92.3]		11 [91.7]	11 [91.7]	
Full Mayo score, median [IQR]	7.0 [5.0–9.0]	7.0 [5.0–9.0]	7.00 [6.0–9.0]	0.975	8.0 [6–9]	7.0 [5.0–9.0]	0.385
Prior anti-TNF therapy, *n* [%]	10 [41.7]	2 [18.2]	8 [61.5]	**0.047**	5 [41.7]	5 [41.7]	1.000
Concomitant drugs, *n* [%]							
5-ASA	16 [66.7]	8 [72.7]	8 [61.5]	0.679	8 [66.7]	8 [66.7]	1.000
Immunomodulators	3 [12.5]	2 [18.2]	1 [7.7]	0.576	1 [8.3]	2 [16.7]	1.000
Anti-TNF-α	3 [12.5]	0 [0.0]	3 [23.1]	0.223	2 [16.7]	1 [8.3]	1.000

	Total group [*n* = 24]	Pre-treatment			Donor		
		Placebo [*n* = 11]	Budesonide [*n* = 13]	*p*-value	D07 [*n* = 12]	D08 [*n* = 12]	*p*-value
Hb concentration, g/L, median [IQR]	8.3 [7.6–9.0]	8.5 [7.7–9.1]	8.1 [7.5–8.6]	0.223	8.2 [7.4–9.0]	8.3 [8.0–9.1]	0.506
White cell count, $\times 10^9$/L, median [IQR]	7.0 [5.5–8.2]	7.1 [4.6–7.4]	6.7 [5.5–9.7]	0.469	7.2 [4.8–8.1]	6.9 [5.5–8.4]	0.840
CRP, mg/L, median [IQR]	1.9 [1.4–5.4]	2.5 [1.4–5.5]	1.8 [1.4–5.2]	0.450	1.9 [1.4–5.8]	2.2 [1.4–5.2]	0.772
FCP, µg/g, median [IQR]	944.0 [369.0–1719.0]	1042.5 [702.8–1814.3]	841.0 [241.0–1610.5]	0.193	989.5 [452.5–1399.8]	795.0 [326.0–1976.0]	0.926

Anti-TNF = anti-tumour necrosis factor, CRP = C-reactive protein, FCP = faecal calprotectin, Hb = haemoglobin, *n* = number.

Bold indicates $p<0.05$.

Total engraftment of donor microbiome

To assess engraftment, one total engraftment score was computed from the weighted sum of three engraftment metrics: SEI, SEF, and Aitchison similarity of recipient microbiome to that of the donor following FMT. The relationship between each metric and donor, pretreatment, week 10 response, and week 14 response are reported in Supplementary Figures 1–4. Engraftment was calculated during FMT treatment [Acute Engraftment, Weeks 4–7; indicated in Figure 2], following treatment [Durable Engraftment, Weeks 8–14], and over the entire study [Overall Engraftment].

First, budesonide pretreatment was not associated with a higher total engraftment score [$p > 0.5$, Figure 3A], as compared with placebo. Further, none of the individual engraftment metrics showed a difference in engraftment between patients pretreated with budesonide or placebo [Supplementary Figure 1].

Second, we compared the total combined engraftment scores for patients who received FMT from donor D07 to those with FMT from donor D08 [Figure 3B]. Donor D07 was associated with higher total engraftment at all studied time frames [$p < 0.04$], which was driven by a higher SEF [Supplementary Figure 2B] and higher Aitchison similarity [Supplementary Figure 2C]. The median SEI was not different between donors D07 and D08 [$p > 0.05$; Supplementary Figure 2A]. To ensure the budesonide impact on engraftment was not trumped by the donor difference, two-way ANOVAs were performed on acute, durable, and overall engraftment outcomes taking into account both pretreatment and donor assignment. The effect of budesonide pretreatment on engraftment remained insignificant in the acute, durable, and overall time windows [$p = 0.3$, $p = 0.5$, $p = 0.3$, respectively] while the effect of donor remained significant given the patient's pretreatment status [Acute $p = 0.02$, Durable $p = 0.01$, Overall $p = 0.01$, respectively, Figure 3E].

Clinical outcomes

The partial Mayo score did decrease from 5.69 [1.0] at baseline to 4.38 [1.3] after 3 weeks of pretreatment with budesonide, and in the placebo group the partial Mayo score did decrease from 5.64 [1.2] to 4.82 [1.7] after 3 weeks [no significant difference between groups].

At week 10, 42% [10/24] achieved clinical remission, 8% [2/24] achieved partial remission, and 50% of the patients [12/24] had no clinical response or progressive symptoms. Pretreatment with budesonide was associated with a non-significant higher response rate compared to placebo [8/12 vs 4/12 patients, $p = 0.5$] and donor randomization [$p = 0.8$] did not influence the clinical Mayo score at week 10 [Table 2].

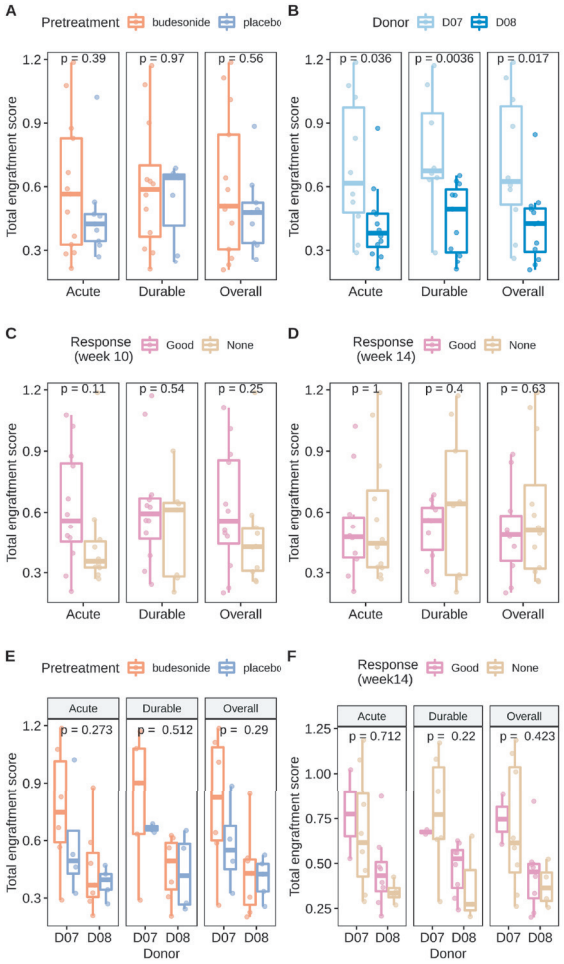


Figure 3. Total engraftment of donor microbiome. To test for an association between bacterial engraftment and pretreatment, donor, or clinical response, we first calculated a single engraftment score, combining metrics using species abundance, number of species, and species composition. We then used the weighted sum of these three metrics as the total engraftment score. [A] Impact of pretreatment on engraftment, [B] impact of donor, [C] association with response at week 10, [D] association with response at week 14, [E] impact of pretreatment while controlling for donor, and [F] association of response at week 14 with engraftment while controlling for donor. Donor D07 is correlated with higher acute, durable, and overall engraftment [$p < 0.05$]. All other associations were not significant [$p > 0.05$]. Panels A–D show p-values from Wilcoxon rank sum exact tests, while panels E and F show p-values from two-way ANOVA; panel E shows values dependent on the pretreatment variable [p-values for donor variable: Acute $p = 0.022$; Durable $p = 0.011$; Overall $p = 0.016$], and panel F shows values dependent on response [p-values for donor variable: Acute $p = 0.018$; Durable $p = 0.0201$; Overall $p = 0.024$]. Acute = 1 week after each FMT; Durable = 2–8 weeks after final FMT; Overall = from 1 week after the first FMT up to 2 months after FMT.

At week 14, 38% [9/24] of patients achieved combined clinical and endoscopic remission, 4% [1/24] achieved partial remission, and 58% [14/24] had no response. Of the ten FMT responders at week 14 [partial remission or remission], four patients were pretreated with placebo versus patients with budesonide [$p = 1.0$]. Of note, at week 14, there was a donor-dependent effect, with 80% [8/10] of the responders receiving donor material from D08 compared with 20% [2/10] receiving donor material from D07 [$p < 0.05$, Table 2]. The 14 non-responders were treated with budesonide [$n = 2$], prednisolone [$n = 4$], anti-TNF-α [$n = 2$], or chose themselves not to start a new induction therapy [$n = 3$]. Three patients were admitted to the hospital due to worsening of their complaints, and responded well on intravenous prednisolone and anti-TNF-α.

Table 2. Overview of clinical outcomes and the effect of pretreatment and donors in the [partial] responder group.

	Week 10	*p*-value	Week 14	*p*-value
Clinical outcomes [*n*]				
No response	50% [12/24]		58% [14/24]	
Partial remission	8% [2/24]		4% [1/24]	
Clinical remission	42% [10/24]		38% [9/24]	
[Partial] remission*	*N* = 12		*N* = 10	
Pre-treatment [*n*]		0.508		1.0
Placebo	4/12		4/10	
Budesonide	8/12		6/10	
Donors [*n*]		0.828		**<0.04**
D07	5/12		2/10	
D08	7/12		8/10	

*The effect of pretreatment/donors in the [partial] responder group.

N = number of patients.

Quality of life

In the responder group [$n = 10$], the median score of the EQ-5D-5L increased significantly at the end of the study compared to baseline, independent of pretreatment [see Supplementary results]. However, assessment of changes in quality of life in non-responders was unfortunately not possible, as some of them did not complete study follow-up due to progressive symptoms.

Adverse events

In total, 24 adverse events were observed [Table 3]. Of these, 16 events were listed as possibly related to FMT, such as abdominal cramps after infusion and constipation. Eight events were registered as not related to FMT, including: high blood pressure and

high glucose level that were considered incidental findings and were treated by the general practitioner, weight gain that was attributed to cessation of smoking at the same time, stasis of food in the stomach resulting in the postponement of FMT by 1 week, and pain in a foot which was treated with paracetamol. Two cases of anaemia were attributed to ongoing inflammation and were treated with iron supplementation. None of the adverse events mentioned in Table 3 led to hospitalization.

Influence of total engraftment on clinical response

Clinical response was not associated with higher total engraftment. Total engraftment did not differ between responders and non-responders at week 10 or week 14 [Figure 3C and D; all $p > 0.05$], although there was a trend towards more engraftment in responders at week 10 [$p = 0.25$]. Separate engraftment metrics were also not found to differ between patients with a good clinical response and those with no response [Supplementary Figures 3 and 4]. To test whether the association between engraftment and response might be donor-dependent, a two-way ANOVA was used [Figure 3F]. This test yielded no significant correlation between response and engraftment [Acute, Durable, and Overall all $p > 0,2$], but only confirmed the significant donor-effect [Acute, Durable, and Overall $p < 0.02$].

Table 3. Overview of adverse events*.

Related to FMT [*n* = 16]	*N*	Not related to FMT [*n* = 8]	*N*
Borborygmus	2	High blood pressure	1
Abdominal cramps	5	Increasing weight	1
Increase of stool frequency	2	High glucose level	1
Vomited after infusion	2	Stasis of food in stomach	1
Constipation	3	Flu-like complaints	1
Nausea	2	Anaemia	2
		Stinging pain in foot	1

FMT = faecal microbiota transplantation, *N* = number.

*None of the adverse events resulted in hospitalization.

Taxonomic associations with FMT response

RFC modelling was used to predict the binary response versus non-response for each patient as a function of the bacterial relative abundance at different time points and aggregated at different taxonomic levels [Supplementary results]. The models considering just baseline samples [Model 1], or samples collected at baseline and during the FMT administration [Model 2], were not found to be predictive of response [Supplementary Figure 5A]. We selected the two models with higher classification accuracy [i.e. Model 3 and Model 4 both at genus-level aggregation] [Supplementary

Figure 5A] to inspect microbiome signatures that are predictive of response. Model 3 considers time points up to week 8, and Model 4 considers time points up to week 14. In both models, the abundance of several genera belonging to Clostridia clusters IV and XIVa [e.g. *Roseburia*, *Dorea*] or Bacteroidales [e.g. *Bacteroides*, *Coprobacter*] post-FMT [Week 8–10, 2–4 weeks after the final FMT] or during FMT dosing [e.g. *Roseburia* FMT 2] were highly predictive of response [Supplementary Figures 5B and C]. Similarly, both models also found that the abundance of *Dialister* immediately following FMT to be discriminatory. In addition, for Model 4, five of the nine highly prognostic predictors of response involved enrichment in abundance of Clostridia members [Eubacteriales: *Agathobaculum*, *Lawsonibacter*, *Roseburia*, *Tyzzerella*, *Phascolarctobacterium*] at weeks 7–10 [1–4 weeks after the final FMT, Supplementary Figure 5B]. Similarly, Model 4 found the enrichment in Coriobacteria representatives [e.g. *Enorma*] at weeks 10–14 to be highly discriminatory between responders and non-responders. SIRUS analysis to gather human-interpretable rules for both models found the enrichment in *Dialister* immediately following FMT to be strongly predictive of no response [Supplementary Figure 6]. Conversely higher colonization by *Roseburia* [for both models], *Coprobacter* [for Model 3, Supplemental Figures 5B and 6A] or *Phascolarctobacterium* [Model 4, Supplemental Figures 5C and 6B] to be strongly indicative of response.

Microbiota community differences among FMT donors

Although donors D07 and D08 were both found to contain high abundances of bacterial families [i.e. Bacteroideceae and Akkermansiaceae] with presumed clinical benefit to UC patients and low in presumed antagonistic bacterial families [i.e. Prevotellaceae and Enterococcaceae, Figure 1], the faecal microbiome of the two rationally selected donors were distinct [Figure 4]. Specifically, the donors differed significantly from each other in terms of richness, measured as number of observed species [D07 = 280; D08 = 220; $p < 0.005$, Figure 4A] and as evenness measured as Shannon Diversity [D07 = 4.3; D08 = 3.89; $p < 0.03$, Figure 4B]. Moreover, stool samples clustered by donor, with the overall community structure showing statistically significant differences [PERMANOVA $p = 0.001$, Figure 4C]. The D07 stool microbiome samples had greater *Prevotella*, *Slackia*, *Paraprevotella*, *Asaccharobacteria*, and *Desulfovibrio*. Collectively, these five genera comprised 0.9–9% of the total microbial composition of samples originating from D07. These genera were largely absent in D08 samples with only *Desulfovibrio* detected in five samples at <0.06% abundance. In contrast, the D08 microbiome samples contained *Enorma*, *Faecalimonas*, *Erysipelatoclostridium*, *Phascolarctobacterium*, *Fournierella*, and *Massiliomicroiota* [Figure 4C]. Collectively, these six genera comprised 0.4–1.3% of the overall microbial composition with a median total relative abundance of the six genera collectively of 0.6%. *Erysipelatoclostridium* was the only genus of the six detected in any D07 samples, where it was detected in one sample at 0.04% abundance.

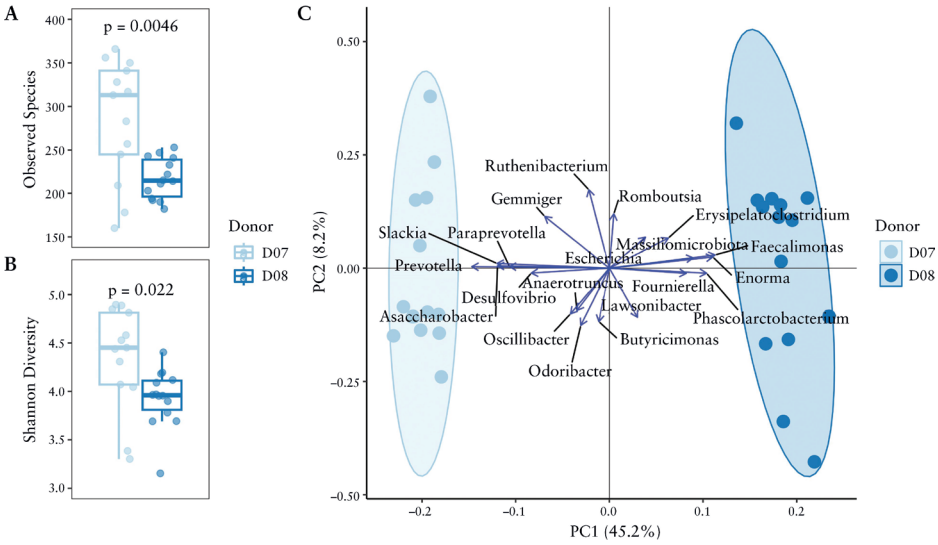


Figure 4. Microbiota community differences among FECBUD FMT donors. Differences in microbial diversity between D07 and D08 samples as number of observed species [A] and Shannon index [B] demonstrate that D07 has higher diversity than D08 [$p < 0.03$]. Thirteen donor samples from D07 and 14 samples from D08 were collected between June 2017 and July 2019. [A] Microbiota community differences between D07 and D08 samples as clusters [C] were significantly different by PERMANOVA [$p = 0.001$].

Discussion

In several randomized controlled trials, microbiota transplantation has been described as a promising induction treatment strategy for patients with UC.[10] However, the optimal selection of donors and patients, and the correlation between engraftment of donor microbes and clinical outcomes is not yet known. In this small, randomized study, overall response or remission after FMT treatment was 42%, which appears similar to previous studies.[10,11] After budesonide pretreatment, the engraftment of donor microbiota and clinical response appeared slightly higher, but these effects were not statistically significant. Also, total engraftment did not predict treatment response significantly. However, our study confirmed the previous observation of a donor-dependent effect of FMT in patients with UC.

A number of studies suggest that FMT success is dependent on the microbial diversity and composition of the stool donor, leading to the proposition of the existence of FMT super-donors.[36] In our study, two donors were rationally selected based on their microbiota profile, the ability to induce Tregs, and the capacity to produce SCFAs *in vivo*. The selected donors had high abundance of bacteria within the families Clostridiaceae, Bifidobacteriaceae, Ruminococcaceae, and Lachnospiraceae, all known

as important components of the gut commensal bacteria in healthy people,7 and Akkermansiaceae, which has been shown to provide additional benefit in UC.[22] In addition, the selected donors had low abundance of Prevotellaceae compared to all other donors considered for the study, as in IBD mouse models colonization with Prevotellaceae seems to be detrimental.[19] Surprisingly, we observed a significant difference in efficacy between those two donors, in favour of donor D08 that showed lower Treg induction in germ-free mice compared to D07. FMT preparations from this donor induced clinical remission in 80% of patients. Similarly, Moayyedi *et al*. found one donor with significantly higher response rates than other donors, with a remission rate of 39% in recipients who received FMT preparations of the successful donor.[11] Of note, the total number of species and the species diversity in our clinically more effective donor were lower. This agrees with recently published observations by Podlesny *et al*.[37] and thereby questions whether microbiome diversity of donors predicts clinical outcome, as suggested previously.[29,30] These findings also suggest that selecting FMT donors based on *in vitro*, microbiome community, or animal model data alone may be insufficient to predict FMT clinical response in patients with IBD. Therefore, more research isolating and identifying bacterial strains responsible for mediating response is required to provide more consistent results for microbiome-mediated therapies in UC patients.

The engraftment of the donor microbiota can be predicted from the abundance of bacteria in the donor and the pre-FMT patient,[37,38] and is thereby considered as an informative endpoint in FMT studies. However, quantifying the degree of bacterial engraftment is not standardized and the factors that promote engraftment of individual strains remain elusive.[39] A study by Rossen *et al.* demonstrated that engraftment of donor species contributed to the success of FMT, since responders who received allogenic faeces had greater similarity to their donor's microbiota.[12] More recently, the results of the comprehensive meta-analysis by Ianiro *et al.* suggests that higher donor strain engraftment is associated with improved clinical outcomes across a variety of clinical indications but not per indication,[40] whereas the results of another comprehensive meta-analysis did not determine an association.[41] Podlesney *et al*. found that the role of engraftment and clinical outcomes is probably disease-dependent and suggested that IBD patients rather benefit from targeted FMT strategies with the goal to induce engraftment of specific donor species or strains instead of maximizing overall engraftment.[37] Each meta-analysis reassessed engraftment from metagenomes collected in clinical trials in various disease indications, specifically analysing sample 'triads' consisting of donor and pre- and post-FMT samples from recipients. This highlights how, even when using strain-level tracking, there is still considerable heterogeneity in outcomes and dependence of strain tracking and modelling methodologies.[42] Additionally, interpretations of results may be hampered due to considerable variability in study methodologies of FMT trials in IBD, including number of individual donors and their relationship to the patient, FMT delivery

method, frequency of and time intervals between FMT treatments, use of antibiotic pretreatments, and differences in primary endpoint definitions.[10] To study engraftment in our relatively small randomized controlled trial, we used three different metrics to provide insights into the displacement of recipient microbiota by donor, the diversity of donor-derived species engrafting, and the recipient's exposure to donor strains during and after FMT treatment. By combining these three measures we represent different facets of engraftment in one comprehensive metric. All outcomes in the total engraftment score are similar to those observed in the separate analyses, demonstrating that in the current study population, total engraftment is not predictive of reaching a state of remission in UC regardless of the calculation method. Although we think our methods take into account the major concepts of engraftment processes and provide a robust result, we cannot exclude the possibility that other engraftment metrics may correlate better with UC remission. In our pilot study, we did not find a significantly higher total engraftment in patients with a good clinical response after FMT and pretreatment with budesonide did not affect engraftment. More interestingly, we found lower total engraftment in the patients treated by a donor who was associated with better clinical outcomes. This suggests that overall bacterial engraftment may not be the sole indicator for clinical success, but that specific bacteria may be involved in the observed clinical effects of FMT in patients with UC.[28] However, the sample size of this study as well as the resolution of taxonomic assignment are notable limitations. For instance, strain-level tracking of donor and recipient microbiomes may facilitate greater differentiation between recipient- and donor-originating taxa.[31]

Taxonomic associations were further investigated with RFC modelling and we found that the abundance of several genera belonging to Clostridia clusters IV and XIVa were most predictive of response, which is consistent with previous studies.[12,43] Health-associated Clostridia, and their produced metabolites, appear to promote several innate and adaptive anti-inflammatory phenotypes *in vivo*[44] and are associated with reduced inflammatory responses in humans.[45] Remarkably, our analysis also identified that a high abundance of *Dialister* immediately following FMT was predictive of non-response. Interestingly, this genus was found at higher abundance at baseline in the non-responder individuals, which agrees with multiple studies observing that higher abundance of *Dialister* correlates with inflammation or with failed response to anti-IBD therapies in humans.[46] However, models that relied on baseline microbiome composition alone were not predictive of response. Overall, these associations could guide a first step towards further research into targeting key [isolated] bacteria and donor–patient matching, which may ultimately result in a personalized FMT treatment approach.

Limitations apply to the present study. First, the number of patients was relatively small and the difference in clinical response between the two rationally selected donors led to an unbalanced dataset and hindered the comparison between treatment arms. Also,

patients randomized to budesonide pretreatment had significantly more often been treated with anti-TNF-α therapy in the past, which suggests that those patients had a more complicated course of UC prior to study participation. This precludes a definite conclusion regarding the presumed beneficial effects of anti-inflammatory treatment prior to FMT. Furthermore, 3 weeks of budesonide 9 mg has limited anti-inflammatory potential, and may not suffice to show a difference in engraftment between treatment groups. Therefore, an effect of pretreatment or induction therapy on FMT donor engraftment cannot convincingly be excluded. Finally, a standard method to assess engraftment still remains to be determined, which means that conclusions should be interpreted considering limitations of the applied method to assess engraftment. Still, although this was a pilot study with a small group of patients, this study provides new insights into the use of FMT for IBD and may help the design of future FMT trials for patients with UC.

In conclusion, in our small randomized study, pretreatment with 3 weeks of budesonide did not significantly influence engraftment of donor microbiota and engraftment was not associated with clinical outcome. However, a significant donor-dependent effect of FMT in UC was seen, where caution is advised when interpretating this results, with the small number of patients included. Interestingly, patients treated with FMT suspensions from the donor who was associated with clinical response had lower engraftment, suggesting that response might be related to specific microorganisms instead of overall engraftment. Future studies must address the optimal selection of both patients and donors, the timing of FMT, and the combination of FMT with certain immunosuppressants.

Supplementary Data

The study was registered in the Netherlands Trial Register, with reference number NL9858.

Funding

The author[s] received a research grant from Vedanta Biosciences.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank E. K. L. Berssenbrugge, A. Costa, M. Mosho, A. Pepicelli, and V. Yeliseyev for their technical support and experimental execution for the donor selection studies.

Author Contributions

All authors have read and agreed to the published version of the manuscript. Conceptualization and design—E.L., E.T., J.N., H.V., A.M., J.K. Investigation, data curation, project administration and formal analysis: E.L., S.N., E.T., E.C., A.P., S.B., A.W., R.M., R.S., S.M., H.V., A.M., J.K. Writing—original draft preparation: E.L., S.N., A.M., J.K. Writing—review and editing: E.L., S.N., E.T., E.C., A.P., S.B., A.W., G.G., R.M., V.B., J.N., S.M., H.V., A.M., J.K. Final approval of the version to be submitted: E.L., S.N., E.T., E.C., A.P., S.B., A.W., G.G., R.M., R.S., V.B., J.N., C.W., S.M., H.V., A.M., J.K. Ethics approval: Not applicable. Consent to participate: All eligible patients provided informed consent at the participating centres at the moment of FMT according to local requirements. Consent for publication: The manuscript, including related data, figures, and tables, has not been previously published and the manuscript is not under consideration elsewhere.

Data Availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Supplementary documents are available online on the publisher's website:

https://academic.oup.com/ecco-jcc/article/18/9/1381/7640395#supplementary-data

References

1. Chang JT. Pathophysiology of inflammatory bowel diseases. *N Engl J Med* 2020;383:2652–64.
2. Harbord M, Eliakim R, Bettenworth D, et al. *European Crohn's and Colitis Organisation* [ECCO]. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. part current management. *J Crohns Colitis* 2017;11:769–84.
3. Parragi L, Fournier N, Zeitz J, et al. Swiss IBD Cohort Study Group. Colectomy rates in ulcerative colitis are low and decreasing: 10-year follow-up data from the Swiss IBD Cohort Study. *J Crohns Colitis* 2018;12:811–8.
4. Eastaff-Leung N, Mabarrack N, Barbour A, Cummins A, Barry S. Foxp regulatory T cells, Th effector cells, and cytokine environment in inflammatory bowel disease. J Clin Immunol 2010;30:80–9.
5. Alam MT, Amos GCA, Murphy ARJ, Murch S, Wellington EMH, Arasaradnam RP. Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. *Gut Pathog* 2020;12:1–8.
6. Zhang SL, Wang S-N, Miao C-Y. Influence of microbiota on intestinal immune system in ulcerative colitis and its intervention. *Front Immunol* 2017;8:1–11.
7. Rinninella E, Raoul P, Cintoni M, et al. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 2019;7:14.
8. Venegas DP, De la Fuente MK, Landskron G, et al. Short chain fatty acids (SCFAs)-Mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol* 2019;10:277
9. van Nood E, Vrieze A, Nieuwdorp M, et al.Duodenal infusion of donor feces for recurrent Clostridium difficile. *N Engl J Med* 2013;368:407–15.
10. Paramsothy S, Paramsothy R, Rubin DT, et al. Faecal microbiota transplantation for inflammatory bowel disease: a systematic review and meta-analysis. *J Crohns Colitis* 2017;11:S380–S1.
11. Moayyedi P, Surette MG, Kim PT, et al. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology* 2015;149:102–9.e6.
12. Rossen NG, Fuentes S, van der Spek MJ, et al. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. *Gastroenterology* 2015;149:110–8.e4.
13. James D Lewis SC, Lisa N, Gar RL, Faten NA, Jonas HF. Use of the non-invasive components of the Mayo score to assess clinical response in ulcerative colitis. *Inflamm Bowel Dis* 2008;1660:6.
14. Julious SA. Sample size of per group rule of thumb for a pilot study. *Pharm Stat* 2005;4:287–91.
15. Terveer EM, van Beurden YH, Goorhuis A, et al. How to: establish and run a stool bank. *Clin Microbiol Infect* 2017;23:924–30.
16. van Lingen E. TEM, van der Meulen-de Jong AE, Vendrik KEW, et al. Advances in stoolbanking. *Microb Health Dis* 2020;1:1–8.

17. Keller JJ, Ooijevaar RE, Hvas CL, et al. A standardised model for stool banking for faecal microbiota transplantation: a consensus report from a multidisciplinary UEG working group. *United Eur Gastroent* 2021;9:229–47.

18. Elinav E, Strowig T, Kau AL, et al. NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 2011;145:745–57.

19. Iljazovic A, Roy U, Gálvez EJC, et al. Perturbation of the gut microbiome by Prevotella spp. enhances host susceptibility to mucosal inflammation. *Mucosal Immunol.* 2021;14:113–24.

20. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol* 2017;19:29–41.

21. Baldelli V, Scaldaferri F, Putignani L, Del Chierico F. The role of Enterobacteriaceae in gut microbiota dysbiosis in inflammatory bowel diseases. *Microorganisms* 2021;9:697.

22. Bian XY, Wu W, Yang L, et al. Administration of *Akkermansia muciniphila* ameliorates dextran sulfate sodium-induced ulcerative colitis in mice. *Front Microbiol* 2019;10:2259.

23. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474:307–17.

24. Hinz A, Kohlmann T, Stöbel-Richter Y, Zenger M, Brähler E. The quality of life questionnaire EQ-5D-5L: psychometric properties and normative values for the general German population. *Qual Life Res* 2014;23:443–7.

25. Dsouza M, Menon R, Crossette E, et al. Colonization of the live biotherapeutic product VE and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host Microbe* 2022;30:583–98.e8.

26. Ott SJ, Waetzig GH, Rehman A, et al. Efficacy of sterile fecal filtrate transfer for treating patients with clostridium difficile infection. *Gastroenterology* 2017;152:799–811.e7.

27. Tamburini FB, Andermann TM, Tkachenko E, Senchyna F, Banaei N, Bhatt AS. Precision identification of diverse bloodstream pathogens in the gut microbiome. *Nat Med* 2018;24:1809–14.

28. Paramsothy S, Nielsen S, Kamm MA, et al. Specific Bacteria and metabolites associated with response to fecal microbiota transplantation in patients with ulcerative colitis. *Gastroenterology* 2019;156:1440–54.e2.

29. Vermeire S, Joossens M, Verbeke K, et al. Donor species richness determines faecal microbiota transplantation success in inflammatory bowel disease. *J Crohns Colitis* 2016;10:387–94.

30. Kump P, Wurm P, Gröchenig HP, et al. The taxonomic composition of the donor intestinal microbiota is a major factor influencing the efficacy of faecal microbiota transplantation in therapy refractory ulcerative colitis. *Aliment Pharmacol Ther* 2018;47:67–77.

31. Smith BJ, Piceno Y, Zydek M, et al. Strain-resolved analysis in a randomized trial of antibiotic pretreatment and maintenance dose delivery mode with fecal microbiota transplant for ulcerative colitis. *Sci Rep* 2022;12:5517.

32. Haran JP, Bhattarai SK, Foley SE, et al. Alzheimer's disease microbiome is associated with dysregulation of the anti-inflammatory P-Glycoprotein pathway. *Mbio* 2019;10:e00632–19.

33. Kursa MB, Rudnicki W. Feature selection with the boruta package. *J Stat Softw* 2010;36:1–13.

34. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26:1340–7.

35. Clément Bénard GB, da Veiga S, Scornet S. Interpretable Random Forests via Rule Extraction 20hal-02557113v1.

36. Wilson BC, et al. The super-donor phenomenon in fecal microbiota transplantation. *Front Cell Infect Mi.* 2019;9:2.

37. Podlesny D, Durdevic M, Paramsothy S, et al. Identification of clinical and ecological determinants of strain engraftment after fecal microbiota transplantation using metagenomics. *Cell Rep Med* 2022;3:100711.

38. Smillie CS, Sauk J, Gevers D, et al. Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal microbiota transplantation. *Cell Host Microbe* 2018;23:229–40.e5.

39. Toit A. Principles of microbiota engraftment. *Nat Rev Microbiol* 2018;16:186.

40. Ianiro G, Punčochář M, Karcher N, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. *Nat Med* 2022;28:1913–23.

41. Schmidt TSB, Li SS, Maistrenko OM, et al. Drivers and determinants of strain dynamics following fecal microbiota transplantation. *Nat Med* 2022;28:1902–12.

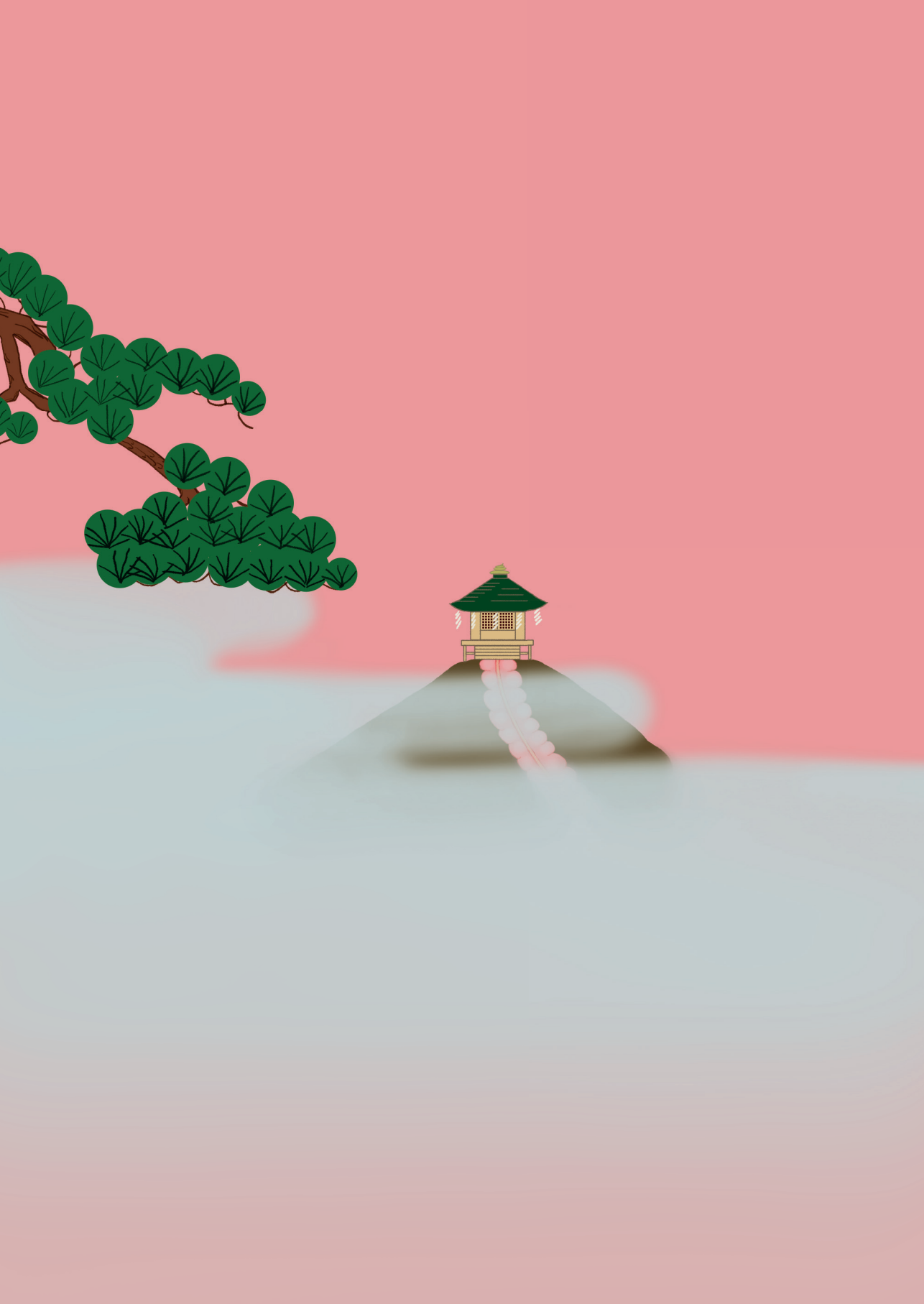
42. Lavelle A, Sokol H. Understanding and predicting the efficacy of FMT. *Nat Med* 2022;28:1759–60.

43. Nishida A, Inoue R, Inatomi O, Bamba S, Naito Y, Andoh A. Gut microbiota in the pathogenesis of inflammatory bowel disease. Clin J *Gastroenterol* 2018;11:1–10.

44. Atarashi K, Tanoue T, Oshima K, et al. T-reg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 2013;500:232–6.

45. Olendzki B, Bucci V, Cawley C, et al. Dietary manipulation of the gut microbiome in inflammatory bowel disease patients: Pilot study. *Gut Microbes* 2022;14:2046244.

46. Ventin-Holmberg R, et al. The gut fungal and bacterial microbiota in pediatric patients with inflammatory bowel disease introduced to treatment with anti-tumor necrosis factor-alpha. *Sci Rep* 2022;12:6654.



Chapter 5

Dynamics of Gut Microbiota after Fecal Microbiota Transplantation in Ulcerative Colitis: Success Linked to Control of *Prevotellaceae*

Susanne Pinto[1], Dominika Šajbenová[1], Elisa Benincà[2], Sam Nooij[3], Elisabeth M. Terveer[3], Josbert J. Keller[4], Andrea E. van der Meulen – de Jong[4], Johannes A. Bogaards[5]*, and Ewout Steyerberg[1]*

[1] Leiden University Medical Center, Biomedical Data Sciences, Leiden, The Netherlands
[2] National Institute for Public Health and the Environment (RIVM), Centre for Infectious Disease Control, Bilthoven, The Netherlands
[3] Leiden University Medical Center, Center for Infectious Diseases (LU-CID), Leiden, The Netherlands
[4] Leiden University Medical Center, Gastroenterology and Hepatology, Leiden, The Netherlands
[5] Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, Amsterdam, Netherlands; Institute for Infection and Immunity (AII), Amsterdam, The Netherlands
* These authors contributed equally

Journal of Crohn's and Colitis, *jjae137, DOI: 10.1093/ecco-jcc/jjae137, 3 September 2024*

Abstract

Background

Fecal microbiota transplantation (FMT) is an experimental treatment for ulcerative colitis (UC). We aimed to study microbial families associated with FMT treatment success.

Methods

We analyzed stools from 24 UC patients treated with four FMTs weekly after randomization for pretreatment during three weeks with budesonide (n = 12) or placebo (n = 12). Stool samples were collected nine times pre-, during, and post-FMT. Clinical and endoscopic response was assessed 14 weeks after initiation of the study using the full Mayo score. Early withdrawal due to worsening of UC symptoms was classified as non-response.

Results

Nine patients (38%) reached remission at week 14, and 15 patients had a partial response or non-response at or before week 14. With a Dirichlet Multinomial Mixture model we identified five distinct clusters based on the microbiota composition of 180 longitudinally collected patient samples and 27 donor samples. A Prevotellaceae-dominant cluster was associated with poor response to FMT treatment. Conversely, the families Ruminococcaceae and Lachnospiraceae were associated with a successful clinical response. These associations were already visible at the start of the treatment for a subgroup of patients and were retained in repeated measures analyses of family-specific abundance over time. Responders were also characterized by a significantly lower Simpson dominance compared to non-responders.

Conclusions

The success of FMT treatment of UC patients appears to be associated with specific gut microbiota families, such as control of Prevotellaceae. Monitoring the dynamics of these microbial families could potentially be used to inform treatment success early during FMT.

Clinical trial registration number

This research project was reviewed and approved by the Medical Ethical Committee in the LUMC, with reference number NL 65976.098.18. The study was registered in the Netherlands Trial Register, with reference number NL9858.

Introduction

Ulcerative colitis (UC) is a chronic inflammatory disorder affecting the colon. Symptoms experienced by patients during disease exacerbation include bloody stools, diarrhea, and abdominal pain.[1] The etiology of UC is multifactorial, involving a complex interplay between the host immune system, gut microbiota, and genetic and environmental factors.[2-4] UC patients exhibit a reduced microbial diversity and alterations in the composition of their gut microbiota compared to healthy individuals.[4, 5] Notably, a decrease in Bacillota (formally Firmicutes), especially Clostridia (such as *Clostridium, Roseburia,* and *Faecalibacterium)*, and Verrucomicrobia, along with an overgrowth of species from the Enterobacteriaceae family (such as *Escherichia coli* or *Klebsiella spp.*), have been observed.[6-9] Studies investigating associations with common Bacteroidota in the human gut, such as the Bacteroidaceae and Prevotellaceae families, have yielded conflicting results.[6, 7, 9-13]

The current approach to treat UC focuses on attenuating the hyperactive immune response using pharmaceutical drugs, such as local immune suppression with 5-aminosalicylates (5-ASA), or systemic immune suppression with prednisolone, thiopurines, biologics, or small molecules.[6] However, many patients do not derive lasting benefits from these interventions and may even experience severe side effects.[14] Fecal microbiota transplantation (FMT) has emerged as a promising alternative treatment for microbiota-associated disorders, particularly in the treatment of recurrent *Clostridioides difficile* infection.[15-17] FMT involves transferring fecal matter from a healthy donor to a patient with the aim of modulating the microbiota composition towards a more favorable state. The effectiveness of FMT in UC is limited, with a lower response rate observed as compared to FMT treatment of *Clostridioides difficile* infection.[18, 19] A recent meta-analysis comprising six randomized controlled trials (RCT) reported a short-term clinical response in only half of the patients with active UC following FMT administration.[19] The specific host factors influencing successful FMT response in UC are still unclear, and the donor characteristics that influence patient response to clinical success after FMT remain uncertain.[20, 21] In this study, we explore differences in gut microbiota dynamics between patients with clinical remission and non-responders following FMT treatment. Recently, our group performed an RCT assessing the effects of pretreatment with budesonide on FMT outcome and engraftment of donor bacteria.[22] Budesonide reduces inflammation, and therefor might promote greater engraftment of the healthy donor microbiota, as inflammation can disrupt microbiota homeostasis. Interestingly, the primary analysis showed no association between pretreatment or overall engraftment with clinical response, but there was a significant donor-dependent effect on engraftment.[22] Although the study was not powered to detect differences with regard to clinical endpoints, here we aimed to further identify longitudinal associations between the microbiota compositions and clinical response to FMT treatment.

Materials and methods

The study population

For the current study we used the stool samples collected of the 24 UC patients included in our previously described FMT trial (Supplementary file 1).[22] Patients were randomly assigned to be pretreated daily for three weeks with oral budesonide (9 mg) or with a placebo, and for treatment with FMT suspensions from donor D07 or D08 (block randomization). Inclusion criteria were: at least 18 years old and a confirmed diagnosis of mild to moderate UC (i.e., a full Mayo score (including partial Mayo score and endoscopic sub score of 1 or 2) ranging from 4 to 9). Exclusion criteria were among others proctitis, antibiotic use or surgery within the last 6 weeks, or received other treatments within 12 weeks prior to study entry.

The following clinical and demographic information was collected for each patient in the study (Supplementary file 1): sex, age at baseline (years), donor ID (D07 or D08), pretreatment (placebo or budesonide) and clinical outcome at week 14. Patients that did not complete the study because of progressive symptoms/disease were considered treatment failures and classified as non-responders. At week 14, nine patients were in clinical and endoscopic remission (hereafter called responders), 14 patients were non-responder, and one patient was a partial responder. We included this last patient in the non-responder group.

