Universiteit
Leiden
The Netherlands

Microbiome-mediated colonization resistance: defense against enteropathogens and multi-drug resistant organisms

Ducarmon, Q.R.

Citation
Ducarmon, Q. R. (2022, March 23). *Microbiome-mediated colonization resistance: defense against enteropathogens and multi-drug resistant organisms*. Retrieved from https://hdl.handle.net/1887/3280022

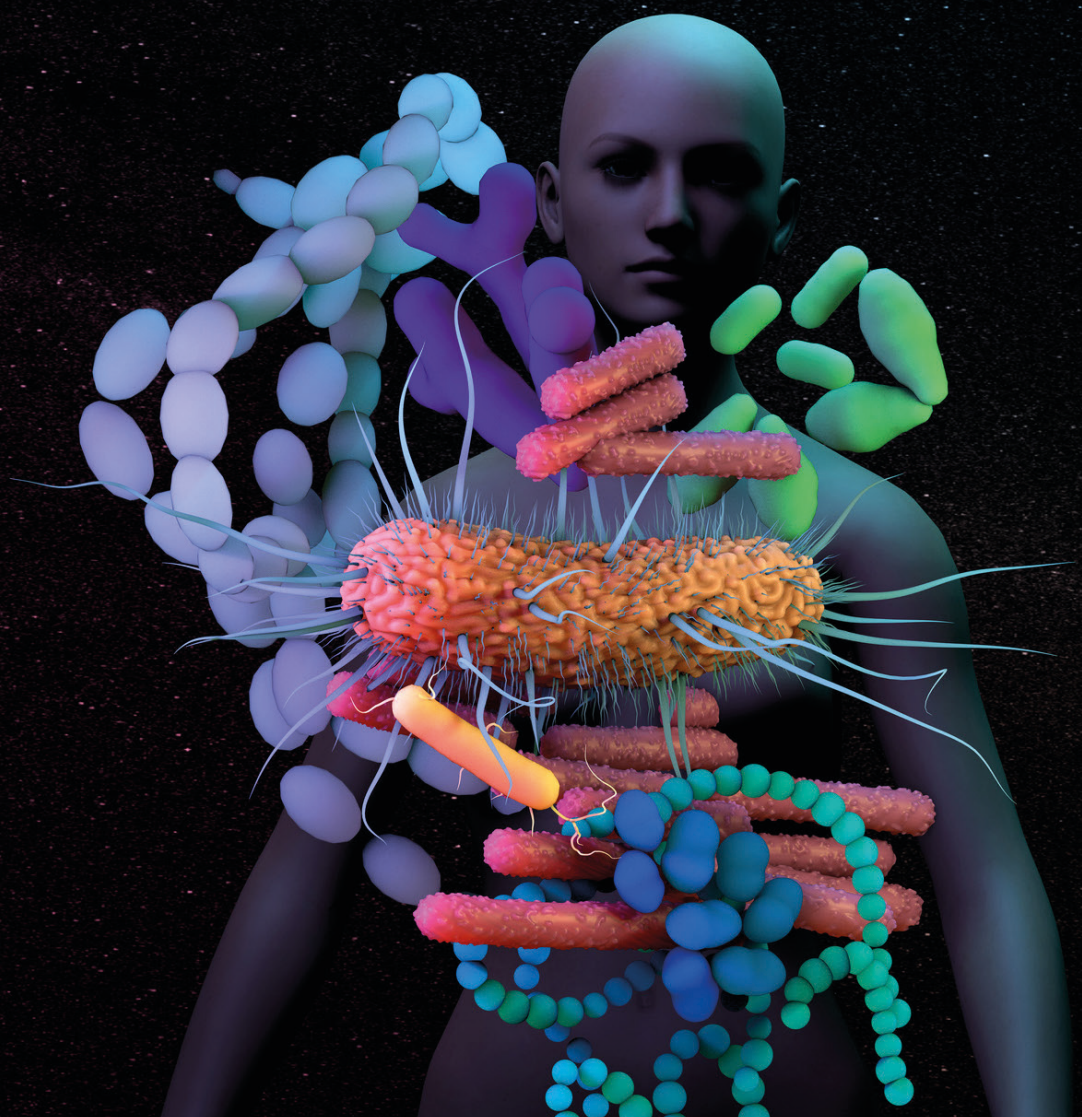
Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3280022

Note: To cite this publication please use the final published version (if applicable).

Microbiome-mediated colonization resistance:

Defense against enteropathogens and multi-drug resistant organisms

Quinten R. Ducarmon



Microbiome-mediated colonization resistance:
defense against enteropathogens and multi-drug resistant organisms

Quinten R. Ducarmon

Colophon

Thesis **Microbiome-mediated colonization resistance:** defense against enteropathogens and multi-drug resistant organisms

Cover Image iStock.com/Design Cells

Layout Ilse Modder (www.ilsemodder.nl)

Printing Gildeprint (https://www.gildeprint.nl)

ISBN 978-94-6419-431-9

Copyright The copyright of the published articles has been transferred to the respective journals or publishers.© 2022, Quinten R. Ducarmon, Oegstgeest, the Netherlands. No part of this publication may be reproduced, stored in retrieval systems, or transmitted in any form or by any means without prior permission of the author, the respective journal or publisher.

Microbiome-mediated colonization resistance:
defense against enteropathogens and multi-drug resistant organisms

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
woensdag 23 maart 2022
klokke: 11:15 uur

door

Quinten Raymond Ducarmon
geboren te Terneuzen
In 1994

Promotor
Prof. Dr. E.J. Kuijper

Co-promotor
Dr. R.D. Zwittink

Leden promotiecommissie
Prof. Dr. P.C.N. Rensen
Prof. Dr. D. Bogaert (University of Edinburgh)
Dr. G. Zeller (European Molecular Biology Laboratory)
Prof. Dr. L.G. Visser

The printing of this thesis was financially supported by the Medical Microbiology department of the Leiden University Medical Center

Table of contents

Chapter 1	General introduction and thesis outline	11
Part I: Mechanisms of microbiome-mediated colonization resistance and how to develop microbiome-based therapies		**29**
Chapter 2	Gut microbiota and colonization resistance against bacterial enteric infection	31
Chapter 3	Opportunities and challenges in development of live biotherapeutic products to fight infections	77
Part II: Optimization and standardization of laboratory and computational procedures for microbiome research		**91**
Chapter 4	Toward standards in clinical microbiota studies: comparison of three DNA extraction methods and two bioinformatic pipelines	93
Chapter 5	Development of a novel computational tool for profiling of carbohydrate-active enzymes in the human gut and its application in colorectal cancer cohorts	119
Part III: Identifying microorganisms associated with colonization resistance in clinical studies		**155**
Chapter 6	The bacterial gut microbiota of adult patients infected, colonized or non-colonized by *Clostridioides difficile*	157
Chapter 7	Dynamics of the bacterial gut microbiota during controlled human infection with *Necator americanus* larvae	177
Chapter 8	Microbiota-associated risk factors for asymptomatic gut colonisation with multi-drug-resistant organisms in a Dutch nursing home	199
Chapter 9	Asymptomatic gut colonization by extended-spectrum beta-lactamase-producing *Escherichia coli* is not associated with an altered gut microbiome or metabolome in Dutch adults	229

Chapter 10 General discussion and future perspectives 253

Nederlandse samenvatting 280
Acknowledgements 286
List of publications 288
PhD Portfolio 290
Curriculum vitae 292

Chapter 1

General introduction and
thesis outline

The human microbiome

1

Humans are colonized by microorganisms at different body sites, such as in the oral cavity, on the skin and in the gut. It is estimated that bacterial cells outnumber somatic cells (approximately by a factor of 1.3) and they certainly contain a much wider repertoire of genes than encoded by the human genome[1]. These numbers do not yet take into account other crucial, but understudied, components of the microbiome such as viruses, archaea, fungi and other eukaryotic microorganisms. Defining and distinguishing 'microbiota' and 'microbiome' remains a somewhat controversial topic and one for which extensive debate will likely remain for the coming years. A consensus statement from 2020 defined microbiota as "the assemblage of living microorganisms present in a defined environment" while microbiome was defined as not only including the community of microorganisms, but also their "theatre of activity"[2]. One of the first large-scale projects to characterize microbial communities at different human body sites was the Human Microbiome Project (HMP), officially launched in 2007[3, 4]. The HMP contributed important biological discoveries such as the notion that functional capacity of the microbiome is very stable within a healthy adult over time, but also highly similar between adults, in contrast to taxonomic composition (Figure 1)[4]. While taxonomic composition varied between individuals, general patterns could still be noticed. Most individuals' gut microbiome was dominated by either Firmicutes or Bacteroidetes, with three other phyla (Actinobacteria, Proteobacteria, Verrucomicrobia) being less abundant, but still prevalent. The notion of functional stability between individuals despite taxonomic differences was not only true for the gut microbiome, but also for all other investigated body sites, which included the buccal mucosa, tongue dorsum and anterior nares, amongst others (Figure 1)[4]. In addition, they also released several freely available computational tools (most notably MetaPhlAn for taxonomic profiling and HUManN for functional profiling) which are used by many researchers to this day[3, 5, 6]. Since the launch of the HMP, the HMP and many other research groups and consortia have uncovered that the microbial communities colonizing humans are crucial for maintaining health, and many diseases have been associated with changes in these communities. For example, the gut microbiome can contribute to human health by producing short-chain fatty acids, synthesis of several vitamins and providing resistance against colonization of incoming pathogens[7, 8]. While microbiome research is performed on different body sites, the gut remains the most intensely studied body site.