Clinical and laboratory procedures

Patients received a weekly FMT for four times (at end of week 3, 4, 5, and 6) from the Netherlands Donor Feces Bank (NDFB), either from donor D07 or donor D08 following standard protocols for donor screening, sample collection, sample preparation, sample storage and FMT infusion.[23] The samples used for the different FMTs came from different donations. Before every FMT the patients fasted for at least six hours. A bowel lavage with two liters of macrogol solution (Kleanprep) was performed one day before the first FMT to cleanse the intestine. No changes in diet or medication were reported during the study.

Stool samples of the patients were collected once at baseline, once after the pretreatment phase (but still before the FMT treatment), one week after every FMT (four times; designated Post-1 to Post-4) and three times as a follow-up, at 8, 10 and 14 weeks after randomization.[22] In total we collected 81 stool samples in the responder group (n = 9) and 99 stool samples in the non-responder group (n = 15). Stool samples of donors D07 and D08 were collected regularly and a total of 27 samples (n = 13 for donor D07 and n = 14 for donor D08) were used for analysis.

Microbiota composition

DNA was extracted from the collected stool samples (both from the donors and recipients) and sequenced by Diversigen (new Brighton, USA) with the Illumina NovaSeq platform (100 bp single-end reads to a median depth of 2.9m reads). Raw reads mapping to the human genome were removed using bowtie2 (version 2.4.2)[24] and the GRCh37 reference genome and reads were quality-trimmed using fastq (version 0.20.1)[25], both of which are part of an in-house workflow (https://git.lumc.nl/snooij/metagenomics-preprocessing). The mOTUs3 workflow (version 3.0.1) was used to generate taxonomic profiles.[26, 27] Unassigned, human-derived, Archaeal, and low-quality reads were removed from the data, which resulted in 93 different families (i.e., 1552 unique mOTUs). The database of mOTUs3 includes taxa based on metagenomic bins that have not formally been classified. These are listed as '*incertae sedis*' or '*i.s.*'. Due to the sparsity of the data and the relatively small number of patients, the analyses performed at taxonomic genus rank lacked the statistical power needed to provide robust and reliable results. For this reason, the data were aggregated to family level prior to the statistical analysis. All analyses were performed using R software (R version 4.2.2).

Differences in relative abundances of specific microbial families amongst responders and non-responders were tested for statistical significance in repeated measurements analyses, as described below in the 'longitudinal models' section. The average relative abundances of the same bacterial families were calculated for each donor from multiple samples, considering the donor samples were not collected at the same timepoints as the patient samples. Differences amongst donor D07 and donor D08 were tested with Pearson's chi-squared test and the p-values were corrected for multiple hypothesis testing with the Bonferroni method.

Principal component analysis

We performed Principal Component Analysis (PCA) on the Aitchison distances calculated between each pair of patient microbiota profiles. The Aitchison distance is often used in microbiota data because it takes into account the compositionality of the data.[28, 29] The Aitchison distance involved each patient sample undergoing the centered-log-ratio (clr) transformation and then obtaining the Euclidean distance between each pair of samples as implemented in the microViz R package.[30]

Dirichlet Multinomial Mixture models

We used the Dirichlet Multinomial Mixtures (DMM) clustering algorithm to identify distinct clusters of samples based on their microbial abundance profiles. DMM assumes that the microbial abundances in each sample follow one of a given number of multinomial distributions, the number of which is determined by the assumed number of clusters in the data. We used the dmn function from the DirichletMultinomial R package to cluster patient and donor samples.[31] The parameters of the different clusters are estimated by maximizing the likelihood of the observed data given the

assumed model, with a Dirichlet prior for relative abundances of the bacterial families to facilitate parameter estimation and prevent overfitting. The prior consisted of a mixture of Dirichlets with $k = 1, ..., K$ to represent the K clusters, with hyperparameters denoting cluster-specific weights and relative abundances. Next, the bacterial families in each cluster were ranked based on the posterior difference between the cluster in a multi-cluster solution versus a one-cluster model. A more detailed description of DMM models is presented elsewhere.[32] Considering that the DMM clustering algorithm uses stochastic likelihood optimization with random initial parameter values, we performed the clustering algorithm 1000 times and chose the model with the lowest Laplace value, indicating a better parsimonious fit of the model to the data.

Data were clustered according to a combination of patient and donor samples. As a sensitivity analysis, we also applied the algorithm in the following situations: patient samples only; patient samples excluding a patient who was placed in a distinct cluster compared to all other patients (patient 102); patient samples excluding patients who both had only two samples available (patients 109 and 117).

Longitudinal models of bacterial relative abundances

Mixed models were used to model the changes over time in relative abundance for each of the 15 most abundant bacterial families in the patient samples. Regarding the distribution of relative abundance, many families had a high proportion of zeros, resulting in right-skewed distributions. All abundances, except for Ruminococcaceae, were therefore transformed with an arcsine square root transformation to approximate normally distributed data. We modelled the relative abundances of the 15 selected bacterial families separately in 15 different longitudinal models with a linear mixed effect model (LMM), possibly augmented with a zero-inflation component (ZILMM). The lme4 package was used for constructing LMMs and the glmmTMB package was used for constructing ZILMMs.[33, 34] To account for the correlation of repeated observations within each patient, both random slopes and random intercepts were considered as potential models for each bacterial family. Note that the dataset was too small for specification of predictors in the zero-inflation component. To incorporate possible non-linearity in relative abundance trajectories over time into the model, a natural cubic spline (with the ns function from the splines package in R) with node at week 8 (the beginning of the follow-up phase) was considered for all models. Model preference was based on the lowest AIC and model diagnostics, judged by QQ-plot and a plot of residuals against predicted values. All choices per family are given in Supplementary file 2.

The longitudinal models further included the variables clinical outcome (non-responder vs. responder), time (possibly with a cubic spline), and an interaction with time and clinical outcome (non-responder vs. responder). The interaction determined whether there was a divergence in relative abundance of a particular family between non-responders and responders, with statistical significance assessed by Wald tests.[35]

The inclusion of the patient specific variables donor (donor D07 vs. D08), pretreatment (budesonide vs. placebo), age and sex (female vs. male) in the model was dependent upon testing their role as confounders or contribution to the model fit. This assessment involved examining whether their inclusion led to a greater than 15% change in the primary coefficients (notable influence on the model's outcome) or a significant Likelihood Ratio Test (contribution of the variable to the model); with flexibility allowed for a variable to meet one of these criteria during the evaluation process.

Simpson dominance

Simpson dominance was used to summarize microbiota diversity of each sample. We calculated this measure (the sum of the squared relative abundances) with the dominance function from the microbiome package.[36] The Simpson dominance estimates the probability that two random entities taken from a sample represent the same bacterial family within a patient's microbiota. Hence, a higher Simpson dominance means a higher concentration of species from the same family in the sample, which corresponds with a less diverse microbiota. To account for the correlation of repeated observations within each patient, the Simpson dominance was modelled with a random-intercepts LMM (with the lme function from the nlme package).[37] A log transformation was applied to the Simpson dominance measure to correct for non-normality. The regression parameter of primary interest was the relation between Simpson dominance and clinical response, either as a main effect (denoting baseline differences in diversity) or in interaction with time (denoting divergence in diversity between responders and non-responders over time). Additional parameters included the effects sex and time. Similar to the longitudinal LMM of bacterial families, time was modelled as a continuous variable with a natural cubic spline (node at week 8). The effects of pretreatment, donor, and age were negligible and therefore not included in the model. Wald tests were performed to test for statistical significance of the clinical response variables jointly in the model.

Results

Microbiota community composition of donors, responders, and non-responders

The fecal microbiota composition between the two donors was distinctly different (-ure 1 and Supplementary file 3). Donor D07 had a significantly higher relative abundance of the families Clostridiaceae, Clostridiales fam. *i.s.* (i.e., an unclassified family within the order Clostridiales), Ruminococcaceae, and Veillonellaceae compared to donor D08, while donor D08 had a significantly higher relative abundance of Bacillota fam. *i.s.* and Lachnospiraceae (Figure 1, Supplementary file 3, and Supplementary file 4).

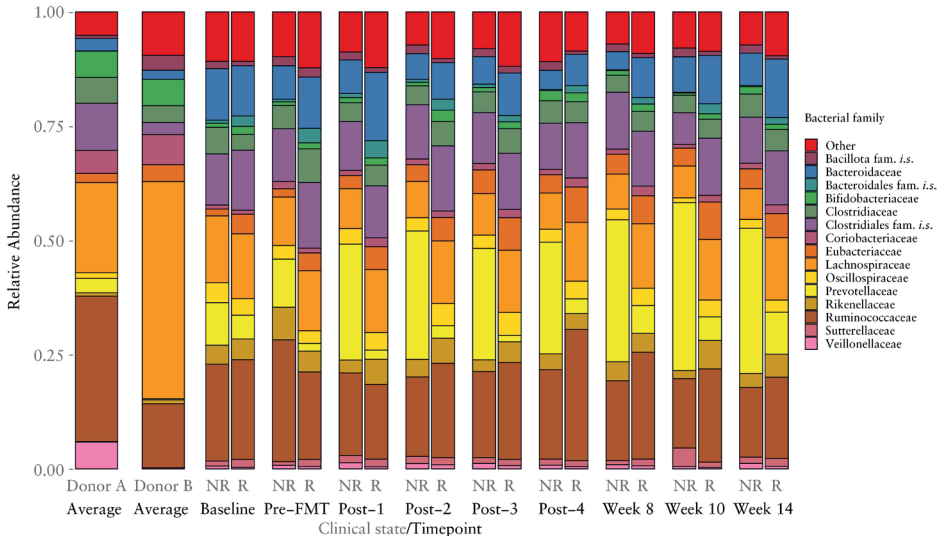


Figure 1. Average microbiota composition of the 15 most abundant bacterial families. Abundances are followed over time for the two donors, non-responders (NR) and responders (R). Here, the 'other' category includes all remaining bacterial families.

Overall, the most abundant bacterial family in the patients was Ruminococcaceae. However, from the second timepoint onwards, the relative abundance of Prevotellaceae continued to increase in the microbiota of the non-responders. Prevotellaceae overtook Ruminococcaceae as the most abundant family for non-responders at Post-1 and remained the most abundant for the remaining timepoints (Figure 1, Supplementary file 3, and Supplementary file 5). Compared to the non-responders, Lachnospiraceae and Oscillospiraceae seemed to become more abundant in the responder group over time (Figure 1, Supplementary file 3, and Supplementary file 5).

PCA results for donors and patients

The first two components in PCA analysis of patient and donor samples, based on the Aitchison distance, explained 24% of the total variation in the data (Figure 2). The samples of donor D08 clustered away from the patients' samples, driven by a difference in the relative abundance of Lachnospiraceae (Figure 2). Patients treated with an FMT from donor D08 showed a higher responder rate than those from donor D07 (Supplementary file 1). The difference in distance between non-responders and responders seemed to be explained by the relative abundance of Prevotellaceae (Figure 2). This applied particularly to the patients who received an FMT from donor D08 (Supplementary file 6). Only a few of the patient samples seemed to traverse considerable Aitchinson distance over time. Notably, the patients whose microbiota became more donor-like over time were more often non-responders (e.g. patients 110 and 111) (Supplementary file 7).

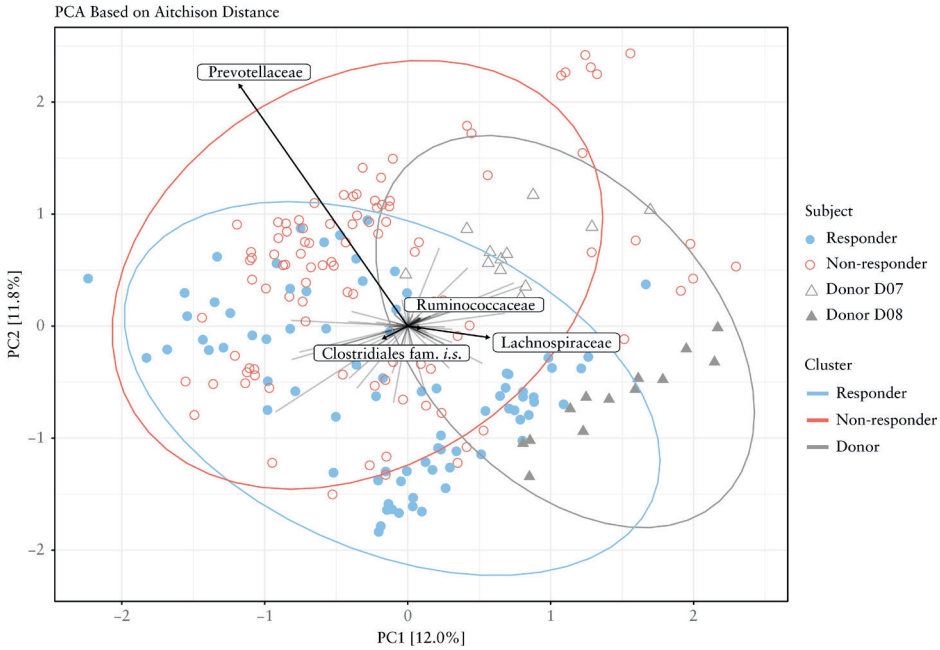


Figure 2. PCA plot with Aitchison distances in microbiota profiles for the distance between sample types. Responders are given in blue, non-responders in green and donors in grey. The PCA plots include data ellipses around the different groups and loading vectors of families to obtain an initial visualization about the extent of separation between non-responder, responder, and donor samples. The different symbols, closed circles, open circles, open triangles, and closed triangles, indicate responder, non-responder patients, donor D07, and donor D08, respectively.

Sample clustering with Dirichlet Multinomial Mixture models

Over 1000 iterations, a five-clusters model was selected as the best-fitting model (i.e., having the lowest Laplace value). Figure 3 and Supplementary file 8 show that Ruminococcaceae was present in all clusters whereas Lachnospiraceae, Bacteroidaceae and Clostridiales fam. *i.s.* were present in four of the five clusters. The relative abundances of those families in each cluster differed: clusters 1 and 4 were dominated by Ruminococcaceae and Lachnospiraceae, whereas clusters 2 and 5 were dominated by Ruminococcaceae and Clostridiales fam. *i.s.*. Prevotellaceae was the only family almost defining an entire cluster (cluster 3). Cluster 1 appeared to be associated with a successful clinical response, while cluster 3 appeared to be associated with non-response (Figure 4). For the patient samples, 56% of responder samples were classified into cluster 1, and 38% into cluster 2, whereas 42% of non-responder samples were classified into cluster 3 (Figure 4B). All donor samples, except for one, were assigned to cluster 4 (Supplementary file 9). Five non-responder patient samples were also assigned to cluster 4 (Figure 4A). This donor-dominated cluster disappeared in sensitivity analysis on patient samples only (Supplementary file 10A), resulting in the re-assignment

of the corresponding patient samples to cluster 2. Patient 102 was responsible for the existence of a separate cluster (cluster 5), with all its measurements belonging to that cluster. Removal of this patient in sensitivity analysis resulted in the deletion of that cluster, with re-assignment of the other corresponding samples to cluster 2 (Supplementary file 10B). Removal of patients with only two measurements (107 and 119) had minor impact on the results (Supplementary file 10C).

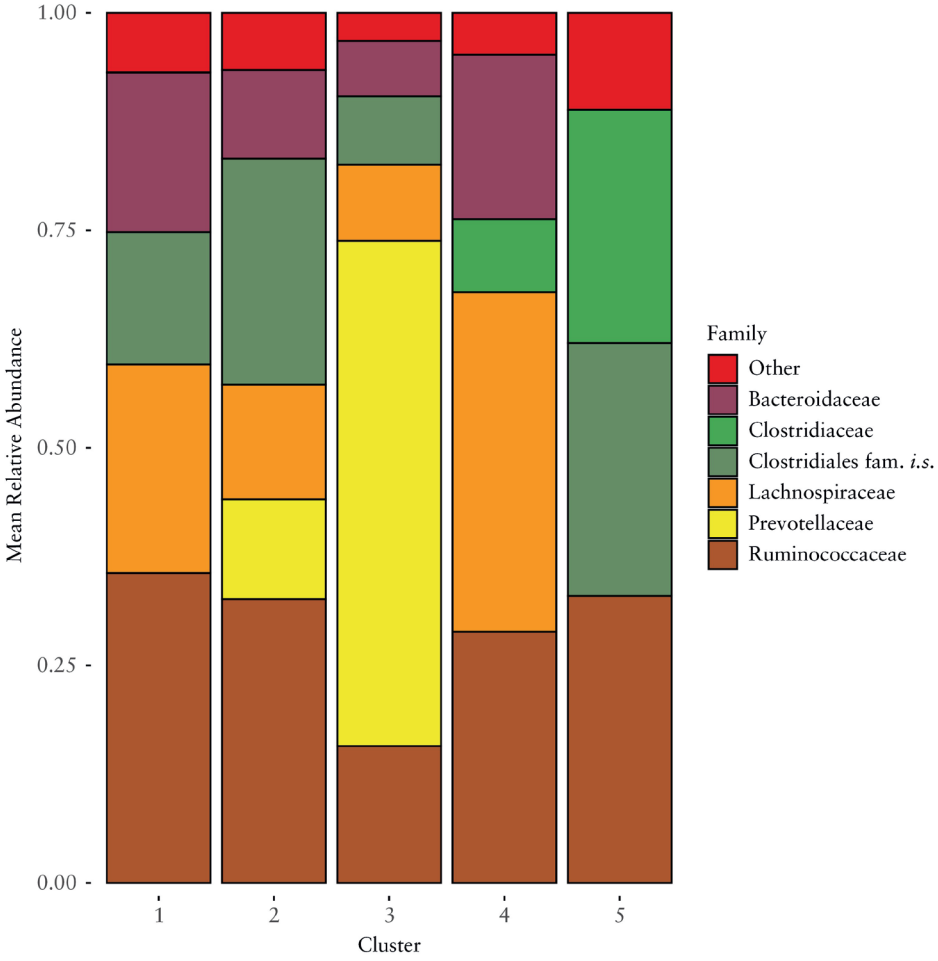


Figure 3. Mean relative abundance of bacterial families in the five clusters. Clusters are detected by the Dirichlet Multinomial Mixture model.

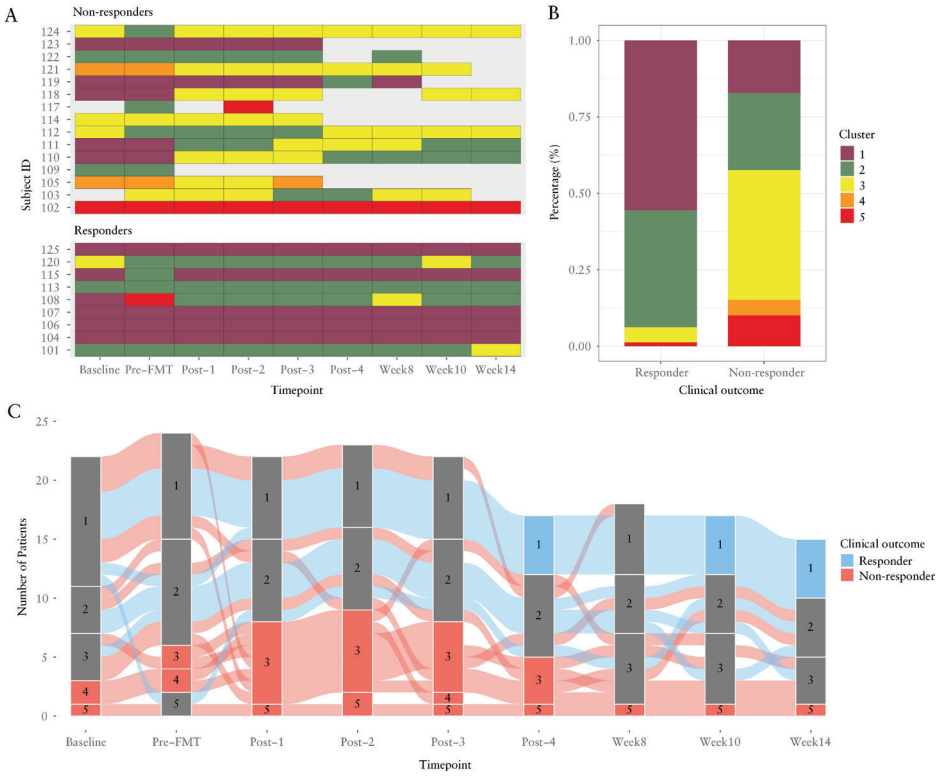


Figure 4. Clustering of donor and patient samples. A) Cluster membership over time per patient for the non-responders (upper facet) and responders (lower facet). Lack of colored bar indicates that no stool sample was collected at that timepoint. B) Percentage of each cluster for non-responders and responders. C) Alluvial plot of patients distributed over the different clusters over time. This plot displays the distribution of clusters per timepoint and whether each cluster is comprised of only one clinical group (e.g., only non-responders) for every timepoint. A grey box means that the cluster at that timepoint contains both samples from responder and non-responder patients, a red box only contains non-responder samples, and a blue box only contains responder samples.

Out of 24 patients, only nine patients (38%) remained in the same cluster for all of their provided samples (Figure 4A). An alluvial plot of patient samples showed the substantial changes in sample membership and cluster size throughout the clinical trial (Figure 4C). There was a mixture of non-responder and responder samples in cluster 1 at the beginning, with most samples at baseline being classified into cluster 1. There was then a shift towards more responder samples in cluster 1 from Pre-FMT onwards. Samples in cluster 1 were exclusively composed of responder samples at timepoints Post-4, Week 10, and Week 14. Cluster 3 was fully composed of non-responder samples after pretreatment and after every FMT treatment (Figure 4C).

Coloring samples by their cluster membership in the Aitchison distances PCA plot showed separation between clusters 1, 2, and 3, with cluster 2 being the intermediate cluster (Supplementary file 9). The Prevotellaceae vector was pointed in the direction of cluster 3, corresponding to a potential association between this cluster and non-response (Supplementary file 9), possibly driven by the donor (Figure 2 and Supplementary file 6). There appeared to be some separation between donor samples, a majority of which were in cluster 4, and patient samples. Donor D08 samples were close to cluster 1 samples. Meanwhile, donor D07 samples were positioned near cluster 2 samples (Supplementary file 9). Finally, samples from cluster 5 were grouped together tightly, likely as a result of belonging to the same patient.

Mixed models of bacterial families

Responders and non-responders showed significantly different trajectories in relative abundance over time for the families Prevotellaceae, Lachnospiraceae, Ruminococcaceae, Oscillospiraceae, and Sutterellaceae (Figure 5, Supplementary file 2, and Supplementary file 5). Prevotellaceae showed the greatest difference in trajectory between non-responders and responders over time. Note that the preferred model for Prevotellaceae had a straightforward linear trajectory and used the original time variable instead of splines. The family Prevotellaceae consisted of four named genera, of which *Prevotella* (especially *Prevotella copri*) was the most abundant (Supplementary file 11).

There were four families with a significant donor effect: Veillonellaceae, Rikenellaceae, Sutterellaceae, and Bifidobacteriaceae (Supplementary file 2). Notably, removal of the donor variable from the model for Sutterellaceae diminished the significance of the main effect related to clinical response. This observation underscores the role of the donor variable in influencing the association between Sutterellaceae and clinical response. Rikenellaceae and Bacillota fam. *i.s.* had a significant sex effect, Veillonellaceae had a significant pretreatment effect (Supplementary file 2). None of these other significant covariates altered the statistical significance of clinical response. This observation suggests that the estimated associations were not confounded by these covariates.

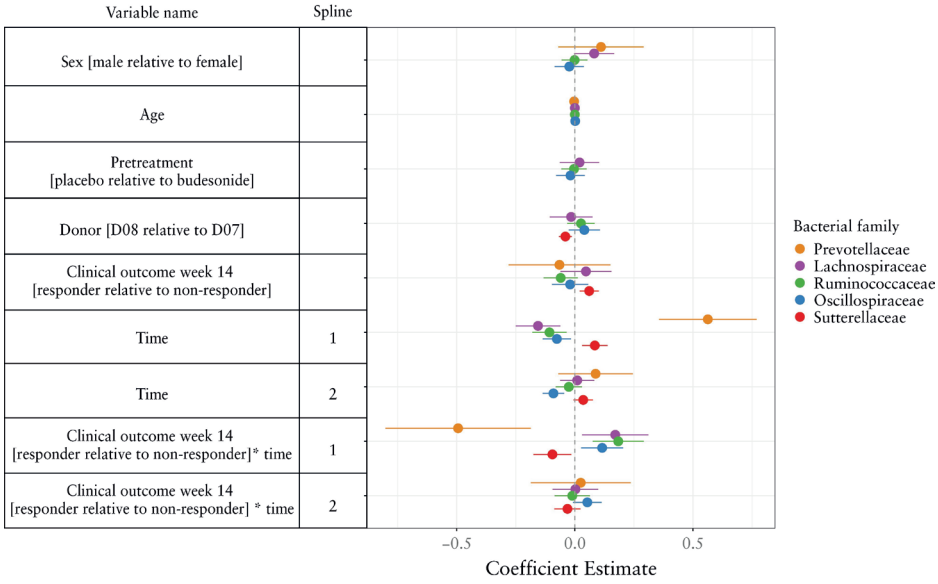


Figure 5. Results of the mixed models. Only the families among the 15 most abundant families (Prevotellaceae, Lachnospiraceae, Ruminococcaceae and Oscillospiraceae) for whom we found a significant effect in relation to clinical response with the Wald test are shown. The point estimates, 95% confidence intervals and a reference line at zero are shown. When the horizontal lines do not cross the vertical reference line, this means that the coefficients are significantly different from 0. P-values are given in Supplementary file 2.

Simpson dominance

The steadily increasing relative abundance of Prevotellaceae in non-responders found before was reflected in the Simpson dominance. Simpson dominance was higher for non-responders compared to responders, especially throughout the follow-up period (Figure 6). There was a significant difference between the Simpson dominance in responder and non-responder patients (Wald test: $p = 0.004$). Our study was too small to determine whether this difference already existed at baseline or developed over time (Supplementary file 12). The LMM random-intercept model suggested that there was also a significant sex effect (Supplementary file 12). However, sex did not alter the statistical significance of clinical response. This observation suggests that the estimated associations were not confounded by the sex of the patients.

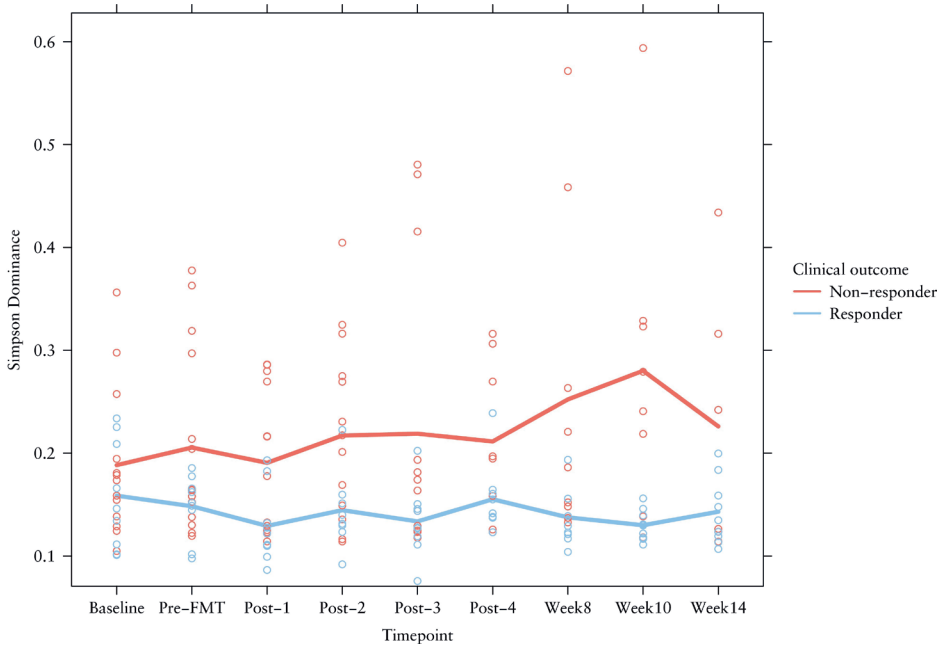


Figure 6. Change in Simpson dominance calculated for non-responders and responders. The points indicate the individual measurements of the patients. The lines are the mean Simpson dominance per group (non-responders in red and responders in blue).

Discussion

Inflammatory bowel diseases (IBD) such as ulcerative colitis (UC) have been linked to alterations in both the composition and metagenomic function of the gut microbiota.[4, 5] In this study, we employed a wide range of analytical techniques to investigate potential associations between microbiota and clinical outcomes following FMT in UC patients. A subgroup of the cohort (9 of 24 patients) reached a successful combined clinical and endoscopic remission after the FMT treatment, and our results suggest that this response may be related to certain gut microbiota families. Specifically, longitudinal models and cluster analysis of repeatedly measured compositional data indicated that the success of FMT treatment of UC patients appears to be associated with control of Prevotellaceae. Conversely, our analyses also highlighted a potentially beneficial role of Lachnospiraceae and Ruminococcaceae in FMT treatment response. Furthermore, we identified several other bacterial families, including Oscillospiraceae, Enterobacteriaceae, and Sutterellaceae, that exhibited associations with clinical remission. The clustering results indicated that differences in the gut microbiota of responders versus non-responders might already be apparent early during the treatment. If this result can be confirmed by larger studies, clinical success may be predicted from early microbiota analysis after the first FMT treatment and mitigating

actions, for example stopping, personalizing, or changing the treatment, might be envisioned.

Donor-related microbiota characteristics may potentially impact the clinical efficacy of FMT.[38] Intriguingly, we observed marked differences between the donors' and the patients' microbiota. Amongst patients who responded well to FMT, gut microbiota composition did not transition fully to resemble that of the donors at the end of follow-up. This contrasts with earlier studies that suggested that a donor-like microbiota is preferred after FMT treatment[20, 21, 38, 39], and suggest that some complementarity in microbiota compositions between donors and recipients is required for successful clinical response.[12, 21] However, additional research is needed to determine whether attaining a donor-like microbiota by the end of FMT affects the ability of patients to achieve remission. The samples of donor D08 clustered closer to cluster 1 (associated with a successful clinical response), and the samples of donor D07 were closer to cluster 3 (indicating non-response). Note that, an FMT from donor D08 resulted in relatively more treatment success in the patients than donor D07. Also, donor D08 seemed to have a more diverse microbiota than donor D07, although not statistically significant. Donor diversity has been associated with a higher clinical response before.[40] In addition, higher post-FMT diversity has been associated with remission, suggesting that the variety of introduced organisms may promote recovery.[20] It was already noted in [22] that donor D08 was the more successful donor; however, intriguingly, this was the donor with the least engraftment. This observation suggests that the persistent transfer of microbes may not be the prime reason for clinical success. Possibly, the transient exposure to an external microbial community might still induce a beneficial change in the recipient's gut environment. It is also possible that patients who received FMT from donor D08 had a more favorable starting state, while those who received FMT from donor D07 required stronger microbiota changes to move to a more favorable state. Further investigations are warranted to unravel the intricate dynamics underlying the observed outcomes.

This study provides novel evidence for a potential association between control of Prevotellaceae at moderate abundance and favorable clinical outcomes following FMT in UC patients. Additionally, the Simpson dominance measure suggests that Prevotellaceae constituted a sizable proportion of the microbiota in non-responsive FMT patients throughout the course of the clinical trial. Screening the patients (and donors) for Prevotellaceae might improve the response rate. However, previous study suggested that higher levels of *Prevotella* (a genus level within Prevotellaceae) may confer health benefits in UC patients after treatment. For instance, studies on UC patients who underwent drug and surgical treatments, excluding FMT, demonstrated that responders had higher baseline levels of *Prevotella* compared to non-responders.[11] Notably, a previous FMT trial on IBD patients did not report any detrimental effects of increased *Prevotella* abundance, despite observing a substantial increase in this

bacterium in their patients after FMT treatment.[12] They classified *Prevotella* as a colonizing bacterium, as its abundance in patients reached a level comparable to that in the donors. Of note, in our study, responders also maintained levels of Prevotellaceae comparable to donors, but in non-responders there was a clear overgrowth. The conflicting role of *Prevotella* in human health has been attributed to the high diversity within the *Prevotella* genus. While the majority of *Prevotella* species are commonly found in healthy individuals, certain strains may be implicated in disease pathogenesis.[41, 42] For instance, *Prevotella intestinalis* has been shown to induce intestinal inflammation upon colonization in mice.[13] *Prevotella melaninogenica* and *Prevotella oralis* have been characterized as tipping elements.[43] This means that *Prevotella* stands out as a bimodal group, with either a high or low abundance state, and can be a pivotal driver in the context of microbial ecosystem stability. This finding was reiterated in a recent investigation into the involvement of gut microbiota families with Crohn's disease activity, where we found that associations with Prevotellaceae were amongst the most heterogeneous across individual patients.[44]

In contrast to Prevotellaceae, other bacterial families have shown associations with positive clinical outcomes. Specifically, the families Lachnospiraceae, Ruminococcaceae, and Oscillospiraceae have also been found to increase following FMT in patients with UC in other studies.[45] Lachnospiraceae and Ruminococcaceae may play a role in modulating the immune response and inflammatory pathways in the colon.[45] Interestingly, contrary to previous literature, the expected increase in Clostridiaceae among responders was not observed in the present study. This discrepancy in Clostridiaceae abundance may be attributed to variations in FMT protocols employed across different clinical trials or the low number of patients in this study.[46] Additionally, in contrast to the present study, previous research has reported increased abundance of Enterobacteriaceae in UC patients who did not respond to drug and surgical interventions, with higher levels being associated with mucosal inflammation.[11] Discrepancies in Enterobacteriaceae abundance may stem from differences in the types of UC treatments employed, for example when FMT was not involved as a treatment modality.[11] In the context of FMT, a study involving IBD patients who underwent FMT revealed the presence of a dysbiotic *Bacteroides* cluster, as well as an Enterobacteriaceae cluster. Donors were subjected to cluster analysis and categorized into *Prevotella* or *Bacteroides* clusters. Interestingly, the clinical outcome of FMT varied depending on the cluster of both the patients and their respective donors.[12]

The longitudinal study design of our trial, with protocolized data collection across all stages of FMT, enabled a uniquely fine-grained view of gut microbiota dynamics after FMT in UC patients. Our study allowed us to assess changes on an almost weekly basis. RCTs with a strong longitudinal component often involve a smaller number of patients with more frequent repeated measurements, as compared to RCTs that focus on clinical outcomes. For example, in a recent clinical trial 42 patients provided

one stool sample for microbiota analysis before FMT, followed by one sample after FMT.[12] Another clinical trial included 12 patients who submitted stool samples weekly throughout their 12-week FMT treatment and at the 18-week follow-up.[47] A limitation of our study is that the results of statistical analyses should be interpreted with caution due to multiple testing in a small number of patients. Yet, most associations found in cluster analysis were retained in repeated measures analyses where we also accounted for the correlation of repeated observations within each patient. Moreover, despite the relatively small number of patients (n = 24) and donors (n = 2), both DMM and PCA clustering utilize all 180 patient samples and 27 donor samples available, rather than considering observations per patient.

Microbiota data is compositional, high-dimensional, and often zero-inflated.[28, 48] Moreover, the intestinal microbiota exhibits complex interactions, including competition and cooperation, forming intricate networks.[49, 50] These characteristics pose challenges to analytical methods, such as mixed models, which are commonly employed to investigate temporal variation and potential differences in bacterial abundance trajectories among clinical groups. Our analysis was limited by the individual modeling of each bacterial family, neglecting the interplay and interactions between families within the microbiota network. However, results obtained by supervised models of family-specific abundance over time were in line with results obtained by unsupervised methods (PCA and DMM clustering) that use community characteristics. Cluster analysis has been widely employed to explore the relationship between gut microbiota and conditions such as child gut development, depression, obesity, and IBD.[12, 51-53] Conventionally, unsupervised methods are suitable for exploratory analyses.[32] If the distinct clusters that we identified are confirmed in further larger-scale longitudinal analyses, this may lead to tailored diagnosis and treatment approaches based on specific cluster characteristics.[54] In our study, this would for example mean that the FMT treatment is stopped or changed to another donor when patients are found to be in the Prevotellaceae-dominated cluster during the treatment. While clustering techniques provide valuable insights, it is important to recognize that they depend on various choices by the modeler, including cutoffs and priors, which may lead to different clustering results.

Our study is admittedly rather exploratory in nature, but consistently revealed indications of a potential association between controlled abundances of Prevotellaceae with successful clinical and endoscopic remission following FMT treatment in UC patients. Moreover, we also highlighted a potential beneficial role of Lachnospiraceae and Ruminococcaceae. This provides a basis for new hypotheses regarding the role of gut microbiota in UC. Therapeutic interventions may be refined in the future, with early prediction of clinical outcomes and more personalized FMT treatments.

List of abbreviations

UC	Ulcerative colitis
FMT	Fecal microbiota transplantation
mOTU	Metagenomic-based Operational Taxonomic Unit
RCT	Randomized Controlled Trial
i.s.	*incertae sedis*
PCA	Principal Component Analysis
DMM	Dirichlet Multinomial Mixtures
AIC	Akaike Information Criterion
LMM	Linear Mixed effect Model
ZILMM	Zero-Inflated Linear Mixed effect Model

Declarations

Acknowledgements
We would like to acknowledge dr. E. van Lingen for her contribution to the data collection process.

Ethics approval and consent to participate
This research project was reviewed and approved by the Medical Ethical Committee in the LUMC, with reference number NL 65976.098.18. The study was registered in the Netherlands Trial Register, with reference number NL9858.

Consent for publication
Not applicable

Availability of data and material
R code is available via GitHub (https://github.com/susannepinto/FECBUD_microbiome.git) and the in-house preprocessing workflow is available via (https://git.lumc.nl/snooij/metagenomics-preprocessing/-/releases/v1.0). We have uploaded the metagenomic sequences to NCBI with: SRA Bioproject PRJNA1071720

Supplementary documents are available online on the publisher's website:

https://academic.oup.com/ecco-jcc/advance-article/doi/10.1093/ecco-jcc/jjae137/7748263?searchresult=1#supplementary-data

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Dutch Organization for Scientific Research (NWO) through the program Complexity in Health and Nutrition (NWO grant 645.001.002; www.nwo.nl/onderzoeksprogrammas/complexiteit), with co-funding by the National Institute for Public Health and the Environment (RIVM) of the Netherlands. The clinical trial was supported by a research grant from Vedanta Bioscience (Boston, US).

Authors' contributions

The authors confirm contribution to the paper as follows: study conception and design: SP, DS, EB, JAB, ES; funding acquisition: JAB, EB; data collection: SN, EMT, JJK, AEvdM-dJ; analysis and interpretation of results: SP, DS, EB, JAB, ES; draft manuscript preparation: SP, DS, EB, JAB, ES. All authors reviewed the results and approved the final version of the manuscript.

References

1. Seyedian SS, Nokhostin F, Malamir MD. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. J Med Life. 2019;12(2):113-122.
2. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011;474(7351):307-317.
3. Molodecky NA, Kaplan GG. Environmental risk factors for inflammatory bowel disease. Gastroenterol Hepatol (N Y). 2010;6(5):339-346.
4. Halfvarson J, Brislawn CJ, Lamendella R, Vazquez-Baeza Y, Walters WA, Bramer LM, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol. 2017;2:17004.
5. Alam MT, Amos GCA, Murphy ARJ, Murch S, Wellington EMH, Arasaradnam RP. Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. Gut Pathog. 2020;12(1).
6. Schierova D, Brezina J, Mrazek J, Fliegerova KO, Kvasnova S, Bajer L, et al. Gut Microbiome Changes in Patients with Active Left-Sided Ulcerative Colitis after Fecal Microbiome Transplantation and Topical 5-aminosalicylic Acid Therapy. Cells. 2020;9(10).
7. Zhang SL, Wang SN, Miao CY. Influence of Microbiota on Intestinal Immune System in Ulcerative Colitis and Its Intervention. Front Immunol. 2017;8:1674.
8. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019;7(1).
9. Kumari R, Ahuja V, Paul J. Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. World J Gastroenterol. 2013;19(22):3404-3414.
10. Sha S, Xu B, Wang X, Zhang Y, Wang H, Kong X, et al. The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. Diagn Microbiol Infect Dis. 2013;75(3):245-251.
11. Nishihara Y, Ogino H, Tanaka M, Ihara E, Fukaura K, Nishioka K, et al. Mucosa-associated gut microbiota reflects clinical course of ulcerative colitis. Sci Rep. 2021;11(1):13743.
12. He R, Li P, Wang J, Cui B, Zhang F, Zhao F. The interplay of gut microbiota between donors and recipients determines the efficacy of fecal microbiota transplantation. Gut Microbes. 2022;14(1):2100197.
13. Iljazovic A, Roy U, Galvez EJC, Lesker TR, Zhao B, Gronow A, et al. Perturbation of the gut microbiome by *Prevotella* spp. enhances host susceptibility to mucosal inflammation. Mucosal Immunol. 2021;14(1):113-124.
14. Maconi G, Camatta D, Cannatelli R, Ferretti F, Carvalhas Gabrielli A, Ardizzone S. Budesonide MMX in the Treatment of Ulcerative Colitis: Current Perspectives on Efficacy and Safety. Ther Clin Risk Manag. 2021;17:285-292.
15. Ianiro G, Puncochar M, Karcher N, Porcari S, Armanini F, Asnicar F, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nat Med. 2022;28(9):1913-1923.

16. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med. 2013;368(5):407-415.

17. Debast SB, Bauer MP, Kuijper EJ, European Society of Clinical M, Infectious D. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for *Clostridium difficile* infection. Clin Microbiol Infect. 2014;20 Suppl 2:1-26.

18. El-Salhy M, Hatlebakk JG, Gilja OH, Brathen Kristoffersen A, Hausken T. Efficacy of faecal microbiota transplantation for patients with irritable bowel syndrome in a randomised, double-blind, placebo-controlled study. Gut. 2020;69(5):859-867.

19. El Hage Chehade N, Ghoneim S, Shah S, Chahine A, Mourad FH, Francis FF, et al. Efficacy of Fecal Microbiota Transplantation in the Treatment of Active Ulcerative Colitis: A Systematic Review and Meta-Analysis of Double-Blind Randomized Controlled Trials. Inflamm Bowel Dis. 2023;29(5):808-817.

20. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630-638.

21. Schmidt TSB, Li SS, Maistrenko OM, Akanni W, Coelho LP, Dolai S, et al. Drivers and determinants of strain dynamics following fecal microbiota transplantation. Nat Med. 2022;28(9):1902-1912.

22. van Lingen E, Nooij S, Terveer E, Crossette E, Prince A, Bhattarai S, et al. Fecal Microbiota Transplantation engraftment after budesonide or placebo in patients with active ulcerative colitis using pre-selected donors: a randomized pilot study. J Crohns Colitis. 2024;jjae043.

23. Terveer EM, Vendrik KE, Ooijevaar RE, Lingen EV, Boeije-Koppenol E, Nood EV, et al. Faecal microbiota transplantation for *Clostridioides difficile* infection: Four years' experience of the Netherlands Donor Feces Bank. United European Gastroenterol J. 2020;8(10):1236-1247.

24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-359.

25. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i890.

26. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.

27. Ruscheweyh HJ, Milanese A, Paoli L, Karcher N, Clayssen Q, Keller MI, et al. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. Microbiome. 2022;10(1):212.

28. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.

29. Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Mathematical Geology. 2000;32(3):271-275.

30. Barnett D, Arts I, Penders J. microViz: an R package for microbiome data visualization and statistics. Journal of Open Source Software. 2021;6(63).

31. Morgan M. Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. 2022.

32. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126.

33. Brooks ME KK, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal. 2017;9(2):378-400.

34. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Usinglme4. Journal of Statistical Software. 2015;67(1).

35. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society. 1943;54(3):426-482.

36. Lahti L, Shetty S. microbiome R package. 2012-2019.

37. Pinheiro J BD, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2022.

38. Moayyedi P, Surette MG, Kim PT, Libertucci J, Wolfe M, Onischi C, et al. Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. Gastroenterology. 2015;149(1):102-109.e106.

39. Rossen NG, Fuentes S, van der Spek MJ, Tijssen JG, Hartman JH, Duflou A, et al. Findings From a Randomized Controlled Trial of Fecal Transplantation for Patients With Ulcerative Colitis. Gastroenterology. 2015;149(1):110-118 e114.

40. Wilson BC, Vatanen T, Cutfield WS, O'Sullivan JM. The Super-Donor Phenomenon in Fecal Microbiota Transplantation. Front Cell Infect Microbiol. 2019;9:2.

41. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. Immunology. 2017;151(4):363-374.

42. Precup G, Vodnar DC. Gut *Prevotella* as a possible biomarker of diet and its eubiotic versus dysbiotic roles: a comprehensive literature review. Br J Nutr. 2019;122(2):131-140.

43. Lahti L, Salojärvi J, Salonen A, Scheffer M, Vos WMd. Tipping elements in the human intestinal ecosystem. Nature Communications 2014;4344.

44. Pinto S, Beninca E, Galazzo G, Jonkers D, Penders J, Bogaards JA. Heterogeneous associations of gut microbiota with Crohn's disease activity. Gut Microbes. 2024;16(1):2292239.

45. Rees NP, Shaheen W, Quince C, Tselepis C, Horniblow RD, Sharma N, et al. Systematic review of donor and recipient predictive biomarkers of response to faecal microbiota transplantation in patients with ulcerative colitis. EBioMedicine. 2022;81:104088.

46. Tian Y, Zhou Y, Huang S, Li J, Zhao K, Li X, et al. Fecal microbiota transplantation for ulcerative colitis: a prospective clinical study. BMC Gastroenterol. 2019;19(1):116.

47. Crothers JW, Chu ND, Nguyen LTT, Phillips M, Collins C, Fortner K, et al. Daily, oral FMT for long-term maintenance therapy in ulcerative colitis: results of a single-center, prospective, randomized pilot study. BMC Gastroenterol. 2021;21(1):281.

48. Kodikara S, Ellul S, Le Cao KA. Statistical challenges in longitudinal microbiome data analysis. Brief Bioinform. 2022;23(4).

49. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. Plos Computational Biology. 2012;8:1-17.

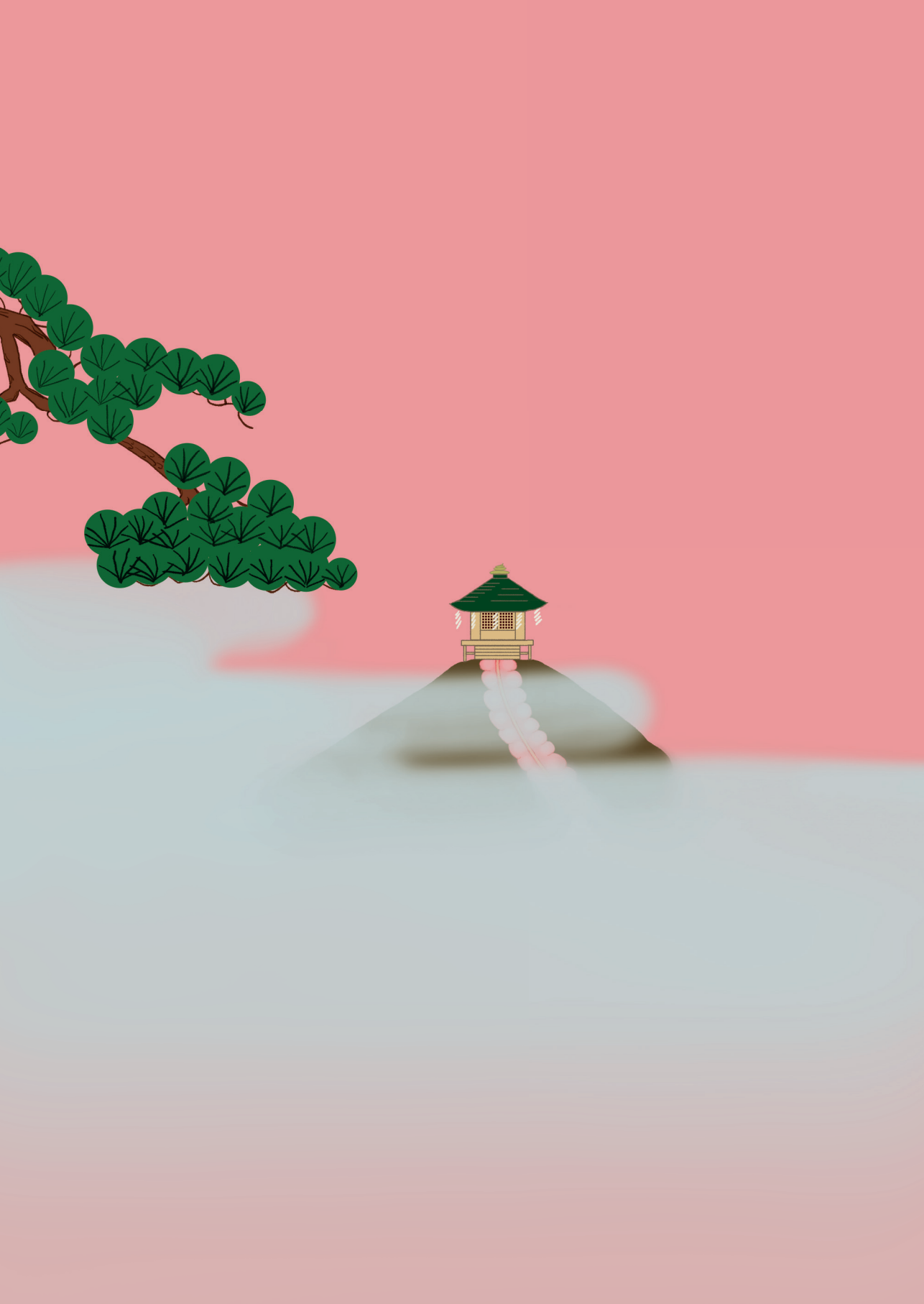
50. Faust K, Raes J. Microbial interactions: from networks to models. Nature Reviews Microbiology volume. 2012;10:538–550.

51. Christensen L, Roager HM, Astrup A, Hjorth MF. Microbial enterotypes in personalized nutrition and obesity management. Am J Clin Nutr. 2018;108(4):645-651.

52. Lee SH, Yoon SH, Jung Y, Kim N, Min U, Chun J, et al. Emotional well-being and gut microbiome profiles by enterotype. Sci Rep. 2020;10(1):20736.

53. Roswall J, Olsson LM, Kovatcheva-Datchary P, Nilsson S, Tremaroli V, Simon MC, et al. Developmental trajectory of the healthy human gut microbiota during the first 5 years of life. Cell Host Microbe. 2021;29(5):765-776 e763.

54. Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. Nat Microbiol. 2018;3(1):8-16.



Chapter 6

Ecological resilience in ulcerative colitis: Microbial dynamics of donor and resident species in a longitudinal fecal microbiota transplantation study

Susanne Pinto[1], Elisa Benincà[2], Sam Nooij[3], Elisabeth M. Terveer[3, 4], Josbert J. Keller[4, 5, 6], Andrea E. van der Meulen – de Jong[5], Ewout Steyerberg[1], and Johannes A. Bogaards[7,8]

[1]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[2]Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands

[3]Leiden University Center for Infectious Diseases (LUCID) Research, Leiden University Medical Center, Leiden, The Netherlands

[4]Netherlands donor Feces bank, LUCID Medical Microbiology & Infection Control, Leiden University Medical Center, Leiden, The Netherlands.

[5]Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands

[6]Department of Gastroenterology, Haaglanden Medisch Centrum, The Hague, The Netherlands

[7]Department of Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, The Netherlands

[8]Amsterdam Institute for Immunology & Infectious Diseases (AI&I), Amsterdam UMC, Amsterdam, The Netherlands

ISME Communications, DOI: 10.1093/ismeco/ycaf119, 16 July 2025

Abstract

Fecal Microbiota Transplantation (FMT) has emerged as promising treatment for the chronic immune-mediated disease ulcerative colitis (UC). However, the ecological dynamics underlying clinical remission remain poorly understood. To investigate these dynamics, we analysed data from 24 UC patients treated with four rounds of FMT donated by two healthy donors. Microbiota samples from patients were collected at nine standardised timepoints before, during, and after treatment, covering a period of 14 weeks. Additionally, 27 donor samples were analysed. Species detected in the recipients' gut microbiota were categorised into ecological categories based on their origin and temporal dynamics: species already present in the host pre-FMT, species derived from the donor, or novel species, i.e. absent in both the pre-FMT host and the donor but detected later. Overdispersed Poisson regression models with random effects were employed to model the number of species within each category over time. Furthermore, we investigated the change in relative abundance for species present in the host pre-FMT. The results revealed that host species with higher relative abundances prior to FMT were more likely to persist following FMT. Notably, patients who achieved combined clinical and endoscopic remission at week 14 retained a significantly higher number of host species compared to non-responders. In contrast, non-responders initially exhibited a higher colonisation of donor species than responders, but their number decreased significantly over time in non-responders. These findings suggest that clinical remission following FMT is associated with a resilient patient gut community, capable of controlled incorporation of donor species, without replacing resident species.

Introduction

Fecal microbiota transplantation (FMT) is the transfer of fecal matter, including gut microorganisms, from the intestine of a healthy donor to a diseased recipient with the goal of modulating the recipient's disturbed microbiota.[1-3] FMT has demonstrated to be effective in recurrent *Clostridioides difficile* infection[2, 3], but success rate is lower for more complex diseases, such as inflammatory bowel disease (IBD).[4, 5] A possible cause for the lower success rate of FMT in complex diseases is the tendency of the recipient's microbiota to revert to its original pre-FMT adverse state.[6] Transition to a healthier state is likely helped by the successful colonisation (engraftment) of donor-derived microorganisms. Therefore, it has been suggested that the success of FMT depends on the donor's microbial diversity and composition.[7, 8] The extent to which shifts in the patient's microbiota towards the donor's microbiota are beneficial for resolving gut disturbances remains unclear.[6, 9-11] This donor-centric view has been challenged, and the importance of the recipient and procedural factors to determine FMT outcomes has been highlighted.[12-15]

In previous analyses of the FMT trial for ulcerative colitis (UC) we examined the engraftment of specific microbial species following FMT, and their associations with clinical remission.[11, 16] For this, we analysed the data from a randomized controlled trial (RCT) involving 24 UC patients treated with four rounds of FMT donated by two healthy donors. Interestingly, we observed that the rate of microbial engraftment did not correlate with successful clinical remission[16], a paradox also noted in a meta-analysis conducted by Schmidt et al. (2022) involving 316 FMT procedures.[12] In their study, clinical success was not correlated with donor strain colonisation or replacement of recipient species. Instead, recipient factors seemed to play a more important role in determining FMT outcomes than donor-related factors.[12] The seemingly limited role of engraftment in predicting clinical outcome of FMT defies the super-donor hypothesis and necessitates deeper investigation into the ecological changes underlying clinical remission.

In this study, the role of donor and host microbial species in determining clinical outcome of FMT is investigated further by applying the conceptual framework introduced by Schmidt et al. (2022)[12] to a longitudinal setting. We capitalize on a randomized controlled trial[16] with dense repeated sampling to map the succession dynamics in the recipient's gut microbiota of UC patients following FMT treatment in relation to clinical remission. Our analysis focuses on ecological dynamics on a species level, categorising all taxa based on their origin and temporal presence: already present in the host before FMT, derived from the donor, or detected during or after the FMT therapies while absent in both the pre-FMT host and the donor.

Methods

The study population

A total of 24 adult patients experiencing mild to moderate exacerbations of UC were included in a double-blind randomized controlled trial conducted at LUMC.[16] Written informed consent was obtained from all study participants prior to their participation. Demographic variables and subject characteristics are provided in Supplementary Table S1, with further details on the study population and clinical characteristics provided by van Lingen et al. (2024) and in Supplementary Information S1.[16]

Following pretreatment with either budesonide (n=12) or placebo (n=12), patients received four fecal transplants at weekly intervals. Donors (D07 and D08) were randomly assigned. FMTs were infused in the duodenum via a nasoduodenal tube or gastroscope.[17] Stool samples were obtained before and after the pretreatment phase, four times before every FMT, and 1 week, 4 weeks, and 8 weeks after treatment. At the end of the study, at week 14, a sigmoidoscopy was performed to assess the endoscopic MAYO score. Clinical remission (i.e. response) was defined as no symptoms (partial MAYO score of 2 with no individual sub score of >2) and an endoscopic MAYO score 0-1. A total of ten patients achieved combined clinical and endoscopic remission (n=9) or partial remission (n=1). Of the 14 non–responders, 10 patients left the study early (in total 2 patients did not finish all 4 FMT treatments) because their symptoms worsened.[11, 16] For this study we defined a responder as a patient in combined clinical and endoscopic remission at week 14 (n=9).

Microbiota data

DNA was extracted from the donor and recipient stool samples and shotgun sequenced with 100 bp single-end reads to a median depth of 2.9m reads by Diversigen (New Brighton, Minneapolis, USA) using the Illumina NovaSeq platform. Raw reads mapping to the human genome were removed using bowtie2 (version 2.4.2)[18] and the GRCh37 reference genome and reads were quality-trimmed using fastq (version 0.20.1)[19], both of which are part of an in-house workflow (https://git.lumc.nl/snooij/metagenomics-preprocessing). The mOTUs3 workflow (version 3.0.1) was used to generate taxonomic profiles.[20, 21] Unassigned, human-derived, archaeal, and low-quality reads were removed from the data, which resulted in 1552 unique mOTUs. For the sake of simplicity, we use the term 'species' to refer to unique mOTUs throughout. The results table was then imported into R for analyzing the data, visualizing the results and performing the statistical tests (R version 4.2.2). R code is available via GitHub (https://github.com/susannepinto/FECBUD_microbiome.git).

Mapping ecological categories

Respectively 13 and 14 samples were available for donor D07 and donor D08. Note that every recipient received FMT material from only one of the donors. We could not

match every recipient sample to a specific donor sample used for the FMT, because not every donor sample used for FMT was sequenced. Therefore, we created a dataset with the core microbiota for each donor. The core donor microbiota was defined as having its relative abundance higher than the detection limit of 0.1% in at least one sample. The core donor microbiota yielded 120 and 84 unique species for donors D07 and D08, respectively.

Subsequently, we created a presence/absence dataset of all species per recipient and per timepoint, and every species was assigned to an ecological category per recipient and per timepoint based on its origin and presence over time, according to the decision tree presented in Figure 1 (detailed explanation Supplementary Information S2). Per recipient, for every species ever present at any timepoint in the recipient, or present in the microbiota of the associated donor, a comparison was made with the recipient's pre-FMT sample and with the microbiota of the corresponding donor. All species present in the recipient's pre-FMT sample were placed into a host category (Resident, Host transient, or Species loss), depending on the pattern of presence over time. If species were unique for the donor relative to the recipient's pre-FMT samples, species were placed into a donor category (Colonisation, Donor transient, or Rejection). If species were not present in the host pre-FMT or in the microbiota of the donor, they were classified as a novel species (Novel, Novel transient, or Novel loss). Within these broad categories, a species was further categorised as a stable (Resident, Colonisation, or Novel), intermittent (Host transient, Donor transient or Novel transient), or previous occupant (Species loss, Rejection, or Novel loss) in the microbiota, depending on the presence at that moment and at the previous timepoints. Because absence in microbiota data can also mean that the abundance was under the detection limit, in the base case we allowed for each species the occurrence of one single absence without direct consequences for categorisation in the rest of the timeseries. Due to the way the categories are defined, some categories cannot occur at the first timepoints. For example, a donor-derived species first had to colonise the gut (colonisation), then be absent for at least two timepoints (NA and Rejection), and then be detected again to be categorised as a Donor transient species (Supplementary Information S2).

In sensitivity analyses, we tested some variations to the base case criteria regarding the temporal information used for categorising the species. In Sensitivity 1 we did not allow the occurrence of any absence when categorising species into either of the host, donor, or novel categories (Fig. 1). In Sensitivity 2 we only considered the presence/absence at the previous timepoint instead of all the previous timepoints (Supplementary Figure S3). In contrast, in Sensitivity 3 the presence of species at all timepoints is considered in categorisation of species at a particular timepoint (Supplementary Figure S4). Sensitivity 4 is the same as Sensitivity 3 but with the added criterion of not allowing the occurrence of any absence (Supplementary Figure S4). In Supplementary Information S2 examples on categorisation of species and the differences between the sensitivity analyses are illustrated.

Modelling the number of species across ecological categories

We modelled the number of species across ecological categories by means of overdispersed Poisson regression models with random effects to accommodate correlation between repeated measurements per recipient. For this, we employed a generalized linear mixed-effects model (GLMM) for the negative binomial family with log-link using the glmer.nb function from the lme4 R package.[22] The temporal evolution of the expected log-number of species in each category was modelled with a spline transformation of the original time variable (in weeks since start of FMT treatment). Estimates from the spline model were compared to those from a linear growth model in sensitivity analysis, by modelling the expected log-number of species as a simple linear function of time. Possible differences in succession dynamics between responders and non-responders were investigated by adding the treatment response variable as a covariate to the model, and through specification of interaction terms with time and ecological category. Patient specific variables, namely, donor (donor D07 vs. D08), pretreatment (budesonide vs. placebo), age and sex (female vs. male), were included based upon their role as possible confounders.

Change in population abundances of host-derived species

To explore the dynamics of host-derived species in response to FMT in more detail, we investigated the relative abundance over time for the species that were already present in the host pre-FMT. Results reveal the distribution of abundance differences at particular timepoints across subjects per ecological category for the species that were already present pre-FMT. In addition, we compared the baseline distributions among species that were later categorized as resident, host transient, and lost among both responders and non-responders. Finally, we also calculated the differences in microbial abundance before and after FMT for all species that were present in the recipients' pre-FMT samples. Because several non-responder patients quitted early during the study, we only included patients who completed all four rounds of FMT (n = 22 patients, of whom 9 were defined as responders) and used the last available post-FMT measurement when calculating the difference in relative abundance before and after FMT. Because the abundance distributions were right-skewed, we used a natural log transformation of the abundances. Consequently, the abundance differences on the log scale can be interpreted as proportional differences on the original scale (in percentage difference). To assess the significance of these differences between responders and non-responders, linear mixed-effects models (LMM) were applied, accounting for the correlation of repeated observations within each patient (using the lmer function from the lme4 R package).[22]

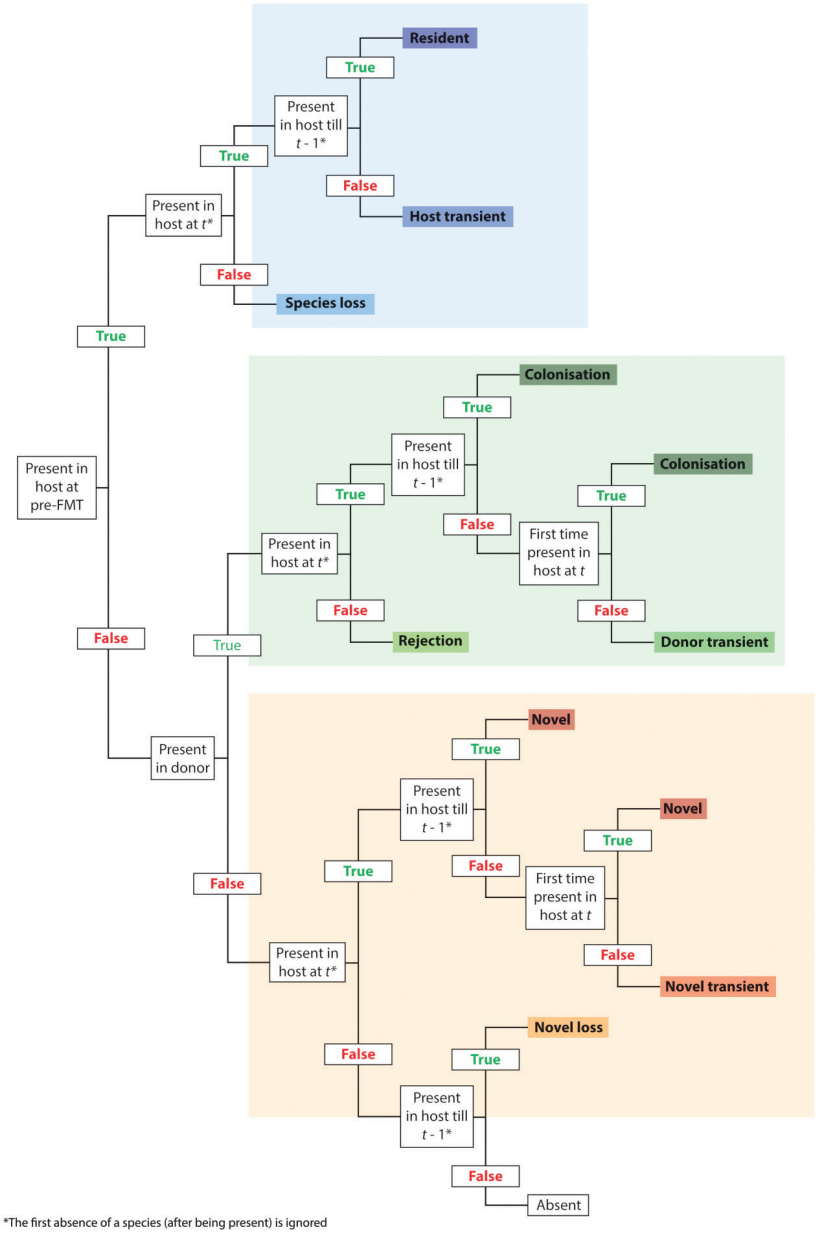


Figure 1. Decision tree used to assign species to ecological categories. The categories are based on the origin and presence of a species over time. First, the species was compared to the pre-FMT host samples, then to the core donor microbiota. Next, the presence/absence at all previous timepoints was considered to assign the species to an ecological category. Note that we ignored the first absence of species when categorising species as lost or as transient upon re-detection. In Sensitivity 1 we evaluated whether this choice had an impact on the results (Supplementary Information S2).

Results

Succession of host-derived, donor-derived, and novel species following FMT

To study the succession dynamics of species during and after FMT in our UC cohort, we modelled the number of species across ecological categories and investigated differences between responders and non-responders (Figure 2). In these models, donor and sex were included as covariates, while pretreatment and age were not relevant as confounders. Supplementary Figure S1 shows the specific parameter estimates of the model depicted in Figure 2.

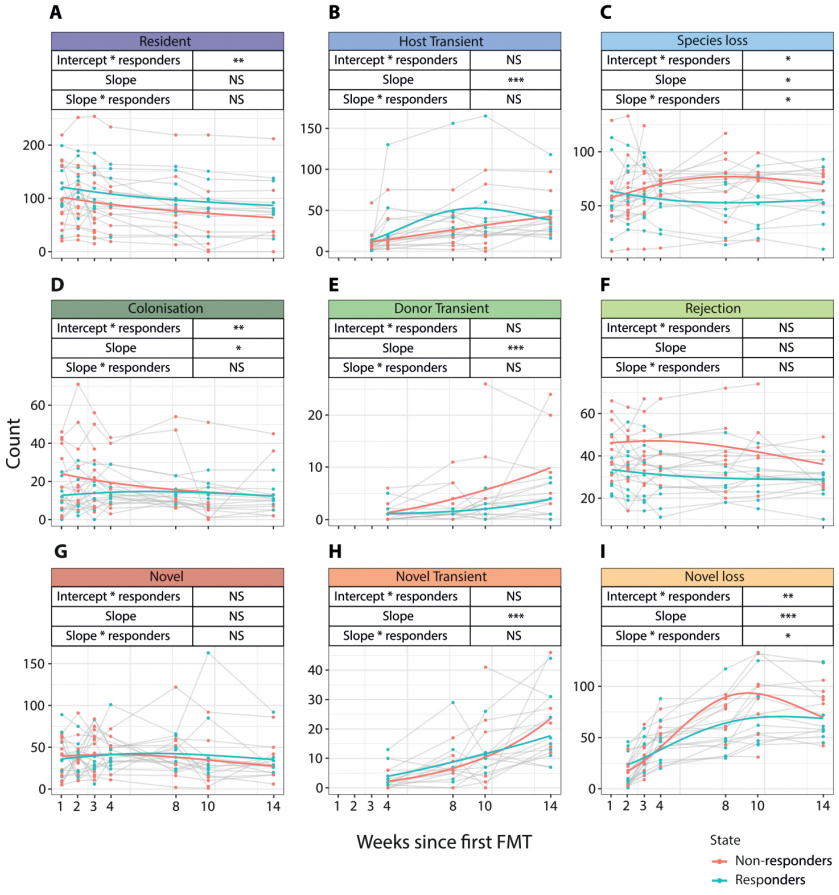


Figure 2. Changes in the number of species per ecological category over time. Average trajectories among responders to the treatment are indicated with blue lines, average trajectories among non-responders with red lines. Individual patient trajectories are shown with grey lines. Note the different scaling of the y-axis per category. The model contained a random intercept per patient to account for repeated measurements. Time was modelled with a spline. The levels of significance are reported above each plot and are indicated by asterisks (***: $p < 0.01$; **: $p < 0.01$; *: $p < 0.05$; NS: not significant).

At the start of the study, we observed a significantly higher number of host species in the resident categories (species that were present in the patient's gut pre-FMT) among responders compared to non-responders, and this difference persisted over time (Figure 2A). Although the number of resident species declined over time in both responders and non-responders this decrease was not statistically significant. In contrast, the number of host transient species increased significantly over time in both patient groups (Figure 2B). Of note, this increase may be partly attributable to the definition of host-derived species being transient upon re-detection after temporary absence. Non-responder patients exhibited a significantly greater loss of host species over time compared to responders, in whom, the number of host species lost decreased significantly over time (Figure 2C).

Conversely, non-responders were initially colonised by a significantly higher number of donor species compared to responders. However, the number of colonising species in non-responders significantly declined over time, whereas it remained constant in responders (Figure 2D). The number of donor transient species was similar between the two patient groups at the start of the study and showed a significant increase over time, especially in non-responders. However, this category remained relatively small and differences according to treatment response were not significant (Figure 2E). The number of rejected donor species was higher at baseline and over time for non-responders compared to responders, however this difference also did not reach statistical significance (Figure 2F).

The number of novel species detected post-FMT was similar for both responders and non-responders and remained constant in time (Figure 2G). The number of novel transient species increased significantly over time, this increase was more or less similar for both the responders and non-responders (Figure 2H). Initially, the responders lost significantly more novel species than the non-responders, but over time the latter group lost significantly more novel species than the responders (Figure 2, panel I).

We also found significant differences between responders and non-responders in the host transient and novel transient categories when applying a linear growth model instead of splines for the temporal evolution of number of species in each category (Supplementary Figure S2). It should be noted that these categories contained relatively few species, and the lack of statistical significance when using splines is likely explained by a reduced statistical power. Importantly, all differences between responders and non-responders identified by the spline model were retained in the linear growth model for category size (Supplementary Figure S2).

Sensitivity analyses

We conducted four different sensitivity analyses concerning the categorisation of the species. To illustrate the effect of categorisation on the rates of change over time, we

generated a plot of the average slope estimates according to each sensitivity analysis (Supplementary Figures S5-S9). Sensitivity analysis 1 resulted in a slightly stronger decline in the number of species for the resident, colonisation, and novel categories (Supplementary Figures S5, S9, and S10). This outcome is a logical consequence of the criterion that a species can no longer be absent for a single time point. Consequently, the likelihood of a species moving to a different category (transient or loss) increased, since it was by definition not possible to return to the categories denoting stable presence over time. This resulted in transient categories having higher intercepts, but the average slopes remained unchanged for all other categories (Supplementary Figures S5, S9, and S10). Similarly, for Sensitivity analysis 2, no substantial differences from the base case were found (Supplementary Figures S6, S9, and S10). The most profound differences were noted in the slopes of the resident and transient categories. The slopes of the transient categories were smaller, especially for the host-derived species among non-responders (Supplementary Figures S6, S9, and S10). Sensitivity analyses 3 and 4 led to more stable patterns over time, especially for the resident category, as compared to both the base case scenario and the other sensitivity analyses (Supplementary Figures S7-S10). This stability can be attributed to the modifications in the category assignment criteria in Sensitivity analyses 3 and 4, where stable presence is defined on all timepoints. Consequently, fewer species were assigned to the resident, colonisation and novel categories and more to the transient categories (Supplementary Information S2).

Relative abundances of host resident species pre- and post-FMT

We further assessed changes in the relative abundance of species present in the gut prior to treatment to investigate whether the relative abundance pre-FMT is indicative of the category that a species will reach post-FMT. Host transient species displayed significantly lower relative abundances at all timepoints compared to resident species (Figure 3A and Supplementary Table S2). In both responders and non-responders host species with higher pre-FMT relative abundances were more likely to become resident species compared to host transient or lost species in both recipient groups (Figure 3B, Supplementary Figure S11, and Supplementary Table S2). Therefore, our findings show that initial microbiota composition is associated with post-FMT composition. The differences in relative abundance of host resident species between the pre-FMT measurement and the last available post-FMT measurement were centered around zero (Figure 3C). A positive difference indicates an increase in the relative abundance of resident species following FMT, while a negative difference denotes a decrease. Thus, approximately equal numbers of resident species showed either a positive or negative response to FMT. No significant differences were found between responders and non-responders in relative abundances of resident species in response to FMT (Figure 3C, Supplementary Figure S12, and Supplementary Table S2).

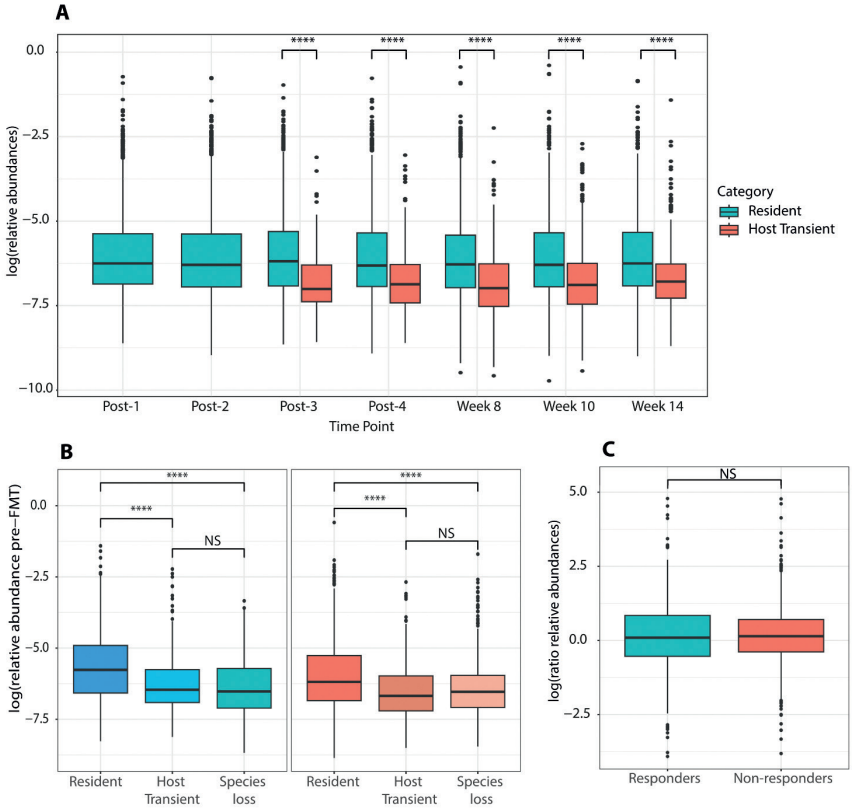


Figure 3. Comparison of relative abundances of species in different categories. A) Relative abundances of Resident and Host Transient species over time. Here, no distinction has been made between responders and non-responders. B) Relative abundance of host species at pre-FMT measurement. The relative abundances in resident, host transient, and species loss species between responders (blue) and non-responders (red) are not significant (Supplementary Table S2). C) Difference in relative abundance in resident species between pre-FMT and last available post-FMT measurement. Significance was tested with linear mixed-models and shown in the plots (***: $p < 0.01$; **: $p < 0.01$; *: $p < 0.05$; NS: not significant).

Discussion

The success of FMT for UC is ultimately determined by whether the patient achieves clinical and endoscopic remission after treatment. It has been suggested that treatment success is related to the extent to which the recipient's microbiota composition shifts towards that of the donor.[7, 23] However, we found no evidence supporting this link, in line with several other studies.[10-13, 16]

We used an ecological framework of succession to investigate microbiota dynamics associated with clinical success of FMT. Microbial species were categorised as pre-existing in the host before FMT, donor-derived, or newly detected. We found that responders retained a higher number of host species compared to non-responders. Although non-responders initially exhibited colonisation by more donor species than responders, this colonisation in non-responders declined over time and eventually became equal to the levels observed in responders. These findings suggest that a successful clinical response to FMT may be facilitated by a microbiota receptive to colonisation without compromising the resident microbiota. Additionally, non-responders lost substantially more novel species over time compared to responders, indicating that newly detected species failed to establish stably within the non-responder gut microbiota. This finding suggests less robust alterations in gut microbiota composition among non-responders. A successful FMT may induce a shift in which the recipient's microbiota integrates donor and novel species, achieving a balanced coexistence to restore the gut microbial ecosystem. This observation aligns with earlier research.[12, 24] Our study expands upon previous analyses by using longitudinal analysis of UC patients, thereby providing a fine-grained view of the ecological dynamics over time of donor and host species following FMT.

FMT can be seen as a perturbation experiment on the gut microbiota, creating a dynamic interplay between donor and recipient communities, which may open ecological niches for other microorganisms.[12, 25] The balance between the engraftment of beneficial microorganisms and competition with deleterious microorganisms in the recipient gut, combined with systemic host processes, such as the modulation of immune responses and the interaction with (external) environmental factors and genetic characteristics, could initiate clinical remission.[4] The process of microbial invasion involves various challenges that incoming microorganisms need to overcome to establish colonisation and influence the existing microbial community. It is important for the invading species to achieve sufficient metabolic activity in the gut to interact with the resident community. This interaction may also be achieved by transient species, indicating that permanent colonisation is not always necessary.[25] Analogous to nurturing an ecosystem such as a crop field through biological control, FMT necessitates the introduction of donor species with healthy functional properties to modify the recipient's system rather than inducing wholesale changes that might lead to the extinction and the replacement of existing microbial inhabitants. Therefore, the recipient microbiota must exhibit a degree of resilience, allowing it to integrate donor species without completely altering its composition. FMTs may also strengthen recipient species by introducing beneficial spores or metabolites, thereby enhancing the stability and functionality of the recipient's own microbiota.[26] The stability of the microbiota is maintained through controlled species loss, ensuring that introduced organisms integrate harmoniously with the pre-existing ecosystem.

The outcome of FMT is influenced by a range of ecological processes, spanning from neutral or stochastic factors (e.g. donor propagule pressure) to adaptive or selective factors (e.g. niche competition and differentiation).[12, 27] This indicates a complex mechanism of action of FMT in patients with UC, necessitating the establishment of a novel homeostasis between the donor and recipient microbiota. This complexity may also explain why prolonged FMT treatment with multiple donor infusions appears necessary in UC, as repeated exposition may be required to achieve an optimal balance between recipient and donor microbiota. This approach contrasts with the FMT treatment of rCDI, which is characterised by a depleted microbiota that can be effectively restored with a single infusion, with a cure rate of about 80%.[1]

The success of FMT may not be reliant on resembling the donor's microbiota, but rather on establishing a complementary relationship, emphasizing the importance of selecting donors whose microbiota optimally aligns with the recipient's specific needs.[7] Unlike the developmental stages of a child's microbiota, the gut microbiota of a UC patient is already an established, independent microbial community. This pre-existing microbiota makes the introduction of new species and the induction of change considerably more challenging.[28, 29] Tailoring the selection of FMT donors to those enriched in taxa capable of restoring disturbed metabolic pathways in the recipient might enhance the effectiveness of FMT, particularly in metabolic dysfunction associated diseases.[6, 7, 10] For example, incoming species that are metabolically complementary to the recipient's community, by introducing novel functions or by occupying previously unfilled niches, may be more likely to colonise the resident community.[30, 31] In addition, a high diversity in the donor and low diversity in the recipient may further influence the success of colonisation.[10, 32]

From an ecological perspective, our findings suggest that donor and recipient species can coexist. We might hypothesise that they occupy distinct metabolic niches. Moreover, we observed that species with a higher abundance prior to FMT (the main 'founders') are more likely to persist during the FMT than species with a lower abundance. This implies that the competitive strength of the resident species is related to their abundance, indicating that within each metabolic niche, communities are built by random winners, driven by stochastic colonisation.[33] This is in line with ecological studies showing that functional differences create opportunities for coexistence (niche theory). However, within each niche functionally similar species can coexist, and communities are structured to random stochastic rules (neutral theory).[34] Within the gut microbiota, species often have overlapping functions, allowing them to replace each other and take over specific functional traits if one species is perturbed or removed.[35]

This study has several limitations. The first concerns the classification of patients into responders and non-responders. Patients that dropped out early due to worsening symptoms were classified as non-responders. Microbiota data were not collected for

these patients, as a result, this potentially introduces bias into the results for the non-responder group. Moreover, the study concerns only 24 UC patients and the time series up to 14 weeks represents only a snapshot of the dynamic process of microbial succession. This sample size is too small to draw definite conclusions and further investigation into longer-term outcomes is necessary to gain a more comprehensive understanding.[36] A third limitation is the sequencing depth (2.9 M 100bp single-end Illumina), which does not allow for definitive determination whether an absent species was actually absent in the host or donor, or simply undetected.[31] Also, the low sequencing resolution makes it impossible to determine whether the same strain present in the donor sample successfully colonised the recipient's gut microbiota or whether the donor and host strains coexisted or were replaced following FMT. Lastly, we did not have data to directly link the unique donor sample used for FMT to the corresponding recipient samples. Therefore, we used the combined microbiota data, which may have led to the misclassification of some low abundant colonising species from the donor as novel species.

By applying an ecological perspective to FMT, our study sheds new light on the importance of ecological principles, such as succession of microorganisms and the resilience of the recipient's system, in shaping therapeutic outcomes. Our study reveals the ecological dynamics of the gut microbiota during and after FMT in patients with UC, with a particular focus on the dynamics of recipient, donor, and novel species. Contrary to some previous studies, the overall engraftment of the donor microbiota did not emerge as the most important factor for FMT success in this study.[7, 13] The key factor influencing the response may not be the overall engraftment of donor species, but rather the recipient's ability to retain resident species while simultaneously enriching specific novel and donor species. Thus, successful FMT hinges on fostering a microbiota shift that complements rather than compromises the existing ecosystem. This ecological interpretation aids in understanding the mechanism through which FMT may induce clinical remission and also underscores the nuanced interplay between donor and recipient microbiota essential for therapeutic efficacy.

Ethics approval and consent to participate

This research project was reviewed and approved by the Medical Ethical Committee of the LUMC, with reference number NL 65976.098.18. The study was registered in the Netherlands Trial Register, with reference number NL9858.

Consent for publication

Not applicable

Availability of data and material

R code is available via GitHub (https://github.com/susannepinto/FECBUD_microbiome.git) and the in-house preprocessing workflow is available via https://git.lumc.nl/snooij/metagenomics-preprocessing. We have uploaded the metagenomic sequences to NCBI with: SRA Bioproject PRJNA1071720.

Competing interests

The authors (SP, EB, SN, ES, and JAB) declare that they have no competing interests. EMT, AEM, and JJK report a research grant and consulting fee from Vedanta Biosciences (Boston, MA, USA US). JJK serves in the scientific advisory board of Microviable Therapeutics (Gyon, Spain).

Funding

This work was supported by the Dutch Organization for Scientific Research (I) through the program Complexity in Health and Nutrition (I grant 645.001.002; www.nwo.nl/onderzoeksprogrammas/complexiteit), with co-funding by the National Institute for Public Health and the Environment (RIVM) of the Netherlands. The clinical trial was supported by a research grant from Vedanta Biosciences.

Authors' contributions

The authors confirm contribution to the paper as follows: study conception and design: SP, EB, JAB; funding acquisition: JAB, EB, ES, EMT, AEM, JJK; data collection: SN, EMT, JJK, AEvdM-dJ; data analysis and interpretation of results of microbiota analysis: SP, EB, JAB; draft manuscript preparation: SP, EB, JAB. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

We would like to acknowledge dr. E. van Lingen for her contribution to the treatment of patients and data collection process. The authors thank Vedanta Biosciences for the collaboration in the initial project, laying a foundation for the current work.

List of Supplementary files

Information S1. Clinical and laboratory procedures
Information S2. Examples illustrating the categorisation of the species in the base case and in the four sensitivity analyses.
Table S1. Clinical and demographic information of non-responders and responders.
Table S2. Model estimates and p-values for the differences in relative abundances.
Figure S1. Results of modelling the number of species per ecological category for the base case.
Figure S2. Results of modelling the number of species per ecological category for the base case.
Figure S3. Decision tree used in Sensitivity 2 to assign species to ecological categories according to different inclusion criteria.
Figure S4. Decision tree to assign species to ecological categories according to different criteria as in the base case analysis (Sensitivity 3 and 4).
Figure S5. Changes in the number of species per ecological category over time for Sensitivity 1.
Figure S6. Changes in the number of species per ecological category over time for Sensitivity 2.
Figure S7. Changes in the number of species per ecological category over time for Sensitivity 3.
Figure S8. Changes in the number of species per ecological category over time for Sensitivity 4.
Figure S9. Average changes in the number of species per ecological category over time for the base case (BC) and all Sensitivity (S1, S2, S3, and S4) analyses.
Figure S10. Distribution of the number of species per ecological category for the base case and all sensitivity analyses estimated by overdispersed Poisson regression models with random effects.
Figure S11. Histograms showing the absolute abundance of host species (Resident, Host transient and Species loss) pre-FMT.
Figure S12. Histograms showing the distribution of the differences in relative abundance of resident species (pre- and post FMT).

Supplementary documents are available online on the publisher's website:

https://academic.oup.com/ismecommun/advance-article/doi/10.1093/ismeco/ycaf119/8203245?searchresult=1#supplementary-data

References

1. van Nood E, Vrieze A, Nieuwdorp M, Fuentes S, Zoetendal EG, de Vos WM, et al. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. N Engl J Med. 2013;368(5):407-415.
2. Debast SB, Bauer MP, Kuijper EJ, Diseases ESoCMaI. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for *Clostridium difficile* infection. Clin Microbiol Infect. 2014;20 Suppl 2:1-26.
3. Ianiro G, Puncochar M, Karcher N, Porcari S, Armanini F, Asnicar F, et al. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nat Med. 2022;28(9):1913-1923.
4. Hanssen NMJ, de Vos WM, Nieuwdorp M. Fecal microbiota transplantation in human metabolic diseases: From a murky past to a bright future? Cell Metab. 2021;33(6):1098-1110.
5. El Hage Chehade N, Ghoneim S, Shah S, Chahine A, Mourad FH, Francis FF, et al. Efficacy of Fecal Microbiota Transplantation in the Treatment of Active Ulcerative Colitis: A Systematic Review and Meta-Analysis of Double-Blind Randomized Controlled Trials. Inflamm Bowel Dis. 2023;29(5):808-817.
6. Sommer F, Anderson JM, Bharti R, Raes J, Rosenstiel P. The resilience of the intestinal microbiota influences health and disease. Nat Rev Microbiol. 2017;15(10):630-638.
7. Wilson BC, Vatanen T, Cutfield WS, O'Sullivan JM. The Super-Donor Phenomenon in Fecal Microbiota Transplantation. Front Cell Infect Microbiol. 2019;9:2.
8. Vermeire S, Joossens M, Verbeke K, Wang J, Machiels K, Sabino J, et al. Donor Species Richness Determines Faecal Microbiota Transplantation Success in Inflammatory Bowel Disease. J Crohns Colitis. 2016;10(4):387-394.
9. Rees NP, Shaheen W, Quince C, Tselepis C, Horniblow RD, Sharma N, et al. Systematic review of donor and recipient predictive biomarkers of response to faecal microbiota transplantation in patients with ulcerative colitis. EBioMedicine. 2022;81:104088.
10. He R, Li P, Wang J, Cui B, Zhang F, Zhao F. The interplay of gut microbiota between donors and recipients determines the efficacy of fecal microbiota transplantation. Gut Microbes. 2022;14(1):2100197.
11. Pinto S, Sajbenova D, Beninca E, Nooij S, Terveer EM, Keller JJ, et al. Dynamics of Gut Microbiota after Fecal Microbiota Transplantation in Ulcerative Colitis: Success Linked to Control of Prevotellaceae. J Crohns Colitis. 2024.
12. Schmidt TSB, Li SS, Maistrenko OM, Akanni W, Coelho LP, Dolai S, et al. Drivers and determinants of strain dynamics following fecal microbiota transplantation. Nat Med. 2022;28(9):1902-1912.
13. Olesen SW, Gerardin Y. Re-Evaluating the Evidence for Faecal Microbiota Transplantation 'Super-Donors' in Inflammatory Bowel Disease. J Crohns Colitis. 2021;15(3):453-461.
14. Danne C, Rolhion N, Sokol H. Recipient factors in faecal microbiota transplantation: one stool does not fit all. Nat Rev Gastroenterol Hepatol. 2021;18(7):503-513.
15. Peri R, Aguilar RC, Tuffers K, Erhardt A, Link A, Ehlermann P, et al. The impact of technical and clinical factors on fecal microbiota transfer outcomes for the treatment of recurrent *Clostridioides difficile* infections in Germany. United European Gastroenterol J. 2019;7(5):716-722.

16. van Lingen E, Nooij S, Terveer E, Crossette E, Prince A, Bhattarai S, et al. Fecal Microbiota Transplantation engraftment after budesonide or placebo in patients with active ulcerative colitis using pre-selected donors: a randomized pilot study. J Crohns Colitis. 2024;jjae043:1381-1393.

17. Terveer EM, Vendrik KE, Ooijevaar RE, Lingen EV, Boeije-Koppenol E, Nood EV, et al. Faecal microbiota transplantation for *Clostridioides difficile* infection: Four years' experience of the Netherlands Donor Feces Bank. United European Gastroenterol J. 2020;8(10):1236-1247.

18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-359.

19. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i890.

20. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh HJ, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.

21. Ruscheweyh HJ, Milanese A, Paoli L, Karcher N, Clayssen Q, Keller MI, et al. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. Microbiome. 2022;10(1):212.

22. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1).

23. Xiao Y, Angulo MT, Lao S, Weiss ST, Liu YY. An ecological framework to understand the efficacy of fecal microbiota transplantation. Nat Commun. 2020;11(1):3329.

24. Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science. 2016;352(6285):586-589.

25. Walter J, Maldonado-Gomez MX, Martinez I. To engraft or not to engraft: an ecological framework for gut microbiome modulation with live microbes. Curr Opin Biotechnol. 2018;49:129-139.

26. Gonze D, Lahti L, Raes J, Faust K. Multi-stability and the origin of microbial community types. ISME J. 2017;11(10):2159-2166.

27. Schmidt TSB, Raes J, Bork P. The Human Gut Microbiome: From Association to Modulation. Cell. 2018;172(6):1198-1215.

28. Bosch A, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, et al. Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. EBioMedicine. 2016;9:336-345.

29. van Best N, Hornef MW, Savelkoul PH, Penders J. On the origin of species: Factors shaping the establishment of infant's gut microbiota. Birth Defects Res C Embryo Today. 2015;105(4):240-251.

30. Grinspan AM, Kelly CR. Fecal Microbiota Transplantation for Ulcerative Colitis: Not Just Yet. Gastroenterology. 2015;149(1):15-18.

31. Darcy JL, Washburne AD, Robeson MS, Prest T, Schmidt SK, Lozupone CA. A phylogenetic model for the recruitment of species into microbial communities and application to studies of the human microbiome. ISME J. 2020;14(6):1359-1368.

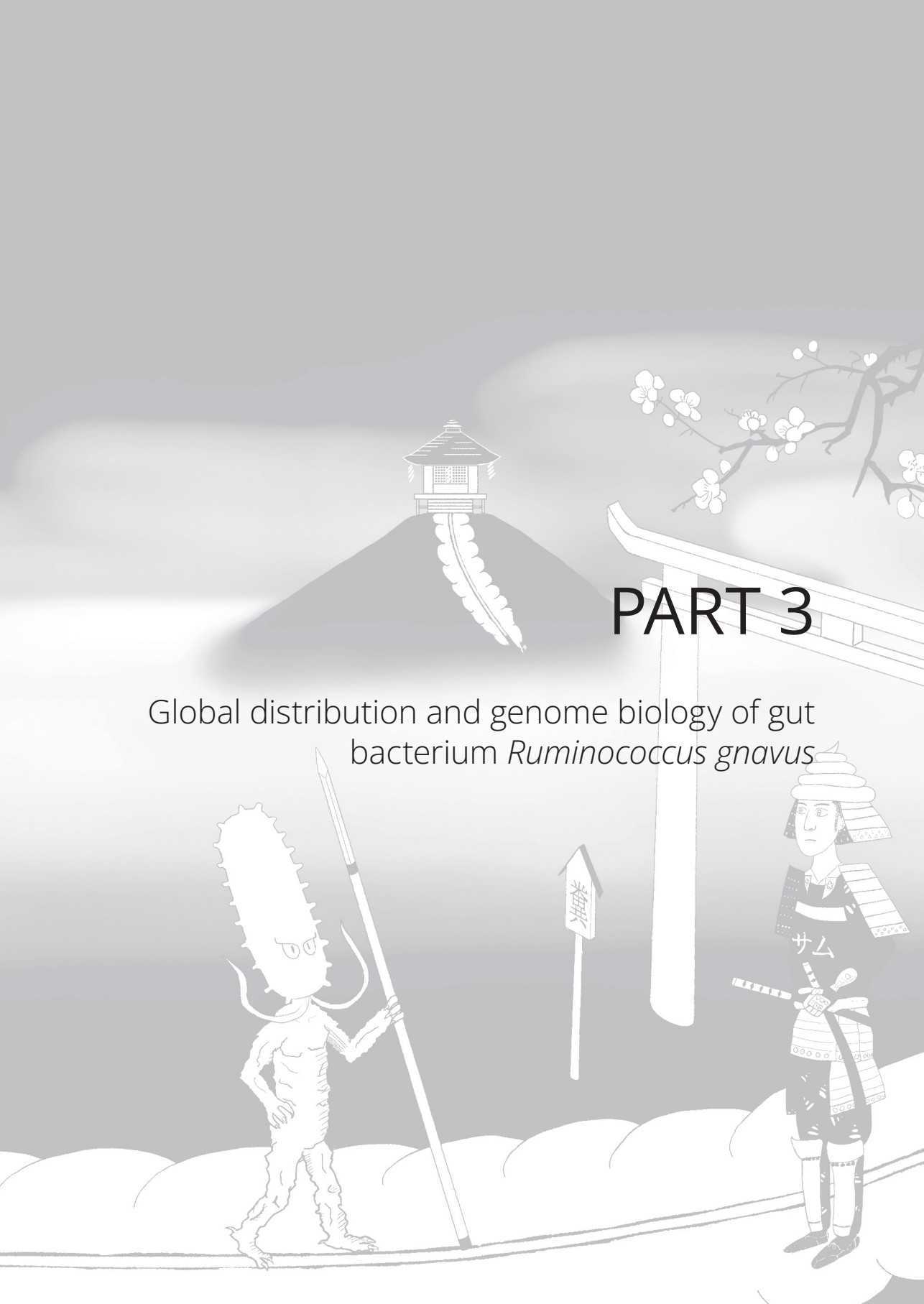
32. Lewin R. Complexity: life at the edge of chaos. 2nd ed: The university of Chicago Press; 1999.

33. Verster AJ, Borenstein E. Competitive lottery-based assembly of selected clades in the human gut microbiome. Microbiome. 2018;6(1):186.

34. Scheffer M, van Nes EH. Self-organized similarity, the evolutionary emergence of groups of similar species. Proc Natl Acad Sci U S A. 2006;103(16):6230-6235.

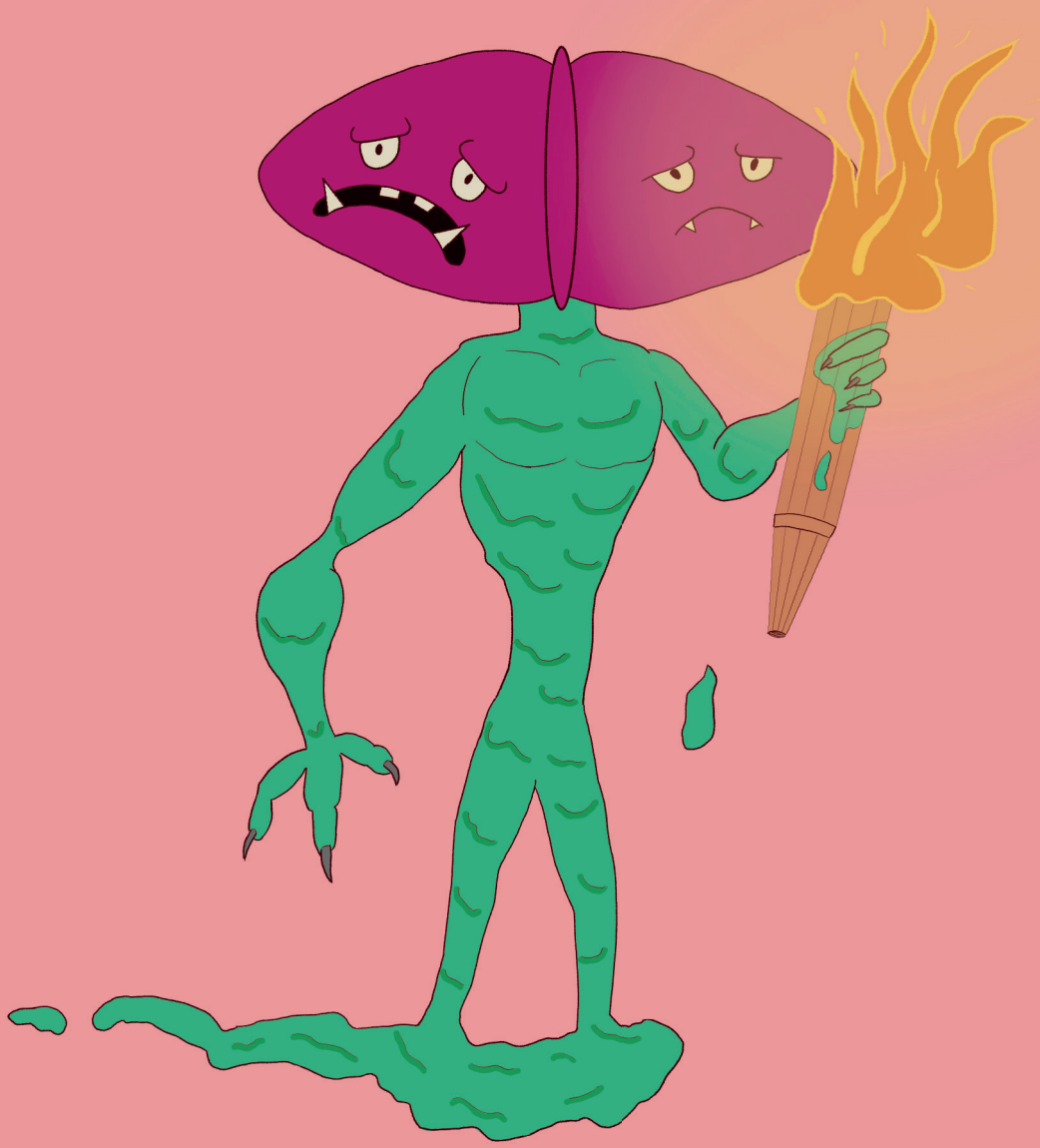
35. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489(7415):220-230.

36. Fuentes S, Rossen NG, van der Spek MJ, Hartman JH, Huuskonen L, Korpela K, et al. Microbial shifts and signatures of long-term remission in ulcerative colitis after faecal microbiota transplantation. ISME J. 2017;11(8):1877-1889.



PART 3

Global distribution and genome biology of gut bacterium *Ruminococcus gnavus*



Chapter 7.1

Metagenomic global survey and in-depth genomic analyses of *Ruminococcus gnavus* reveal differences across host lifestyle and health status

S. Nooij[1,2,3], N. Plomp[4], I.M.J.G. Sanders[1], L. Schout[1,2], A.E. van der Meulen[5], E.M. Terveer[1,2,3], J.M. Norman[6], N. Karcher[7], M.F. Larralde[1], R.H.A.M. Vossen[8], S.L. Kloet[8], K.N. Faber[9], H.J.M. Harmsen[4], G.F. Zeller[1,2,7], E.J. Kuijper[1,2], W.K. Smits[1,2], Q.R. Ducarmon[1,2,7]

[1]Leiden University Center of Infectious Diseases (LUCID), Leiden University Medical Center, Leiden, The Netherlands

[2]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands

[3]Netherlands Donor Feces Bank (NDFB), Leiden University Medical Center, Leiden, the Netherlands

[4]Department of Medical Microbiology and Infection Prevention, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[5]Department of Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands

[6]Vedanta Biosciences, Inc., Cambridge, Massachusetts, USA

[7]Molecular Systems Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

[8]Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands

[9]Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Nature Communications ***16***, *1182 (2025). DOI: 10.1038/s41467-025-56449-x*

Summary

Ruminococcus gnavus is a gut bacterium found in >90% of healthy individuals, but its increased abundance is also associated with chronic inflammatory diseases, particularly Crohn's disease. Nevertheless, its global distribution and intraspecies genomic variation remain understudied. By surveying 12,791 gut metagenomes, we recapitulated known associations with metabolic diseases and inflammatory bowel disease. We uncovered a higher prevalence and abundance of *R. gnavus* in Westernized populations and observed bacterial relative abundances up to 83% in newborns. Next, we built a resource of *R. gnavus* isolates (N=45) from healthy individuals and Crohn's disease patients and generated complete *R. gnavus* genomes using PacBio circular consensus sequencing. Analysis of these genomes and publicly available high-quality draft genomes (N=333 genomes) revealed multiple clades which separated Crohn's-derived isolates from healthy-derived isolates. Presumed *R. gnavus* virulence factors could not explain this separation. Bacterial genome-wide association study revealed that Crohn's-derived isolates were enriched in genes related to mobile elements and mucin foraging. Together, we present a large *R. gnavus* resource that will be available to the scientific community and provide novel biological insights into the global distribution and genomic variation of *R. gnavus*.

Introduction

The human gut microbiome is a topic of intense research interest and many bacterial species have been associated with specific diseases[1]. One such species is *Ruminococcus gnavus*, for which associations with human health have been reported in the context of various ailments[2-7]. Officially, its taxonomic status has been revised and *R. gnavus* is now member of the genus *Mediterraneibacter*, but it has also been termed *Faecalicatena gnavus*[8]. Here, we will designate the species as *Ruminococcus gnavus*. *R. gnavus* is a non-spore forming Gram-positive member of the bacterial phylum Bacillota (formerly Firmicutes) and was first described in 1976[9]. It is considered a prevalent member of the human gut microbiome (present in > 90% of healthy European and North-American adults), but can also be found in the gastrointestinal tract of a variety of animal species[10,11]. Its median relative abundance in humans is reported to be approximately 0.1% - 0.3%, although it should be noted that these estimates were based on small and geographically restricted studies[12,13].

In microbiome association studies, increases in *R. gnavus* relative abundance have consistently been linked to diseases including metabolic syndrome, type 2 diabetes mellitus and Crohn's disease (CD, a form of inflammatory bowel disease (IBD))[2,3,14]. Furthermore, its relative abundance increased concomitantly with symptomatic flares in CD, where it reached up to 69.5% of the gut microbiome[2]. While it remains unknown if *R. gnavus* causally contributes to disease development or whether the increased abundance is a result of the changing intestinal environment, several molecular mediators have been identified that potentially contribute to disease. For instance, the cell-surface exposed polysaccharide glucorhamnan has been described as pro-inflammatory, with a strain-dependent effect, depending on whether the *R. gnavus* isolate carried a capsular polysaccharide that promoted a more tolerogenic response[15,16]. However, these observations are limited by the fact that they were made using one or few isolates and strain variation remains underexplored in many gut microbes, including *R. gnavus*.

Not only mechanistic, but also genomic studies of *R. gnavus* have suffered from a limited scope. One study divided *R. gnavus* into two clades based on genome sequences and noted that one was enriched in IBD patients[2]. However, this study was limited by a low number of draft isolate genomes (N = 11) and a scarcity of knowledge on experimentally verified virulence factors of *R. gnavus* at the time[15-20]. A more recent study based on 152 draft genomes identified three major lineages, but genomes of different host organisms were mixed and this study did not investigate associations of genetic features with metadata[11]. Therefore, an important outstanding question remains whether proposed *R. gnavus* virulence factors are enriched in IBD-derived isolates, or whether different genes and functions could separate IBD-derived *R. gnavus* isolates from controls.

In this work, we surveyed global *R. gnavus* prevalence and abundance across thousands of gut metagenomes to provide a more nuanced picture across human lifespan, different lifestyles, and disease, thereby revealing striking differences. Next, through extensive culturing efforts we established a resource of 45 *R. gnavus* isolates and applied PacBio circular consensus sequencing (CCS) to generate complete genomes. This collection of isolates and their complete genomes provides ample scope for targeted experimental follow-up work and will be available as a community resource for the scientific community. We complemented this unique collection with publicly available (short-read draft) genomes, which allowed us to perform large-scale comparative genomics at both the level of phylogeny and predicted gene functions.

Results

Intestinal colonization with *R. gnavus* is associated with age, health, geography and lifestyle

In order to provide a nuanced view of *R. gnavus* prevalence and abundance across health and disease, geography, and lifestyle, we screened 12,791 publicly available metagenomes from all over the world with manually curated metadata (Fig. 1, Supplementary Data 1; full per-sample metadata are available through https://waldronlab.io/curatedMetagenomicData/)[21]. We observed *R. gnavus* in 50.58% of all included subjects and the prevalence in 9,126 healthy individuals was 43.09% (Fig. 1a). As *R. gnavus* has been robustly associated with disease, especially with metabolic disease and IBD[2,3], we compared *R. gnavus* prevalence and abundance between patients with these diseases and healthy subjects (or asymptomatic control subjects) in a meta-analysis. *R. gnavus* was approximately 1.6 times more prevalent in IBD patients (70.2%; logistic regression, $p < 2.2 \times 10^{-16}$, odds ratio (OR [95% confidence interval]) = 3.1 [2.6-3.7]), 1.3 times more with hypertension (58.0%; p = 0.00127, OR = 1.8 [1.3-2.5]), 1.5 times with type-2 diabetes (T2D; 62.9%; $p = 1.52 \times 10^{-9}$, OR = 2.2 [1.8-2.8]), and 2.2 times with atherosclerotic cardiovascular diseases (ACVD; 96.2%; $p < 2.2 \times 10^{-16}$, OR = 33.4 [17.6-74.1]) compared to healthy subjects. Furthermore, the relative abundance of *R. gnavus* was also higher in these conditions as compared to healthy (Fig. 1b; healthy: median = 0%, 1st-3rd quartile [0-0.08%]; IBD: median = 0.11% [0-1.04%], linear model, $p < 2.2 \times 10^{-16}$; T2D: median = 0.027% [0.0-0.22%], $p = 1.9 \times 10^{-10}$; ACVD: median = 0.78% [0.09-3.14%], $p < 2.2 \times 10^{-16}$), except hypertension (median = 0.01% [0-0.07%], p = 0.399). Together, we thus recapitulated that *R. gnavus* occurs more frequently and in higher abundances in the gut microbiome of patients suffering from IBD, hypertension and T2D. Additionally, our analysis uncovered a striking novel enrichment in ACVD, which had the highest prevalence and abundance of any disease group.

Subsequently, we investigated prevalence and relative abundance of *R. gnavus* across countries (Fig. 1c, 1d and Supplementary Fig. 1a). We show only healthy

individuals to exclude possible confounding by diseases such as IBD and metabolic disease. We observed large differences in prevalence, which ranged between 10-90% across countries (overall median: 41%) and mean relative abundance per country ranged between 0.0078-4.05% (overall mean = 0.67% ±3.20 standard deviation; Supplementary Fig. 1a). This variation could be partly explained by Westernization status; this binary classification of Westernized / non-Westernized lifestyles is based on, among others, access to medical care and pharmaceuticals, livestock exposure and diet[22]. Westernized individuals had higher prevalence and abundance of *R. gnavus* compared to non-Westernized individuals (Fig. 1c-e; prevalence: logistic regression, $p < 2.2 \times 10^{-16}$; abundance: linear model, $p < 2.2 \times 10^{-16}$). As these data were generated in multiple studies, we cannot exclude effects of technical differences (e.g., DNA extraction method). To partially check for this, we investigated sequencing depth and found that higher prevalence and abundance were not the result of higher sequencing depth in Westernized countries as non-Western samples were sequenced deeper (Supplementary Fig. 1b; t-test $p = 6.9 \times 10^{-20}$). These differences hold true for any 10% quantile of sequencing depth (Supplementary Fig. 1c, Methods). We also checked for possible correlations between sequencing depth and *R. gnavus* abundance and found a weakly negative correlation in both Westernized and non-Westernized metagenomes (Supplementary Fig. 1d). In conclusion, *R. gnavus* colonization is vastly different between countries, and Westernization (lifestyle) may be a major factor contributing to these differences.

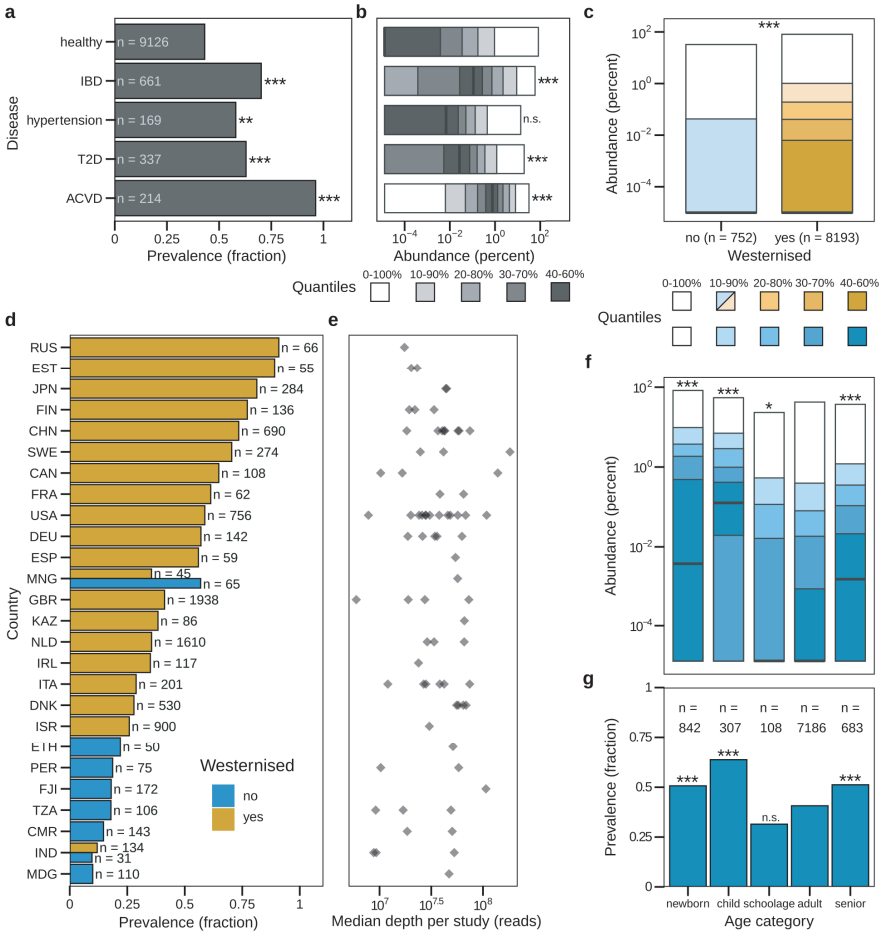


Fig. 1. Intestinal colonization with *R. gnavus* is associated with age, health, geography, and lifestyle. a We queried the public resource curatedMetagenomicData for relative abundances of *R. gnavus* in human stools to conduct a meta-analysis of global prevalence and abundance. Prevalence is shown as fraction of subjects with *R. gnavus* abundance > 0, grouped by selected health conditions. IBD: inflammatory bowel diseases, T2D: type-2 diabetes, ACVD: atherosclerotic cardiovascular diseases. Each disease is compared to healthy. Each disease group is compared to healthy using logistic regression. IBD: $p < 2.2 \times 10^{-16}$; hypertension: $p = 0.00127$; T2D: $p = 1.52 \times 10^{-9}$; ACVD: $p < 2.2 \times 10^{-16}$. **b** Relative abundance of *R. gnavus* in the same groups as **a**, shown as quantile plots, using quantiles ranging from 0 to 100% in increments of 10 with the median shown as a thick black line and quantiles closer to the median shown as darker shades of the same color (see Methods). Each disease is compared to healthy using linear regression. IBD: $p < 2.2 \times 10^{-16}$; hypertension: $p = 0.399$; T2D: 1.9×10^{-10}; ACVD: $p < 2.2 \times 10^{-16}$. **c** Comparison of *R. gnavus* abundance between healthy people from Westernized and non-Westernized societies as quantile plot. $P < 2.2 \times 10^{-16}$, calculated using linear regression. **d** Prevalence of *R. gnavus* grouped per country and colored by Westernization, only showing results from countries from which at least 50 samples were collected. (Countries are abbreviated by ISO 3166-1 alpha-3 codes.) **e** Sequencing depth control per country (same as **d**). Each diamond represents a study that col-

lected samples from the corresponding country. Sequencing depth is shown as median number of reads generated per country in the study. **f** Relative abundance of *R. gnavus* in different age categories (newborn: < 1 year, child: 1-11 years, school age: 12-18 years, adult: 19-65, senior: 65+ years) shown as quantile plots. Age categories are listed in **g**. Each age category is compared to adult using linear regression. Newborn: $p < 2.2 \times 10^{-16}$; child: $p < 2.2 \times 10^{-16}$; schoolage: $p = 0.0164$; senior: $p = 1.37 \times 10^{-6}$. **g** Prevalence of *R. gnavus* among different age categories. Each category is compared to adult using logistic regression. Newborn: $p = 1.14 \times 10^{-6}$; child: $p < 2.2 \times 10^{-16}$; schoolage: $p = 0.0797$; senior: $p = 2.92 \times 10^{-4}$. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. not significant. In **b**, **c**, and **f**, a pseudocount of 1.3×10^{-5} is added to all abundances to enable visualization on a logarithmic scale. Source data are provided as a Source Data file.

Subsequently, we investigated prevalence and relative abundance of *R. gnavus* across countries (Figures 1C, 1D and S1A). In this analysis, we included only healthy individuals to account for the possibility of confounding by diseases such as IBD and metabolic disease. We observed large differences in prevalence, which ranged between 10-90% across countries (overall median: 41%) and mean relative abundance per country ranged between 0.0078-4.05% (overall mean = 0.67% ±3.20 standard deviation; Figure S1A). This variation could be partly explained by Westernization status; this binary classification of Westernized / non-Westernized lifestyles is based on, among others, access to medical care and pharmaceuticals, livestock exposure and diet[21]. Westernized individuals had higher prevalence and abundance of *R. gnavus* compared to non-Westernized individuals (Figure 1C-E; prevalence: Chi-square $p = 2.6 \times 10^{-47}$). As these data were generated in multiple studies, we cannot exclude effects of technical differences (e.g., DNA extraction method). To partially check for this, we investigated sequencing depth and found that higher prevalence and abundance were not the result of higher sequencing depth in Westernized countries as non-Western samples were sequenced deeper (Figure S1B; t-test $p = 6.9 \times 10^{-20}$). These differences hold true for any 10% quantile of sequencing depth (Figure S1C, Methods). We also checked for possible correlations between sequencing depth and *R. gnavus* abundance and found a weakly negative correlation in both Westernized and non-Westernized metagenomes (Figure S1D). In conclusion, *R. gnavus* colonization is vastly different between countries, and Westernization (lifestyle) may be a major factor contributing to these differences.

We noted extremely high *R. gnavus* abundance values in healthy people, up to a relative abundance of 83%. Metagenomes with the highest abundances were often samples collected from newborns and children up to age 2, most of whom were recorded not to have received antibiotics. This motivated a further analysis of age-related patterns of *R. gnavus* colonization (Fig. 1f)[21]. *R. gnavus* abundances were higher in newborns (linear model, $p < 2.2 \times 10^{-16}$), children up to 11 years old ($p < 2.2 \times 10^{-16}$), and adolescents between 12 and 18 years old ('schoolage'; $p = 0.0164$) as compared to adults. Abundances were also higher in seniors (65-92 years old) than in adults ($p = 1.37 \times 10^{-6}$). We observed similar patterns regarding *R. gnavus* prevalence (Figure 1G), where newborns (logistic regression, $p = 1.14 \times 10^{-6}$), children aged 1-11 ($p < 2.2 \times 10^{-16}$) and

seniors ($p = 2.92\times 10^{-4}$) were more likely to carry *R. gnavus* than adults. Adolescents and adults did not have different prevalence of *R. gnavus* ($p = 0.0797$).

The high abundances of *R. gnavus* in infants instigated a closer inspection of abundance over age and in correlation to breastfeeding, as breastfeeding was recently reported to have a strong impact on *R. gnavus* colonization[23]. Looking at *R. gnavus* abundance in the first ten years of life (Supplementary Fig. 2a), we see a rapid increase after the first half year, followed by a decline and rebound around 8 years. We found the shift in the first half year to strongly correlate with feeding practice (Chi-square, $p = 1.49 \times 10^{-15}$). Specifically, infants that were breastfed had lower *R. gnavus* abundance than children that received no breastfeeding (linear model; exclusive breastfeeding, p = 6.49 x 10-11; mixed feeding, $p = 4.69 \times 10^{-4}$; Supplementary Fig. 2b). To exclude possible confounding and identify other associated factors, we also tested for associations between *R. gnavus* abundance with feeding practice (n = 184), mode of delivery (n = 170) and antibiotics use (n = 94) in infants of age up to two years, for whom feeding practice data had been recorded, using multivariable linear modelling. This indicated that only feeding practice was significantly associated with *R. gnavus* abundance in infants (Chi-square of total variable effect, $p = 8.93 \times 10^{-6}$). In summary, we find evidence that indicates that breastfeeding delays of *R. gnavus* colonization in infants, corresponding with previous reports[23].

In summary, colonization with *R. gnavus* appears to be dynamic across the lifespan in healthy individuals, with the highest abundances observed in newborns. While these metagenomic analyses provide important insight into the global distribution of *R. gnavus*, in-depth genomic analyses are required to investigate whether genomic content differs across described factors such as disease and geography.

Newly generated complete genomes have superior assembly characteristics and cover phylogenetic diversity

For our large-scale genomic analysis of *R. gnavus*, we first established an isolate collection through extensive culturing efforts and by collecting available isolates, from which we sequenced the genome of 45 isolates using PacBio circular consensus sequencing (CCS) to yield complete, circular genomes and potential extrachromosomal elements (Fig. 2, Methods; Supplementary Data 2). We next complemented these with 208 available MAGs for which sufficient metadata could be retrieved and short-read genome data of an additional 79 isolates (Methods). To obtain assemblies of optimal quality, we tested five long-read *de novo* assemblers and selected the result with the longest contig (Methods, Supplementary Fig. 3, Supplementary Data 3). We also comprehensively analysed methylation patterns for the sequenced isolates (Supplementary Methods, Supplementary Fig. 4, Supplementary Data 4). Comparing the quality of these genomes, we observed that MAGs were worse in every aspect of genome assembly when compared to isolate assemblies (Fig. 2a). While total length

and number of genes were lower for MAGs as expected, GC content clearly differed between MAGs and isolate genomes, suggesting that current MAG binning techniques may fail to capture AT-rich regions. We further observed that isolates that underwent PacBio CCS were often assembled into single circular contigs, in contrast to a mean of 107 (±58.4 standard deviation) contigs per short-read isolate genome. Additionally, we found four circular extrachromosomal elements predicted to be plasmids with 99.9% confidence (Fig. 2a,b, Supplementary Fig. 5), demonstrating the added value of PacBio CCS. These four putative plasmids comprise two different large sequences of 191kb and 164kb, which derived from two distinct isolates from healthy individuals (i.e., QRD006, QRD009 and QRD010 contain one plasmid, QRD011 the other), and have not been described in *R. gnavus* to date. The plasmids are modular and highly related, that is, they are identical except for one gene cluster that is missing from the shorter 164kb plasmid (Supplementary Fig. 5a). They do not contain evident predicted antibiotic resistance or virulence genes (Methods). The plasmids are likely conjugative or mobilizable based on identified putative transposase genes which is consistent with their geographically distinct origins (USA and Japan). The plasmids contain a putative ParABS segregation system, annotated as 'Soj' (ParA) and 'ParB domain containing protein' (ParB). A key feature is a (hypothetical) non-ribosomal protein synthesis (NRPS) cluster with no known homologs (Supplementary Fig. 5b). However, upstream of it we identified with moderate confidence a transcription factor binding site for CatR, an H_2O_2-responsive repressor.

Leveraging our large genome collection, we then investigated the phylogenetic diversity of *R. gnavus* (Fig. 2c). This revealed no continent or genome source-specific clustering, but importantly, demonstrated that our *R. gnavus* isolate collection captures the full breadth of phylogenetic diversity across the tree (Fig. 2c).

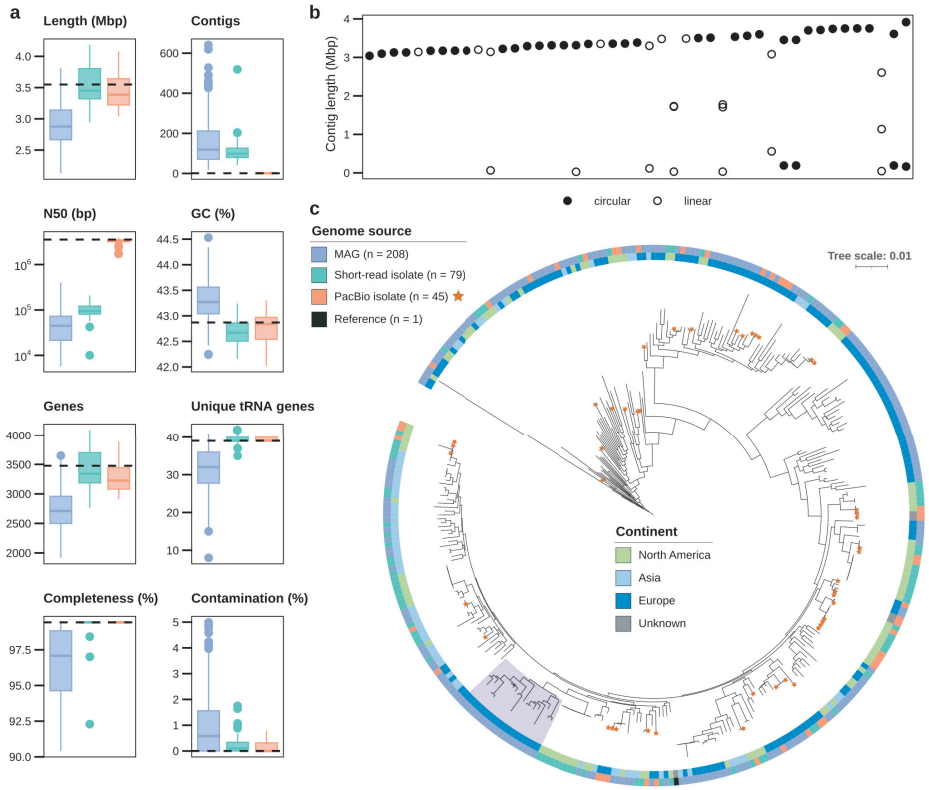


Fig. 2. Newly generated complete genomes have superior assembly characteristics and cover phylogenetic diversity. a We collected both publicly available short-read-based genomes from isolates and metagenome-assembled genomes (MAG), as well as long-read genomes generated from isolates in this study using PacBio HiFi sequencing and compared them to the one reference genome from NCBI GenBank (accession number GCF_009831375.1). Assembly statistics of each group of genomes are compared to the reference genome, shown as dashed line. Thick lines indicate medians, boxes represent first and third quantile and whiskers indicate the rest of the data excluding outliers; outliers are shown as separate dots. Color legend is shared with **c**. **b** Length and circularity of *de novo* assembled contigs from PacBio HiFi reads. **c** Maximum likelihood phylogenetic tree based on concatenated core genes. Each genome is annotated with its corresponding genome source and continent of origin. Stars mark genomes sequenced with PacBio newly added in this work. The gray shaded area marks the infant-associated clade that contains 8/10 MAGs with flagellum genes. Source data are provided as a Source Data file.

R. gnavus motility possibly restricted to infant-derived strains

In order to characterize the functional capacity of *R. gnavus*, we annotated our genomes with functional orthologs, modules and pathways (from KEGG[24]) and used linear modelling to identify associations between microbial functions and metadata. Using this methodology, we observed flagellum biosynthesis exclusively in newborns and infants up to 1 year of age, and this association was also statistically significant ($p = 0.008$). We further investigated flagellum biosynthesis together with chemotaxis, as these are

functionally closely related, and found both pathways in ten out of 333 genomes. These ten genomes are all MAGs originating from newborns and infants up to 1 year of age (Supplementary Fig. 6a) and contained (almost) complete operons (Supplementary Fig. 6b). To ensure this finding was not a technical assembly artefact, we traced the origin of these genomes, which revealed that these MAGs derive from infants sampled in three studies and five geographically separated locations (Estonia, Finland, Italy, Russia, and Sweden). Eight out of ten genomes with flagellum genes belong to a phylogenetic clade that is associated with newborns and infants (17/19 genomes in that clade derive from infants of 1 year old or younger; Fig. 2c, clade highlighted in gray), suggesting that motility might be associated with a specific infant-associated clade of *R. gnavus*. The absence of isolates in this clade precludes experimental verification of flagellum functionality, but strain differences in flagella and motility have been described[9].

We also screened all genomes for antibiotic resistance genes and found that resistance against tetracycline is the most common among *R. gnavus* (75/125 isolates; Supplementary Fig. 7). A minority of genomes contains resistance genes against aminoglycosides (n = 19), chloramphenicol (n = 8), trimethoprim (n = 11), lincosamide/macrolide (n = 24), and one and two genomes contain genes related to beta-lactamase and streptothricin resistance, respectively. For selective culturing of *R. gnavus* we therefore deem tetracycline the most helpful and *in vitro* validation confirmed that at least isolates containing the *tet(O)* and/or *tet(40)* genes, which account for the majority of the observed tetracycline resistance determinants, indeed have increased minimum inhibitory concentrations compared to isolates without *tet* gene (Supplementary Data 5).

Genomic differences between isolates from healthy and Crohn's indicates a Crohn's-specific subspecies

To evaluate whether CD-derived *R. gnavus* isolates genomically differ from healthy-derived isolates, we first placed our genomes into a core genome-based phylogenetic tree (Fig. 3a). As this tree contains practically identical isolates derived from the same person, we also constructed a tree of deduplicated genomes to facilitate statistical testing (Supplementary Fig. 8). This revealed three main clades with a strong enrichment of Crohn's-derived isolates in the two more basal clades (Fisher's exact test, $p = 3.7 \times 10^{-4}$, OR = 12.1 [2.5-69.8]). As our phylogenetic tree was reconstructed from only the core genome, we next performed whole-genome ANI analysis and accessory genome comparisons to also assess differences in the other genomic loci, which resulted in a highly similar clustering (Fig. 3a,b). As all *R. gnavus* genomes included here share at least 95% similarity with one another, which is often considered the species boundary[25,26], we consider that these clades represent subspecies. Together, these results demonstrate that at *R. gnavus* isolates from CD patients are often but not always genomically distinct from isolates from healthy controls based both on their core and accessory genome. The phylogeny indicates that most healthy-derived isolates form a

monophyletic subspecies clade, while the CD isolates appear polyphyletic and may be categorized into multiple groups.

Host phenotypes cannot be explained by previously identified putative virulence factors in *R. gnavus*

A previous study established that *R. gnavus* can secrete a glucorhamnan polysaccharide with pro-inflammatory properties[15]. However, another study found the putative gene cluster encoding the production machinery for this polysaccharide varied between strains, but direct comparison was not possible with short-read sequencing data[27]. Such insights into genomic variations may be crucial to understand immunogenicity of different isolates, motivating a more detailed analysis of this gene cluster and other genes with similar putative functions. We therefore tested whether previously suggested *R. gnavus* virulence factors could explain the association with CD (Supplementary Fig. 8). First, we observed that four genes or gene clusters (superantigens, tryptophane decarboxylase, bilirubin reductase and selenium-dependent xanthine dehydrogenase) were present in all complete *R. gnavus* genomes and are therefore part of the core genome (Fig. 3a). While we saw variation in several other gene clusters (glucorhamnan-producing gene cluster, Fisher's exact test, $p = 0.19$; and the *nan* gene cluster, $p = 0.35$), only one, namely the capsular polysaccharide gene (*cps*) cluster was associated with the distinction and was detected exclusively in isolates from the healthy-associated clade ($p = 8 \times 10^{-4}$). In conclusion, only the *cps* cluster, that leads to a more tolerogenic immune response[16], could distinguish host phenotype groups.