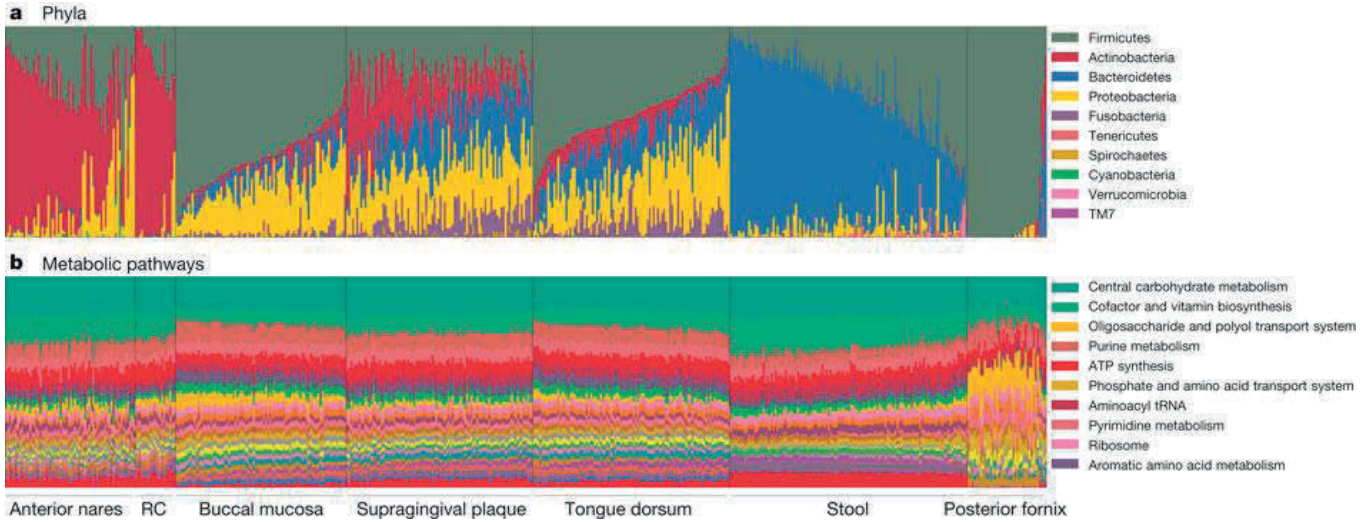


Figure 1: Carriage of microbial taxa varies while metabolic pathways remain stable within a healthy population. Vertical stacked bar charts represent relative abundance from operational taxonomic units (OTUs) obtained from 16S data (a) and metabolic modules obtained from metagenomic shotgun sequencing data (b). Legends indicate the, on average, most abundant phyla and pathways over the different body habitats. RC: retroauricular crease. Retrieved from the HMP consortium publication[4].

Development of the gut microbiome research field

The gut microbiome research field has emerged as an independent research field over the last fifteen to twenty years, even though researchers have already hypothesized about the role of the gut microbiome for far longer. For example, Theodor Escherich (after whom *Escherichia coli* was named) stated that it was crucial to study the microorganisms in the gut to understand (patho)physiological processes in the intestine in a publication from 1885[9]. The major breakthroughs in this field have been made possible by the advent of next-generation sequencing (NGS) technologies, which became more widely adopted in the first decade of this century due to the decreasing costs. NGS allows for characterization of entire microbial communities, which was not possible with traditional microbiological methods (e.g. culturing) or older sequencing methods such as Sanger sequencing. It should be noted that shotgun metagenomic sequencing (one variant of NGS) was already applied on environmental samples (by Craig Venter, amongst others) before it was widely adopted by the human microbiome field[10].

One of the early milestone papers in the gut microbiome field is a study by Turnbaugh et al. Here, the authors showed that obese mice had a gut microbiome with increased capability for energy harvest from the diet and causally linked the gut microbiome to the pathophysiology of obesity through a series of elegant experiments[11]. This included transplanting feces from obese mice into gnotobiotic mice, which led to a greater increase in body fat than when gnotobiotic mice received a fecal microbiota transplantation from lean mice. This study was one of the first to not only find a correlation between the gut microbiome and disease, but to causally link the two, and subsequently triggered a global interest in the role of the gut microbiome in human health and disease.

Most studies in the early days of microbiome research were observational studies where 16S rRNA gene amplicon sequencing of hypervariable regions was performed to compare patient groups, and differences in gut microbiota composition would be associated with disease or health parameters. Using 16S rRNA as an evolutionary marker for classifying bacteria was proposed by Carl Woese and George Fox for the first time in 1977 and preceded the first efficient sequencing technique for the 16S rRNA gene by almost ten years[12, 13]. While 16S rRNA gene amplicon sequencing can be highly valuable for understanding differences in microbiota composition, sequencing of the 16S rRNA gene only provides accurate taxonomic classification up to the genus level and does not provide functional information. Deeper resolution, or application of different methods, is necessary to obtain a more systemic image of the composition and function of a microbial community[14, 15]. In recent years there has been an increase in the number of studies employing metagenomics (sequencing of all DNA in a sample), metabolomics (measuring the metabolites in a sample) and to a smaller extent metatranscriptomics

(sequencing all RNA in a sample) and metaproteomics (measuring all proteins in a sample). Metagenomics allows for accurate taxonomic classification at species level, and sometimes strain level, and for profiling functional potential. While presence of a gene can be detected using metagenomics, this does not necessarily mean that the gene is expressed. This is why metatranscriptomics and metaproteomics are becoming increasingly important, as these techniques directly measure transcripts and proteins[15]. In addition, the metabolome is viewed as a functional readout of microbial metabolism and provides an important link between composition and function (Figure 2)[16].

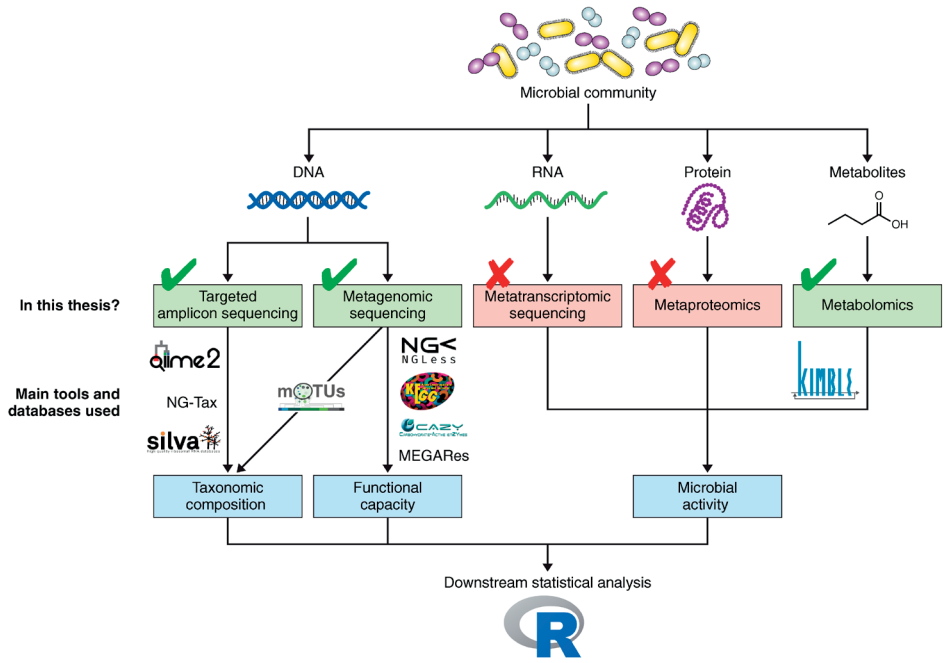


Figure 2: Multi-omics to investigate microbial communities. Each method provides specific information about the community and methods are generally complementary. The computational tools used in this thesis to process raw data are indicated, as well as the main tool used for statistical analysis (R). The reason for choosing these tools is further explained in the section below and in the respective chapters where they are employed. Logos of the tools are obtained from their respective publications and corresponding material[17-26].