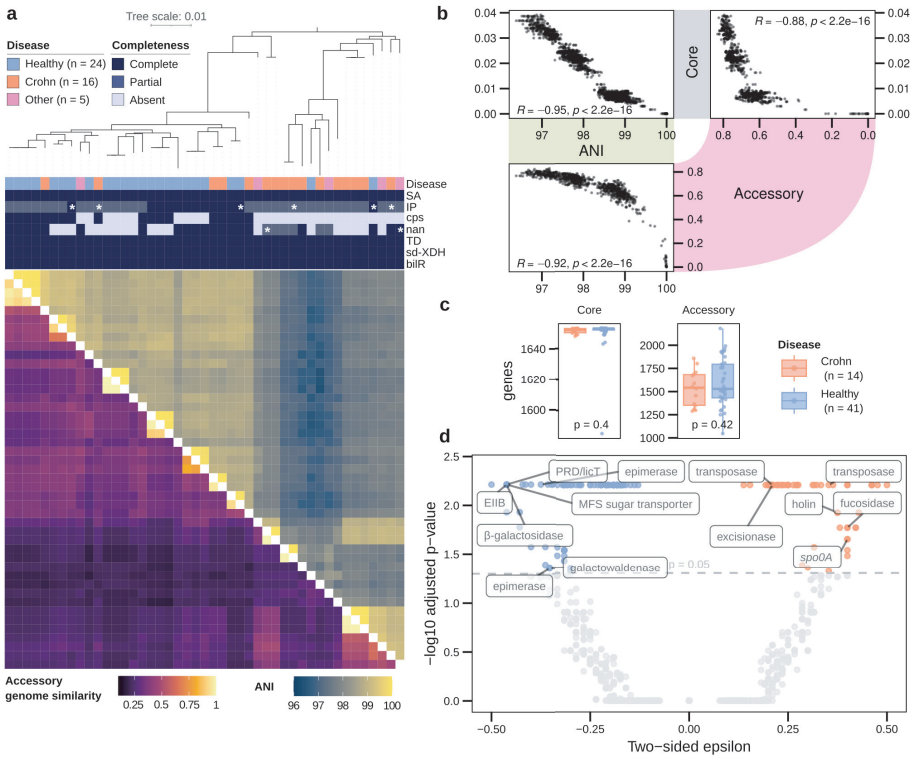


Fig. 3. Genomic differences between isolates from healthy and Crohn's indicates a Crohn's-specific subspecies. a Using our newly generated PacBio genomes, we compared genomes of isolates from healthy people to isolates from CD patients. Maximum likelihood phylogenetic tree of PacBio isolate genomes using concatenated core genes, with annotation of disease status and genes and gene clusters described previously in literature. Asterisks indicate gene clusters from genomes that are highlighted in Supplementary Fig. 9. Below are heatmaps of pairwise average nucleotide identity (ANI) and accessory genome similarity (calculated as 1 / binary distance). SA: superantigen (2 genes), IP: inflammatory polysaccharide (23 genes, 'partial' = 20 or 21 genes), cps: capsular polysaccharide (20 genes), nan: sialic acid metabolic cluster (11 genes, 'partial' = 6 genes), TD: tryptophane decarboxylase (1 gene), sd-XHD: selenium-dependent xanthine dehydrogenase (1 gene), bilR: bilirubin reductase (1 gene). **b** Comparison of genome comparison metrics core genome phylogenetic distance, average nucleotide identity and accessory genome binary distance tested with Spearman correlations. $P < 2.2 \times 10^{-16}$. **c** Comparison of core and accessory genome size between deduplicated isolate genomes with a CD or healthy phenotype, derived from short-read or long-read sequencing. Box plots represent median values with first and third quartile, whiskers indicate the rest of the data excluding outliers, and overlayed dots (jitter) show individual values. P-values were calculated using two-sided Wilcoxon rank-sum test. Core genome: p = 0.4, accessory genome: p = 0.42. **d** We compared accessory genomes of isolates from healthy people and CD patients using a bacterial GWAS to identify genes associated with disease phenotype. Results are expressed as false discovery rate-adjusted p-value (using the Benjamini-Hochberg correction) and epsilon, which is a measure of association strength between phenotype and genotype based on the (maximum likelihood) phylogenetic tree. The gray dashed line indicates a p-value of 0.05, anything above the line is

considered statistically significant. Positive values of epsilon correspond to an enrichment in CD and negative epsilon values are associated with a healthy host phenotype. P- and epsilon-values are adapted from the synchronous GWAS model as implemented in Hogwash. Source data are provided as a Source Data file.

Genomic architecture of gene cluster producing the proinflammatory polysaccharide glucorhamnan reveals genomic variations

Previous studies have highlighted the relevance and genomic architecture of the gene cluster producing inflammatory glucorhamnan based on complete, intermediate, or limited short-read coverage[27]. Here, we re-examined in our diverse collection of complete genomes if these clusters derive from the same genomic locus and are likely to be homologous (Supplementary Fig. 9). Compared to the isolate in which the gene cluster was experimentally verified (QRD039 = RJX 1121)[15], we saw variations in multiple genes, including several glycosyltransferases (Supplementary Fig. 9a). We observed 13 out of 45 long-read genomes to have the complete original cluster as identified in RJX1121, while 30 genomes had 20/23 genes as annotated in NZ_AAYG02000032.1 and two had 21/23 genes (those with 20 or 21 hits are subsequently called 'partially complete')[15,27]. These genomes lacked the same genes: a glycosyltransferase (RUMGNA_03519; present in the two genomes with 21 genes found), a transporter (RUMGNA_03522) and a polyphosphoglycerol synthesis gene (RUMGNA_03523). These partially complete cluster variants lack the genes in positions that were reported to have low coverage and we think they are therefore the same as those described in Sorbara *et al.*, 2020 as 'intermediate coverage'. To elucidate whether these genomes contain a truly different gene cluster at a different genomic location, the flanking genes were determined to map the genomic neighborhood. All investigated genomes had the same neighboring genes, thereby revealing a conserved genomic locus. (The 3' and 5'-flanking genes are annotated as 'HPr family phosphocarrier protein' and 'glutamine-fructose-6-phosphate transaminase'.) By closer inspection of the genomic loci, we found that the operon lacking RUMGNA_03519, RUMGNA_03522 and RUMGNA_03523 had other genes inserted instead (Supplementary Fig. 9a). Moreover, the variability at protein level compared to the reference gene (30-70% identity) suggests that this whole locus may be subject to positive selection or adaptation pressure. Nevertheless, based on similarity in genomic architecture we expect that all these strains still produce polysaccharides, although it remains to be established whether all of them induce pro-inflammatory effects.

A similar comparative genomics analysis for the *nan* gene cluster, responsible for releasing 2,7-anhydro-Neu5Ac from mucin[20], showed some genomes with *nan*-like genes in a different locus (Supplementary Fig. 9b). All these alternative *nan*-like clusters had the same genomic architecture, which importantly lacked the *nanH* (intramolecular trans sialidase) gene, suggesting that this partial cluster does not confer the same function. Together, these data show that strain differences across functionally relevant gene clusters are common, indicating that statements regarding virulence of R. gnavus based on single isolates should be interpreted with caution. Our collection of well-

characterized isolates allows researchers to assess the relevance of strain differences in future experiments.

GWAS reveals genes related to healthy or Crohn's-associated phenotype

In order to find genes that could explain differences in genomic repertoire of Crohn's-derived versus healthy-derived isolates, we conducted a bacterial GWAS using Hogwash, which incorporates genomic relatedness information (Methods). On a technical note, we confirmed high correlation between core and accessory genomes (Fig. 3b), and high pangenome size similarity between the Crohn's-associated and healthy-associated groups (Fig. 3c). We deemed including MAGs for this analysis to be inappropriate, as both the core and accessory genome of MAGs are substantially smaller than that of isolates (Supplementary Fig. 10, $p < 2 \times 10^{-16}$). Thus, their inclusion may increase false negatives or otherwise lead to spurious results.

Our bacterial GWAS analysis revealed 163 genes that were robustly associated with Crohn's isolates (FDR < 0.05, stricter synchronous model) through a high epsilon value, which quantifies the correlation between genotype and phenotype (Fig. 3d)[28]. We visualised and counted the presence of these genes in all *R. gnavus* genomes to better understand their possible correlation with host phenotype (Supplementary Fig. 11,12). Among the genes enriched in Crohn's-derived isolates we found nineteen genes related to mobile genetic elements (transposases and excisionases), a predicted fucosidase which might be involved in cleaving off the terminal fucose residue on mucin, a response regulator that Bakta annotated as '*spo0A*', and a holin gene (Supplementary Data 6). We screened the consensus sequence of this putative fucosidase gene for CAZyme domains (Methods) to gain more functional insight and indeed found a GH29 domain encoding a fucosidase. We also compared fucosidase domains between Crohn's and healthy isolates using CAZyme annotations for GH29 and GH95 (CAZymes with known fucose-cleaving functionality off mucin molecules), but found no significant differences (Wilcoxon rank sum test, $p = 0.098$ and $p = 0.39$, respectively; Supplementary Fig. 13). On the other hand, healthy-derived isolates were especially enriched for galactosidases and other genes involved in sugar metabolism (Fig. 3d, Supplementary Data 5). Taken together, we find novel gene-phenotype associations and provide a set of candidate genes for follow-up research on the role of *R. gnavus* in CD.

Discussion

Host phenotype-microbe association studies are often restricted to single diseases, age groups and geographic regions, which has also been the case for *R. gnavus*[12,13]. In this work we provide a detailed, global image of both the relative abundance and prevalence of *R. gnavus*, while we also investigate genomic variation within *R. gnavus*

isolates in depth. In both aspects, this is to our knowledge the largest investigation to date. Key findings are the remarkably high relative abundance in newborns and young infants (Fig. 1f), which is inversely associated with breastfeeding (Supplementary Fig. 2), and the increased prevalence and abundance of *R. gnavus* in Westernized populations (Fig. 1c,d). Given the robust associations of increased relative abundance of *R. gnavus* with several inflammatory diseases and allergies, many of which have high incidence in high-income countries and have their incidences rapidly increasing in newly industrialized countries[29-32], this begs the question of whether *R. gnavus* can have detrimental immunogenic effects on the host and whether this is strain-dependent. We show extensive genetic variation between strains in immunomodulating gene clusters, and our genetically well-characterized isolate resource can be used for experimental validation of differences in immunogenicity. The high prevalence of *R. gnavus* across both healthy and diseased individuals suggests that the consequences of being colonized with *R. gnavus* per se are unlikely exclusively negative, prompting the question if disease-associations become apparent when distinguishing *R. gnavus* strains. This hypothesis is in line with what we observed in clustering of our isolate genomes (Fig. 3a), where we see that isolates deriving from healthy individuals generally cluster apart from those isolated from Crohn's patients. Indeed, there have also been examples in literature of a positive health influence of *R. gnavus*, for example with healthy weight gain in undernourished children[33]. It would therefore be crucial that future intervention studies using *R. gnavus* determine if the used isolates belong to a healthy-associated or disease-associated clade.

In the past decade MAGs have been increasingly used in large-scale gut bacterial genomics studies[34-38], especially because culturing of specific gut bacteria can be highly laborious and challenging. While these MAGs have led to important biological advances, we show here that even high-quality MAGs (as defined by international standards[39]) remain of substantially worse quality than isolate genomes in multiple aspects (lower genome size and missing genes, higher GC content, amongst others, Figure 2A)[40]. In case of bacterial GWAS analyses, which aims to associate bacterial genes or genomic features with a phenotype of interest, including MAGs may therefore lead to biases and spurious associations caused by (non-)randomly missing genes due to binning and assembly artifacts. Extrachromosomal elements such as plasmids are generally not represented in MAGs, as they cannot be confidently binned, while these may be the most relevant in connection to disease and treatment options[41,42].

Through bacterial culture combined with PacBio CCS, we have generated high-quality genome data that lead to novel insights into *R. gnavus* biology. Two aspects that highlight this are the identification of large plasmids and a conserved methylated sequence motif. To date, only one 7kb-long plasmid of *Ruminococcus gnavus* is described in GenBank (accession number NZ_CP084015.1)[43]. The two related novel plasmids we identified in the present study are much larger (164kb and 191kb; Supplementary Fig. 5) and likely conjugative,

indicating a diversity of plasmids in *R. gnavus* that is of yet underexplored. The methylated DNA motifs that are identified here are different from those known so far (http://rebase.neb.com/cgi-bin/pacbioget?10929; Supplementary Fig. 4)[44], in line with the high variability in motifs we found per genome. Nevertheless, we find a single m^4C-methylated motif that is almost universally conserved across *R. gnavus* genomes (VNNVNCTGVNCAN). These results are reminiscent of those described for *Clostridioides difficile*[45].

We demonstrated that *R. gnavus* is a polyphyletic species, divided into multiple (genotypically and phenotypically distinct) subspecies clades. Notably, Crohn's-derived isolates were overrepresented in specific phylogenetic groups, while previously suggested virulence factors could not explain this separation. This suggests that these virulence factors may not play a significant role in CD symptomatology. Instead, by bacterial GWAS we identified 163 genes that could be targets for experimental validation of their role in CD development (Fig. 3, Supplementary Data 5). Among these genes are 56 that we find overrepresented in CD. However, we advise further validation of these genes in larger numbers of Crohn's-derived *R. gnavus* genomes before conducting laborious in vitro or in vivo experiments. Validations with the currently available data indicate that some presumable Crohn's-associated genes are also common among *R. gnavus* derived from healthy people. We listed the more noticeable candidates for which functions could be predicted. The most striking candidate is a putative fucosidase gene, as this could be directly involved in relevant cellular processes such as cell adhesion and immune system regulation[46]. Secondly, we hypothesize that genomic rearrangements and horizontal gene transfer may play an important role in the evolution of CD-associated *R. gnavus*, given the enrichment of predicted transposase and excisionase genes. Thirdly, we find a predicted holin gene which, although highly speculative, might play a role in suppressing competing bacteria[47]. A previous study identified 199 IBD-specific genes[2], based on a pangenome of 17 draft genomes. Those draft genomes include multiple IBD-related strains and genomes from the type strain, which we find to be phylogenetically distant in our core genome phylogeny based on a pangenome of 333 genomes. This increase in genome number in the current work particularly expands the accessory genome, where the largest differences in functionality are expected. Both the previous report and our results indicate predicted functional differences in e.g. mobile elements such as transposases and (putative) mucus utilization genes underscoring the robustness of the results and narrowing down the set of target genes for IBD-specific research[2]. Furthermore, IBD research on *R. gnavus* could benefit from considering the host and possible complex host-microbe interplay for the proposed virulence factors. For example, in antibiotic-treated mice the genetic background determined whether *R. gnavus* would ameliorate or exacerbate colitis[48].

In conclusion, we present one of the largest collections of complete genomes and associated extrachromosomal elements of any gut microbe not usually causing acute infection[49], and provide important novel biological insight into the global epidemiology

and genomic variation of *R. gnavus*. *R. gnavus* has an ambiguous relationship with human health[50], and different strains may exert different effects on their host. Our resource of complete genomes and isolates opens promising avenues for experimental validation and further bioinformatic scrutiny, and we expect this to be valuable to the broad gut microbiome research community.

Materials and methods

Assessing prevalence and abundance of *R. gnavus* across human populations
We used the publicly available 'curatedMetagenomicData' (version 3.6.2) resource to screen 21,030 fecal metagenomes from 86 studies on all habitable continents for the prevalence and abundance of *R. gnavus*[21]. We used R (version 4.0.2; https://www.R-project.org/) to interrogate this dataset and calculate statistical parameters. We focused our analyses on metagenomes with a sequencing depth of at least five million reads and retained only the first sample per subject ID, after which 12,791 samples remained. We used the accompanying curated metadata to assess prevalence and abundance among healthy individuals across age, geography, lifestyle, and health states (Supplementary Data 1). Prevalence of *R. gnavus* was compared using logistic regression. Relative abundances were compared after adding a pseudocount of 1.3×10^{-5}, followed by log-transformation and multivariable linear modelling. To identify suitable variables for logistic and linear models, we calculated collinearity between variables using Variance Inflation Factors (VIF) using the 'vif()' function from the 'car' package. VIF values above 2 were excluded by removing age (in years) and country from the models, leaving disease, age category, gender and westernization included as informative variables. Rows with missing values were discarded when building the models. For the final models, the association with each variable to *R. gnavus* prevalence or abundance was tested with Chi-square using the 'drop1()' R function. For infants, linear models were built using the same approach, including the variables feeding_practice, born_method and antibiotics_current_use. Correlation between feeding practice and age under or over half a year were tested using Chi-square. Sequencing depth (number of reads) was also log-transformed and compared using parametric t-test. P-values ≤ 0.05 were considered significant. To compare differences in *R. gnavus* prevalence in relation to sequencing depth, we divided all Westernized and non-Westernized metagenomes in ten equal groups (quantiles) based on sequencing depth (number of reads). Relative abundances of *R. gnavus* are shown as quantiles, as adapted from previous publications[51,52].

Mapping the distribution of *R. gnavus* across environments

To map the spread of *R. gnavus* across different environments, we searched publications and online resources that link the presence of *R. gnavus* to an environment or biome. *R. gnavus* has been described to reside in the intestinal tract of different animals: cats and dogs[10], chickens[53], lambs[54], rodents and pigs[11], and cattle[55]. Furthermore, we have

downloaded and screened the dataset related to the 2022 Microbiome publication by Ruscheweyh and colleagues to visualize prevalence and abundance of *R. gnavus* (Supplementary Fig. 14)[56].

Collection and curation of publicly available genome datasets

To compose a collection of *R. gnavus* metagenome-assembled genomes (MAGs) and isolate genomes, we queried a large, recent collection of gut MAGs[34]. Here, we specifically selected high-quality (HQ) MAGs annotated as *Ruminococcus gnavus* or its synonym *Faecalicatena gnavus* (with completeness > 90% and contamination < 5%)[39]. As the metadata from Almeida *et al.* does not contain curated information on disease status of the individual and this is of prime interest to our study[34], we matched identifiers to those present in the curatedMetagenomicData package. HQ-MAGs were only included if at least both disease status and geographic origin of the original sample could be traced back. This led to a collection of 201 HQ *R. gnavus* MAGs with associated metadata.

In order to obtain additional isolate genomes to complement the MAG collection, we queried the NCBI database in December 2021 and associated metadata to retrieve at least information on disease status and geographic origin of the isolate, like the HQ-MAGs. This yielded an additional 65 *R. gnavus* isolate genomes, which all originated from China or the USA. Furthermore, we included the type strain as reference genome (ATCC 29149, accession number GCA_009831375.1)[2,27,57].

Metagenome-assembled genome generation from fecal metagenomes derived from multiple recurrent *Clostridioides difficile*-infected patients

We used an in-house metagenomic dataset of multiple recurrent *Clostridioides difficile*-infected patients to generate seven additional HQ *R. gnavus* MAGs – the metagenomic data of which are available in the European Nucleotide Archive under project number PRJEB44737[58]. To produce high-quality metagenome-assembled genomes (MAGs), we adapted a previously published protocol[59].

The workflow is available as Snakemake[60] on Zenodo (https://doi.org/10.5281/zenodo.14628195) and works as follows. Raw metagenomics sequencing reads, from which human reads had already been removed, were preprocessed using fastp (version 0.20.1, parameters: '--cut_right --cut_window_size 4 --cut_mean_quality 20 -l 75 --detect_adapter_for_pe -y') to trim low-quality ends, remove reads shorter than 75 bases, remove adapter sequences and remove low-complexity reads[61]. (Note: preprocessing is not part of the workflow as described on Zenodo.) Remaining, high-quality reads were assembled into scaffolds using metaSPAdes (version 3.15.4, parameters: '--only-assembler')[62]. Scaffolds were binned with metaWRAP[63] (version 1.3.2) using three binning tools: MaxBin2[64] (version 2.2. 6), MetaBAT2[65] (version 2.12.1) and CONCOCT[66] (version 1.0.0) using a minimum contig length of 2500bp ('-l' option). Bins were then

refined using metaWRAP's 'bin_refinement' function, which uses CheckM[67] (version 1.0.12) to assess bin quality, setting completeness and contamination cut-offs of 75% and 10%, respectively ('-c' and '-x' options). After refinement, bins were reassembled using metaWRAP's 'reassemble_bins' function with assemblers MEGAHIT[68] (version 1.1.3) and metaSPAdes (version 3.13.0), again setting the minimum completeness to 75% and contamination to 10%, and the minimum length to 2000 ('-l' option). The resulting refined and reassembled bins were classified with the Genome Taxonomy Database toolkit (GTDB-Tk; version 2.1.0)[69]. Bins classified as *Ruminococcus gnavus* with >90% completeness and <5% contamination were included for further analyses.

Culturing of *R. gnavus* from feces of healthy donors and patient material

We ordered *R. gnavus* strain H2_28 (DSM number 108212) from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany), resuspended it in Brain Heart Infusion broth (bioMérieux, Marcy-l'Étoile, France) and streaked it on Tryptic Soy agar +5% Sheep blood (TSS; bioMérieux) to isolate pure cultures. Two unique cultures (QRD001-QRD002) were isolated from feces by streaking on Columbia Naladixic acid Agar (bioMérieux; Supplementary Data 2). These were all cultured in an anaerobic cabinet (Whitley A35, Don Whitley Scientific Limited, UK) with an anaerobic gas mixture (10% H_2, 10% CO_2, 80% N_2) at 37°C. These samples were cultured from two different sample collections. First, healthy-derived isolates were obtained from donor faecal samples of Netherlands Donor Feces Bank donors and written informed consent was obtained for using these and clinical data, and approved by the Medical Ethics Committee at Leiden University Medical Center (P15.145). Second, CD-derived isolates from LUMC were obtained from fecal samples of patients aged above 18 years with a planned fistula surgery at LUMC and material was collected between July 2019 and June 2021. The study was approved by the Central Committee on Research involving Human Subjects and the local Medical Ethical Committee of the Leiden University Medical Center (study number P18.069). All patients gave written informed consent.

To further expand our *R. gnavus* genome collection, we cultured fourteen *R. gnavus* isolates from fecal samples of healthy feces donors that were available at Vedanta Biosciences (Supplementary Data 2). Human donor samples were obtained from both university hospitals and commercial sources. In all instances, informed consent language was reviewed and approved by the local ethics and regulatory authorities. Consent for the use of the sample was obtained from each subject. These were isolated and identified as follows: *R. gnavus* strains were isolated from various healthy donor stools by generating spore and non-spore fractions. Briefly, the non-spore fraction was generated by resuspending 1g of fecal material in 10mL sterile, pre-reduced PBS. The spore fraction was generated by adding 100% ethanol to the PBS fecal suspension to achieve a 50% (v/v) ethanol concentration. The fecal ethanol suspension was incubated at 25°C for 1hr while shaking. Following incubation, the fecal ethanol

suspension was centrifuged at 3400×g for 20 minutes and the cell pellet resuspended in 1mL of sterile, reduced PBS. Serial dilutions of the spore and non-spore fraction were plated on either Eggerth-Gagnon + 5% horse blood agar, Brucella Blood Agar (Anaerobe Systems, Inc., Morgan Hill, California, USA), MSAT (Anaerobe Systems), or chocolate agar and incubated at 37°C anaerobically for 72hr. Isolated colonies were identified by Sanger sequencing of the 16S amplicon using 8F and 1492R primers and Illumina shotgun sequencing. Isolated colonies were inoculated into 1.2mL of Peptone Yeast Extract Broth with Glucose (PYG; Anaerobe Systems) in a 96-deep well plate and incubated at 37°C anaerobically for 48hr. After incubation, colony identity was determined by performing PCR from 200µL of the culture using universal 16S primers 8F and 1492R. Selected isolates were then sub-cultured from the 96-deep well plate onto the appropriate agar medium and incubated at 37°C anaerobically for 72hr. An isolated colony from this plate was inoculated into 5mL of PYG and incubated at 37°C anaerobically for 24hr. 1mL of the culture was pelleted by centrifuging at 10000×g for 5 minutes. DNA was extracted from the pellet using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) following the manufacturer instructions. Colony identity was determined again by Sanger sequencing of the 16S gene amplicon using 8F and 1492R primers and Illumina shotgun sequencing.

Furthermore, fourteen isolates were cultured and collected at the University Medical Center Groningen as follows. Brucella blood agar medium (Mediaproducts BV, Groningen, The Netherlands) was used to cultivate the *R. gnavus* strains QRD024, QRD025 and QRD028 from human clinical specimens (Supplementary Data 2). QRD024, QRD025 and QRD028 were obtained from clinical samples and isolated bacteria were used for research purposes as no objections were raised by patients and no patient data was used. The plates were transferred to an anaerobic workstation (Whitley A45) after inoculation and incubated for one to three days at 37°C. The anaerobic medium YCFA supplemented with either apple pectin or porcine mucin type III (4.5 g/l) was used for the isolation of QRD026, QRD027, and QRD029-QRD031 as described earlier[70].. Fecal samples of healthy volunteers were used for inoculation on pre-reduced medium and the plates were incubated at 37^{o}C in an anaerobic chamber (Whitley A35 Workstation) with an anaerobic gas mixture (10% H_2, 10% CO_2, 80% N_2). The strains QRD032-QRD037 were isolated from fecal samples of IBD patients on either phenylethyl alcohol agar (Mediaproducts BV, Groningen, The Netherlands), brain heart infusion agar (Oxoid Limited, Cheshire, UK) supplemented with yeast (2,5 g/l), hemin (0,001% w/v) and cysteine (1 g/l) or YCFA medium supplemented with glucose (4.5 g/l). Ethical approval for collecting and using biological material was obtained as previously described for QRD026, QRD027 and QRD029-QRD037 (local ethics committee of the University Medical Center Groningen METc2014.236 and METc2014.291, respectively)[71]. Additional details on logistics and sample collection can be found in Plomp et al for QRD026, QRD027 and QRD029-QRD031[72], and in von Martels et al. (study was registered on ClinicalTrials.gov under NCT02538354) for QRD032-QRD037[71].

Moreover, isolates as cultured in their respective publications were obtained from the Broad Institute[15], and Sanger Institute[73]. All cultures from outside the Leiden University Medical Center (LUMC) were sent to the LUMC as frozen glycerol stocks and anaerobically cultured on TSS. After obtaining pure colonies, all isolates were independently confirmed to be *R. gnavus* in our laboratory using matrix-assisted laser desorption/ionization coupled to a time-of-flight mass spectrometer (MALDI-TOF; Bruker Daltonics GmbH, Bremen, Germany). All isolates were able to grow on TSS, CNA and Chocolate agar PolyViteX (bioMérieux) and the colony morphology appeared on plates as round, glassy white colonies with a bright white center. Sometimes colonies displayed concentric circles, reminiscent of checker game pieces.

Data processing of Illumina-sequenced *R. gnavus* isolates

The fourteen isolates cultured at Vedanta Biosciences were sequenced on the Illumina NextSeq platform using 150bp paired-end reads. These data were included with the isolate short-read-based genomes, increasing the number to 79 short-read isolates. Raw Illumina sequence data was cleaned and trimmed using fastp (v0.23.2) and sequence quality was inspected using Fastqc (v0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Multiqc[74] (v1.8). Cleaned reads were assembled by first using SKESA[75] (v2.4.0) and subsequently SPAdes (v3.15.3) with "--untrusted-contigs" and "--isolate" parameters.

Quality control and annotation of short-read-based genome collection

We have collected a total of 287 short-read-based genomes of *R. gnavus*, consisting of 79 assembled whole-genome sequences from cultured isolates and 208 metagenome-assembled genomes (MAGs). We also added the one available reference sequence in our analyses (NCBI GenBank accession number GCF_009831375.1). We filtered out contigs shorter than 1,000 bp using BBtools' reformat.sh (version 37.62; https://sourceforge.net/projects/bbmap/). We estimated completeness and contamination of all genomes using CheckM (version 1.0.13) and verified that all genomes taxonomically classify as *R. gnavus* using GTDB-Tk (version 2.1.0). Assembly length statistics were determined using QUAST[76] (version 5.0.2). Finally, genomes were annotated using Bakta[77] (version 1.6.1), which also provides the number of open reading frames, or predicted genes, per genome.

DNA isolation of *R. gnavus* isolates and generation of complete genomes using PacBio circular consensus sequencing

To generate complete genomes, 45 isolates were subjected to long read sequencing on the Pacific Biosciences (PacBio, Menlo Park, California, USA) Sequel IIe platform at the Leiden Genome Technology Center. To prepare high molecular weight total DNA, isolates were cultured anaerobically overnight in 10 mL BHI at 37°C. Cells from 5 mL of culture were pelleted and processed using the Qiagen Genomic-tip 100/G, according

to the manufacturer's instructions. SMRTbell® libraries were generated as follows. Genomic DNA was sheared with the Megaruptor 3 system (Diagenode LLC, Denville, New Jersey, USA) using 35 cycles. Libraries were generated according to the following manufacturer's procedure and checklist: Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0 (PN 102-166-600 REV02 MAR2023), thereby using barcoded adapters. Size-selection was performed on library sub-pools using either diluted AMPure PBbeads (PacBio, 35% beads, 3.1x v/v ratio) or Blue Pippin (Sage Science, Beverly, Massachusetts, USA), depending on the insert-size of the libraries. The libraries were sequenced on a PacBio Sequel IIe platform with a 30 hour movie time using Sequel II Binding Kit 3.2 and Sequel II sequencing kit 2.0.

Long-read assembler mini-benchmark

Given the relative infancy of assembly algorithms for PacBio CCS data of microbial genomes, we performed a mini-benchmark of five long-read *de novo* assemblers: Canu[78,79] (version 2.2), Flye[80] (version 2.9.2), Raven[81] (version 1.8.1), Hifiasm[82] (version 0.19.6-r595) and IPA (version 1.8.0; https://github.com/PacificBiosciences/pbipa). In this benchmark, each assembler was provided 8 processor threads on the Shark high-performance computing cluster of the Leiden University Medical Center. Shark runs on Rocky Linux 8.7, with SLURM version 23.02.7. The available processors include Intel Xeon E5-2697, E5-2690 and E5-4650. Each assembler was provided as much memory as it needed to complete the assembly. The tools exhibited clear differences in number of contigs generated, processing time and memory use (Supplementary Fig. 3). Note that sample QRD034 was sequenced much deeper than the rest and subsampled to 30% of reads (= 277X coverage) to facilitate assembly. Contigs were taxonomically classified using the Contig Annotation Tool (CAT version 5.2.3)[83] to verify if they derived from *R. gnavus*. Canu, Flye, Hifiasm and IPA report if assembled contigs are linear or circular. From the different assemblies, we selected the assembly that yielded the longest contig and the longest total assembly length (all exceeding 3 Mb), giving Flye precedence as it provides the most extensive statistics (Supplementary Data 3). This resulted in 38 assemblies from Flye, 3 from Hifiasm, and 2 each from IPA and Raven. All contigs from selected assemblies were reoriented using dnaapler[84] (version 0.3.0) to start at the *dnaA*, *repA* or *terL* gene for chromosomes, plasmids and bacteriophages, respectively. Raven and Hifiasm produce assembly graphs, which were viewed to assess if contigs were linear or circular. Assemblies with a smaller secondary circular contig were analyzed with geNomad[85] (version 1.7.4) to predict the probability of it being a plasmid, using the built-in score calibration module with aggregated results from both the marker-based and neural net-based classifications.

We included two isolates derived from the strain DSMZ 108212, of which one we obtained directly from the DSMZ (QRD005) and the other was cultured at the Sanger Institute (QRD022). Assembly with Hifiasm yielded a 3.3Mb contig and a 28kb contig for QRD022, while QRD005 could not be resolved to less than three contigs, with

the longest being 2.4Mbp. These two assemblies were not completely identical and we decided to use a reference-based assembly of the unresolved one against the 3.3Mb contig using minimap2[86] (version 2.29) and samtools[87] consensus (version 1.19; parameters: '--min-MQ 5 --min-depth 10') to generate an improved assembly of QRD005. This resulted in two contigs of 3.3Mb and 178bp. We manually removed the 178bp fragment and use the single 3.3Mb contig assembly as representative of the 'DSMZ-108212' = QRD005 isolate (Supplementary Data 3).

Final genome assemblies were annotated with DNA methylation information from the PacBio SMRT Link Microbial Genome Analysis platform.

Antibiotic resistance screening of isolate genomes

To assess the genotypic antibiotic resistances in isolate genomes, we screened 79 short-read genome sequences of isolates, the 45 newly generated long-read genomes, and the one reference genome for the presence of antibiotic resistance genes using ABRicate (version 0.8.13; https://www.github.com/tseemann/abricate) with NCBI's AMRFinderPlus database (downloaded 11 November 2022, containing 5,735 sequences)[88]. Genes were assumed present if at least 95% of the gene matched with at least 95% identity to the gene in the database. For *in vitro* validation, ten isolates – five with *tet* tetracycline resistance genes and five without – were assessed for tetracycline minimum inhibitory concentrations (MIC at 48h) using an ETEST (bioMérieux) on TSS medium at 37°C in a Whitley A35 anaerobic cabinet. However, since we managed to isolate *R. gnavus* without the use of antibiotic selection and tetracycline resistance is also common among other human gut commensals, we did not pursue this further.

Search for previously described inflammatory factors of *R. gnavus*

Several *R. gnavus* genes have previously been associated with intestinal inflammation. We screened our collection of genomes for the presence of two superantigen genes (accession numbers WP_105084811.1 and WP_105084812.1)[17], 23 genes encoding the machinery to produce a proinflammatory (glucarhamnan) polysaccharide (NZ_AAYG02000032.1)[15], one tryptophane decarboxylase gene (RUMGNA_01526 from UniProt)[18], and 20 genes encoding a capsule polysaccharide (RUMGNA_02411 – RUMGNA_02392 from UniProt)[16]. We used protein BLAST[89] (blastp; version 2.13.0) to screen the genomes for the presence of each of these genes. Only hits that covered at least half of the gene of interest ('-qcov_hsp_perc 50') with an E-value of 1×10^{-20} or smaller ('-evalue 1e-20') were considered for further analysis. Gene clusters were considered present when all the genes were detected.

Using the same method, we also screened genomes for the presence of the bilirubin reductase gene (*bilR,* WP_009244284.1)[90], selenium-dependent xanthine dehydrogenase (*sd-XDH*, QHB24869.1)[19], and the *nan* cluster for sialic acid metabolism

(RUMGNA_02691 through RUMGNA_02701 from UniProt)[20]. Gene operons were visualized using clinker[91].

Annotation of functional pathway genes

We annotated carbohydrate-active enzymes (CAZymes) by comparing the genomes to dbCAN[92] (version 10) using HMMer[93] (version 3.3.2). Within the CAZyme families, we focused on two glycosyl hydrolase families that include fucosidases, GH29 and GH95, which have been described as important for mucus utilization[94], a main feature of *R. gnavus*. Genomes were also annotated using KEGG-Decoder[95]. Pathways for chemotaxis and flagellum biosynthesis were annotated using the KOALA definitions available online[24]. Moreover, genomes were screened for the presence of annotated biosynthetic gene clusters (BGC) using antiSMASH[96] (version 6.1.1).

Comparison of whole genomes to find clusters of genomic variants

Whole genomes were compared to one another using average nucleotide identity (ANI) with fastANI (version 1.33)[26]. Furthermore, genomes were subjected to a pangenome analysis using Panaroo (version 1.3.0; parameters '--clean-mode strict -a core --aligner mafft --core_threshold 0.95')[97]. For the pangenome, we considered genes that occur in at least 95% of genomes core genes as recommended when including MAGs[98]. The core genes were concatenated and using MAFFT[99] (version 7.505) a core genome multiple sequence alignment was generated, which was automatically trimmed using trimAl[100] (version 1.4.1). A maximum likelihood phylogeny was inferred from the trimmed multiple alignment using IQ-tree[101] (version 2.2.0.3), including ModelFinder Plus[102] to automatically select the best fitting evolutionary model and ultrafast bootstrap (1000 replicates) to calculate branch support[103]. The selected models were: short-read genomes GTR+F+I+I+R9; long-read genomes GTR+F+R7; all genomes GTR+F+R10. Trees were visualized in iTOL[104].

Bacterial genome-wide association study (GWAS)

To identify genes that are putatively associated with CD, we subjected genomes of *R. gnavus* isolates to a bacterial genome-wide association study using Hogwash (version 1.2.6; parameters: 'fdr = 0.05, bootstrap = 0.875, grouping_method = "post-ar" ')[28]. Hogwash implements a more stringent version of the homoplasy-based PhyC method introduced in 2013[105]. Hogwash reconstructs the evolutionary history of the genomes of interest using a phylogenetic tree and predicts where genotype and phenotype transitions occurred to assess where genotype and phenotype transitions coincide. We made use of the high correlation between core and accessory genome to use these two as input, together with phenotype of either CD or healthy. Genomes were assigned healthy or CD phenotype based on available metadata on health status from the person from whom the *R. gnavus* isolate was cultured. We included short-read sequencing isolate draft genomes as well as our in-house generated PacBio complete genomes. If multiple sequences of the same isolate existed, we deduplicated based

on ANI > 99.9%. Of these duplicates, we picked the first based on alphabetic order as representative, and we preferentially select long-read-based genomes when available. This resulted in fourteen *R. gnavus* isolate genomes derived from CD patients and 41 from healthy people (total N = 55). We used a matrix of (accessory) gene presence and absence generated by Panaroo as input for Hogwash. As phylogenetic tree, we pruned the tree of all *R. gnavus* genomes inferred by IQ-tree to include only this set of 55 deduplicated genomes and midpoint rooted the tree. Associations between genotype and phenotype are evaluated both by p-value indicating statistical significance, and epsilon value, which calculates the strength of genotype-phenotype association on a 0-1 scale (Supplementary Data 6).

To further validate the genes found to be significantly associated with either a healthy or Crohn's host phenotype, we counted the prevalence of each group of genes in both healthy-derived (n = 123) and IBD-derived MAGs (Crohn's n=8; ulcerative colitis n = 1; Supplementary Fig. 11). Furthermore, we visualised the prevalence of these genes among genomes, annotated by their host disease phenotype, as a heatmap to visually inspect the predicted gene associations (Supplementary Fig. 12).

Statistical analyses

All tools were run with default parameters unless stated otherwise. Statistical analyses and visualization were done in R (version 4.0.2) using RStudio (https://posit.co/). A p-value of 0.05 or smaller was considered significant. Data were visualised using the R package ggplot2 (version 3.5.0)[106], with the publication theme from ggembl (version 0.1.2; https://git.embl.de/grp-zeller/ggembl). Figures were polished manually using Inkscape (version 0.92.5; https://inkscape.org/).

Data availability

The long-read whole-genome sequencing data generated in this study and corresponding assemblies of isolates presented in this study are available from the European Nucleotide Archive under accession number PRJEB76407 (https://www.ebi.ac.uk/ena/browser/view/PRJEB76407). Raw metagenomic data used for additional MAG building, and the MAGs themselves, are available under accession number PRJEB44737 (https://www.ebi.ac.uk/ena/browser/view/PRJEB44737). A complete set of short and long-read genomes together with metadata, along with processed data, is available through Zenodo, https://doi.org/10.5281/zenodo.13907031. Isolates will be made available upon request to the corresponding author (q.r.ducarmon@lumc.nl). The use of biological materials for research purposes generated in this study by Vedanta Biosciences can be made available under a material transfer agreement. Correspondence should be sent to jnorman@vedantabio.com and legal@vedantabio.com and will addressed within 2 weeks. Source data are provided with this paper.

Code Availability

Scripts of both the whole-genome annotation and comparative genomics analyses, as well as further downstream and statistical analyses are available on Zenodo, https://doi.org/10.5281/zenodo.14628203. The code used to generate MAGs is also available on Zenodo (https://doi.org/10.5281/zenodo.14628195).

Acknowledgements

We thank all members of the scientific community that generously provided us with *R. gnavus* isolates. This study was supported by the Leiden University Fund / Dr. F.F. Hofman Fonds, (www.luf.nl) to QD. Further funding was provided by the LUMC (LUMC Fellowship to G.Z.), the Health + Life Science Alliance Heidelberg Mannheim through state funds approved by the State Parliament of Baden-Württemberg (postdoctoral fellowships to Q.D. and N.K.) and EMBO postdoctoral fellowship (ALTF 1030-2022 to Q.D.). The Graduate School of the Medical Sciences of the University of Groningen provided a grant to N.P. The NDFB (S.N. and E.M.T.) received an unrestricted research grant from Vedanta Biosciences.

Author contributions statement

Conceptualization, EK, WKS and QD; methodology, NP, IS, LS, AvdM, ET, JN, NK, RV, SK, HH, KF; investigation, SN, NP, RV, IS, LS, ML and QD; formal analysis, SN and QD; writing – original draft, SN and QD; writing – review & editing, all authors; supervision, SK, GZ, EK, WKS and QD; funding acquisition: QD

Competing interest statement

JN is an employee of Vedanta Biosciences Inc. The other authors report no competing interests.

Supplementary Material

Supplementary figures, methods and data are available online on the publisher's website:

https://www.nature.com/articles/s41467-025-56449-x#Sec28

References

1. VanEvery, H., Franzosa, E. A., Nguyen, L. H. & Huttenhower, C. Microbiome epidemiology and association studies in human health. *Nature Reviews Genetics* **24**, 109-124 (2023). https://doi.org:10.1038/s41576-022-00529-x

2. Hall, A. B. *et al.* A novel Ruminococcus gnavus clade enriched in inflammatory bowel disease patients. *Genome Medicine* **9**, 103 (2017). https://doi.org:10.1186/s13073-017-0490-5

3. Grahnemo, L. *et al.* Cross-sectional associations between the gut microbe Ruminococcus gnavus and features of the metabolic syndrome. *The Lancet Diabetes & Endocrinology* **10**, 481-483 (2022). https://doi.org:10.1016/S2213-8587(22)00113-9

4. De Filippis, F. *et al.* Specific gut microbiome signatures and the associated pro-inflamatory functions are linked to pediatric allergy and acquisition of immune tolerance. *Nature Communications* **12**, 5958 (2021). https://doi.org:10.1038/s41467-021-26266-z

5. Wirbel, J., Essex, M., Forslund, S. K. & Zeller, G. Evaluation of microbiome association models under realistic and confounded conditions. *bioRxiv*, 2022.2005.2009.491139 (2022). https://doi.org:10.1101/2022.05.09.491139

6. Berland, M. *et al.* Both Disease Activity and HLA-B27 Status Are Associated With Gut Microbiome Dysbiosis in Spondyloarthritis Patients. *Arthritis & Rheumatology* **75**, 41-52 (2023). https://doi.org:10.1002/art.42289

7. Watanabe, N. *et al.* Clinical and microbiological characteristics of Ruminococcus gnavus bacteremia and intra-abdominal infection. *Anaerobe* **85**, 102818 (2024). https://doi.org:10.1016/j.anaerobe.2024.102818

8. Togo, A. H. *et al.* Description of Mediterraneibacter massiliensis, gen. nov., sp. nov., a new genus isolated from the gut microbiota of an obese patient and reclassification of Ruminococcus faecis, Ruminococcus lactaris, Ruminococcus torques, Ruminococcus gnavus and Clostridium glycyrrhizinilyticum as Mediterraneibacter faecis comb. nov., Mediterraneibacter lactaris comb. nov., Mediterraneibacter torques comb. nov., Mediterraneibacter gnavus comb. nov. and Mediterraneibacter glycyrrhizinilyticus comb. nov. *Antonie Van Leeuwenhoek* **111**, 2107-2128 (2018). https://doi.org:10.1007/s10482-018-1104-y

9. Moore, W. E. C., Johnson, J. L. & Holdeman, L. V. Emendation of Bacteroidaceae and Butyrivibrio and Descriptions of Desulfomonas gen. nov. and Ten New Species in the Genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus. *International Journal of Systematic and Evolutionary Microbiology* **26**, 238-252 (1976). https://doi.org:https://doi.org/10.1099/00207713-26-2-238

10. Branck, T. *et al.* Comprehensive profile of the companion animal gut microbiome integrating reference-based and reference-free methods. *ISME Journal* (2024). https://doi.org:10.1093/ismejo/wrae201

11. Abdugheni, R. *et al.* Comparative genomics reveals extensive intra-species genetic divergence of the prevalent gut commensal Ruminococcus gnavus. *Microbial Genomics* **9** (2023). https://doi.org:10.1099/mgen.0.001071

12. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010). https://doi.org:10.1038/nature08821

13. Kraal, L., Abubucker, S., Kota, K., Fischbach, M. A. & Mitreva, M. The prevalence of species and strains in the human microbiome: a resource for experimental efforts. *PLoS One* **9**, e97279 (2014). https://doi.org:10.1371/journal.pone.0097279

14. Xu, T. *et al.* Microbiome Features Differentiating Unsupervised-Stratification-Based Clusters of Patients with Abnormal Glycometabolism. *mBio* **14**, e0348722 (2023). https://doi.org:10.1128/mbio.03487-22

15. Henke, M. T. *et al.* Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences* **116**, 12672-12677 (2019). https://doi.org:10.1073/pnas.1904099116

16. Henke, M. T. *et al.* Capsular polysaccharide correlates with immune response to the human gut microbe Ruminococcus gnavus. *Proceedings of the National Academy of Sciences* **118** (2021). https://doi.org:10.1073/pnas.2007595118

17. Bunker, J. J. *et al.* B cell superantigens in the human intestinal microbiota. *Science Translational Medicine* **11** (2019). https://doi.org:10.1126/scitranslmed.aau9356

18. Zhai, L. *et al.* Ruminococcus gnavus plays a pathogenic role in diarrhea-predominant irritable bowel syndrome by increasing serotonin biosynthesis. *Cell Host & Microbe* **31**, 33-44 e35 (2023). https://doi.org:10.1016/j.chom.2022.11.006

19. Yan, Y. *et al.* Commensal bacteria promote azathioprine therapy failure in inflammatory bowel disease via decreasing 6-mercaptopurine bioavailability. *Cell Reports Medicine* **4**, 101153 (2023). https://doi.org:10.1016/j.xcrm.2023.101153

20. Bell, A. *et al.* Elucidation of a sialic acid metabolism pathway in mucus-foraging Ruminococcus gnavus unravels mechanisms of bacterial adaptation to the gut. *Nature Microbiology* **4**, 2393-2404 (2019). https://doi.org:10.1038/s41564-019-0590-7

21. Pasolli, E. *et al.* Accessible, curated metagenomic data through ExperimentHub. *Nature Methods* **14**, 1023-1024 (2017). https://doi.org:10.1038/nmeth.4468

22. Valles-Colomer, M. *et al.* The person-to-person transmission landscape of the gut and oral microbiomes. *Nature* **614**, 125-135 (2023). https://doi.org:10.1038/s41586-022-05620-1

23. Shenhav, L. *et al.* Microbial colonization programs are structured by breastfeeding and guide healthy respiratory development. *Cell* **187**, 5431-5452 e5420 (2024). https://doi.org:10.1016/j.cell.2024.07.022

24. Tully, B. J. *KEGGDecoder KOALA definitions*, <https://github.com/bjtully/BioData/blob/master/KEGGDecoder/KOALA_definitions.txt> (2021).