Application of -omics techniques (metagenomics, metabolomics, metatranscriptomics, metaproteomics) pose bioinformatic and computational challenges. Expert knowledge is generally necessary to process raw data obtained from these techniques and extensive computational infrastructure can be required. At the Leiden University Medical Center, researchers are fortunate enough to have the luxury of working on a high-performance computing cluster, which allows for processing of large amounts of (sequencing) data.

Shotgun metagenomics and 16S rRNA gene amplicon sequencing data have been processed in this thesis using a variety of techniques (Figure 2). For 16S rRNA gene amplicon data we have used and evaluated two different tools (QIIME 2 and NG-Tax)[17, 18] and concluded that both methods work very well in combination with the SILVA database, which is the most often used database for 16S rRNA gene amplicon data[19, 27]. An enormous variety of tools exist for taxonomic profiling of shotgun metagenomic data[28]. In this thesis we opted for the mOTUs tool, as it uses single-copy marker genes for taxonomic profiling and thereby allows (as one of the very few, if not the only tool) for accurate estimation of bacterial cell numbers[20]. For functional profiling, we used the golden standard databases for metabolism (Kyoto Encyclopedia of Genes and Genomes, KEGG) and carbohydrate-active enzymes (CAZy)[22, 23]. Lastly, for resistome profiling we opted for the MEGARes 2.0 database, as it has manually curated hierarchical annotation from antimicrobial resistance genes to antimicrobial resistance mechanisms which greatly facilitate interpretation of results[24].

After data pre-processing, when matrices of e.g. bacterial species or metabolites are obtained, these matrices generally contain hundreds to thousands of features. This requires the use of specialized statistical software such as R[26] and advanced statistical techniques which can deal with the 'curse of dimensionality', whereby more features than samples are present. It needs to be emphasized however that integration of multiple -omics techniques should not be the endpoint of a microbiome study, but that findings should be taken back into the wet lab. Before taking findings back into the wet lab, it is important to be as confident as possible about computational findings, and ideally these would be confirmed by re-using data from previously conducted studies on a similar topic. This is currently often hampered by the use of different methods between research groups, which by itself can induce large variation in outcomes.

Technical opportunities and challenges for the (gut) microbiome field

Standardization of sample processing methods

Performing a clinical microbiome study typically involves multiple steps including sample collection, sample processing and choice of DNA extraction method and sequencing method[27, 29]. The use of different methods at each step in the workflow of a microbiome study complicates comparing results from different studies, as these technical factors affect the obtained profiles[29]. Research consortia have been set up to identify an optimal workflow for processing fecal samples, but this has not led to its widespread adoption across the research community[29]. This is unfortunate, as this would allow for more efficient re-use of (sequence) data from studies. Re-use of data becomes

crucial when researchers perform meta-analyses to identify robust disease-associated microbial signatures. While meta-analyses have been highly informative despite technical differences between studies[30], such an approach may fail when the disease or variable of interest is associated with more subtle changes in the microbiome. In such cases, technical variation may overshadow the biological signal. Standardization of sample processing methods would facilitate meta-analyses to allow for identification of both prominent and subtle disease-associated microbial signatures.

Biological samples with low bacterial biomass: contamination versus biological signal

In recent years, the concept of contamination (the occurrence of sequence reads in a sample which belonged to a microbe not originally present in the sample) has gained recognition[31]. When conducting a microbiome study using biological samples with a low bacterial biomass, like tumor tissue or urine, contamination can pose huge challenges. Nowadays, an increasing number of researchers is including positive and negative controls into their microbiome studies, which is an encouraging trend. At the Center for Microbiome Analyses and Therapeutics, we always include positive controls in the form of mock communities (both cell-based and DNA-based) and negative controls in the form of blank DNA extractions and blank samples for sequencing. For low-biomass samples, it may also be important to include negative controls during sample collection, although it should be noted that this is not always feasible. The inclusion of such controls in other studies has, amongst others, led to debunking of the claim of the existence of both a placental microbiome and a brain microbiome[32, 33]. In table 1 studies that investigated low-biomass samples but did not include appropriate controls are listed (which, importantly, does not necessarily mean that results are not valid)[34].

Table 1: An overview of ten studies which did not report the use of appropriate controls, thereby making it impossible to properly judge the reported results. As can be seen in the last column, a variety of low-bacterial biomass samples is studied for containing a potential microbiome.

Authors	Year	Journal	Investigated location
Aagaard et al.	2014	*Science*	Placenta
Schierwagen et al.	2019	*Gut*	Blood
Al Alam et al.	2020	*The American Journal of Respiratory and Critical Care Medicine*	Fetal lung tissue
Branton et al.	2013	*PloS One*	Brain
Gosiewksi et al.	2017	*European Journal of Clinical Microbiology & Infectious Diseases*	Blood
Willis et al.	2020	*Scientific Reports*	Eye tears
Hieken et al.	2016	*Scientific Reports*	Breast tissue
Borewicz et al.	2013	*FEMS Microbiology Letters*	Bronchoalveolar lavage fluid
Cavarretta et al.	2017	*European Urology*	Prostate tissue
Fouts et al.	2012	*Journal of Translational Medicine*	Urine

Another debated topic is the existence of a tumor microbiome and its function, which potentially has strong clinical relevance[35]. A recent and very extensive study has looked into the tumor microbiome at several locations in the human body[36]. While this paper took all possible precautions during sample collection, processing and data analysis to avoid and exclude contaminations, it is not unlikely that some contamination signal has ended up in their final data[36]. It must be noted here that it remains unclear for now what to do with information provided by positive and negative controls for potentially correcting microbiota profiles. Nevertheless, important conclusions can be drawn from these controls. For positive controls, it can be judged whether different steps in the workflow (e.g. DNA extraction and sequencing) can induce technical variation. Negative controls are especially valuable for interpretation of low-bacterial biomass samples, as negative control profiles can be compared with those of the low-biomass samples. In case these are highly similar, this suggests that the microbiota profile of the low-bacterial biomass sample is not reflecting a biological profile, but may rather be a result of contamination. Some methods have been developed to 'clean' potential contaminants from microbiome data based on control data, but no consensus has been reached in the scientific community on how to exactly deal with contamination in low-biomass samples[37-39]. Therefore, at this point, controls mainly serve to verify whether DNA extraction, sequencing and bioinformatic processing have been conducted successfully. In conclusion, it remains highly challenging to separate contamination signals from biological signals in samples with a low bacterial biomass and an important future challenge of the microbiome field is to discover what represents real biology in these cases.

State-of-the-art computational methods to profile microbiomes

Traditionally, the first step of a microbiome study after obtaining sequence data involves the accurate identification and estimation of relative abundance (taxonomic profiling) of the microorganisms in a sample. The most often applied technique for this purpose is 16S rRNA gene amplicon sequencing. This usually involves amplifying a short hypervariable region of the 16S rRNA gene and hereby profiles the bacterial fraction of the microbiota and provides accurate identification up to genus level. Advantages of this method are the relatively low costs and lower complexity as compared to metagenomic sequencing. Species level classification through 16S rRNA gene sequencing could possibly be achieved by the advent of long-read sequencing techniques, but this is not commonly implemented yet[40]. However, the most often used technique for obtaining species level resolution in a microbiome is metagenomic shotgun sequencing.

Metagenomics allows for deep resolution (accurate classification of bacterial species, and sometimes strains) and for insight into the functional potential of the microbiome. By sequencing all DNA in a sample, information is also obtained about other microorganisms than bacteria, although in feces, this is usually only a minor fraction of

reads as compared to bacteria. Therefore, tools for taxonomic profiling of metagenomes are currently largely focused on profiling the bacterial fraction of the microbiome.

Many different tools exist for taxonomic profiling of metagenomes, but they can broadly be divided into assembly-based methods (assembling short reads into larger contigs and classifying these larger contigs to a reference database) and read-based methods (assign reads to taxa by using e.g. specific marker genes)[28]. The selection of specific marker genes for taxonomic profiling is not trivial, but ideally they are universal single-copy markers and phylogenetically informative. A major advantage of single-copy markers is that no correction for genome size of each microbe is required and a closer value to the 'real' relative abundance of (bacterial) cell counts can obtained. As for 16S rRNA gene sequencing, the advent of long-read sequencing techniques may become an important tool in metagenomics, as it may allow for achieving circular bacterial genomes using assembly-based methods[41]. Functional profiling of metagenomes is a more complex and computationally intensive task than taxonomic profiling, as one needs to take all reads into account and can only focus on a subset of genes when specific functionalities are searched for. One method for functional profiling is mapping reads to a specific gene catalog relevant for the sample under investigation, for example the Integrated Gene Catalog for the gut microbiome[42], although other options exist[6, 43]. These genes can then be grouped into more informative functional groups, for example KEGG orthology (KO) groups or into carbohydrate-active enzyme (CAZymes) groups. In the context of colorectal cancer (CRC), functional profiling allowed for detecting a shift from carbohydrate degradation in a healthy microbiome towards amino acid degradation in CRC[30]. After obtaining taxonomic and functional profiles, statistics should be performed on the obtained matrices to answer the relevant research question and to link the microbiome to health or disease.

Lack of golden standards for statistical analysis

After having processed raw sequencing data, researchers are faced with the challenge of analyzing complex microbiome data. This usually involves, among others, testing differences in relative abundance of microbial taxa between groups or associating clinical variables with microbiota composition. However, there are no clear guidelines or golden standards for performing such analyses. For example, for a relatively common procedure such as differential abundance testing, many different tests are available and expert opinions differ about which tests are optimal[44, 45]. It is probably not possible to define one optimal test for differential abundance testing, as it is likely that the ideal test will depend on the dataset under study. To define an optimal test, one should have simulated data where a ground truth is known (is a taxon differentially abundant or not). However, the question here is how to define a ground truth, as in when is a taxon defined to be differentially abundant? While for standard differential abundance analysis

a wide variety of tools is available, this is not the case for all analyses that researchers wish to perform. Currently, microbiome studies are moving towards longitudinal data collection and towards causality instead of correlations. However, tools for longitudinal analysis specifically adapted to microbiome data (e.g. taking into account zero-inflation) are scarce and currently available ones are probably insufficient to capture the full complexity of the dynamics of e.g. the gut microbiome. Some tools (e.g. MetaLonDA and MetaDprof) are available which test for differences in microbial taxa over time between different groups[46, 47], but there are no such tools available that also allow for incorporation of covariates, which would be an important next step in development of statistical methods for the microbiome field.

Microbiome-mediated colonization resistance

The gut microbiome has a myriad of functions which are important for maintaining human health, and among these functions is providing colonization resistance against incoming, potentially pathogenic, microorganisms[8]. The notion that the gut microbiota can defend against enteric bacterial pathogens is far from new. For example, a paper from 1962 described that when mice are given streptomycin prior to oral administration of *Salmonella enteridis*, the resistance against this pathogen became 100,000 fold lower, with less than 10 *S. enteridis* cells being able to cause an infection in 63% of mice. In contrast, when no prior antibiotics were administered, a dose of approximately one million *S. enteridis* cells was required to infect the same percentage of mice[48].

I have previously defined colonization resistance as the ability of the microbiome to prevent colonization by exogenous microorganisms[8]. While in literature this mostly refers to incoming bacterial pathogens[49, 50], in my opinion colonization resistance to incoming commensal bacteria or other microorganisms such as viruses, fungi and even parasitic worms should also be considered. Gut microbiome-mediated colonization resistance can be conferred through several mechanisms, including nutrient competition and production of antimicrobial compounds[8]. However, the complete set of mechanisms through which microbiome-mediated colonization resistance is conferred is not completely clear yet and it is very likely that required mechanisms are different against different (pathogenic) microorganisms. It is critical to make a distinction between asymptomatic colonization by a potentially pathogenic microorganism and actual infection with enteropathogenic microorganisms whereby the pathogen causes intestinal disease. It has been shown that colonization of pathogenic bacteria often precedes overt infection[51]. Therefore, this stage might be the ideal period for intervention to prevent infection, especially in vulnerable populations such as hospitalized patients and nursing home residents. These vulnerable populations usually receive a wide array of medication, including antibiotics, which are

able to profoundly impact the gut microbiome and decrease colonization resistance[8, 52]. This concept is especially well established in the case of *Clostridioides difficile*, as *C. difficile* infection (CDI) is often observed after an antibiotic treatment. The altered microbial environment through antibiotic administrations can allow *C. difficile* to outgrow and cause infection[53]. In a landmark paper from 2013 it was shown that restoring the gut microbiome through fecal microbiota transplantation (FMT) is highly effective for curing recurrent CDI, with cure rates up to 89%[54, 55]. This publication has paved the way for development of microbiome-based therapeutics. However, before development of such products can start, fundamental research is necessary for generating insight into which commensal microorganisms can provide protection against enteropathogens.

Thesis aim

The research described in this thesis aims at identifying bacteria with potential antagonistic properties against pathogenic microorganisms and antibiotic resistant bacteria, and to address and contribute to technical challenges and opportunities in the microbiome research field.

Research questions and thesis outline

The research described in this thesis can be divided into three parts. First, we aimed to summarize the current knowledge of microbiome-mediated colonization resistance against enteropathogens (**Chapter 2**) and to provide an overview of opportunities and challenges in development of microbiome therapeutics against such pathogens (**Chapter 3**). In the second part, we focused on method optimization for microbiome research, both for wet-lab and dry-lab procedures (**Chapter 4** and **Chapter 5**). The final part describes changes in the human gut microbiota during infection or asymptomatic colonization by potentially pathogenic enteropathogens, including hookworm (**Chapter 6**), *C. difficile* (**Chapter 7**) and multidrug-resistant bacteria (**Chapters 8 and 9**). We hypothesized that we could identify bacteria or bacterial metabolites that are involved in providing microbiome-mediated colonization resistance against these pathogens. Specific research questions that we aimed to answer in this thesis were the following:

1. What is the current knowledge on microbiome-mediated colonization resistance against enteropathogenic bacteria?
2. What are the current opportunities and challenges in development of microbiome therapeutics against enteropathogenic and antibiotic-resistant bacteria, and how can we translate these into well-designed studies?
3. What is the impact of different DNA extraction procedures and different

bioinformatic pipelines for the obtained microbiota profile? How do positive and negative controls affect interpretation of microbiota profiles for low-bacterial biomass samples?

4. How do we optimize detection of carbohydrate-active enzymes (CAZymes) from (meta)genomic data using Hidden Markov models? What is the difference in CAZyme repertoire in colorectal cancer patients and are these differences independent of the geographical area of the study?
5. How does an infection with the helminth *Necator americanus* affect temporal dynamics of the human gut microbiota?
6. Can we identify bacteria that are associated with protection from asymptomatic colonization by *C. difficile*? Is it possible to understand, based on gut microbiota composition, why some individuals develop *C. difficile* infection but others only remain asymptomatically colonized?
7. Is there a role for microbiome-mediated colonization resistance against asymptomatic gut colonization of MDROs in nursing home residents? Is there spread of MDROs in this nursing home?
8. Is the gut microbiome involved in providing resistance against colonization by extended-spectrum beta-lactamase producing *Escherichia coli* in the general Dutch population?

In more detail:

Chapter 2 aimed to summarize the main mechanisms by which the gut microbiome can provide colonization resistance against enteric bacterial pathogens (nutrient competition, production of antibacterial compounds, maintenance of a healthy mucus layer and bacteriophage deployment). An important research field developing over the last few years is the effect of medication on gut microbiome function, and this chapter therefore also describes the effects of non-antibiotic medication on impacting the ability of the gut microbiome to provide colonization resistance. Lastly, it is explained how eight of the most common enteric bacterial pathogens have developed mechanisms to subvert microbiome-mediated defensive mechanisms, so that they are able to colonize the gut and cause infection.