25. Goris, J. *et al.* DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**, 81-91 (2007). https://doi.org:10.1099/ijs.0.64483-0

26. Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* **9**, 5114 (2018). https://doi.org:10.1038/s41467-018-07641-9

27. Sorbara, M. T. *et al.* Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity. *Cell Host & Microbe* **28**, 134-146 e134 (2020). https://doi.org:10.1016/j.chom.2020.05.005

28. Saund, K. & Snitkin, E. S. Hogwash: three methods for genome-wide association studies in bacteria. *Microbial Genomics* **6** (2020). https://doi.org:10.1099/mgen.0.000469

29. Collaborators, G. B. D. I. B. D. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Gastroenterology and Hepatology* **5**, 17-30 (2020). https://doi.org:10.1016/S2468-1253(19)30333-4

30. Fogarty, A. W. What have studies of non-industrialized countries told us about the cause of allergic disease? *Clinical & Experimental Allergy* **45**, 87-93 (2015). https://doi.org:10.1111/cea.12339

31. Tian, J., Zhang, D., Yao, X., Huang, Y. & Lu, Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Annals of the Rheumatic Diseases* **82**, 351-356 (2023). https://doi.org:10.1136/ard-2022-223035

32. Gacesa, R. *et al.* Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732-739 (2022). https://doi.org:10.1038/s41586-022-04567-7

33. Blanton, L. V. *et al.* Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* **351** (2016). https://doi.org:10.1126/science.aad3311

34. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology* **39**, 105-114 (2021). https://doi.org:10.1038/s41587-020-0603-3

35. Karcher, N. *et al*. Analysis of 1321 Eubacterium rectale genomes from metagenomes uncovers complex phylogeographic population structure and subspecies functional adaptations. *Genome Biology* **21**, 138 (2020). https://doi.org:10.1186/s13059-020-02042-y

36. Karcher, N. *et al.* Genomic diversity and ecology of human-associated Akkermansia species in the gut microbiome revealed by extensive metagenomic assembly. *Genome Biology* **22**, 209 (2021). https://doi.org:10.1186/s13059-021-02427-7

37. Tett, A. *et al.* The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. *Cell Host & Microbe* **26**, 666-679 e667 (2019). https://doi.org:10.1016/j.chom.2019.08.018

38. Blanco-Miguez, A. *et al.* Extension of the Segatella copri complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell Host & Microbe* **31**, 1804-1819 e1809 (2023). https://doi.org:10.1016/j.chom.2023.09.013

39. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* **35**, 725-731 (2017). https://doi.org:10.1038/nbt.3893

40. Meziti, A. *et al.* The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample. *Applied Environmental Microbiology* **87** (2021). https://doi.org:10.1128/AEM.02593-20

41. Castañeda-Barba, S., Top, E. M. & Stalder, T. Plasmids, a molecular cornerstone of antimicrobial resistance in the One Health era. *Nature Reviews Microbiology* **22**, 18-32 (2024). https://doi.org:10.1038/s41579-023-00926-x

42. Zorea, A. *et al.* Plasmids in the human gut reveal neutral dispersal and recombination that is overpowered by inflammatory diseases. *Nature Communications* **15**, 3147 (2024). https://doi.org:10.1038/s41467-024-47272-x

43. Tourlousse, D. M. *et al.* Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements. *Microbiology Spectrum* **10**, e0191521 (2022). https://doi.org:10.1128/spectrum.01915-21

44. Blow, M. J. *et al.* The Epigenomic Landscape of Prokaryotes. *PLoS Genetics* **12**, e1005854 (2016). https://doi.org:10.1371/journal.pgen.1005854

45. Oliveira, P. H. *et al.* Epigenomic characterization of Clostridioides difficile finds a conserved DNA methyltransferase that mediates sporulation and pathogenesis. *Nature Microbiology* **5**, 166-180 (2020). https://doi.org:10.1038/s41564-019-0613-4

46. Li, J., Hsu, H. C., Mountz, J. D. & Allen, J. G. Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. *Cell Chemical Biology* **25**, 499-512 (2018). https://doi.org:10.1016/j.chembiol.2018.02.005

47. Backman, T. *et al.* A phage tail-like bacteriocin suppresses competitors in metapopulations of pathogenic bacteria. *Science* **384**, eado0713 (2024). https://doi.org:10.1126/science.ado0713

48. Yu, S. *et al.* Paneth Cell-Derived Lysozyme Defines the Composition of Mucolytic Microbiota and the Inflammatory Tone of the Intestine. *Immunity* **53**, 398-416 e398 (2020). https://doi.org:10.1016/j.immuni.2020.07.010

49. Bartlett, A., Padfield, D., Lear, L., Bendall, R. & Vos, M. A comprehensive list of bacterial pathogens infecting humans. *Microbiology* **168** (2022). https://doi.org:10.1099/mic.0.001269

50. Crost, E. H., Coletto, E., Bell, A. & Juge, N. Ruminococcus gnavus: friend or foe for human health. *FEMS Microbiology Reviews* **47** (2023). https://doi.org:10.1093/femsre/fuad014

51. Wirbel, J. *et al.* Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine* **25**, 679-689 (2019). https://doi.org:10.1038/s41591-019-0406-6

52. Wirbel, J. *et al.* Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. *Genome Biology* **22**, 93 (2021). https://doi.org:10.1186/s13059-021-02306-1

53. Li, Z. *et al.* Effects of herbal dregs supplementation of Salvia miltiorrhiza and Isatidis Radix residues improved production performance and gut microbiota abundance in late-phase laying hens. *Frontiers in Veterinary Science* **11**, 1381226 (2024). https://doi.org:10.3389/fvets.2024.1381226

54. Xiao, H. *et al.* The effect of early colonized gut microbiota on the growth performance of suckling lambs. *Frontiers in Microbiology* **14**, 1273444 (2023). https://doi.org:10.3389/fmicb.2023.1273444

55. Chen, Z. *et al.* Differences in meat quality between Angus cattle and Xinjiang brown cattle in association with gut microbiota and its lipid metabolism. *Frontiers in Microbiology* **13**, 988984 (2022). https://doi.org:10.3389/fmicb.2022.988984

56. Ruscheweyh, H. J. *et al.* Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* **10**, 212 (2022). https://doi.org:10.1186/s40168-022-01410-z

57. Zou, Y. *et al.* 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nature Biotechnology* **37**, 179-185 (2019). https://doi.org:10.1038/s41587-018-0008-8

58. Nooij, S. *et al.* Fecal Microbiota Transplantation Influences Procarcinogenic Escherichia coli in Recipient Recurrent Clostridioides difficile Patients. *Gastroenterology* **161**, 1218-1228 e1215 (2021). https://doi.org:10.1053/j.gastro.2021.06.009

59. Saheb Kashaf, S., Almeida, A., Segre, J. A. & Finn, R. D. Recovering prokaryotic genomes from host-associated, short-read shotgun metagenomic sequencing data. *Nature Protocols* **16**, 2520-2541 (2021). https://doi.org:10.1038/s41596-021-00508-2

60. Molder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021). https://doi.org:10.12688/f1000research.29032.2

61. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560

62. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Research* **27**, 824-834 (2017). https://doi.org:10.1101/gr.213959.116

63. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018). https://doi.org:10.1186/s40168-018-0541-1

64. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2016). https://doi.org:10.1093/bioinformatics/btv638

65. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015). https://doi.org:10.7717/peerj.1165

66. Alneberg, J. *et al.* Binning metagenomic contigs by coverage and composition. *Nature Methods* **11**, 1144-1146 (2014). https://doi.org:10.1038/nmeth.3103

67. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25**, 1043-1055 (2015). https://doi.org:10.1101/gr.186072.114

68. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3-11 (2016). https://doi.org:10.1016/j.ymeth.2016.02.020

69. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927 (2019). https://doi.org:10.1093/bioinformatics/btz848

70. Lopez-Siles, M. *et al.* Cultured representatives of two major phylogroups of human colonic Faecalibacterium prausnitzii can utilize pectin, uronic acids, and host-derived substrates for growth. *Applied Environmental Microbiology* **78**, 420-428 (2012). https://doi.org:10.1128/AEM.06858-11

71. von Martels, J. Z. H. *et al.* Riboflavin Supplementation in Patients with Crohn's Disease [the RISE-UP study]. *J Crohns Colitis* **14**, 595-607 (2020). https://doi.org:10.1093/ecco-jcc/jjz208

72. Plomp, N. *et al.* A convenient and versatile culturomics platform to expand the human gut culturome of Lachnospiraceae and Oscillospiraceae. *Benef Microbes*, 1-16 (2024). https://doi.org:10.1163/18762891-bja00042

73. Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nature Biotechnology* **37**, 186-192 (2019). https://doi.org:10.1038/s41587-018-0009-7

74. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016). https://doi.org:10.1093/bioinformatics/btw354

75. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biology* **19**, 153 (2018). https://doi.org:10.1186/s13059-018-1540-z

76. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075 (2013). https://doi.org:10.1093/bioinformatics/btt086

77. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics* **7** (2021). https://doi.org:10.1099/mgen.0.000685

78. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* **27**, 722-736 (2017). https://doi.org:10.1101/gr.215087.116

79. Nurk, S. *et al.* HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Research* **30**, 1291-1305 (2020). https://doi.org:10.1101/gr.263566.120

80. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540-546 (2019). https://doi.org:10.1038/s41587-019-0072-8

81. Vaser, R. & Sikic, M. Time- and memory-efficient genome assembly with Raven. *Nature Computational Science* **1**, 332-336 (2021). https://doi.org:10.1038/s43588-021-00073-4

82. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170-175 (2021). https://doi.org:10.1038/s41592-020-01056-5

83. von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology* **20**, 217 (2019). https://doi.org:10.1186/s13059-019-1817-x

84. Bouras, G., Grigson., S., Papudeshi., B., Mallawaarachchi V., Roach, M. J. *Dnaapler: A tool to reorient circular microbial genomes* <https://github.com/gbouras13/dnaapler> (2023).

85. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nature Biotechnology* (2023). https://doi.org:10.1038/s41587-023-01953-y

86. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018). https://doi.org:10.1093/bioinformatics/bty191

87. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). https://doi.org:10.1093/bioinformatics/btp352

88. Feldgarden, M. *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports* **11**, 12728 (2021). https://doi.org:10.1038/s41598-021-91456-0

89. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). https://doi.org:10.1186/1471-2105-10-421

90. Hall, B. *et al.* BilR is a gut microbial enzyme that reduces bilirubin to urobilinogen. *Nature Microbiology* **9**, 173-184 (2024). https://doi.org:10.1038/s41564-023-01549-x

91. Gilchrist, C. L. M. & Chooi, Y. H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473-2475 (2021). https://doi.org:10.1093/bioinformatics/btab007

92. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46**, W95-W101 (2018). https://doi.org:10.1093/nar/gky418

93. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Computational Biology* **7**, e1002195 (2011). https://doi.org:10.1371/journal.pcbi.1002195

94. Berkhout, M. D., Plugge, C. M. & Belzer, C. How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding. *Glycobiology* **32**, 182-200 (2022). https://doi.org:10.1093/glycob/cwab105

95. Graham, E. D., Heidelberg, J. F. & Tully, B. J. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME Journal* **12**, 1861-1866 (2018). https://doi.org:10.1038/s41396-018-0091-3

96. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **49**, W29-W35 (2021). https://doi.org:10.1093/nar/gkab335

97. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology* **21**, 180 (2020). https://doi.org:10.1186/s13059-020-02090-4

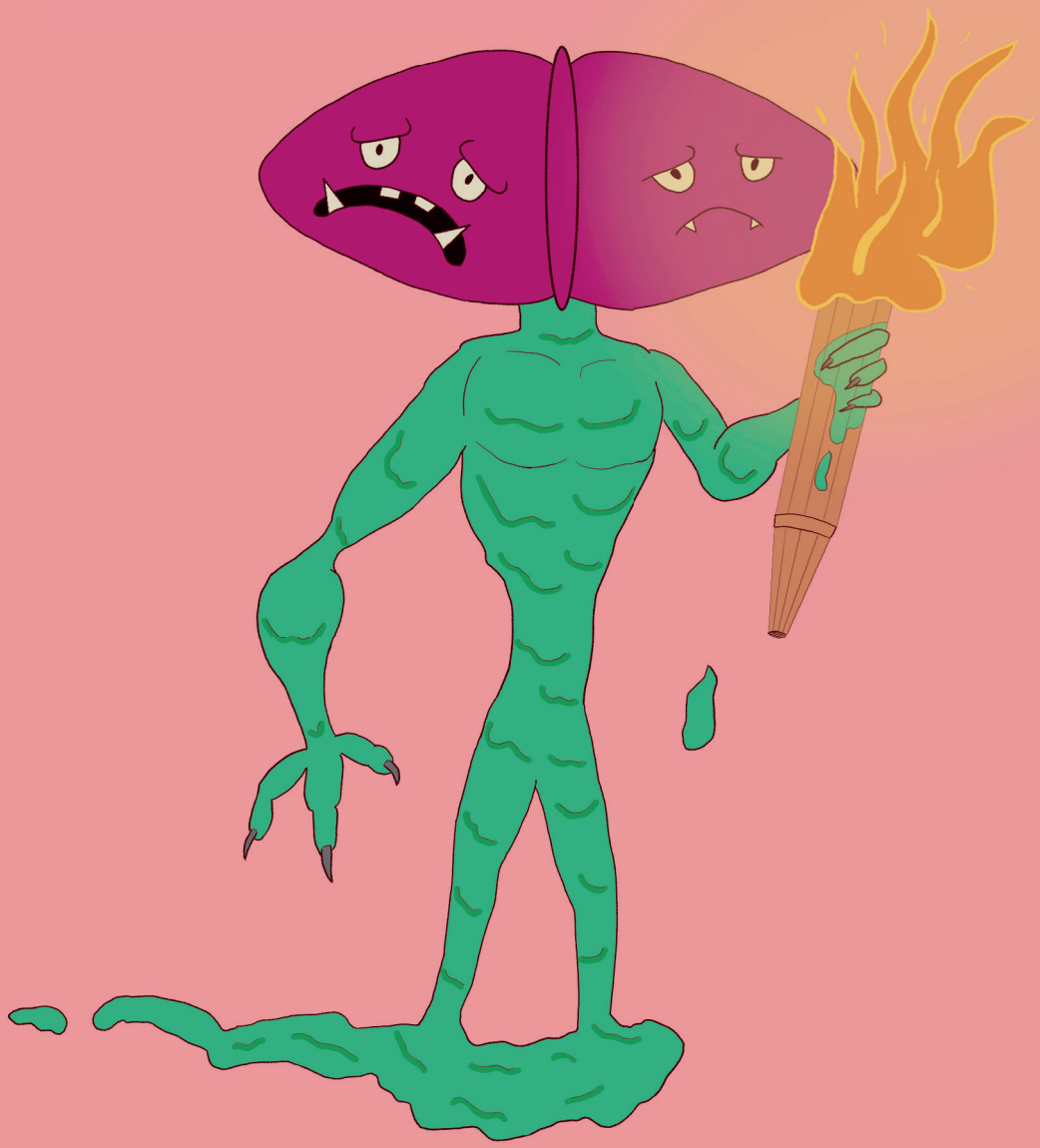
98. Li, T. & Yin, Y. Critical assessment of pan-genomic analysis of metagenome-assembled genomes. *Briefings in Bioinformatics* **23** (2022). https://doi.org:10.1093/bib/bbac413

99. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066 (2002). https://doi.org:10.1093/nar/gkf436

100. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009). https://doi.org:10.1093/bioinformatics/btp348

101. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020). https://doi.org:10.1093/molbev/msaa015

102. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587-589 (2017). https://doi.org:10.1038/nmeth.4285



Chapter 7.2

Draft and complete genome sequences of 17 *Streptococcus* species

Sam Nooij[a], Ingrid M.J.G. Sanders[a], Lesley Schout[a], Rolf H.A.M. Vossen[b], Susan L. Kloet[b], Wiep Klaas Smits[a], Quinten R. Ducarmon[a]

[a]Leiden University Center of Infectious Diseases (LUCID), Leiden University Medical Center, Leiden, The Netherlands
[b]Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands

Microbial Resource Announcements, Accepted 7 July 2025

Abstract

We present seventeen near-complete and complete genomes of *Streptococcus* species obtained from eight mixed cultures of presumed *Ruminococcus gnavus* isolates. The genomes are classified as eight different *Streptococcus* species and three are unclassified (currently have no species representative available in databases). We provide these high-quality genomes to the scientific community for further scrutinization.

Announcement

In a project focusing on isolating and sequencing genomes of *Ruminococcus gnavus* (reference *R. gnavus* manuscript), we sent presumed pure *R. gnavus* DNA for long-read PacBio circular consensus sequencing (CCS) to the Leiden Genome Technology Center (Leiden University Medical Center, Leiden, Netherlands). The DNA libraries were prepared using the SMRTbell prep kit 3.0 kit and sequenced on the Sequel II platform (Pacific Biosciences of California, Inc., CA, USA). Raw reads were assembled with Flye[1] (version 2.9.2) and resulting contigs were classified using the Contig Annotation Tool[2] (CAT; version 5.2.3). Contigs were reoriented to start at the *dnaA* for putative chromosomes using dnaapler[3] (version 0.3.0). To our surprise, eight of the samples returned complete *Streptococcus* genomes and short (7-55kbp), predicted *R. gnavus* contigs with an estimated depth of coverage of 4X. As we suppose these Streptococci are lab contaminants, we cannot be certain of their exact origin. Nonetheless, we aimed to generate complete genomes and improved assemblies of *Streptococcus* genomes by using a metagenome assembly approach with metaFlye[4] (version 2.9.2). This yielded up to 194 contigs per sample with a median length of 23kbp. We separated the contigs classified as *Streptococcus* using seqkit[5] (version 0.16.0) and did a quality control consisting of: 1) a length and GC-content check by QUAST[6] (version 5.0.2), 2) completeness and contamination estimation by CheckM[7] (version 1.0.13), CheckM2[8] (version 1.0.1) and BUSCO[9] (version 5.4.3; using flag '--auto-lineage-prok'), 3) gene annotation with Bakta[10] (version 1.6.1), and 4) taxonomic classification with GTDB-Tk[11] (version 2.3.2) using the Genome Taxonomy database (GTDB) version r207_v2. All tools were run with default parameters unless stated otherwise. Relevant statistics of each genome are listed in table 1. Furthermore, to determine the relatedness between the *Streptococcus* genomes and infer if they might originate from a common source, we used fastANI[12] (version 1.33) to calculate pairwise average nucleotide identity (ANI; figure 1). We found five *Streptococcus equinus* (NCBI taxonomy; GTDB: *Streptococcus sp001556435*) highly similar genomes (ANI ≥ 99.9%) from five different samples, four *S. oralis* (GTDB: 2 × identical *S. oralis_S* and 2 × *S. oralis_Y*) from three different samples, and six genomes with slightly different classifications that are all close to *S. parasanguinis* (ANI ≥ 93.8%) from three different samples (1×3, 1×2, 1×1 genome/sample). We found one genome classified as *S. sp902363395* (GTDB), which is unknown

to NCBI, that remotely resembles (ANI ~86.4%) the *parasanguinis*-like genomes. The final genome was classified as *S. sanguinis* SK49 (GTDB: *S. sanguinis_C*) and is only remotely similar to the other genomes (ANI ≤ 80%).

All raw reads and assembled genomes are available through the European Nucleotide Archive under project number PRJEB76410.

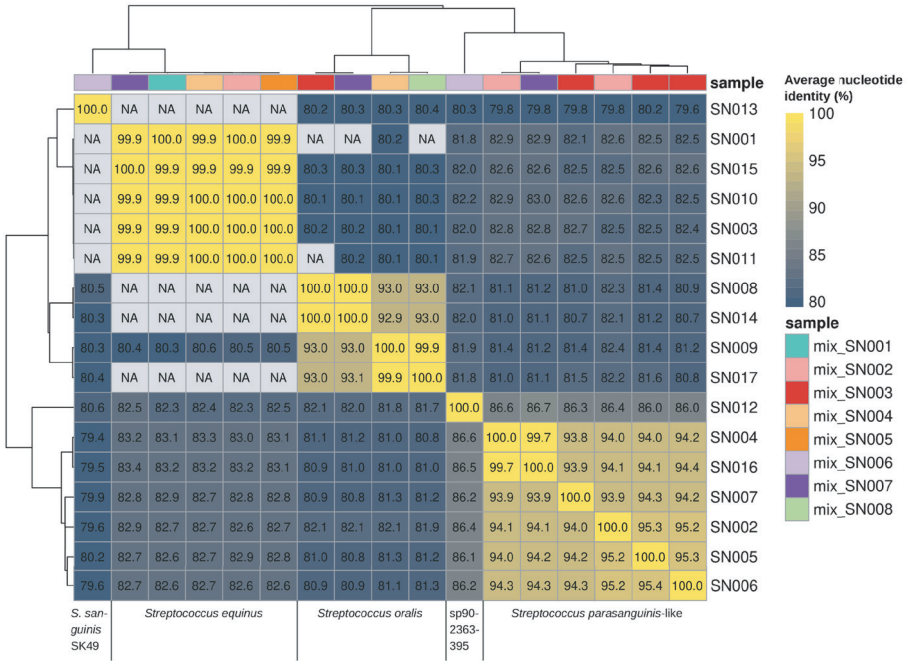


Figure 1. Whole-genome comparison of 17 *Streptococcus* genomes. Whole-genome sequences of 17 *Streptococcus* bacteria were compared using Average Nucleotide Identity (ANI). Values in cells indicate percentage ANI. Genomes are annotated with the sample they derive from.

7.2

Acknowledgements

This study was sponsored by the Leiden University Fund / Dr. F.F. Hofman Fonds, www.luf.nl.

References

1. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol *37*, 540-546. 10.1038/s41587-019-0072-8.

2. von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol *20*, 217. 10.1186/s13059-019-1817-x.

3. Bouras, G., Grigson., S., Papudeshi., B., Mallawaarachchi V., Roach, M. J. (2023). Dnaapler: A tool to reorient circular microbial genomes https://github.com/gbouras13/dnaapler.

4. Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., and Pevzner, P.A. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods *17*, 1103-1110. 10.1038/s41592-020-00971-x.

5. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962. 10.1371/journal.pone.0163962.

6. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072-1075. 10.1093/bioinformatics/btt086.

7. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res *25*, 1043-1055. 10.1101/gr.186072.114.

8. Chklovski, A. (2024). CheckM2. https://github.com/chklovski/CheckM2.

9. Manni, M., Berkeley, M.R., Seppey, M., Simao, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol *38*, 4647-4654. 10.1093/molbev/msab199.

10. Schwengers, O., Jelonek, L., Dieckmann, M.A., Beyvers, S., Blom, J., and Goesmann, A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom *7*. 10.1099/mgen.0.000685.

11. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics *38*, 5315-5316. 10.1093/bioinformatics/btac672.

12. Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun *9*, 5114. 10.1038/s41467-018-07641-9.

Table 1. Streptococcus genome assembly characteristics and quality measures. Initial assemblies were made with Flye. In case of multiple linear contigs, a secondary assembly was made using metaFlye. The most complete assembly was selected from either approach.

Genome	Length (bp)	Predicted genes	GC%	Circular?	Species	CheckM completeness	CheckM contamination	Assembly method	Depth of coverage
SN001	2140162	1914	40.12	Y	*Streptococcus equinus*	99.69	0.15	metaFlye	17
SN002	2223664	2093	41.52	N	*Streptococcus* sp. HMSC072G04	100	0.09	Flye	13
SN003	2141365	1913	40.11	Y	*Streptococcus equinus*	99.69	0.15	Flye	21
SN004	2160498	2021	42.06	N	*Streptococcus* sp.	100	0.17	metaFlye	18
SN005	2125957	2016	41.72	Y	*Streptococcus parasanguinis*	99.66	0.11	Flye	120
SN006	2081198	1934	42.11	Y	*Streptococcus parasanguinis*	100	0.07	Flye	120
SN007	2118644	2017	41.95	Y	*Streptococcus* sp.	100	0.23	Flye	158
SN008	2068017	1940	41.06	Y	*Streptococcus oralis*	99.87	0.04	Flye	25
SN009	2117116	2018	40.83	Y	*Streptococcus oralis*	99.87	0.06	metaFlye	24
SN010	2147719	1922	40.12	Y	*Streptococcus equinus*	99.69	0.15	metaFlye	25
SN011	2141236	1914	40.11	N	*Streptococcus equinus*	99.69	0.15	metaFlye	14
SN012	2025938	1881	42.21	Y	*Streptococcus* sp.	100	0.17	metaFlye	21
SN013	2316732	2233	43	N	*Streptococcus sanguinis*	100	0	metaFlye	20
SN014	2068026	1939	41.06	Y	*Streptococcus oralis*	99.87	0.04	metaFlye	117
SN015	2141124	1914	40.12	Y	*Streptococcus equinus*	99.69	0.15	metaFlye	87
SN016	2166044	2028	42.02	Y	*Streptococcus* sp.	100	0.17	Flye	153
SN017	2119531	2017	40.83	Y	*Streptococcus oralis*	99.87	0.6	metaFlye	48


サム

Chapter 8

General Discussion

Summary of main findings

Part 1: Concomitant microbiota impacts after faecal microbiota transplantation for recurrent *Clostridioides difficile* infections

Clostridioides difficile is a Gram-positive bacterial pathogen that causes gastrointestinal infections and forms spores that are resistant to antibiotic treatment. Besides, *C. difficile* thrives in an antibiotic-depleted gut microbial environment, leading to a high incidence of recurrent infections. Recurrence rates after first infection are 15-30%, and 40-65% after one or more recurrences[1]. Faecal microbiota transplantation (FMT) has proven to be an effective method to restore the gut microbiota and prevent further recurrences of *C. difficile* infections (CDI). Furthermore, FMT serves as an interesting model for studying the transfer of human gut microbiota. This first part of the thesis leverages FMT data from multiple recurrent CDI patients to study effects on other potentially harmful bacteria.

Chapter 2 focuses on a recently discovered microbial risk factor associated with the development of colorectal cancer: colibactin-producing (*pks*$^+$) *Escherichia coli*. We screened stool metagenomic data from FMT recipients and their respective healthy donors to assess the presence of *E. coli* and the *pks* operon. The objective was to determine both the presence and relative abundance of *E. coli* in general, and the subpopulation capable of producing colibactin. Thereby we also assessed the effect of FMT on this putative carcinogenic bacterium. We found that *pks*$^+$ *E. coli* was present in donors and FMT-treated patients, and decreased in patients after FMT, which was correlated with the absence of *pks*$^+$ *E. coli* in the donor. We found no evidence supporting the transmission of *pks*$^+$ *E. coli* from donor to patient. Consequently, we conclude that FMT affects putatively carcinogenic *E. coli* by reducing their presence and abundance, particularly when the donor is free of the *pks*$^+$ *E. coli* variant. Our results suggest that FMT is more likely to mitigate the risk of colonisation by carcinogenic bacteria rather than elevate the risk through bacterial transfer. This proof of concept may inspire the development of microbiota-based therapies aiming to prevent the development of colorectal cancer.

In **chapter 3**, we studied more FMT sample triads (comprising donor, patient before FMT, and patient after FMT) along with long-term follow-up samples. Through a combination of traditional bacterial culture, whole-genome sequencing (WGS) and metagenomic sequencing, we extensively characterised antibiotic resistance in this cohort and evaluated the effect that FMT exerted on antibiotic-resistant bacteria. Our results indicated that FMT reduced the prevalence of multidrug-resistant (MDR) bacteria in patients. WGS combined with metagenomics suggested that MDR bacteria may persist post-FMT, although their relative abundances were significantly reduced, making them harder to detect. Further analysis of the metagenomic data showed that the relative abundance of antibiotic resistance genes (ARG) decreased after FMT,

while the richness of these genes remained unchanged. Computational predictions identified which ARG-containing contigs were plasmid-derived and we found their quantity was not significantly affected by FMT. In summary, we hypothesise that FMT may effectively reduce MDR bacteria and ARG abundance, potentially decreasing the risk of infection. However, it appears that FMT may not completely eradicate antibiotic-resistant bacteria from the gut microbiome. Thus, FMT could serve as an additional strategy for combatting the spread of antibiotic resistance within a patient population that is at increased risk of bacterial infections.

Part 2: Microbiota alterations following faecal microbiota transplantation for ulcerative colitis

Ulcerative colitis (UC) is a form of inflammatory bowel disease characterised by chronic inflammation of the colon, which is associated with alterations in gut microbial composition and function. A randomised clinical trial was conducted to assess the safety of FMT in these patients, and data were collected to evaluate bacterial colonisation of the recipients' gastrointestinal tracts, both with and without anti-inflammatory pretreatment. **Chapters 4-6** describe analyses of this dataset and the changes in gut microbiota following repeated FMT.

Chapter 4 describes the clinical trial and bacterial engraftment using a combination of microbiota diversity metrics. We hypothesised that anti-inflammatory pretreatment would facilitate colonisation of foreign bacteria. However, our findings indicated that the donor had a more pronounced effect on engraftment. Therefore, we conclude that donor selection and an improved understanding of how to identify optimal donors may be crucial for the success of FMT in UC.

Chapter 5 builds upon the findings of **chapter 4** by delving deeper into microbial ecology. The main objective was to identify changes in the microbial communities correlated with clinical remission. Using a computational modelling approach, we defined clusters of bacterial profiles. Among these clusters, we found one associated with poorer clinical outcomes, specifically a failure to achieve remission. Conversely, a cluster characterised by high relative abundances of *Ruminococcaceae* and *Lachnospiraceae* was associated with treatment success. Using these broad microbiota characteristics, we hypothesised that it may be possible to predict treatment success of FMT early after the procedure.

Chapter 6 continues the investigation from **chapter 4 and 5**, with a combination of engraftment analyses with microbiota dynamics of the bacteria present in the patients before FMT. This adapted methodology addresses a slightly different question: how do the dynamics of donor and patient species correlate with FMT success? We observed that patient species present at high relative abundances often persisted after FMT. In terms of donor-derived species, moderate and stable colonisation of donor species

was correlated with treatment success, while initial high engraftment followed by loss of donor species was indicative of failure to achieve or maintain remission. Taken together, we predict that FMT treatment success in UC depends on resilience of the recipient's resident microbiota, combined with the ability to stably incorporate a moderate amount of healthy donor-derived microbiota.

Part 3: Global distribution and genome biology of gut bacterium *Ruminococcus gnavus*

Chapter 7 revolves around the gut bacterium *Ruminococcus gnavus*, which is strongly associated with inflammatory bowel disease, specifically Crohn's disease. At the same time, it has been described to be present in around 90% of healthy adults. Our objective was to re-evaluate these correlations and elucidate the genomic basis of the contrasting host-microbe interactions. We discovered that *R. gnavus* is more prevalent and abundant in conditions such as IBD, type-2 diabetes, hypertension and atherosclerotic cardiovascular disease compared to healthy people. This is also true for Westernised compared to non-Westernised societies and more in infants and young children compared to adults. Based on a large collection of complete and draft genomes, we found that bacterial motility may be overrepresented in infant-derived strains, and isolates derived from healthy people are phylogenetically and functionally different from isolates from Crohn's disease patients. We conclude that there may be distinct subspecies of *R. gnavus* that co-evolved and adapted to environments or lifestyles. Our work shows that within a species there may be several genomically and phenotypically distinct variants. Therefore, more nuance is warranted when attributing disease to bacterial species or describing host-microbe relations in general.

Finally, **chapter 7.2** is a spin-off of **chapter 7.1,** in which we describe complete genomes derived from presumed laboratory contaminants. Our analysis revealed that the sequence data from supposed *R. gnavus* isolates contained DNA from different species, prompting us to computationally separate the sequences belonging to these various species. This resulted in the identification of seventeen *Streptococcus* genomes, one genome from *Bacteroides fragilis* and one *Staphylococcus capitis*. The latter two genomes have not been described in the publication; their genomes have also been deposited in the European Nucleotide Archive public repository. Although the description of these bacteria does not fit the scope of **chapter 7**, we recognise that their genome information could be valuable to the scientific community. As strong supporters of open science and the FAIR data principles, we have made these genomes publicly available for further research.

A broader view of FMT and gut microbiota

The further discussion covers the following topics, which extend from the work presented here to the broader developments in the scientific field:

Unpredictability of FMT: for better or worse

FMT is a promising therapy for various diseases in which the gastrointestinal microbiota plays a role. For example, it can be used to 1) prevent recurrence of *C. difficile* infections[2,3], 2) induce remission in IBD[4,5], and 3) help relieve symptoms of immune checkpoint inhibitor-induced colitis[6,7] and gastrointestinal acute graft-versus-host disease[8-10]. However, as with any therapeutic intervention there is a risk of side-effects[11,12]. The most reported adverse events relate to the procedure itself and the diseases being treated, including increased stool frequency, abdominal pain, nausea, and diarrhoea[13]. The most severe described adverse event is the death of an immunocompromised FMT recipient caused by antibiotic-resistant bacteraemia[14]. The bacterial pathogen was obtained from the donor, who had not been screened for antibiotic-resistant bacteria. Additionally, there have been reports of self-administered FMT from family members providing relief from gastrointestinal symptoms but also resulting in the emergence of a seemingly unrelated condition of acne[15]. Conversely, previous studies, including **chapters 2 and 3** of this thesis, have shown that FMT can also yield beneficial side-effects[16-22]. Next to resolving CDI recurrences, FMT can modulate the gut microbiota by reducing antibiotic-resistant bacteria and putative carcinogenic bacteria. This underscores the therapeutic potential of FMT, while also

stressing the need for rigorous donor screening, and safe, standardised administration methods[23,24]. Furthermore, the effects of FMT are usually studied on the timescale of weeks, while some effect may only arise after years' time. Therefore, the use of registries for both FMT donors and recipients would be useful for monitoring and evaluating these potential long-term effects. FMT is a versatile treatment for various human microbiome-related health issues that merits further investigation. It should only be administered in a professional and controlled setting, with careful monitoring of outcomes.

FMT for IBD: inducing and maintaining remission remains challenging

FMT has been repeatedly tested for its potential to treat IBD, particularly ulcerative colitis, with variable outcomes[5,25]. FMT is capable of inducing remission in UC patients, but success rates are significantly lower compared to those for multiple recurrent CDI, with a mean clinical remission rate of 42% for IBD[25], compared to a 92% clinical resolution rate for CDI[26]. To compare, placebo treatment achieved a 22.6% mean remission in the IBD trials[25]. Next to a different aetiology, the difference in treatment success may partly be explained by the antibiotic pretreatment used for FMT in multiple recurrent CDI (rCDI). The rationale of antibiotic pretreatment is that reducing the patients' existing gut microbiota may facilitate stable introduction of the healthy donor microbiota[27]. While antibiotics are standard therapy for rCDI, they are not typically employed in the treatment of IBD. This is but one of the main factors complicating one-to-one comparisons between these two conditions and FMT.

It has been hypothesised that a gut microbiota that is more similar to the donors', achieved by high bacterial engraftment, should lead to a healthier and less inflamed gastrointestinal tract in IBD patients. To stimulate engraftment, FMT for IBD might benefit from similar pretreatment as rCDI, as successfully demonstrated in previous studies[28]. Contrastingly, **chapter 4** illustrates that anti-inflammatory pretreatment, which is commonly used for symptom management in IBD, does not enhance colonisation of donor-derived bacteria. An alternative strategy to increase engraftment is to optimally match the donor and recipient microbiota, as suggested in **chapter 6** and other studies[29,30]. Matching may be done based on gut microbiota enterotypes or microbiota distance between donor and recipient[30], or based on recipient microbiota composition and dysbiosis[29]. The methodologies to calculate these parameters are not standardised, may be dataset-specific, and depend on code and instructions provided by the authors. Using microbiota parameters for matching donors to patients requires more thorough preparation to gain detailed knowledge of both donor and patient microbiota, thereby increasing the resources and expertise needed to administer FMT to IBD patients effectively. Nonetheless, how to achieve maximum bacterial transfer and colonisation, and whether that is beneficial to the recipient's health, remains to be established.

Furthermore, IBD is a chronic disease that is unlikely to be cured by a single intervention. The disease itself and the proposed treatment mechanisms are entirely different from CDI and no immunosuppressive therapy achieves response rates as high as those for FMT in rCDI. FMT may be used to reach a state of remission, but to maintain remission repeated FMTs[28], or alternative microbiota modulating therapies such as prebiotics are probably necessary[31]. On the other hand, if IBD is conceptualised as a chronic overreaction of the immune system, like an allergy, a possible treatment approach may also be inspired by modern allergy immunotherapies. Several approaches have been proposed that target the intestinal immune system and trigger anti-inflammatory pathways using small molecules or biological agents[32]. Combining the immune system-targeting approach with techniques from vaccine development[33], one could envision re-training of the immune system by administering microbial surface molecules which may be engineered to manipulate the immune response. Or looking further, it might be possible to design an mRNA vaccine that helps protect against pro-inflammatory bacteria, similar to what has been demonstrated for *C. difficile*[34]. These lines of therapeutic research may complement one another and necessitate a deep understanding of both the gut microbiota and immune system. Which therapy or combination of therapies is most efficacious is probably patient-dependent, with a variety of host-microbe characteristics being relevant. More trials and multidisciplinary research into the interactions between the immune system and gut microbiota are needed to assess effectiveness of the different treatment options for IBD.

Finding the balance

The goal of FMT is to restore a functional gut microbiome, a system in which interactions between human and microbiota are in balance: a state of homeostasis[35-37]. It is thought that finding a proper match between donor and patient microbiota is necessary to re-establish a balanced gut microbiota in the recipient patient. To find this match, researchers have experimented using machine learning models to predict optimal donor-recipient combinations based on microbiota composition, pretreatment and clinical outcome measures[27,29,30,38]. By using such high-dimensional and rich data, solutions may be found that go beyond our current understanding of gut microbiome biology.

The optimal complementary donor microbiota for a patient may depend on the disease and other factors. In the case of rCDI, the main goal is preventing outgrowth and toxin production of *C. difficile*. Other diseases may likewise be treated by inhibiting specific deleterious species, or by shifting the microbiome's collective metabolism, such as with the drug levodopa in Parkinson's patients[39-41]. As for other factors that influence engraftment and thereby possibly treatment success, the microbiota themselves need to adapt to their new environment: the recipient's colon. The colons of donor and recipient will be different in various aspects relevant to the microbiome, including their physiology and immunology[42-44]: they are different ecosystems. Furthermore,

the compatibility rate of microbial communities – that is, the extent to which two microbiotas can co-exist – may also be influenced by lifestyle factors such as diet or physical exercise. These factors have not yet been extensively studied in connection with FMT, because the main disease for which it is used, rCDI, has had high success rates irrespective of the exact microbiota composition[26]. In fact, it is striking that although it is known that diet and exercise impact the gut microbiota[45-47], there appear to be few guidelines to provide recipients with advice or support as to what to eat and do after FMT. Would it be beneficial for recipients if they adopt their donor's lifestyle? For how long, and what aspects matter most? A 2020 survey among mostly gastroenterologists indicated that most FMT experts find it important to consider the diet of both donors and recipients[48]. Moreover, there have been studies on the combined effect of diet and FMT[49-51], and the Australian Centre for Digestive Diseases has published dietary requirements for stool donors prior to donation[52]. Taken together, there appear to be opportunities to include more disciplines into FMT research and thereby further optimise the treatment. However, added complexity should not impose barriers on stool banks, donors, treating physicians and patients.

Monitoring of long-term FMT outcomes

FMT is not a frequently applied therapy, and its variable composition necessitates meticulous monitoring of outcomes. This monitoring is primarily conducted by the treating physician and recorded by the stool banks and registries of the FMT providing centre. Stool samples may be collected in the weeks following FMT to evaluate microbial composition. If the FMT indication pertains a non-chronic illness and the patient reports no further complaints, follow-up with the physician typically ends. Consequently, should new complaints arise, and the patient turns to another physician, a potential link to the FMT may be overlooked. This complicates the long-term evaluation of FMT and identification of potential deleterious side-effects, such as carcinogenesis or infection with MDR bacteria. To accommodate longer-term monitoring of FMT, initiatives for maintaining registries are developed at both national[53,54], and international[55] levels. These registries document the relevant patient history leading to FMT and any subsequent events that can be linked to the FMT, i.e., the introduction of donor-derived bacteria that cause harm over an extended time period. Examples include colorectal carcinogenesis and multidrug-resistant bacterial infections as discussed in **chapters 2 and 3**. As more patients are recorded in the registries, the likelihood of identifying rare occurrences increases, helping FMT researchers enhance FMT safety. Thus, these registries serve as an invaluable tool for advancing our understanding of the long-term effects of FMT.