Chapter 3 was written with the purpose of reviewing the practical aspects for development of live biotherapeutic products (LBP)s to protect against and/or cure bacterial enteric infection or colonization by multi-drug resistant organisms (MDROs). This type of medication offers an excellent alternative to conventional antibiotic therapy, as it does not damage the native microbiota and does not contribute to development of antibiotic resistance. Emergence of highly antibiotic-resistant pathogens are of ever increasing clinical importance, and solutions are urgently required for this, with LBPs being a promising option. For this chapter, we collaborated with experts from Vedanta

Biosciences Inc., a biotech company that developed several LBPs for treatment of inflammatory bowel diseases and CDI that are currently tested in phase II and phase III clinical trials.

Chapter 4 was designed with the rationale to investigate how variations in the workflow of a microbiome study can impact the obtained microbiota profiles. To this end, three different DNA extraction protocols were compared in combination with two bioinformatic pipelines. In addition, we included positive and negative controls in the workflow, an often overlooked matter in microbiome research. We hypothesized that different extraction methods and bioinformatic pipelines would lead to technical variation, but that the biological conclusions would remain the same.

Chapter 5 describes the development of a novel bioinformatic tool which profiles CAZymes in the human gut from shotgun metagenomic data. To this end, we aimed to extensively optimize settings of Hidden Markov models, annotate the Integrated Gene Catalog with CAZymes and design a novel annotation scheme for substrate specificity. The rationale for designing a novel annotation scheme was that it can be confusing to deal with a large list of different CAZymes, rather than informative functional annotation (e.g. dietary fiber metabolism). Lastly, we applied this tool in metagenomes of colorectal cancer cohorts to identify colorectal cancer-specific CAZyme signatures. With regard to these cohorts, we expected to see a decrease in fiber-degrading CAZymes in the colorectal cancer patients, as epidemiological studies strongly suggest a link between dietary fiber consumption and colorectal cancer development.

For **Chapter 6** we aimed to investigate the effect of a helminth (*Necator americanus*) infection on the bacterial gut microbiota and vice-versa. This helminth is highly prevalent in third-world countries and resides in the duodenum. We used a controlled human infection model, in which human volunteers were infected with this helminth and followed longitudinally. This helminth is highly prevalent in third-world countries and resides in the duodenum. We hypothesized that colonization and infection rates of *N. americanus* would be associated with gut microbiota composition.

Chapter 7 describes a cross-sectional study in which the bacterial gut microbiota of three groups was compared, namely CDI patients, hospitalized patients asymptomatically colonized with *C. difficile* and a control group of hospitalized patients without *C. difficile*. The aim of this study was to investigate whether specific bacterial signatures were associated with resistance against asymptomatic *C. difficile* colonization and against development of CDI. We hypothesized that patients asymptomatically colonized with *C. difficile* would have a different microbiota as compared to patients who were not.

For **Chapter 8** we conducted a point-prevalence study with four time points performed in a Dutch nursing home, where we analyzed microbiota-associated risk factors for asymptomatic MDRO colonization in a cross-sectional and longitudinal manner. In addition, we aimed to identify clinical risk factors for MDRO colonization, to investigate MDRO spread within the nursing home using whole-genome sequencing and we further investigated unexpected findings from 16S rRNA sequencing of the gut microbiota using metagenomic sequencing.

Chapter 9 describes a study which aimed to elucidate whether the microbiome provides resistance against asymptomatic gut colonization by ESBL-producing *E. coli* in adults in the general Dutch population. To this end, we collected paired fecal metagenomics and metabolomics data from individuals, who were, or were not, colonized by this bacterium. This study is unique in the sense that we were able to select samples from a large Dutch population cohort (PIENTER-3). In this way, we could exclude many common confounding factors encountered in gut microbiome research and match colonized individuals to non-colonized individuals on several clinic variables (age, sex, travel history and ethnicity).

Chapter 10 contains the general discussion of the research presented in this thesis, and describes future research directions which are crucial for advancing the microbiome field in the author's opinion.

References

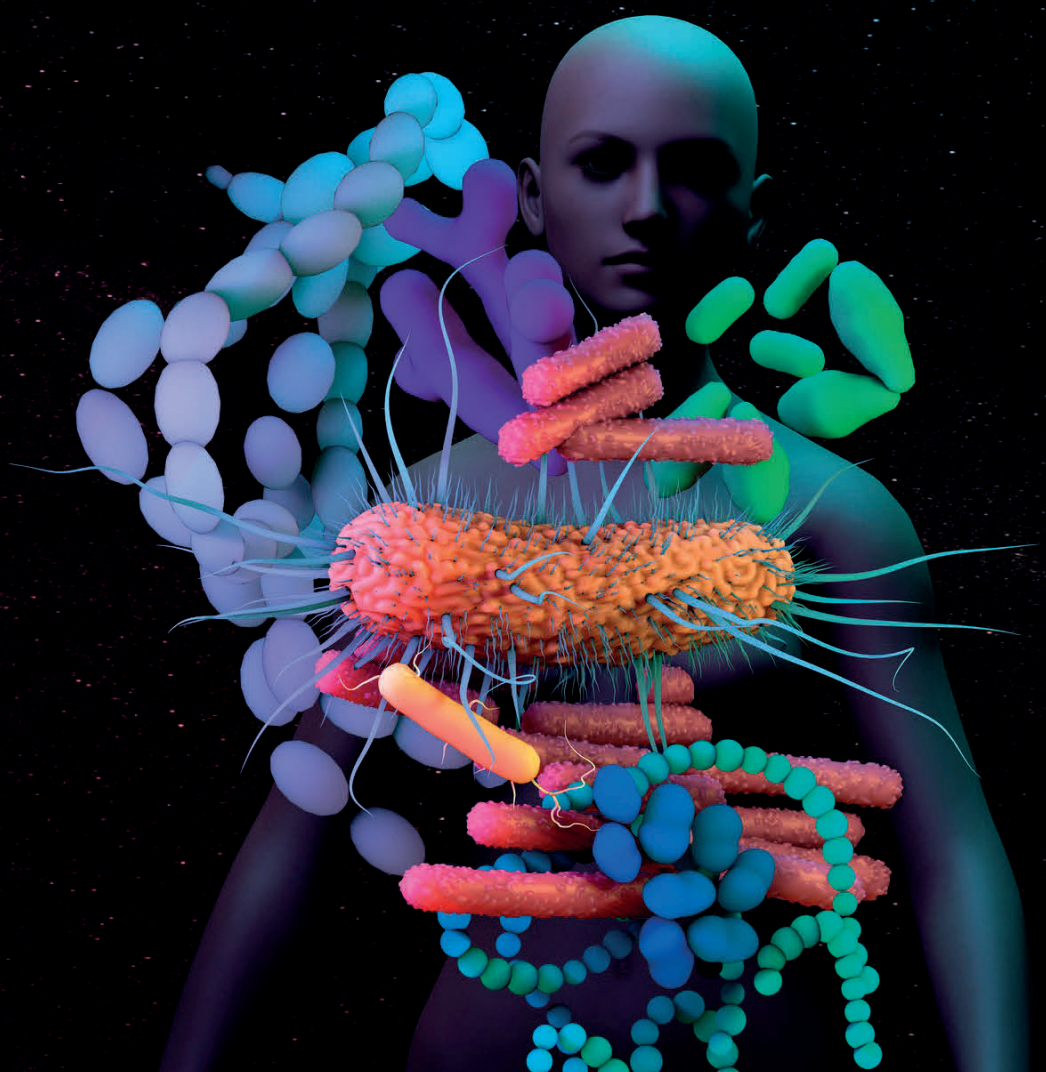
1. Sender R, Fuchs S, Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell. 2016;164(3):337-40.
2. Berg G, Rybakova D, Fischer D, Cernava T, Vergès MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. Microbiome. 2020;8(1):103.
3. Team NHMPA. A review of 10 years of human microbiome research activities at the US National Institutes of Health, Fiscal Years 2007-2016. Microbiome. 2019;7(1):31.
4. Consortium HMP. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.
5. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015;12(10):902-3.
6. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018;15(11):962-8.
7. Sommer F, Backhed F. The gut microbiota--masters of host development and physiology. Nat Rev Microbiol. 2013;11(4):227-38.
8. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev. 2019;83(3).
9. Escherich T. Die darmbakterien des neugeborenen und säuglings. Fortschr Med. 1885;3(515-522):547-54.
10. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004;304(5667):66-74.
11. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31.
12. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A. 1977;74(11):5088-90.
13. Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A. 1985;82(20):6955-9.
14. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. Nat Rev Microbiol. 2018;16(7):410-22.
15. Zhang X, Li L, Butcher J, Stintzi A, Figeys D. Advancing functional and translational microbiome research using meta-omics approaches. Microbiome. 2019;7(1):154.
16. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018;50(6):790-5.
17. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852-7.
18. Ramiro-Garcia J, Hermes G, Giatsis C, Sipkema D, Zoetendal E, Schaap P, et al. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes [version 2; peer review: 2 approved, 1 approved with reservations, 1 not approved]. 2018;5(1791).
19. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
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. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7(1):84.
22. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457-62.
23. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490-5.
24. Doster E, Lakin SM, Dean CJ, Wolfe C, Young JG, Boucher C, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. Nucleic Acids Res. 2020;48(D1):D561-d9.
25. Verhoeven A, Giera M, Mayboroda OA. KIMBLE: A versatile visual NMR metabolomics workbench in KNIME. Anal Chim Acta. 2018;1044:66-76.
26. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. https://www.r-project.org/2020.