Controlling FMT quality

Quality control for FMT has two primary aims: 1) to prevent deleterious side-effects (risk management), and 2) achieve the best possible results (efficacy). Both demand a thorough understanding of the faecal microbiota, which can be estimated using

different methods. Donor screening is mostly used to mitigate risks associated with FMT, but can also help inform about the potential efficacy. The NDFB employs a rigorous donor screening protocol, utilising questionnaires addressing risk factors for transmissible diseases and conditions associated with a perturbed microbiome. This primary screening excludes approximately 70% of volunteers that applied to become donors[13,23]. Exclusion criteria include gastrointestinal complaints, recent antibiotic use, travel to countries with endemic gastrointestinal pathogens or MDR bacteria, and a family history of IBD[23]. To predict treatment efficacy, microbial profiling may be used. A Chinese research group has published guidelines that incorporate 16S rRNA gene sequencing for microbial profiling and a machine learning model that uses a random forest approach to match patients with donors based on predefined microbial clusters, so-called enterotypes[30]. Although these methods effectively maintain high success rates for rCDI with minimal side-effects, they can only prevent transfer of known harmful microorganisms. Microbial profiling before FMT may not be necessary for the effectiveness against rCDI, as success rates have been consistently high (~90%) regardless of bacterial composition[26]. It is not technically and financially feasible to determine the exact composition of donor faeces and classify all bacteria, fungi, viruses, metabolites, and food particles. Besides, if the specific microorganisms and metabolites responsible for the therapeutic effects were fully understood, it would be more practical to isolate them and compose a conventional medicine administered as a capsule. Such a drug is referred to as a live biotherapeutic product (LBP; figure 1) and currently under development by several companies[56,57]. LBPs are discussed in more detail below. To conclude, there are several established ways of controlling FMT quality and when knowledge regarding the mechanisms of action accumulates for all the target diseases, FMT will likely be superseded by human-designed drugs.

Next generation of microbiota therapies

The future of microbiota therapy is anticipated to be centred around LBPs (Figure 1). LBPs consist of live microbial organisms selected or developed to treat or prevent human diseases[56]. The primary advantage of LBPs is that their composition is known and stable, unlike donor stool, which facilitates regulation as a drug and should provide more robust and predictable outcomes. This should also make LBPs safer than FMT. However, LBPs as treatment require a deeper understanding of the human gut microbiome obtained through research, and requires significant investments in trials for regulatory approval. Therefore, LBPs are initially quite expensive, posing challenges in competing with FMT. Nevertheless, given their advantages LBPs are likely to eventually replace FMT for routine treatments in the long term.

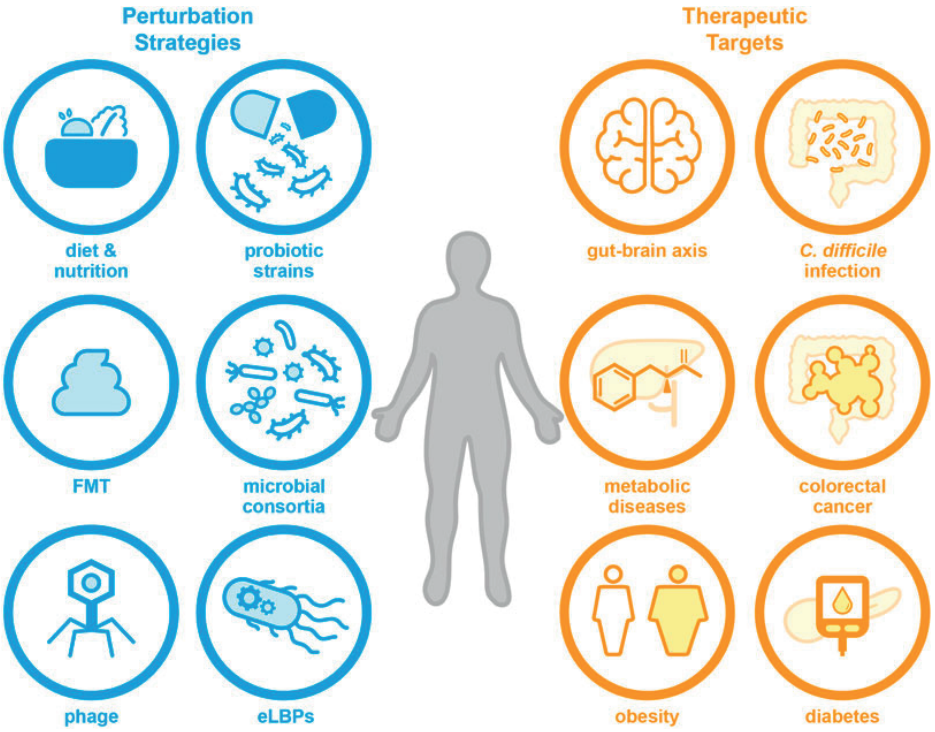


Figure 1. Techniques and aims of microbiome therapies. On the left side, different intervention strategies are listed in blue. Probiotic bacterial strains, microbial consortia and engineered bacteria are examples of live biotherapeutic products. The right-side lists in orange targets for which microbiota therapies are developed or have been tested successfully. FMT: faecal microbiota transplantation, eLBP: engineered live biotherapeutic product. Figure from reference[58].

Companies have developed LBPs using various methods and with different aims of resolving gastrointestinal complaints. For the treatment of rCDI there are at least five products from three different companies: RBX2660[59], RBX7455[60], VOWST (formerly SER-109)[61], VE303[62,63], MET-2[64], and NTCD-M3[65]. These can be broadly categorised in three groups. The first two (RBX2660 and RBX7455 from Ferring Pharmaceuticals) are derived directly from donor stool and are thus practically identical to 'traditional' FMT. RBX2660 is a suspension for rectal administration and RBX7455 is a freeze-dried (lyophilised) product that is administered as oral capsule[59,60]. Next, MET-2 (from NuBiyota), VE303 (from Vedanta Biosciences) and VOWST (from Seres Therapeutics) are encapsulated consortia of well-characterised bacteria. RBX2660 and VOWST have passed phase 3 clinical trials and have been approved for use against rCDI by the U.S. Food & Drug Administration (FDA[66,67]). RBX7455 and MET-2 have successfully completed a phase 1 trial and VE303 is now in a phase 3 trial. NTCD-M3 (from Destiny Pharma) is a non-toxigenic strain of *C. difficile*, which has been administered through oral liquid formulation in a phase 2 trial. Furthermore, LBPs for treating IBD are under development (GUT-103 and GUT-108[68], and VE202[69-71]). These also consist of well-

defined bacterial consortia. VE202 (From Vedanta Biosciences) is now at stage 2 clinical trials, while GUT-103 and GUT-108 (from Gusto Global) have yet to be tested in a phase 1 trial. Finally, LBPs have been described that consist of a single bacterial strain, which are often genetically engineered to have a specific effect in the gut microbiome[56]. This includes NTCD-M3[65], *E. coli* CWG308[72], *E. coli* Nissle 1917[73-75], and ADS024[76]. As an example application, a strain of *E. coli* Nissle 1917 engineered to overproduce microcin I47 has been found to be effective against multidrug-resistant bacteria[56,73]. It is important to note that this *E. coli* strain also poses a health risk as discussed in **chapter 2**: *E. coli* Nissle 1917 produces colibactin and may contribute to carcinogenesis[77,78]. This illustrates the need for proper assessment and regulation of probiotics and especially the use of genetically modified organisms. This poses challenges for their development and market release, particularly in Europe, where approval processes are generally more stringent. To summarise, the development of LBPs is costly and time-consuming. Until LBPs are approved for use and readily available, FMT remains a valuable experimental treatment for gut microbiota-related conditions.

Deeper understanding of host-microbiota interactions

When studying the microbiota, it is common to refer to individual members as species. However, as shown in **chapters 2, 3 and 7**, a microbial species is a group of possibly divergent organisms under one umbrella name. Evolution has produced a spectrum of different phenotypes, and pathogenicity is not necessarily linked to species (Figure 2). A comprehensive understanding of the microbiota requires consideration of this variation. Several research groups have studied intraspecies variants from metagenomics data in the context of FMT[79-88]. They have used different bioinformatics approaches to differentiate variants based on genomic features, which are summarised below. One group of methods relies on pre-built databases to identify and quantify variants of known species:

- *StrainPhlAn* relies on a custom database of species-specific markers to identify the dominant strain for each species and compare their consensus sequence between samples[82]. StrainPhlAn has been developed further together with MetaPhlAn and is now at version 4 with an extended database to capture a greater diversity of prokaryotes[89].
- *Strainer* utilises genomes of cultured isolates to create a database of unique k-mers for each strain and then matches metagenomic reads to these k-mers to quantify strains[80].
- *StrainPanDA* uses a pangenome-based approach to reconstruct gene content variation per strain[84].
- *SameStr* uses species-specific marker genes to identify single-nucleotide variants but is flexible with databases and can analyse multiple strains per species[85].
- *SNV-FEAST* introduces a new single-nucleotide variant signature identification method based on read mapping to a custom bacterial reference genome database and provides relative abundances of variants in source and sink samples[87].

- Having a pre-built database facilitates the application of these tools to datasets of well-studied microbiomes. Other tools make use of the metagenomic data or user-defined references to call variants of any gene or genome, including:
- *Strain Finder* uses a machine learning approach to estimate the optimal number of strains from reference genome alignments, and then calculates frequencies of these strains across metagenomes[79].
- *inStrain* relies on paired-end reads mapped to any reference sequence, which can be a metagenome-assembled genome from the same dataset and calls single-nucleotide variants to distinguish strains in metagenomes[81].
- *STRONG* uses *de novo* assembly of metagenomic reads to identify strains based on assembly graphs[83].
- *StrainGE* specialises in low-abundance variants and uses read mapping to reference genomes, which may be supplied separately by the user[86].
- *ChronoStrain* focuses on low-abundant taxa and models variant abundances over time with uncertainty quantification using a read-based approach and a custom database of user-specified reference marker sequences[88].

These tools require additional user input but can handle any microbiome regardless of reference databases. While these tools allow for identification of intraspecies variants across samples, and some enable studying lowly abundant variants, only StrainPanDA explicitly incorporates functional information, which is essential for elucidating host-microbe interactions.

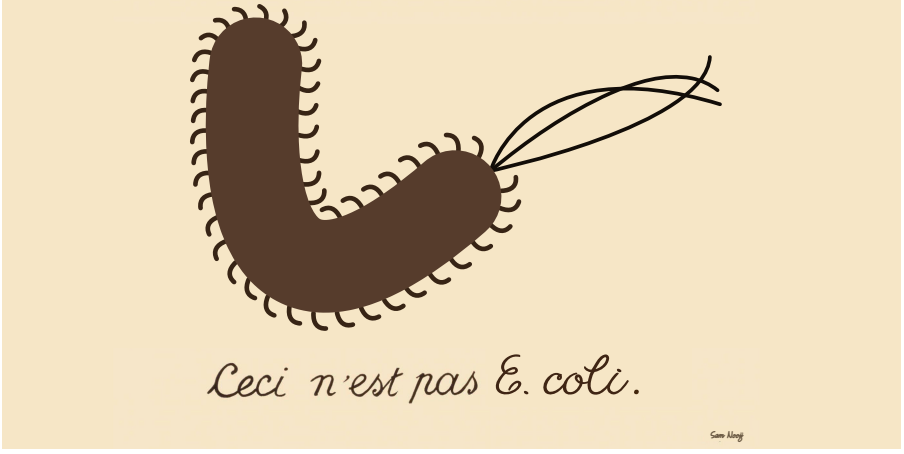


Figure 2. This is not *E. coli*. What appears to be a harmless commensal may actually be a source of infection or harbinger of cancer. A detailed scrutiny of the genome and phenotype is necessary to accurately assess the implications of bacteria and the potential effects on its host. After René Magritte, La Trahison des Images (1929).

The next step in understanding gut bacterial function is to examine the genomes' encoded functionality. This may be partially achieved by using metagenomics

sequencing, which allows for identification of complete genes, representing the functional potential. For example, the recently published tool microSLAM takes into account population structure within metagenomes and the presence of genes to assess associations between species and their genes with diseases[90]. However, to really understand what is going on in the gut microbiome, one should sample not only the DNA but also look at mRNA, proteins, and metabolites. The combination of these techniques in high-throughput fashion is known as 'multi-omics' and discussed further below.

Gut bacteria as double-edged sword

Continuing the discussion of intraspecies variation and differences in pathogenicity; numerous gut bacteria have been characterised as either beneficial or detrimental to the health of their human host (Figure 2). Examples of such ambivalent species include *Clostridioides difficile*[91], *Escherichia coli*[92], *Akkermansia muciniphila*[93], *Prevotella* spp.[94-96], and *Ruminococcus gnavus*[97]. Accurate diagnosis, treatment, and prevention of disorders require methods to distinguish harmful or pathogenic bacteria from commensals. Also, whether or not a bacterium is seen as pathogen depends on host conditions such as immune response. The same bacterium may be harmless when in one person and induce severe symptoms in another, for example with immunocompromised patients, and this may change over time. Therefore, we need to acknowledge that 'pathogen' and 'commensal' describe behavioural traits rather than fixed entities. Proper context is essential to assess these behaviours. For example, *C. difficile* may reside in the gut unnoticed by its host and cause disease upon depletion of the gut microbiota by antibiotics. Or particular variants of *E. coli* may be used as probiotic[98], while other (shiga toxin-producing) strains are severe pathogens[99]. A bacterium's capabilities are primarily dictated by its genetic makeup, which can be determined through genomic analysis. Exceptions such as horizontal gene transfer aside, the genome determines the range of possible behaviours or lifestyles. Therefore, genomics and pangenome analyses are valuable tools for initial screening of potential phenotypes.

Genomic flexibility and horizontal gene transfer

A major driver of bacterial evolution and associated genomic diversity is horizontal gene transfer (HGT). Bacteria may take up DNA from their environment through various mechanisms, which may then be incorporated into their chromosome. Evidence of HGT can be found in most bacterial genomes[100], and over half of a bacterium's genes may be mobilisable[101]. These mobile genes may be clinically relevant and encode antibiotic resistance[102], virulence factors[103], and genotoxins like colibactin in *E. coli*[104,105]. As a result, bacterial strains can rapidly adapt their functions when the opportunity arises. This underscores the importance of detailed characterisation methods, and several bioinformatics approaches have been devised to study HGT[101]. Below, I briefly summarise those relevant to bacterial genomics and metagenomics data.

Extrachromosomal mobile genetic elements such as plasmids can be recovered from metagenomes, although the short-read lengths (100-300 bp) typically used pose significant challenges to their accurate reconstruction. The presence of sequence repeats and sequences shared across different genomes, as well as variable read depths make *de novo* assembly particularly challenging, often resulting in fragmented contigs[101]. Nonetheless, identification of mobile element-derived contigs is possible through specific markers by aligning them to a database or using machine learning approaches[106-109]. This method was also applied in **chapter 3**.

Long-read sequencing, as used in **chapter 7**, facilitates reconstruction of extrachromosomal elements. PacBio's single-molecule, real-time platform, and Oxford Nanopore's platform generate sequence reads of 10,000 bp and longer. This is sufficient to bridge complex insertions and may enable recovery of full plasmids[110]. However, if not combined with culture isolation, linking a complete plasmid to its host organism remains difficult. To associate mobile genetic elements with their host, single-cell sequencing[111,112], and proximity ligation (hi-C) are attractive approaches[102,113-115]. Both can be combined with metagenomics to examine mobile elements in a large number of bacteria. These techniques will likely see increased use in the future, and uncover many more cases of HGT.

Beyond microbial genomics: what do gut bacteria do?

Back to the topic of omics methods. This thesis relies on genomics data, with a focus on bacterial DNA. Genomic methods are ideal for microbial profiling, that is, identifying which bacteria are present, essentially answering the question 'who is there?' Whole-genome sequencing techniques can infer gene presence by employing gene prediction tools and known gene databases, and answer 'what can they do?' To delve deeper into the processes of the microbiome and interactions with the host one needs different methods. These include metatranscriptomics to measure and identify RNA, metaproteomics to measure the enzymes that are produced and metabolomics to measure the products of metabolic activity (Figure 3). This additional information is invaluable for understanding the human gut microbiome to the level of molecular pathways and for designing effective interventions. Combining several of these methods, called multi-omics, is the most powerful approach in decoding the gut microbiota and requires specialised computational tools for integration and interpretation[116,117].

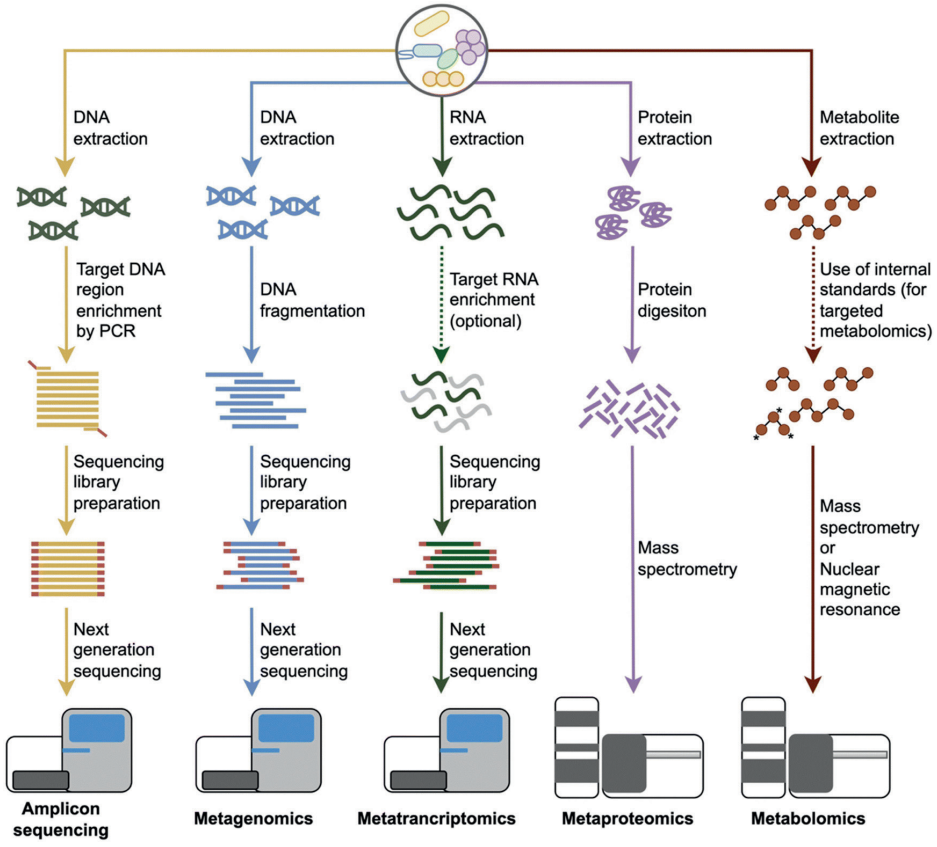


Figure 3. Sample preparation for multi-omics experiments. DNA, RNA, proteins, and metabolites maybe extracted from microbial communities and measured using indicated experimental workflows. Afterwards, data may be integrated using computational methods (not shown). Figure from reference[116].

Obstacles in high-throughput biological science

Modern high-throughput biological omics methods have proven to be very powerful in finding host-microbe interactions and formulating new hypotheses[118-121]. However, they are not without limitations. One major limitation is the inevitability of introducing sampling bias during sample collection, storage, and biochemical processing[122,123]. Despite being theoretically random and relatively unbiased, different experiments have shown that omics methods tend to introduce subtle biases, meaning the resulting data may not be entirely random or fully representative[124]. This also pertains to the identifiable taxa from metagenomic data: while metagenomics does not actively select for specific organisms, sampling kits and processing methods optimised for stool samples tend to predominantly yield bacterial DNA. The taxonomic biases are partly attributable to DNA extraction and library preparation protocols[125-127], and rely on computational databases that tend to reflect well-studied pathogens better than other

microorganisms and may contain biases or errors in the related taxonomy[128]. Therefore, to get a comprehensive view of the microbiome, it is still necessary to undertake multiple experiments and computational analyses. It may be more worthwhile to understand each method's inherent limitations, find the optimal approach for a given research question and report openly on methodological choices[122].

Metagenomics tends to rely on species-rank classifications, which may not be detailed enough to discriminate between disease-associated bacteria and harmless intraspecies variants. Fortunately, commonly used databases and classification methods are constantly improving[89,129], alleviating this problem. Also, the problem can be circumvented by combining the analyses with methods that include gene identification or similar functional inference, like employed in **chapters 2 and 3**.

A significant limitation of metagenomics includes the complexity of generating metagenome-assembled genomes (MAGs) from deeply sequenced, short-read metagenomes. MAGs depend on *de novo* assembly and binning tools to reconstruct as much of an organism's genome as can be predicted based on genetic information and specialised algorithms. However, this approach is liable to underestimate the genome due to 1) missing information because of (quasi-)random sequencing, 2) assembly artifacts or failure to correctly piece together all sequencing reads, and 3) binning errors, where the algorithm may incorrectly predict which assembled contigs belong together[130]. Within a metagenomic analysis, MAGs may be the best possible descriptors of the bacterium from which they derive, but they are typically less complete and more error-prone than genomes from cultured isolates. This we illustrated in **chapter 7**, where we show differences in length, completeness and GC content between MAGs and isolate genomes of *R. gnavus*, even though MAGs were predicted to be of high quality. Therefore, MAGs suffice for some research questions, but complete genomes from isolates are preferred when available.

An important obstacle to the adoption of multi-omics experiments is the high cost[117]. Multi-omics experiments are labour-intensive, require specialised equipment and consumables, and are also data-intensive, necessitating specialised computational expertise. As a result, only larger and well-funded laboratories have the resources to conduct such studies. Besides, depending on the research question, a simpler, more cost-effective experimental approach may be more appropriate. For example, when the possible targets are known detection by PCR may suffice and is faster and cheaper. High-throughput multi-omics experiments have been instrumental to advance our knowledge of the human microbiome, but are not the answer to all questions, nor can they fully replace traditional methods such as bacterial culture.

The proof is in the microbiological pudding

Bacterial culture has been the gold standard to identify and characterise bacteria. Next to facilitating whole-genome sequencing, cultured bacteria allow further experimentation, for example to study interactions with human tissue and the immune system. This enables researchers to validate hypotheses generated using computational experiments, helping to decode the complexities of the human microbiome. Advances in bacterial culturing and the development of high-throughput methods, known as culturomics (Figure 4), have greatly expanded the ability to isolate and study a broader range of previously unculturable gut microbiota[131]. Technologies that have accelerated culturing of many bacteria to become high-throughput include: 1) image recognition software and artificial intelligence (AI) models to identify different colonies on a plate, 2) robot systems that automatically isolate the bacteria, 3) taxonomic identification systems using DNA sequencing or matrix-assisted laser desorption/ionisation-time-of-flight mass spectrometry (MALDI-TOF). Combined with developments of artificial human gut tissue systems[132-136], this provides excellent opportunities for studying host-microbe interactions.

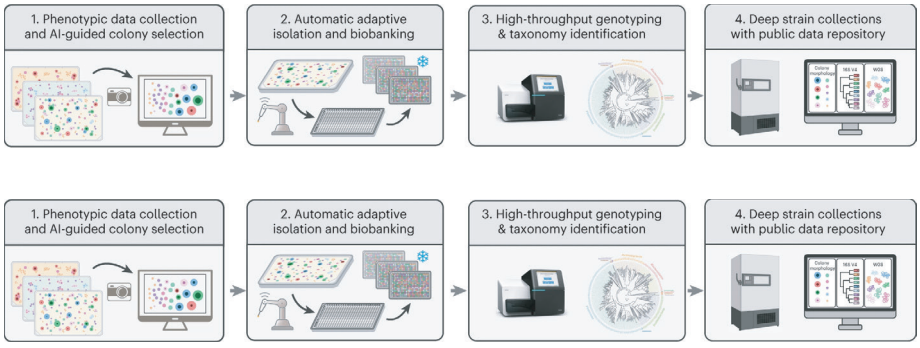


Figure 4. A state-of-the-art culturomics workflow. 1: The process starts by dividing a biological (microbiome) sample and growing the microbiota on different media. Colonies are selected based on phenotype, supported by Artificial Intelligence. 2: Selected colonies are automatically isolated using a robot and stored separately in a biobank. 3: Strains in the biobank are identified using high-throughput genotyping methods such as 16S rRNA sequencing. 4: Isolates are stored in the freezer and their information is made public to facilitate their findability and accessibility. Figure adapted from reference[137].

Developments in computational biology and open data

Omics experiments generate a lot of data which require computers to process and interpret to produce biologically meaningful results. This in turn brings its own set of opportunities and challenges. One major challenge is the need for specialised researchers with expertise in both the biological context and the know-how of the computational methods. This requires specialised training, which is fortunately already integrated in Dutch universities where all biology programmes offer courses with some

form of bioinformatics. The Leiden University Medical Center is taking a next step and will include AI training into its medical programme[138].

Within the broader context of science, a compelling upside to computation compared to traditional wet-lab work is that computers facilitate full reproduction of experiments. Theoretically, rerunning a computer program is simpler than reproducing manual laboratory protocols. Reproducibility has been underappreciated in scientific research, even though it is crucial for validating results and strengthening the foundation of derivative works[139]. This is especially urgent in an era of fake news and a growing distrust in science[140]. Various initiatives aim to improve science's credibility and promote a more transparent evaluation process. For example, there is a broad movement of FAIR data: findable, accessible, interoperable, and reusable[141]. These principles are widely adopted, with many funding bodies demanding data to be made publicly available in grant applications, and scientific journals requiring data sharing as a condition for publication of manuscripts. On a similar note, the FAIR principles have been adapted to software[142], stimulating the publication of research software and thus improving reproducibility. Furthermore, there is a growing interest and appreciation of manuscripts preprints, accelerating knowledge dissemination[143]. These initiatives, and more, can give a boost to the credibility of science[144], while also stimulating different talents within science, a development I sincerely hope will be broadly adopted and valued in the scientific community.

Outlook: solving microbiota in health and disease questions

Connecting the pieces laid out here in the discussion, I envision further developments of host-microbe interaction studies and Open Science, opening avenues for collaborations between researchers of diverse expertise. If bacteriologists, medical microbiologists, gastroenterologists, immunologists, biochemists, bioinformaticians, statisticians and ecologists put their heads together and integrate their knowledge we can learn how the human gut microbiome functions and behaves under varying conditions. This approach should also include fungi and viruses as integral components of the microbiome[121,145-147]. A broader perspective is already exemplified in the national and international One Health movement[148,149], and is further solidified through initiatives such as the recently funded Holomicrobiome[150]. These initiatives aim to unravel all the intricacies of the ecosystems inside and around us, providing insights into the biological processes leading to disease. Understanding the biological processes helps us prevent or control diseases and improve treatment options. This may take the shape of new FMT strategies or live biotherapeutic products and lead to a new era of disease control and preventive medicine. Also, the timing of treatment may shift to prevent symptoms in people whose microbiota indicate an increased risk of developing disease. In this future view, fewer people develop disease, and those who do receive optimal therapies tailored to their needs.

References

1. Song, J. H. & Kim, Y. S. Recurrent Clostridium difficile Infection: Risk Factors, Treatment, and Prevention. *Gut Liver* **13**, 16-24 (2019). https://doi.org/10.5009/gnl18071
2. Minkoff, N. Z. *et al.* Fecal microbiota transplantation for the treatment of recurrent Clostridioides difficile (Clostridium difficile). *Cochrane Database Systematic Reviews* **4**, CD013871 (2023). https://doi.org/10.1002/14651858.CD013871.pub2
3. van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent Clostridium difficile. *New England Journal of Medicine* **368**, 407-415 (2013). https://doi.org/10.1056/NEJMoa1205037
4. Moayyedi, P. *et al.* Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. *Gastroenterology* **149**, 102-109 e106 (2015). https://doi.org/10.1053/j.gastro.2015.04.001
5. Lopetuso, L. R. *et al.* The first international Rome consensus conference on gut microbiota and faecal microbiota transplantation in inflammatory bowel disease. *Gut* **72**, 1642-1650 (2023). https://doi.org/10.1136/gutjnl-2023-329948
6. Groenewegen, B., Terveer, E. M., Joosse, A., Barnhoorn, M. C. & Zwittink, R. D. Fecal Microbiota Transplantation for Immune Checkpoint Inhibitor-Induced Colitis Is Safe and Contributes to Recovery: Two Case Reports. *Journal of Immunotherapy* **46**, 216-220 (2023). https://doi.org/10.1097/CJI.0000000000000474
7. Wang, Y. *et al.* Fecal microbiota transplantation for refractory immune checkpoint inhibitor-associated colitis. *Nature Medicine* **24**, 1804-1808 (2018). https://doi.org/10.1038/s41591-018-0238-9
8. Malard, F. *et al.* Pooled allogeneic faecal microbiota MaaT013 for steroid-resistant gastrointestinal acute graft-versus-host disease: a single-arm, multicentre phase 2 trial. *EClinicalMedicine* **62**, 102111 (2023). https://doi.org/10.1016/j.eclinm.2023.102111
9. van Lier, Y. F. *et al.* Donor fecal microbiota transplantation ameliorates intestinal graft-versus-host disease in allogeneic hematopoietic cell transplant recipients. *Science Translational Medicine* **12** (2020). https://doi.org/10.1126/scitranslmed.aaz8926
10. Youngster, I. *et al.* Fecal microbiota transplantation in capsules for the treatment of steroid refractory and steroid dependent acute graft vs. host disease: a pilot study. *Bone Marrow Transplant* **59**, 409-416 (2024). https://doi.org/10.1038/s41409-024-02198-2
11. Dailey, F. E., Turse, E. P., Daglilar, E. & Tahan, V. The dirty aspects of fecal microbiota transplantation: a review of its adverse effects and complications. *Current Opinion in Pharmacology* **49**, 29-33 (2019). https://doi.org/10.1016/j.coph.2019.04.008
12. Hallowell, H. A., Gao, A. L. & Suez, J. Double-edged sword: impact of fecal microbiome transplants on the gut resistome. *Current Opinion in Gastroenterology* **39**, 16-22 (2023). https://doi.org/10.1097/MOG.0000000000000894
13. Terveer, E. M. *et al.* Faecal microbiota transplantation for Clostridioides difficile infection: Four years' experience of the Netherlands Donor Feces Bank. *United European Gastroenterology Journal* **8**, 1236-1247 (2020). https://doi.org/10.1177/2050640620957765
14. DeFilipp, Z. *et al.* Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. *New England Journal of Medicine* **381**, 2043-2050 (2019). https://doi.org/10.1056/NEJMoa1910437

15. Schewitz, K. in *Business Insider* https://www.businessinsider.com/woman-diy-poop-transplants-got-brother-acne-fecal-microbiota-transplant-2024-2025 (Insider Inc, A woman gave herself poop transplants using her brother's feces to treat debilitating IBS. Then she started getting acne just like him, 2024).

16. Drewes, J. L. *et al.* Transmission and clearance of potential procarcinogenic bacteria during fecal microbiota transplantation for recurrent Clostridioides difficile. *JCI Insight* **4** (2019). https://doi.org/10.1172/jci.insight.130848

17. Jouhten, H., Mattila, E., Arkkila, P. & Satokari, R. Reduction of Antibiotic Resistance Genes in Intestinal Microbiota of Patients With Recurrent Clostridium difficile Infection After Fecal Microbiota Transplantation. *Clinical Infectious Diseases* **63**, 710-711 (2016). https://doi.org/10.1093/cid/ciw390

18. Manges, A. R., Steiner, T. S. & Wright, A. J. Fecal microbiota transplantation for the intestinal decolonization of extensively antimicrobial-resistant opportunistic pathogens: a review. *Infectious Diseases* **48**, 587-592 (2016). https://doi.org/10.1080/23744235.2016.1177199

19. Leung, V., Vincent, C., Edens, T. J., Miller, M. & Manges, A. R. Antimicrobial Resistance Gene Acquisition and Depletion Following Fecal Microbiota Transplantation for Recurrent Clostridium difficile Infection. *Clinical Infectious Diseases* **66**, 456-457 (2018). https://doi.org/10.1093/cid/cix821

20. Millan, B. *et al.* Fecal Microbial Transplants Reduce Antibiotic-resistant Genes in Patients With Recurrent Clostridium difficile Infection. *Clinical Infectious Diseases* **62**, 1479-1486 (2016). https://doi.org/10.1093/cid/ciw185

21. Hourigan, S. K. *et al.* Fecal Transplant in Children With Clostridioides difficile Gives Sustained Reduction in Antimicrobial Resistance and Potential Pathogen Burden. *Open Forum Infectious Diseases* **6**, ofz379 (2019). https://doi.org/10.1093/ofid/ofz379

22. Bajaj, J. S. *et al.* Fecal Microbiota Transplant in Cirrhosis Reduces Gut Microbial Antibiotic Resistance Genes: Analysis of Two Trials. *Hepatology Communications* **5**, 258-271 (2021). https://doi.org/10.1002/hep4.1639

23. Terveer, E. M. *et al.* How to: Establish and run a stool bank. *Clinical Microbiology and Infection* **23**, 924-930 (2017). https://doi.org/10.1016/j.cmi.2017.05.015

24. Zhang, S. *et al.* Donor Screening for Fecal Microbiota Transplantation in China: Evaluation of 8483 Candidates. *Gastroenterology* **162**, 966-968 e963 (2022). https://doi.org/10.1053/j.gastro.2021.11.004

25. Narula, N. *et al.* Systematic Review and Meta-analysis: Fecal Microbiota Transplantation for Treatment of Active Ulcerative Colitis. *Inflammatory Bowel Diseases* **23**, 1702-1709 (2017). https://doi.org/10.1097/MIB.0000000000001228

26. Quraishi, M. N. *et al.* Systematic review with meta-analysis: the efficacy of faecal microbiota transplantation for the treatment of recurrent and refractory Clostridium difficile infection. *Alimentary Pharmacology & Therapeutics* **46**, 479-493 (2017). https://doi.org/10.1111/apt.14201

27. Schmidt, T. S. B. *et al.* Drivers and determinants of strain dynamics following fecal microbiota transplantation. *Nature Medicine* **28**, 1902-1912 (2022). https://doi.org/10.1038/s41591-022-01913-0

28. Haifer, C. *et al.* Lyophilised oral faecal microbiota transplantation for ulcerative colitis (LOTUS): a randomised, double-blind, placebo-controlled trial. *The Lancet Gastroenterology and Hepatology* **7**, 141-151 (2022). https://doi.org/10.1016/S2468-1253(21)00400-3

29. Podlesny, D. *et al.* Identification of clinical and ecological determinants of strain engraftment after fecal microbiota transplantation using metagenomics. *Cell Reports Medicine* **3**, 100711 (2022). https://doi.org/10.1016/j.xcrm.2022.100711

30. He, R. *et al.* The interplay of gut microbiota between donors and recipients determines the efficacy of fecal microbiota transplantation. *Gut Microbes* **14**, 2100197 (2022). https://doi.org/10.1080/19490976.2022.2100197

31. Kaplan, G. G. & Ng, S. C. Understanding and Preventing the Global Increase of Inflammatory Bowel Disease. *Gastroenterology* **152**, 313-321.e312 (2017). https://doi.org/10.1053/j.gastro.2016.10.020

32. Neurath, M. F., Sands, B. E. & Rieder, F. Cellular immunotherapies and immune cell depleting therapies in inflammatory bowel diseases: the next magic bullet? *Gut*, gutjnl-2024-332919 (2024). https://doi.org/10.1136/gutjnl-2024-332919

33. Super, M. *et al.* Biomaterial vaccines capturing pathogen-associated molecular patterns protect against bacterial infections and septic shock. *Nature Biomedical Engineering* **6**, 8-18 (2022). https://doi.org/10.1038/s41551-021-00756-3

34. Alameh, M. G. *et al.* A multivalent mRNA-LNP vaccine protects against Clostridioides difficile infection. *Science* **386**, 69-75 (2024). https://doi.org/10.1126/science.adn4955

35. Gupta, V. K. *et al.* A predictive index for health status using species-level gut microbiome profiling. *Nature Communications* **11**, 4635 (2020). https://doi.org/10.1038/s41467-020-18476-8

36. Shi, N., Li, N., Duan, X. & Niu, H. Interaction between the gut microbiome and mucosal immune system. *Military Medical Research* **4**, 14 (2017). https://doi.org/10.1186/s40779-017-0122-9

37. Dixit, K., Chaudhari, D., Dhotre, D., Shouche, Y. & Saroj, S. Restoration of dysbiotic human gut microbiome for homeostasis. *Life Sciences* **278**, 119622 (2021). https://doi.org/10.1016/j.lfs.2021.119622

38. Ianiro, G. *et al.* Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. *Nature Medicine* **28**, 1913-1923 (2022). https://doi.org/10.1038/s41591-022-01964-3

39. Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* **364** (2019). https://doi.org/10.1126/science.aau6323

40. van Kessel, S. P. *et al.* Gut bacterial tyrosine decarboxylases restrict levels of levodopa in the treatment of Parkinson's disease. *Nature Communications* **10**, 310 (2019). https://doi.org/10.1038/s41467-019-08294-y

41. Vendrik, K. E. *et al.* Safety and feasibility of faecal microbiota transplantation for patients with Parkinson's disease: a protocol for a self-controlled interventional donor-FMT pilot study. *BMJ Open* **13**, e071766 (2023). https://doi.org/10.1136/bmjopen-2023-071766

42. Becattini, S. *et al.* Rapid transcriptional and metabolic adaptation of intestinal microbes to host immune activation. *Cell Host & Microbe* **29**, 378-393 e375 (2021). https://doi.org/10.1016/j.chom.2021.01.003

43. Pral, L. P., Fachi, J. L., Correa, R. O., Colonna, M. & Vinolo, M. A. R. Hypoxia and HIF-1 as key regulators of gut microbiota and host interactions. *Trends in Immunology* **42**, 604-621 (2021). https://doi.org/10.1016/j.it.2021.05.004

44. Litvak, Y., Byndloss, M. X. & Bäumler, A. J. Colonocyte metabolism shapes the gut microbiota. *Science* **362**, eaat9076 (2018). https://doi.org/doi:10.1126/science.aat9076

45. Wastyk, H. C. *et al.* Gut-microbiota-targeted diets modulate human immune status. *Cell* **184**, 4137-4153 e4114 (2021). https://doi.org/10.1016/j.cell.2021.06.019

46. Ross, F. C. *et al.* The interplay between diet and the gut microbiome: implications for health and disease. *Nature Reviews Microbiology* **22**, 671-686 (2024). https://doi.org/10.1038/s41579-024-01068-4

47. Clarke, S. F. *et al.* Exercise and associated dietary extremes impact on gut microbial diversity. *Gut* **63**, 1913-1920 (2014). https://doi.org/10.1136/gutjnl-2013-306541

48. Clancy, A. K., Gunaratne, A. W. & Borody, T. J. Dietary Management for Faecal Microbiota Transplant: An International Survey of Clinical and Research Practice, Knowledge and Attitudes. *Frontiers in Nutrition* **8**, 653653 (2021). https://doi.org/10.3389/fnut.2021.653653

49. Kedia, S. *et al.* Faecal microbiota transplantation with anti-inflammatory diet (FMT-AID) followed by anti-inflammatory diet alone is effective in inducing and maintaining remission over 1 year in mild to moderate ulcerative colitis: a randomised controlled trial. *Gut* **71**, 2401-2413 (2022). https://doi.org/10.1136/gutjnl-2022-327811

50. Lai, Z. L. *et al.* Fecal microbiota transplantation confers beneficial metabolic effects of diet and exercise on diet-induced obese mice. *Scientific Reports* **8**, 15625 (2018). https://doi.org/10.1038/s41598-018-33893-y

51. Rinott, E. *et al.* Effects of Diet-Modulated Autologous Fecal Microbiota Transplantation on Weight Regain. *Gastroenterology* **160**, 158-173 e110 (2021). https://doi.org/10.1053/j.gastro.2020.08.041

52. Centre for Digestive Diseases. *The CDD FMT Donor Diet*, <https://centrefordigestivediseases.com/the-cdd-donor-diet/> (2020).