27. 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).
28. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019;178(4):779-94.
29. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017;35(11):1069-76.
30. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25(4):679-89.
31. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.
32. de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, et al. Human placenta has no microbiome but can contain potential pathogens. Nature. 2019;572(7769):329-34.
33. Bedarf JR, Beraza N, Khazneh H, Özkurt E, Baker D, Borger V, et al. Much ado about nothing? Off-target amplification can lead to false-positive bacterial brain microbiome detection in healthy and Parkinson's disease individuals. Microbiome. 2021;9(1):75.
34. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. Sci Transl Med. 2014;6(237):237ra65.
35. Zhao K, Hu Y. Microbiome harbored within tumors: a new chance to revisit our understanding of cancer pathogenesis and treatment. Signal Transduct Target Ther. 2020;5(1):136.
36. Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science. 2020;368(6494):973-80.
37. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, et al. Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. mSystems. 2019;4(4).
38. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):226.
39. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. microDecon: A highly accurate read□subtraction tool for the post□sequencing removal of contamination in metabarcoding studies. J Environmental DNA. 2019;1(1):14-25.
40. Earl JP, Adappa ND, Krol J, Bhat AS, Balashov S, Ehrlich RL, et al. Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. Microbiome. 2018;6(1):190.
41. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. 2020;38(6):701-7.
42. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32(8):834-41.
43. Niu SY, Yang J, McDermaid A, Zhao J, Kang Y, Ma Q. Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes. Brief Bioinform. 2018;19(6):1415-29.
44. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. NPJ Biofilms Microbiomes. 2020;6(1):60.
45. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol. 2020;21(1):191.
46. Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome. 2018;6(1):32.
47. Luo D, Ziebell S, An L. An informative approach on differential abundance analysis for time-course metagenomic sequencing data. Bioinformatics. 2017;33(9):1286-92.
48. Bohnhoff M, Miller CP. Enhanced susceptibility to Salmonella infection in streptomycin-treated mice. J Infect Dis. 1962;111:117-27.
49. Sorbara MT, Pamer EG. Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. Mucosal Immunol. 2019;12(1):1-9.
50. Buffie CG, Pamer EG. Microbiota-mediated colonization resistance against intestinal pathogens. Nat Rev Immunol. 2013;13(11):790-801.
51. Cassone M, Mody L. Colonization with Multi-Drug Resistant Organisms in Nursing Homes: Scope, Importance, and Management. Curr Geriatr Rep. 2015;4(1):87-95.
52. Kim S, Covington A, Pamer EG. The intestinal microbiota: Antibiotics, colonization resistance, and enteric

pathogens. Immunol Rev. 2017;279(1):90-105.
53. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. Clostridium difficile infection. Nat Rev Dis Primers. 2016;2:16020.
54. 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):407-15.
55. 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. 2020:2050640620957765.

Part I

Mechanisms of microbiome-mediated colonization resistance and how to develop microbiome-based therapies



Chapter 2

Gut microbiota and colonization resistance against bacterial enteric infection

Gut microbiota, colonization resistance and infection
Microbiology and Molecular Biology Reviews, 2019

Q.R. Ducarmon[a,b], R.D. Zwittink[a,b], B.V.H. Hornung[a,b], W. van Schaik[c], V.B. Young[d,e], E.J. Kuijper[a,b,f,g]

[a]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, Netherlands
[b]Experimental Bacteriology, Department of Medical Microbiology, Leiden University Medical Center, Leiden, Netherlands
[c]Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK
[d]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, USA
[e]Department of Internal Medicine/Infectious Diseases Division, University of Michigan Medical Center, Ann Arbor, Michigan, USA
[f]Clinical Microbiology Laboratory, Department of Medical Microbiology, Leiden University Medical Center, Leiden, Netherlands
[g]Netherlands Donor Feces Bank, Leiden, Netherlands

Microbiol Mol Biol Rev. 2019 Jun 5;83(3):e00007-19

Summary

The gut microbiome is critical in providing resistance against colonization by exogenous microorganisms. The mechanisms via which the gut microbiota provides colonization resistance (CR) have not been fully elucidated, but include secretion of antimicrobial products, nutrient competition, support of gut barrier integrity and bacteriophage deployment. However, bacterial enteric infections are an important cause of disease globally, indicating that microbiota-mediated CR can be disturbed, and become ineffective. Changes in microbiota composition, and potential subsequent disruption of CR, can be caused by various drugs, such as antibiotics, proton pump inhibitors, antidiabetics and antipsychotics, thereby providing opportunities for exogenous pathogens to colonize the gut and ultimately cause infection. In addition, the most prevalent bacterial enteropathogens, including *Clostridioides difficile*, *Salmonella enterica* serovar Typhimurium, enterohemorrhagic *Escherichia coli*, *Shigella flexneri*, *Campylobacter jejuni, Vibrio cholerae*, *Yersinia enterocolitica* and *Listeria monocytogenes*, can employ a wide array of mechanisms to overcome colonization resistance. This review aims to summarize current knowledge on how the gut microbiota can mediate colonization resistance against bacterial enteric infection, and on how bacterial enteropathogens can overcome this resistance.

Introduction

The human gastrointestinal tract is colonized by an enormous number of microbes, collectively termed gut microbiota, including bacteria, viruses, fungi, archaea and protozoa. Bacteria achieve the highest cell density, estimated to be approximately 10^{11} bacteria/ml in the colon [1]. Research has long focused on pathogenicity of microbes and not on their potential beneficial roles for human health. Beneficial roles include aiding in immune system maturation, production of short-chain fatty acids (SCFAs), vitamin synthesis and providing a barrier against colonization with potential pathogens [2]. Additionally, the gut microbiota has extensive interactions with our immune system and it has been associated with many immune-mediated diseases both in and outside of the gut [3-5]. Over the last ten years, there has been an increased interest in elucidating the bidirectional relationship between gut microbiota and human health and disease. This has been partly propelled by improved sequencing technologies, allowing the profiling of entire microbial communities at high efficiency and low costs [6].

Hundreds of different bacterial species inhabiting the healthy human gut have been identified [7, 8]. Initial studies seeking to elucidate the relationship between human microbiota and health and disease were largely observational; gut microbiota composition would be compared between diseased and healthy groups and subsequently associated with clinical markers [9]. Currently, the field is moving towards more functional and mechanistic studies by including other –omics techniques.

In healthy individuals, the gut microbiota provides protection against infection by deploying multiple mechanisms including secretion of antimicrobial products, nutrient competition, support of epithelial barrier integrity, bacteriophage deployment, and immune activation. Together, these mechanisms contribute to resistance against colonization of exogenous microorganisms (colonization resistance, CR) [10]. However, also in absence of a fully functional immune system, the gut microbiota can provide a crucial and nonredundant protection against a potentially lethal pathogen [11]. This review will discuss the mechanisms used by gut microbiota to provide CR, the impact of various drugs on gut microbiota and thereby CR, and the strategies of specific bacterial pathogens to overcome CR and ultimately cause enteric infection.

Mechanisms providing colonization resistance

The gut microbiota produces various products with antimicrobial effects, including SCFAs, secondary bile acids and bacteriocins. Each of these contribute to CR in a product-specific manner. The following section describes their general mechanisms of

action. The contribution of the immune system in conferring CR has been extensively reviewed elsewhere and is outside the scope of this review [12, 13].