53. Kelly, C. R., Kim, A. M., Laine, L. & Wu, G. D. The AGA's Fecal Microbiota Transplantation National Registry: An Important Step Toward Understanding Risks and Benefits of Microbiota Therapeutics. *Gastroenterology* **152**, 681-684 (2017). https://doi.org/10.1053/j.gastro.2017.01.028

54. Yau, Y. K. *et al.* Long-Term Safety Outcomes of Fecal Microbiota Transplantation: Real-World Data Over 8 Years From the Hong Kong FMT Registry. *Clinical Gastroenterology and Hepatology* **22**, 611-620.e612 (2024). https://doi.org/10.1016/j.cgh.2023.09.001

55. Cammarota, G. *et al.* International consensus conference on stool banking for faecal microbiota transplantation in clinical practice. *Gut* **68**, 2111-2121 (2019). https://doi.org/10.1136/gutjnl-2019-319548

56. Charbonneau, M. R., Isabella, V. M., Li, N. & Kurtz, C. B. Developing a new class of engineered live bacterial therapeutics to treat human diseases. *Nature Communications* **11**, 1738 (2020). https://doi.org/10.1038/s41467-020-15508-1

57. Navalkele, B. & Chopra, T. Clinical Application of Live Biotherapeutic Products in Infectious Diseases. *Frontiers in Microbiomes* **3** (2024). https://doi.org/10.3389/frmbi.2024.1415083

58. Rutter, J. W., Dekker, L., Owen, K. A. & Barnes, C. P. Microbiome engineering: engineered live biotherapeutic products for treating human disease. *Frontiers in Bioengineering and Biotechnology* **10**, 1000873 (2022). https://doi.org/10.3389/fbioe.2022.1000873

59. Khanna, S. *et al.* Efficacy and Safety of RBX2660 in PUNCH CD3, a Phase III, Randomized, Double-Blind, Placebo-Controlled Trial with a Bayesian Primary Analysis for the Prevention of Recurrent Clostridioides difficile Infection. *Drugs* **82**, 1527-1538 (2022). https://doi.org/10.1007/s40265-022-01797-x

60. Khanna, S. *et al.* RBX7455, a Non-frozen, Orally Administered Investigational Live Biotherapeutic, Is Safe, Effective, and Shifts Patients' Microbiomes in a Phase 1 Study for Recurrent Clostridioides difficile Infections. *Clinical Infectious Diseases* **73**, e1613-e1620 (2021). https://doi.org/10.1093/cid/ciaa1430

61. Feuerstadt, P. *et al.* SER-109, an Oral Microbiome Therapy for Recurrent *Clostridioides difficile* Infection. *New England Journal of Medicine* **386**, 220-229 (2022). https://doi.org/doi:10.1056/NEJMoa2106516

62. Louie, T. *et al.* VE303, a Defined Bacterial Consortium, for Prevention of Recurrent Clostridioides difficile Infection: A Randomized Clinical Trial. *JAMA* **329**, 1356-1366 (2023). https://doi.org/10.1001/jama.2023.4314

63. Dsouza, M. *et al.* Colonization of the live biotherapeutic product VE303 and modulation of the microbiota and metabolites in healthy volunteers. *Cell Host & Microbe* **30**, 583-598. e588 (2022). https://doi.org/10.1016/j.chom.2022.03.016

64. Kao, D. *et al.* The effect of a microbial ecosystem therapeutic (MET-2) on recurrent Clostridioides difficile infection: a phase 1, open-label, single-group trial. *The Lancet Gastroenterology and Hepatology* **6**, 282-291 (2021). https://doi.org/10.1016/S2468-1253(21)00007-8

65. Gerding, D. N. *et al.* Administration of Spores of Nontoxigenic Clostridium difficile Strain M3 for Prevention of Recurrent C difficile Infection: A Randomized Clinical Trial. *JAMA* **313**, 1719-1727 (2015). https://doi.org/10.1001/jama.2015.3725

66. Kaslow, D. C. Approval Letter - REBYOTA. (US Food & Drug Administration, STN: 125739, 2022).

67. Mendoza, M. K., David C. Approval Letter - VOWST. (US Food & Drug Administration, STN: 125757, 2023).

68. van der Lelie, D. *et al.* Rationally designed bacterial consortia to treat chronic immune-mediated colitis and restore intestinal homeostasis. *Nature Communications* **12**, 3105 (2021). https://doi.org/10.1038/s41467-021-23460-x

69. Atarashi, K. *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232-236 (2013). https://doi.org/10.1038/nature12331

70. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450 (2013). https://doi.org/10.1038/nature12721

71. Atarashi, K. *et al.* Induction of Colonic Regulatory T Cells by Indigenous *Clostridium* Species. *Science* **331**, 337-341 (2011). https://doi.org/doi:10.1126/science.1198469

72. Paton, A. W. *et al.* Recombinant probiotics for treatment and prevention of enterotoxigenic Escherichia coli diarrhea. *Gastroenterology* **128**, 1219-1228 (2005). https://doi.org/10.1053/j.gastro.2005.01.050

73. Sonnenborn, U. Escherichia coli strain Nissle 1917-from bench to bedside and back: history of a special Escherichia coli strain with probiotic properties. *FEMS Microbiology Letters* **363** (2016). https://doi.org/10.1093/femsle/fnw212

74. Duan, F. & March, J. C. Engineered bacterial communication prevents Vibrio cholerae virulence in an infant mouse model. *Proceedings of the National Academy of Sciences* **107**, 11260-11264 (2010). https://doi.org/10.1073/pnas.1001294107

75. Hwang, I. Y. *et al.* Engineered probiotic Escherichia coli can eliminate and prevent Pseudomonas aeruginosa gut infection in animal models. *Nature Communications* **8**, 15028 (2017). https://doi.org/10.1038/ncomms15028

76. O'Donnell, M. M. *et al.* Identification of ADS024, a newly characterized strain of Bacillus velezensis with direct Clostridiodes difficile killing and toxin degradation bio-activities. *Scientific Reports* **12**, 9283 (2022). https://doi.org/10.1038/s41598-022-13248-4

77. Nougayrede, J. P. *et al.* A Toxic Friend: Genotoxic and Mutagenic Activity of the Probiotic Strain Escherichia coli Nissle 1917. *mSphere* **6**, e0062421 (2021). https://doi.org/10.1128/mSphere.00624-21

78. Olier, M. *et al.* Genotoxicity of Escherichia coli Nissle 1917 strain cannot be dissociated from its probiotic activity. *Gut Microbes* **3**, 501-509 (2012). https://doi.org/10.4161/gmic.21737

79. Smillie, C. S. *et al.* Strain Tracking Reveals the Determinants of Bacterial Engraftment in the Human Gut Following Fecal Microbiota Transplantation. *Cell Host Microbe* **23**, 229-240 e225 (2018). https://doi.org/10.1016/j.chom.2018.01.003

80. Aggarwala, V. *et al.* Precise quantification of bacterial strains after fecal microbiota transplantation delineates long-term engraftment and explains outcomes. *Nature Microbiology* **6**, 1309-1318 (2021). https://doi.org/10.1038/s41564-021-00966-0

81. Olm, M. R. *et al.* inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nature Biotechnology* **39**, 727-736 (2021). https://doi.org/10.1038/s41587-020-00797-0

82. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Research* **27**, 626-638 (2017). https://doi.org/10.1101/gr.216242.116

83. Quince, C. *et al.* STRONG: metagenomics strain resolution on assembly graphs. *Genome Biology* **22**, 214 (2021). https://doi.org/10.1186/s13059-021-02419-7

84. Hu, H. *et al.* StrainPanDA: Linked reconstruction of strain composition and gene content profiles via pangenome-based decomposition of metagenomic data. *iMeta* **1**, e41 (2022). https://doi.org/10.1002/imt2.41

85. Podlesny, D. *et al.* Metagenomic strain detection with SameStr: identification of a persisting core gut microbiota transferable by fecal transplantation. *Microbiome* **10**, 53 (2022). https://doi.org/10.1186/s40168-022-01251-w

86. van Dijk, L. R. *et al.* StrainGE: a toolkit to track and characterize low-abundance strains in complex microbial communities. *Genome Biology* **23**, 74 (2022). https://doi.org/10.1186/s13059-022-02630-0

87. Briscoe, L., Halperin, E. & Garud, N. R. SNV-FEAST: microbial source tracking with single nucleotide variants. *Genome Biology* **24**, 101 (2023). https://doi.org/10.1186/s13059-023-02927-8

88. Kim, Y. *et al.* Strain tracking with uncertainty quantification. *bioRxiv*, 2023.2001.2025.525531 (2024). https://doi.org/10.1101/2023.01.25.525531

89. Blanco-Miguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology* (2023). https://doi.org/10.1038/s41587-023-01688-w

90. Goldman, M., Zhao, C. & Pollard, K. S. Improved detection of microbiome-disease associations via population structure-aware generalized linear mixed effects models (microSLAM). *bioRxiv*, 2024.2006.2027.600934 (2024). https://doi.org/10.1101/2024.06.27.600934

91. Martin, J. S., Monaghan, T. M. & Wilcox, M. H. Clostridium difficile infection: epidemiology, diagnosis and understanding transmission. *Nature Reviews Gastroenterology & Hepatology* **13**, 206-216 (2016). https://doi.org/10.1038/nrgastro.2016.25

92. Croxen, M. A. *et al.* Recent advances in understanding enteric pathogenic Escherichia coli. *Clinical Microbiology Reviews* **26**, 822-880 (2013). https://doi.org/10.1128/CMR.00022-13

93. Wolter, M. *et al.* Diet-driven differential response of Akkermansia muciniphila modulates pathogen susceptibility. *Molecular Systems Biology* **20**, 596-625 (2024). https://doi.org/10.1038/s44320-024-00036-7

94. Abdelsalam, N. A., Hegazy, S. M. & Aziz, R. K. The curious case of Prevotella copri. *Gut Microbes* **15**, 2249152 (2023). https://doi.org/10.1080/19490976.2023.2249152

95. Blanco-Miguez, A. *et al.* Extension of the Segatella copri complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell Host & Microbe* **31**, 1804-1819 e1809 (2023). https://doi.org/10.1016/j.chom.2023.09.013

96. Tett, A., Pasolli, E., Masetti, G., Ercolini, D. & Segata, N. Prevotella diversity, niches and interactions with the human host. *Nature Reviews Microbiology* **19**, 585-599 (2021). https://doi.org/10.1038/s41579-021-00559-y

97. Crost, E. H., Coletto, E., Bell, A. & Juge, N. Ruminococcus gnavus: friend or foe for human health. *FEMS Microbiology Reviews* **47** (2023). https://doi.org/10.1093/femsre/fuad014

98. Wassenaar, T. M. Insights from 100 Years of Research with Probiotic E. Coli. *European Journal of Microbiology and Immunology* **6**, 147-161 (2016). https://doi.org/10.1556/1886.2016.00029

99. Kaper, J. B., Nataro, J. P. & Mobley, H. L. Pathogenic Escherichia coli. *Nature Reviews Microbiology* **2**, 123-140 (2004). https://doi.org/10.1038/nrmicro818

100. Arnold, B. J., Huang, I. T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nature Reviews Microbiology* **20**, 206-218 (2022). https://doi.org/10.1038/s41579-021-00650-4

101. Brito, I. L. Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology* **19**, 442-453 (2021). https://doi.org/10.1038/s41579-021-00534-7

102. Kent, A. G., Vill, A. C., Shi, Q., Satlin, M. J. & Brito, I. L. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. *Nature Communications* **11**, 4379 (2020). https://doi.org/10.1038/s41467-020-18164-7

103. Bohm, M. E., Huptas, C., Krey, V. M. & Scherer, S. Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of Bacillus cereus enterotoxin operons hbl, cytK and nhe. *BMC Evolutionary Biology* **15**, 246 (2015). https://doi.org/10.1186/s12862-015-0529-4

104. Putze, J. *et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infection and Immunity* **77**, 4696-4703 (2009). https://doi.org/10.1128/IAI.00522-09

105. Chagneau, C. V. *et al.* The pks island: a bacterial Swiss army knife? Colibactin: beyond DNA damage and cancer. *Trends in Microbiology* **30**, 1146-1159 (2022). https://doi.org/10.1016/j.tim.2022.05.010

106. Jiang, X., Hall, A. B., Xavier, R. J. & Alm, E. J. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level niche-adaptive gene pools. *PLoS One* **14**, e0223680 (2019). https://doi.org/10.1371/journal.pone.0223680

107. Arredondo-Alonso, S., Willems, R. J., van Schaik, W. & Schurch, A. C. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics* **3**, e000128 (2017). https://doi.org/10.1099/mgen.0.000128

108. Krawczyk, P. S., Lipinski, L. & Dziembowski, A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research* **46**, e35 (2018). https://doi.org/10.1093/nar/gkx1321

109. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nature Biotechnology* (2023). https://doi.org/10.1038/s41587-023-01953-y

110. Suzuki, Y. *et al.* Long-read metagenomic exploration of extrachromosomal mobile genetic elements in the human gut. *Microbiome* **7**, 119 (2019). https://doi.org/10.1186/s40168-019-0737-z

111. Labonte, J. M. *et al.* Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology* **6**, 349 (2015). https://doi.org/10.3389/fmicb.2015.00349

112. Brito, I. L. *et al.* Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435-439 (2016). https://doi.org/10.1038/nature18927

113. Press, M. O. *et al.* Hi-C deconvolution of a human gut microbiome yields high-quality draft genomes and reveals plasmid-genome interactions. *bioRxiv*, 198713 (2017). https://doi.org/10.1101/198713

114. Yaffe, E. & Relman, D. A. Tracking microbial evolution in the human gut using Hi-C reveals extensive horizontal gene transfer, persistence and adaptation. *Nature Microbiology* **5**, 343-353 (2020). https://doi.org/10.1038/s41564-019-0625-0

115. McCallum, G. E. *et al.* Noise reduction strategies in metagenomic chromosome confirmation capture to link antibiotic resistance genes to microbial hosts. *bioRxiv*, 2022.2011.2005.514866 (2022). https://doi.org/10.1101/2022.11.05.514866

116. Arıkan, M. & Muth, T. Integrated multi-omics analyses of microbial communities: a review of the current state and future directions. *Molecular Omics* **19**, 607-623 (2023). https://doi.org/10.1039/d3mo00089c

117. Sauceda, C. *et al.* Stool multi-omics for the study of host–microbe interactions in inflammatory bowel disease. *Gut Microbes* **14**, 2154092 (2022). https://doi.org/10.1080/19490976.2022.2154092

118. Zepeda-Rivera, M. *et al.* A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature* **628**, 424-432 (2024). https://doi.org/10.1038/s41586-024-07182-w

119. Fukugaiti, M. H. *et al.* High occurrence of Fusobacterium nucleatum and Clostridium difficile in the intestinal microbiota of colorectal carcinoma patients. *Brazilian Journal of Microbiology* **46**, 1135-1140 (2015). https://doi.org/10.1590/S1517-838246420140665

120. Ning, L. *et al.* Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. *Nature Communications* **14**, 7135 (2023). https://doi.org/10.1038/s41467-023-42788-0

121. Liu, N. N. *et al.* Multi-kingdom microbiota analyses identify bacterial-fungal interactions and biomarkers of colorectal cancer across cohorts. *Nature Microbiology* **7**, 238-250 (2022). https://doi.org/10.1038/s41564-021-01030-7

122. Walker, A. W. & Hoyles, L. Human microbiome myths and misconceptions. *Nature Microbiology* **8**, 1392-1396 (2023). https://doi.org/10.1038/s41564-023-01426-7

123. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* **35**, 833-844 (2017). https://doi.org/10.1038/nbt.3935

124. Rajilic-Stojanovic, M., Smidt, H. & de Vos, W. M. Diversity of the human gastrointestinal tract microbiota revisited. *Environmental Microbiology* **9**, 2125-2136 (2007). https://doi.org/10.1111/j.1462-2920.2007.01369.x

125. Costea, P. I. *et al.* Towards standards for human fecal sample processing in metagenomic studies. *Nature Biotechnology* **35**, 1069-1076 (2017). https://doi.org/10.1038/nbt.3960

126. Szostak, N. *et al.* The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. *Scientific Reports* **12**, 8470 (2022). https://doi.org/10.1038/s41598-022-12037-3

127. Fernandez-Pato, A. *et al.* Choice of DNA extraction method affects stool microbiome recovery and subsequent phenotypic association analyses. *Scientific Reports* **14**, 3911 (2024). https://doi.org/10.1038/s41598-024-54353-w

128. Margos, G. *et al.* Evidence of taxonomic bias in public databases: The example of the genus Borrelia. *Ticks and Tick-borne Diseases* **13**, 101994 (2022). https://doi.org/10.1016/j.ttbdis.2022.101994

129. Ruscheweyh, H. J. *et al.* Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* **10**, 212 (2022). https://doi.org/10.1186/s40168-022-01410-z

130. Setubal, J. C. Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophysical Reviews* **13**, 905-909 (2021). https://doi.org/10.1007/s12551-021-00865-y

131. Lagier, J. C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* **1**, 16203 (2016). https://doi.org/10.1038/nmicrobiol.2016.203

132. Zhang, J. *et al.* An immune-competent human gut microphysiological system enables inflammation-modulation by Faecalibacterium prausnitzii. *NPJ Biofilms Microbiomes* **10**, 31 (2024). https://doi.org/10.1038/s41522-024-00501-z

133. Xiang, Y. *et al.* Gut-on-chip: Recreating human intestine in vitro. *Journal of Tissue Engineering* **11**, 2041731420965318 (2020). https://doi.org/10.1177/2041731420965318

134. Min, S., Kim, S. & Cho, S. W. Gastrointestinal tract modeling using organoids engineered with cellular and microbiota niches. *Experimental Molecular Medicine* **52**, 227-237 (2020). https://doi.org/10.1038/s12276-020-0386-0

135. Schuren F, A. V., Keijser B, Abeln E, der Vossen J, Montijn R. The i-screen: A Versatile Preclinical Platform for Gut Microbiota Studies. *J Prob Health* **7** (2019). https://doi.org/10.35248/2329-8901.19.7.212

136. Williamson, I. A. *et al.* A High-Throughput Organoid Microinjection Platform to Study Gastrointestinal Microbiota and Luminal Physiology. *Cellular and Molecular Gastroenterology and Hepatology* **6**, 301-319 (2018). https://doi.org/10.1016/j.jcmgh.2018.05.004

137. Huang, Y. *et al.* High-throughput microbial culturomics using automation and machine learning. *Nature Biotechnology* **41**, 1424-1433 (2023). https://doi.org/10.1038/s41587-023-01674-2

138. Leiden University. *LUMC first medical programme to include AI in curriculum*, <https://www.universiteitleiden.nl/en/news/2024/08/lumc-first-medical-programme-to-include-ai-in-curriculum> (2024).

139. Sutter, P. M. *Rescuing Science: Restoring Trust in an Age of Doubts*. (Rowman & Littlefield Publishers, 2024).

140. Gerbina, T. V. Science Disinformation: On the Problem of Fake News. *Scientific and Technical Information Processing* **48**, 290-298 (2021). https://doi.org/10.3103/S0147688221040092

141. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

142. Barker, M. *et al.* Introducing the FAIR Principles for research software. *Scientific Data* **9**, 622 (2022). https://doi.org/10.1038/s41597-022-01710-x

143. ASAPbio. *Biology preprints over time*, <https://asapbio.org/preprint-info/biology-preprints-over-time> (2020).

144. Korbmacher, M. *et al.* The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology* **1**, 3 (2023). https://doi.org/10.1038/s44271-023-00003-2

145. Haak, B. W. *et al.* Integrative Transkingdom Analysis of the Gut Microbiome in Antibiotic Perturbation and Critical Illness. *mSystems* **6** (2021). https://doi.org/10.1128/mSystems.01148-20

146. Kazemian, N. *et al.* The trans-kingdom battle between donor and recipient gut microbiome influences fecal microbiota transplantation outcome. *Scientific Reports* **10**, 18349 (2020). https://doi.org/10.1038/s41598-020-75162-x

147. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nature Reviews Microbiology* **19**, 77-94 (2021). https://doi.org/10.1038/s41579-020-0438-4

148. Netherlands Center for One Health. *Homepage - NCOH*, <https://ncoh.nl> (2024).

149. World Health Organization. *One Health*, <https://www.who.int/health-topics/one-health> (2024).

150. Holomicrobiome Initiative. *Holomicrobiome Initiative*, <https://holomicrobioom.nl/en/home-en/> (2024).


幸

Appendix

Nederlandse samenvatting

Inleiding

Het darmmicrobioom van de mens speelt een belangrijke rol in ziekte en gezondheid en wordt uitgebreid bestudeerd. Dit microbioom bestaat uit micro-organismen met uiteenlopende levenswijzen die bijvoorbeeld helpen bij de voedselvertering of toxines kunnen produceren. Deze micro-organismen, en met name bacteriën, worden doorgaans onderzocht door middel van DNA-sequencing. De DNA-sequencingtechnieken zijn de afgelopen decennia sterk verbeterd, waardoor het mogelijk is om sneller en nauwkeuriger DNA te analyseren. Hierbij wordt zoveel data gegenereerd, dat analyse alleen mogelijk is met behulp van computers. Dit wordt gedaan door gespecialiseerde 'computationele biologen' of bioinformatici: onderzoekers die kennis van de biologie combineren met computerwetenschappen. Met specialistische software verwerken zij steeds groter wordende datasets, veelal om nieuwe theoretische inzichten te verkrijgen die vervolgens in het laboratorium worden getoetst. Dergelijke technieken hebben er bijvoorbeeld toe geleid dat bepaalde bacteriën in het microbioom konden worden geassocieerd met darmziekten zoals colitis ulcerosa en de ziekte van Crohn (samen bekend als *inflammatory bowel diseases* of IBD), met dikkedarmkanker, obesitas en de ziekte van Parkinson. Dit heeft meerdere aanknopingspunten geboden voor vervolgonderzoek naar de precieze rol van deze bacteriën en de biochemische mechanismen die daaraan ten grondslag liggen.

Voor een goede gezondheid is het belangrijk dat binnen het darmmicrobioom een zekere balans wordt gehandhaafd. Bij het gebruik van antibiotica raakt het microbioom verstoord, wat in ongunstige gevallen tot ziekte kan leiden. Dit is bijvoorbeeld het geval bij een (meervoudig) terugkerende *Clostridioides difficile*-infectie. *C. difficile* is een bacterieel pathogeen dat ernstige darminfecties kan veroorzaken en waar patiënten uiteindelijk zelfs aan kunnen overlijden. Deze *C. difficile*-bacterie produceert sporen die resistent zijn tegen antibiotica, waardoor een terugval (recidief) na een antibioticakuur mogelijk is. Dit kan zelfs meerdere keren achter elkaar gebeuren, waardoor een soort cyclus van antibioticabehandeling en infectie ontstaat. Na meerdere recidieven is een alternatieve therapie effectiever die juist gericht is op het falende microbioom: fecesmicrobiotatransplantatie (FMT). Door de darmbacteriën van een gezonde (ontlastings-)donor in de darmen van een patiënt met recidiverende *C. difficile*-infectie te brengen, wordt de balans in het darmmicrobioom hersteld, waardoor *C. difficile* niet opnieuw kan uitgroeien en geen toxines produceert. De cyclus stopt dan en er treedt geen nieuwe *C. difficile*-infectie op. Deze therapie is zo succesvol dat momenteel wordt onderzocht of FMT ook effectief is bij andere aandoeningen waarvan verondersteld wordt dat een verstoord microbioom een belangrijke rol heeft.

Het onderzoek in dit proefschrift is grotendeels gericht op FMT en de toepassing ervan bij terugkerende *C. difficile*-infecties en colitis ulcerosa. Het doel is om de bacteriële

gemeenschappen binnen donoren en patiënten in kaart te brengen om te bepalen welke invloed de transplantatie van het darmmicrobioom heeft gehad. Het proefschrift is opgedeeld in drie delen. Deel 1 gaat over FMT bij *C. difficile*-patiënten en de gunstige effecten van FMT op andere mogelijk schadelijke bacteriën. Deel 2 onderzoekt FMT bij colitis ulcerosa en hoe de bacteriële gemeenschap van de donor wordt opgenomen in de patiënt en hoe veranderingen in de samenstelling van deze bacteriën verband houden met de verlichting van ontstekingssymptomen. Deel 3 belicht *Ruminococcus gnavus*: een darmbacterie die veel voorkomt en waarvan verhoogde hoeveelheden geassocieerd worden met verschillende ziektebeelden, waaronder de ziekte van Crohn.

Deel 1: Aanvullende effecten van fecesmicrobiotatransplantatie

Naast het voorkomen van terugkerende infectie door *C. difficile*, veroorzaakt FMT veranderingen in het darmmicrobioom die ook op langere termijn gevolgen kunnen hebben. FMT voor recidiverende *C. difficile*-infectie is daardoor een interessant model om de gevolgen van het aanpassen van de microbiota te bestuderen. In dit eerste deel van het proefschrift worden twee studies beschreven die kijken naar de invloed van FMT op andere potentieel schadelijke bacteriën.

Hoofdstuk 2 richt zich op een bacterie die in verband is gebracht met de ontwikkeling van dikkedarmkanker: colibactine-producerende (*pks*$^+$) *Escherichia coli*. Wij hebben deze *E. coli*-variant gekwantificeerd in zowel gezonde fecesdonoren als in de met FMT behandelde patiënten. Door metagenoomdata te gebruiken konden wij zowel de aanwezigheid van deze specifieke bacterie bepalen als het hele microbioom in kaart brengen en zo het effect van FMT op deze bacterie beoordelen. Wij vonden *pks*$^+$ *E. coli* in de patiënten en in mindere mate ook in donoren. Na FMT kwam deze *E. coli* minder voor in patiënten, met name als de bijbehorende donor geen drager was van deze bacterievariant. Voor de overdracht van de *E. coli*-variant van donor naar patiënt hebben wij geen aanwijzingen gevonden. Samenvattend vonden wij dus een gunstig effect van FMT voor de patiënt door het dragerschap van een mogelijk kankerverwekkende bacterie te verlagen. Mogelijk kan deze bevinding in de toekomst bijdragen aan de ontwikkeling van preventieve microbiota-therapieën tegen darmkanker.

In hoofdstuk 3 hebben wij de invloed van FMT op antibioticaresistente bacteriën onderzocht. Hierbij hebben wij gebruik gemaakt van dezelfde metagenoomtechnieken als in hoofdstuk 2, aangevuld met het selectief opkweken van multiresistente bacteriën en het sequencen van hun DNA. Op deze manier konden wij de aanwezigheid van antibioticaresistentiegenen en resistente bacteriestammen in kaart brengen in patiënten die FMT ontvingen en hun donoren. Op basis van de kweekresultaten bleek de prevalentie van multiresistente bacteriën omlaag te gaan na FMT. Middels *sequencing* konden wij echter aantonen dat deze bacteriën toch nog aanwezig kunnen zijn in het darmmicrobioom, maar slechts in zeer lage hoeveelheden. Ook andere antibioticaresistente bacteriën namen in aantal af na FMT, hoewel de diversiteit van

resistentiegenen onveranderd bleef. Van deze resistentiegenen hebben wij voorspeld welke afkomstig zijn van chromosomen en welke van plasmiden. Chromosomale resistentiegenen, die minder makkelijk worden uitgewisseld tussen bacteriën, verminderden in aantallen, terwijl resistentiegenen op plasmiden onverminderd aanwezig bleven. Dit leidt ons tot de conclusie dat FMT een effectieve methode is om antibioticaresistente bacteriën in aantallen te verlagen, wat mogelijk de kans op een moeilijk te bestrijden infectie met deze bacteriën vermindert. De resistente bacteriën verdwijnen echter niet volledig uit de darm. Mogelijk kan FMT worden ingezet bij patiënten met een verhoogd risico op infectie om de verspreiding van antibioticaresistentie in te dammen.

Deel 2: Microbiotaveranderingen na fecesmicobiotatransplantatie voor colitis ulcerosa

Colitis ulcerosa is een vorm van IBD (*inflammatory bowel diseases,* of ontstekingsziekten van de darm) die zich uit in chronische ontstekingen in de dikke darm, waarbij ook een verstoring in de microbiota wordt gezien. In een gerandomiseerde klinische studie hebben wij onderzocht of het veilig is om FMT toe te dienen aan colitis ulcerosapatiënten en in hoeverre de donorbacteriën koloniseren in de darm van de patiënt; een proces dat wordt aangeduid met 'engraftment'. De patiënten werden voorbehandeld met ofwel de ontstekingsremmer budesonide, of met een placebo. Middels metagenomics van ontlastingsmonsters is wederom de darmmicrobiota gekarakteriseerd.

Hoofdstuk 4 beschrijft de klinische studie en bacteriële engraftment met een zelf ontwikkelde maat op basis van verschillende microbiotadiversiteitsparameters. In tegenstelling tot onze verwachting vonden wij dat niet budesonide de engraftment bevorderde, maar dat er een verschil in engraftment was tussen microbiota van twee donoren. Dit benadrukt het belang van verder onderzoek naar donorselectie om colitis ulcerosa effectief met FMT te kunnen behandelen.

Hoofdstuk 5 duikt dieper in de microbiële ecologie om de bevindingen van hoofdstuk 4 in een ander perspectief te plaatsen. Dit onderzoek heeft gepoogd verschillende samenstellingen en veranderingen van darmmicrobiota te correleren aan klinische remissie: het verminderen van colitis ulcerosa-ontstekingen. Middels computermodellen hebben wij clusters van microbiotaprofielen geanalyseerd en vonden één cluster dat sterk samenhing met een slechte uitkomst: het uitblijven van remissie. Een ander cluster, gekenmerkt door hoge hoeveelheden *Ruminococcaceae* en *Lachnospiraceae*, kwam juist vaker voor bij patiënten bij wie de behandeling aansloeg. Deze modellen kunnen helpen om al in een vroeg stadium van de behandeling te voorspellen hoe effectief deze zal zijn.

Hoofdstuk 6 borduurt voort op de bevindingen van de voorgaande twee hoofdstukken en combineert een analyse van engraftment en ecologische dynamiek van de

microbiota bij patiënten vóór de FMT. Op deze manier zochten wij het antwoord op de vraag: hoe hangt de dynamiek tussen bacteriën van donor en patiënt samen met behandelsucces? Bacteriën die al in hoge mate aanwezig waren in de patiënten persisteerden vaak na FMT. De mate van engraftment van donorbacteriën bleek een indicator voor behandelsucces. Een lage mate van stabiele engraftment hing samen met klinisch succes, terwijl hoge mate van kortstondige kolonisatie van donorsoorten vaak voorkwam in patiënten die opnieuw ontstekingen kregen. Samenvattend suggereren onze bevindingen dat het behandelsucces afhangt van 1) de veerkracht van het darmmicrobioom van de patiënt, én van 2) de capaciteit om microbiota van de gezonde donor stabiel op te kunnen nemen.

Deel 3: Wereldwijde verspreiding en genoombiologie van darmbacterie *Ruminococcus gnavus*

Hoofdstuk 7.1 draait om *Ruminococcus gnavus*: een darmbacterie die nauw geassocieerd is met IBD en met name de ziekte van Crohn. Tegelijkertijd komt deze bacteriesoort voor in de darmen van een groot deel van de gezonde bevolking. Ons doel was om deze schijnbare tegenstelling te onderzoeken in een meta-analyse (een samenvatting en heranalyse van reeds gepubliceerde studies) over de wereldwijde verspreiding van *R. gnavus*. Daarnaast hebben wij gepoogd de genomische achtergrond van deze tweedeling te ontrafelen om meer inzicht te krijgen in de interacties tussen gastheer en microbe. Uit onze data blijkt dat *R. gnavus* inderdaad meer en vaker voorkomt bij IBD patiënten, maar ook bij andere aandoeningen zoals type-2 diabetes, hoge bloeddruk, en hart- en vaatziekten. Bovendien zagen we dat *R. gnavus* meer voorkomt in Westerse gemeenschappen dan in niet-Westerse, en vaker bij baby's en jonge kinderen dan bij volwassenen. We vonden dat de genomen van *R. gnavus*-stammen uit gezonde mensen zijn te onderscheiden van stammen die geïsoleerd zijn uit Crohnpatiënten. Dit duidt erop dat er verschillende ondersoorten bestaan die zich elk hebben aangepast aan, of samen zijn geëvolueerd met, een andere gastheerniche. Ons onderzoek onderstreept dat er binnen een enkele bacteriesoort belangrijke verschillen kunnen bestaan. Bij microbioomonderzoek is het daarom belangrijk om deze nuances in acht te nemen bij het koppelen van bacteriën aan ziektebeelden.

Hoofstuk 7.2, tot slot, vloeit voort uit het werk in hoofdstuk 7.1, en beschrijft de reconstructie van volledige genomen van bacteriën die vermoedelijk per ongeluk in onze monsters van gezuiverde *R. gnavus* terecht zijn gekomen (contaminanten). Deze onverwachte bevinding leidde tot de reconstructie van zeventien *Streptococcus*-genomen, één genoom van *Bacteroides fragilis* en één van *Staphylococcus capitis*. Door de gebruikte sequencingmethode met lange DNA-fragmenten zijn deze genomen buitengewoon compleet. Ondanks de onduidelijke herkomst van deze bacteriën willen wij, als aanhangers van FAIR data en Open Science, deze hoogwaardige data graag delen met de wetenschappelijke gemeenschap.

Conclusies en toekomst van microbiotatherapieën

Het onderzoek in dit proefschrift heeft geleid tot nieuwe inzichten in FMT, de effecten op het darmmicrobioom van haar ontvangers en hoe dit samenhangt met de behandeling van recidiverende *C. difficile*-infecties en colitis ulcerosa. Daarnaast hebben wij nieuwe kennis opgedaan over de veelbesproken maar weinig onderzochte darmbacterie *R. gnavus*. De opgedane kennis kan bijdragen aan een verbeterd begrip van microbiotabehandelingen en creëert een groter bewustzijn van mogelijke gezondheidsrisico's op de lange termijn. Dit benadrukt dan ook het belang van het bijhouden van gegevens over FMT in zogenaamde registers, zodat zeldzame bijwerkingen kunnen worden gedetecteerd, begrepen en in de toekomst wellicht voorkomen.

Ondanks onze toegenomen kennis, blijft het een uitdaging om precies te bepalen welke bacteriën sleutelrollen vervullen in de ontwikkeling of juist behandeling van de verschillende ziekten die samenhangen met het darmmicrobioom. Voor *C. difficile* zijn al verschillende *live biotherapeutic products* (LBP) op de Amerikaanse markt of in een ver stadium van klinische studies. Deze producten bestaan vaak uit mengsels van darmbacteriën van gezonde donoren, al dan niet geïsoleerd en in consortia samengebracht. Voor complexere ziekten zoals IBD zijn soortgelijke LPB's nog in ontwikkeling. De verwachting is dat deze als geneesmiddel geregistreerde producten op termijn FMT kunnen vervangen. Echter, tot die tijd zijn studies met FMT zeer waardevol om ons begrip van het darmmicrobioom te vergroten en de rol daarvan in gezondheid en ziekte.

List of publications

1. **Nooij S**, Sanders IMJG, Schout L, Vossen RHAM, Kloet S, *et al*. Draft and complete genome sequences of 17 *Streptococcus* species. *Microbiology Resource Announcements*, 2025. Accepted 7 Jul
2. Pinto, S., Benincà E, **Nooij S**, Terveer EM, Keller JJ, *et al*. Ecological resilience in ulcerative colitis: Microbial dynamics of donor and resident species in a longitudinal fecal microbiota transplantation study. *ISME Communications*. 2025 Jul 16. doi:10.1093/ismeco/ycaf119.
3. Pinto S, Šajbenová D, Benincà E, **Nooij S**, Terveer EM, Keller JJ, et al. Dynamics of Gut Microbiota After Fecal Microbiota Transplantation in Ulcerative Colitis: Success Linked to Control of *Prevotellaceae*. *Journal of Crohn's and Colitis*. 2025 Feb 1;19(2):jjae137.
4. **Nooij S**, Plomp N, Sanders IMJG, Schout L, van der Meulen AE, Terveer EM, et al. Metagenomic global survey and in-depth genomic analyses of *Ruminococcus gnavus* reveal differences across host lifestyle and health status. *Nature Communications*. 2025 Jan 30;16(1):1182.
5. Schmitz D, Zwagemaker F, **Nooij S**, Janssens TKS, Cremer J, Verhagen R, et al. Accessible viral metagenomics for public health and clinical domains with Jovian. *Scientific Reports*. 2024 Oct 29;14(1):26018.
6. Van Lingen E, **Nooij S**, Terveer EM, Crossette E, Prince AL, Bhattarai SK, et al. Faecal Microbiota Transplantation Engraftment After Budesonide or Placebo in Patients With Active Ulcerative Colitis Using Pre-selected Donors: A Randomized Pilot Study. *Journal of Crohn's and Colitis*. 2024 Apr 4;jjae043.
7. **Nooij S**, Vendrik KEW, Zwittink RD, Ducarmon QR, Keller JJ, Kuijper EJ, et al. Long-term beneficial effect of faecal microbiota transplantation on colonisation of multidrug-resistant bacteria and resistome abundance in patients with recurrent *Clostridioides difficile* infection. *Genome Medicine*. 2024 Feb 28;16(1):1–16.
8. Terlouw D, Boot A, Ducarmon QR, **Nooij S**, Suerink M, van Leerdam ME, et al. Enrichment of colibactin-associated mutational signatures in unexplained colorectal polyposis patients. *BMC Cancer*. 2024 Jan 18;24(1):1–7.
9. Terlouw D, Boot A, Ducarmon QR, **Nooij S**, Jessurun MA, van Leerdam ME, et al. Colibactin mutational signatures in tumor syndrome and associated polyposis patients. *Genes, Chromosomes and Cancer*. 2023 Oct 5;63(1):e23208.
10. Ducarmon QR, Zwittink RD, Willems RPJ, Verhoeven A, **Nooij S**, Klis FRM van der, et al. Gut colonisation by extended-spectrum β-lactamase-producing *Escherichia coli* and its association with the gut microbiome and metabolome in Dutch adults: a matched case-control study. *The Lancet Microbe*. 2022 Apr 20;3(6):E443–E451.
11. **Nooij S**, Terveer EM. Reply - Rational donor selection for fecal microbiota transplantation. *Gastroenterology*. 2022 Mar 1;162(3):994–5.
12. Konstantinovski MM, Veldkamp KE, Lavrijsen APM, Bosch T, Kraakman MEM, **Nooij S**, et al. Hospital transmission of borderline oxacillin-resistant *Staphylococcus aureus* evaluated by whole-genome sequencing. *Journal of Medical Microbiology*. 2021 Jul 19;70(7):001384.

13. Boekhoud IM, Sidorov I, **Nooij S**, Harmanus C, Bos-Sanders IMJG, Viprey V, et al. Haem is crucial for medium-dependent metronidazole resistance in clinical isolates of *Clostridioides difficile*. *Journal of Antimicrobial Chemotherapy*. 2021 Jul 1;76(7):1731–40.

14. **Nooij S**, Ducarmon QR, Laros JFJ, Zwittink RD, Norman JM, Smits WK, et al. Faecal microbiota transplantation influences procarcinogenic *Escherichia coli* in recipient recurrent *Clostridioides difficile* patients. *Gastroenterology*. 2021 Jun 11;161(4):1218–1228.e5.

15. Ducarmon QR, Terveer EM, **Nooij S**, Bloem MN, Vendrik KEW, Caljouw MAA, et al. Microbiota-associated risk factors for asymptomatic gut colonisation with multi-drug-resistant organisms in a Dutch nursing home. *Genome Medicine*. 2021 Apr 7;13(1):1–17.

16. Vendrik KEW, Terveer EM, Kuijper EJ, **Nooij S**, Boeije-Koppenol E, Sanders IMJG, et al. Periodic screening of donor faeces with a quarantine period to prevent transmission of multidrug-resistant organisms during faecal microbiota transplantation: a retrospective cohort study. *The Lancet Infectious Diseases*. 2021 May 1;21(5):711–21.

17. Vilsker M, Moosa Y, **Nooij S**, Fonseca V, Ghysens Y, Dumon K, et al. Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics*. 2019 Mar 1;35(5):871–3.

18. **Nooij S**, Schmitz D, Vennema H, Kroneman A, Koopmans MPG. Overview of Virus Metagenomic Classification Methods and Their Biological Applications. *Frontiers in Microbiology*. 2018 Apr 23;9.

About the author

Sam Nooij was born in Amsterdam on 12 September 1990. He received his primary education at De Zevende Montessorischool and secondary education at the Cartesius Lyceum (VWO-Gymnasium), both in Amsterdam. From childhood, Sam has been fascinated by the living world around us, in particular smaller organisms such as insects. This was fostered by the film Microcosmos from 1996. Another fascination, that came to flourish after school, is Japan. Sam has been training traditional Japanese sword fighting (Katori Shinto Ryu) with Katsujinken in Amsterdam since 2008, and has been studying Japanese as a hobby. In 2009, he went to study Biology at the University of Amsterdam (UvA) with elective courses in molecular biology and microbiology. From the start, Sam enjoyed the largely abstract and theoretical side of biology, especially molecular processes and DNA. In his Bachelor's internship, he isolated, cultivated and manipulated the gut bacteria of moth larvae at the Institute for Biodiversity and Ecosystem Dynamics. He continued with the UvA Master's programme Green Life Sciences and conducted two internships: 1) about the interactions between tomato plants and the pathogenic fungus Fusarium oxysporum at the Swammerdam Institute for Life Sciences; and 2) a collaborative project on discovering micro RNAs (miRNA) in carrot plants with Bejo Zaden B.V. and the Green Student Lab. In this lab, Sam got to work with high-throughput DNA sequencing and learned computer programming.

Having graduated from the UvA in 2015, Sam started as PhD candidate at the Dutch National Institute for Public Health and the Environment (RIVM) in Bilthoven. Sam worked on bioinformatics analyses methods for virus genome sequencing, attended scientific conferences and training courses, and spent a week in Leuven to develop the popular virus bioinformatics tool Genome Detective, together with collaborators from Belgium, South Africa and Brazil. Sam participated in the first Dutch Data Carpentry workshop and has since been a part of the community, teaching programming and data analysis skills throughout the Netherlands. Sam also joined the board of the RIVM network for PhD students (Proneri), and expanded his professional network. When the project ended prematurely, it was through this network that Sam was offered a position at the Leiden University Medical Center (LUMC) in 2019. Here, Sam started working with the Center for Microbiome Analyses and Therapeutics (CMAT) and the Netherlands Donor Feces Bank (NDFB), within the Department of Medical Microbiology. This would later turn into the PhD research project on the human gut microbiome and faecal microbiota transplantation (FMT) presented in this thesis. Sam won the NCOH Science Café Pitch Award in 2021 and the Abstract Award of the KNVM Division Microbial Genomics in 2023. He also attended several national and international conferences, and was involved in different bioinformatics teaching courses and workshops. In 2022, he was awarded a travel grant by the Novo Nordisk Foundation to attend an FMT symposium in Copenhagen. Sam lives in Rotterdam together with his partner Demelza Gudde and their son Yukio. They enjoy going out together and watch birds,

trees, insects and whatever else Mother Nature offers. Sam is continuing his academic career as postdoctoral researcher in Utrecht, at the Faculty of Veterinary Medicine, Department of Clinical Infectiology.

Dankwoord

Het proefschrift ligt er en dat is alleen mogelijk geweest met de steun van velen die ik daarvoor hartelijk wil bedanken.

Mijn PhD-avontuur begon in 2015 bij het RIVM. Het kantoor was altijd gezellig en vooral met Dennis heb ik veel opgetrokken. Dennis, wij hebben prettig samengewerkt en je humor was een welkome afleiding. Gaandeweg heb ik veel opgestoken over virussen van Annelies, Harry en Marion.

Voor mijn overstap naar het LUMC wist Jeroen Laros mij te introduceren bij Ed, die mij direct het vertrouwen gaf om in zijn groep Bastian te vervangen. *Bastian, thank you so much for getting me up to speed in the MM department!*

Ik werd geïntegreerd in CMAT en dat klikte meteen. Onder begeleiding van Ed kon ik samen met Quinten, Liz, Romy, Karuna en Anoe meedoen in lopende microbioomprojecten en nieuwe onderzoeken starten. Daarnaast was het fijn samen te werken met de Experimentele Bacteriologiegroep.

Bij de NDFB werd ik onderdeel van de studiegroep en mocht ik aanschuiven bij het wekelijkse overleg waarin ook patiëntenaanvragen voor fecestransplantaties werden besproken. Ook bij zaken waar ik minder in thuis was, werd naar mijn mening geluisterd en ik heb veel geleerd van de multidisciplinaire overleggen met Liz, Ed, Josbert, Hein, Bas, Vlada, Emilie, Joffrey, Karuna, Eline en Demi.

I have been fortunate to collaborate with our research partners from Vedanta Biosciences in the USA and publish several papers together. Thank you Jason, Emily, Manon, and Jeremiah!

In de loop van de tijd veranderde een aantal dingen. Mensen gingen en kwamen: Britt, Lesley, Jannie, Alexarae, Jan, Georg and his team; thank you so much for the fun interactions and being such nice colleagues! En mijn aanstelling werd omgezet naar dat van een promotieonderzoeker. Hierbij wil ik vooral Ed bedanken voor de steun op de achtergrond, Louis voor de bereidheid de taak van promotor op zich te nemen en Jeroen Laros en Joffrey voor extra begeleiding in het proces. Daarnaast kwam ik in het kantoor te zitten met geweldige virologie-promovendi. *Thank you all for being such nice office mates!*

Ook wil ik graag verdere collega's bedanken. Binnen de afdeling MM en ook daarbuiten heb ik mogen deelnemen in verschillende overleggen en samenwerkingen. Dat was altijd leerzaam en ben ik erg blij mee. Mijn dank gaat uit naar Susanne, Igor, Marja, Diantha, Leon en het team van SASC, Susan en Rolf van het LGTC.

Na deze min of meer chronologische opsomming zijn er nog mensen tot wie ik een apart dankwoord wil richten. Ten eerste, Ed. Jij gaf me van het begin het vertrouwen en de vrijheid om al dan niet opnieuw voor een PhD te gaan. Je kwam regelmatig even langs mijn kantoor om gewoon even te kijken hoe het ging en een kort praatje. Ik herinner me nog goed tijdens de covid-lockdown dat we allemaal thuis moesten werken en ik het na een aantal maanden zat was. Ik meldde me ziek en heel snel belde je en heb je geregeld dat ik weer naar kantoor kon komen om te werken. Jij was de promotor die ik nooit heb gehad.

Wiep Klaas, jij bent altijd vriendelijk en behulpzaam geweest. Met werkoverleggen had je scherpe en constructieve opmerkingen en hielp je mijn werk weer een stap verder. Bovendien was je altijd bereid om mijn teksten kritisch na te lezen en becommentariëren, of je nou coauteur was of niet. Ik heb veel van je opgestoken en dat heeft me een betere wetenschapper gemaakt.

Dan mijn toffe copromotoren, met wie ik mijzelf zeer gelukkig prijs.

Liz, jij was er altijd voor me, ondanks dat je vaak bezig was met honderd andere dingen. Je maakte dan tijd voor een kopje thee om bij te praten, over werk en privé en ik ben erg blij met je begeleiding en dat je zo'n fijne collega was.

Quinten, wat waren wij een goed team! We hebben samen mooie projecten bedacht, slimme analysemethoden uitgezocht en daar een paar prachtpublicaties uit gesleept.

Felice, John, Bos, Doris. Jullie snapten misschien niet altijd waar ik mee bezig was, maar hebben me vaak aangehoord en altijd gesteund.

En tot slot, Demelza en Yukio. Jullie zijn mijn lieve, stabiele thuisfront. Jullie zorgen voor een lach na een lange werkdag, steun als iets even niet lukt, en relativering. En voor inspiratie voor de tekening voor- (en achter)op! Laten we samen vieren dat dit ei gelegd is en kijken wat de toekomst voor ons in petto heeft.