Short-chain fatty acids

SCFAs are mainly produced by bacteria through fermentation of non-digestible carbohydrates (Fig. 1) [14]. The three main SCFAs are acetate, propionate and butyrate, constituting 90-95% of the total SCFA pool [15]. During homeostatic conditions, butyrate is the main nutrient for enterocytes and is metabolized through β-oxidation. Hereby, an anaerobic milieu inside the gut can be maintained [16]. SCFAs can impair bacterial growth by affecting intracellular pH and metabolic functioning. SCFA concentrations have been shown to inversely relate to pH throughout different regions of the gut [17]. At lower pH, SCFAs are more prevalent in their non-ionized form and these non-ionized acids can diffuse across the bacterial membrane into the cytoplasm. Within the cytoplasm they will dissociate, resulting in a build-up of anions and protons leading to a lower intracellular pH [18].

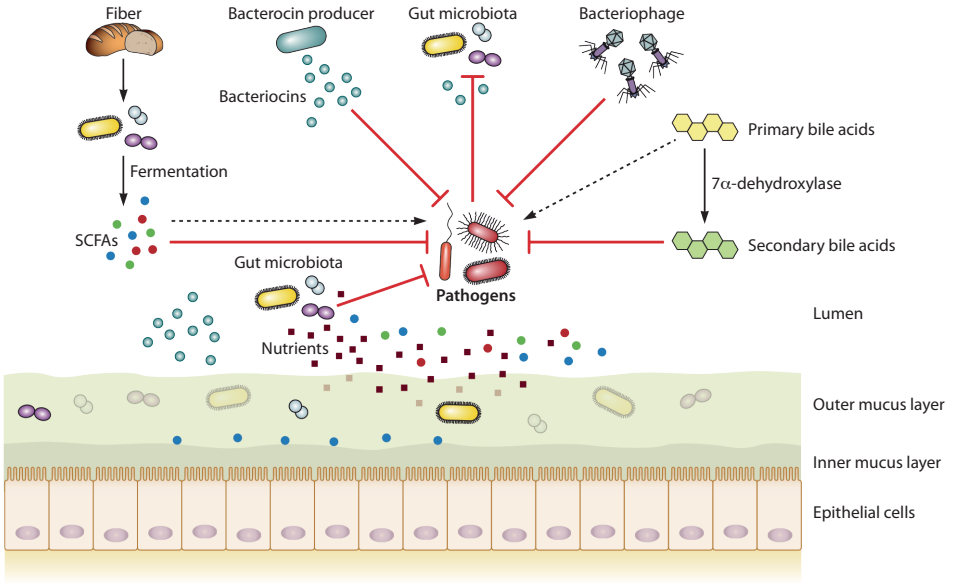


Figure 1: Outline of gut microbiota-mediated colonization resistance mechanisms. Fiber obtained from the diet is fermented by gut microbiota into short-chain fatty acids (SCFAs). Bacteriocin producers produce bacteriocins capable of targeting a specific pathogen. Primary bile acids can be converted by a very select group of gut microbiota into secondary bile acids, which generally have antagonistic properties against pathogens. Nutrient competition of native microbiota can limit access to nutrients for a pathogen. Specific organisms can use SCFAs, bacteriocins and primary bile acids to increase their virulence, as will be discussed in later sections.

In presence of acetate, metabolic functioning of *Escherichia coli* could be impaired by preventing biosynthesis of methionine, leading to accumulation of toxic homocysteine and growth inhibition. Growth inhibition was partly relieved by supplementing the growth medium with methionine, showing that this metabolic dysfunction is one of the factors by which SCFAs impair bacterial growth [19].

Bile acids

Bile acids, possessing antimicrobial properties, are produced by the liver and excreted in the intestinal tract to aid in the digestion of dietary lipids. After production of primary bile acids in the liver, they are subsequently conjugated with glycine or taurine, to increase solubility [20]. These are then stored in the gallbladder, and upon food intake, are released into the duodenum to increase solubilization of ingested lipids. A large part of conjugated primary bile acids is reabsorbed in the distal ileum (50-90%), while the remainder can be subjected to bacterial metabolism in the colon [20]. Here, conjugated bile acids can be deconjugated by bile salt hydrolases (BSH), which are abundantly present in the gut microbiome [21]. Deconjugated primary bile acids can subsequently be converted into the two main secondary bile acids, deoxycholic acid and lithocholic acid, by few bacteria, mostly *Clostridium* species, via 7α-dehydroxylation through a complex biochemical pathway [21-23] (Fig. 1). A crucial step during the conversion is encoded by the *baiCD* gene, which is found in several *Clostridium* strains, including *Clostridium scindens* [24]. Deoxycholic acid is bactericidal to many bacteria, including *Staphylococcus aureus, Bacteroides thetaiotaomicron*, *Clostridioides difficile*, bifidobacteria and lactobacilli by membrane disruption and subsequent leakage of cellular content [25-28].

The importance of bacteria for conversion of primary bile acids was demonstrated by investigating bile acid profiles in germ-free mice, where no secondary bile acids could be measured [29]. Very few colonic bacteria, less than 0.025% of total gut microbiota, are capable of performing 7α-dehydroxylation [23, 30]. One of these bacteria, *C. scindens*, is associated with colonization resistance against *C. difficile* through secondary bile acid production [22, 31]. A follow-up *in vivo* study demonstrated that *C. scindens* provided CR in the first day post infection (p.i), but protection and secondary bile acid production was lost at 72 p.i [32]. *C. scindens* on its own was also not sufficient to inhibit *C. difficile* outgrowth in humans [33]. Together, these studies suggest that *C. scindens* either requires cooperation with other secondary-bile acid producing bacteria or that other mechanisms were involved in providing CR. The secondary bile acid lithocholic acid may exert its antimicrobial effects, and potentially its effects on CR, in an indirect manner. Lithocholic acid has been shown to enhance transcription for the antimicrobial peptide LL-37, in gut epithelium using a HT-29 cell line [34]. However, no increased mRNA transcription nor protein translation of LL-37 was observed in another study using a Caco2 cell line [35].

Bacteriocins

Bacteriocins are short, toxic peptides produced by specific bacterial species that can inhibit colonization and growth of other species [36] (Fig. 1). Their mechanisms of action are multifold and include disturbing RNA and DNA metabolism, and killing cells through pore formation in the cell membrane [37-40]. Bacteriocins can be divided into those produced by Gram-positive bacteria, and those produced by Gram-negative bacteria. Further classification of bacteriocins has been extensively discussed elsewhere [41, 42]. Bacteriocins produced by Gram-positive bacteria are mostly produced by lactic acid bacteria (e.g. *Lactococcus* and *Lactobacillus*) and some *Streptococcus* species, and are further subdivided into three major classes on the basis of the molecular weight of the bacteriocins and the presence of post-translational modifications [42]. Bacteriocins produced by Gram-negative bacteria, mostly by *Enterobacteriaceae*, can be broadly divided into high molecular weight proteins (colicins) and lower molecular weight peptides (microcins) [41].

The lantibiotic nisin is the best studied bacteriocin and is produced by *Lactococcus lactis* strains. It has potent activity against many Gram-positive bacteria but has much less intrinsic activity against Gram-negative organisms [43-45]. By itself, nisin does not induce growth inhibition of Gram-negative bacteria, since binding to lipid II – the main target – is prevented by the outer bacterial membrane [46]. Therefore, studies have used different methods to overcome this problem by combining nisin with chelating agents like EDTA, antibiotics and engineered nisin peptides [47-52]. These compounds can destabilize the outer membrane, allowing nisin to exert its damaging effect [53, 54].

Several *in vivo* models have confirmed the potency of bacteriocins in providing CR. *Lactobacillus salivarius* UCC 118, which produces the bacteriocin Abp118, was able to significantly protect mice from infection by direct killing of *Listeria monocytogenes*, while an UCC 118 mutant could not, confirming the protective role of Abp118 against this food-borne pathogen [55].

Another example is *Bacillus thuringiensis* DPC 6431, which produces the bacteriocin thuricin [36]. Thuricin targets several *C. difficile* strains, including the highly virulent PCR ribotype 027. *In vitro*, its activity was more potent than metronidazole, the common treatment for *C. difficile* infection [56]. In a colon model system, metronidazole, vancomycin and thuricin all effectively reduced *C. difficile* levels. However, thuricin has the advantage of conserving gut microbiota composition. This is highly relevant, as a disturbed microbiota is associated with increased susceptibility to infection [57, 58].

Enterobacteriaceae members can produce specific bacteriocins called colicins and one example, colicin F_Y, is encoded by the *Yersinia frederiksenii* Y27601 plasmid.

Recombinant *E. coli* strains, capable of producing colicin $F_{Y,}$ were shown to be highly effective against *Yersinia enterocolitica in vitro* [59]. *In vivo* experiments were performed by first administering the recombinant *E. coli* strains, after which mice were infected with *Y. enterocolitica*. In mice with a normal gut microbiota the recombinant strains did not inhibit *Y. enterocolitica* infection, while infection was effectively reduced in mice pre-treated with streptomycin [59]. This was most probably the result of increased colonization capacity of recombinant *E. coli* in the inflamed gut, while the normal gut microbiota provided sufficient CR to prevent *E. coli* colonization [59].

Microcins are also produced by *Enterobacteriaceae*, but differ from colicins in several ways [60]. For example, microcins are of much smaller size (<10 kDa) and microcin production is not lethal to the producing bacterium, in contrast to colicin production [60]. *E. coli* Nissle 1917, capable of producing microcin M and microcin H47, could significantly inhibit *Salmonella enterica* serovar Typhimurium *in vitro* and *in vivo* [61]. This inhibition was however only seen during intestinal inflammation, during which *S.* Typhimurium expresses siderophores to scavenge iron from an iron-depleted environment. As microcins are able to conjugate to siderophores and *S.* Typhimurium takes up the siderophore during iron scavenging, microcins are introduced into the bacterial cell in a Trojan-horse like manner [62].

In silico identification of bacteriocin gene clusters shows that much remains to be discovered in this area, as 74 clusters were identified in the gut microbiota [63]. Not all of these clusters may be active *in vivo*, but it illustrates the potential relevance of bacteriocin production by the gut microbiota to provide colonization resistance.

Nutrient competition

Bacteria have to compete for nutrients present in the gut. This is especially relevant for bacterial strains belonging to the same species, as they will often require similar nutrients. The importance of nutrient competition in providing CR has been shown in multiple studies using multiple *E. coli* strains [64-67]. Indigenous *E. coli* strains compete with pathogenic *E. coli* O157:H7 for the amino acid proline [64]. In fecal suspensions, depletion of the proline pool by high-proline-utilizing *E. coli* strains inhibited growth of pathogenic *E. coli*. This inhibition could be reversed by adding proline to the medium, thereby confirming nutrient competition between the strains [64]. In addition to amino acids, different *E. coli* strains use distinct sugars present in the intestinal mucus [65]. When two commensal *E. coli* strains were present in the mouse gut that together utilize the same sugars as *E. coli* O157:H7, *E. coli* O157:H7 was unable to colonize after it was administered to these mice. However, *E. coli* O157:H7 successfully colonized when only one of these commensals was present. This indicated that the two commensals complement each other to sufficiently deplete all sugars used by this pathogenic *E.*

coli strain [66]. Nutrient competition is not limited to macronutrients, but can extend to micronutrients such as iron. *S.* Typhimurium is known to take up large amounts of iron from the inflamed gut during infection [67]. Upon a single administration of the probiotic *E. coli* Nissle 1917, which was proposed to scavenge iron very efficiently, *S.* Typhimurium levels were reduced more than two log-fold during infection via the limitation of iron availability. Administration of *E. coli* Nissle 1917 prior to infection with *S.* Typhimurium led to a 445-fold lower colonization [67].

Finally, genome-scale metabolic models have been used to reconstruct microbiome-wide metabolic networks, which could partly predict which species utilize specific compounds from their environment [68]. These models have been used to study nutrient utilization by *C. difficile*, which will be described in the section on this organism below. Together, these studies show that colonization resistance by nutrient competition is most effective when microbiota take up key nutrients that are required by the pathogen (Fig. 1). Future strategies could therefore aim at administrating probiotic strains that are able to outcompete pathogens for specific nutrients. This is especially relevant at times of gut microbiota disturbances, e.g. during and following an antibiotic treatment, as this is the time window where it is easiest for exogenous bacteria to colonize the GI tract.

Mucus layers

The gut barrier consists of the inner and outer mucus layer, the epithelial barrier and its related immune barrier. It is out of the scope of this review to discuss the full immunological characteristics of the epithelial barrier, the highly complex host-microbe interactions occurring at the mucus layer and host-associated genetic polymorphisms associated with mucus layer composition, as these have been extensively described elsewhere [12, 13, 69, 70]. Instead, a general description with various examples of how the mucus layer provides CR will be given.

The inner mucus layer is impenetrable and firmly attached to the epithelium, forming a physical barrier for bacteria thereby preventing direct interaction with the epithelial layer and a potential inflammatory response [71, 72]. Commensal gut microbes reside and metabolize nutrients in the nonattached outer mucus layer. Thinning of the mucus layer leads to an increased susceptibility for pathogen colonization, which can result from a Western-style diet deficient in microbiota-accessible-carbohydrates (MACs) [58]. When MACs were scarce, mucus-degrading bacteria (*Akkermansia muciniphila* and *Bacteroides caccae*) fed on the outer mucus layer in a gnotobiotic mouse model, resulting in closer proximity of bacteria to the epithelial layer [58]. The host adapts by increasing *muc2* expression, the main producer of intestinal mucin glycans, but fails to sufficiently do so. Inner mucus layer damage could however be reversed by administration of *Bifidobacterium longum*, perhaps due to stimulation of mucus generation [73].

The composition of the microbiota is thus a contributing factor to the integrity of the mucus barrier. Genetically identical mice housed in different rooms at the same facility showed a distinct microbiota composition, with one group of mice showing a more penetrable barrier [74]. When fecal-microbiota transplant (FMT) was performed on germ-free mice, they displayed the same barrier function as their respective donor. No specific microbes were identified to be responsible for the change in observed barrier function [74].

In conclusion, the mucus layers provide a first barrier of defense against colonization of exogenous microorganisms. Diet has been shown to be an important factor for proper functioning of this layer, suggesting that dietary intervention, or specific pro- and prebiotics, may be a future therapeutic option.

Bacteriophages

Bacteriophages are the most abundant microorganisms on our planet and are also highly present in the human gut [75, 76]. Bacteriophages have been proposed as potential alternatives to antibiotics, as they are highly specific, only targeting a single or a few bacterial strains thereby minimizing the impact on commensal members of the microbiota [75, 77] (Fig. 1). Their complex interactions in the intestine with both host immunity and bacterial inhabitants are starting to be explored, but much remains to be elucidated [76]. Here, we will focus on their relationship with bacterial enteropathogens.

Vibrio cholerae infection could be controlled using a prophylactic phage cocktail in mice and rabbits [78]. This prophylactic cocktail killed *V. cholerae in vitro*, reduced colonization of *V. cholerae* in the mouse gut and prevented cholera-like diarrhea in rabbits. Importantly, the authors suggest that the concentration of phages in the gut is an important criterion for successful prevention of infection, as timing between phage cocktail administration and *V. cholerae* inoculation was associated with treatment outcome [78]. Similar findings have been demonstrated for *Campylobacter jejuni* colonization in chickens, where a phage cocktail reduced *C. jejuni* levels several orders of magnitude [79].

Bacteriophages can also confer a competitive advantage for commensals. *Enterococcus faecalis* V583 harbors phages that infect and kill other *E. faecalis* strains, thereby creating a niche for *E. faecalis* V583 [80].

Phages play an important role in excluding specific gut bacteria and can thereby contribute to CR. Therapeutic use in humans is not yet performed at a wide scale in the Western world, as sufficient evidence for their safety and efficacy is still lacking [81]. However, recent case reports indicate that bacteriophage treatment has definite future potential for treating multi-drug resistant bacteria [82, 83].

Effects of various non-antibiotic drugs on gut colonization resistance

Antibiotics are long known for their deleterious effect on gut microbiota. Recently, various other drugs have come to attention for their impact on our microbial ecosystem. As effects of antibiotics have been extensively reviewed elsewhere [84, 85], the focus in the current review will be on non-antibiotic drugs, namely proton-pump inhibitors (PPIs), antidiabetics and antipsychotics.

Proton-pump inhibitors

PPIs inhibit gastric acid production and are among the most prescribed drugs in Western countries [86]. A significant association between long-term use of PPIs and the risk on several bacterial enteric infections has been demonstrated in multiple systematic reviews [87-90].

Several studies have associated PPI use with microbiota alterations that may specifically predispose to *C. difficile* infection and to small intestinal bacterial outgrowth [91-95]. Especially taxa prevalent in oral microbiota (e.g. *Streptococcus*) were associated with PPI use, likely resulting from increased gastric pH and thereby allowing for colonization of these bacteria further down the gastrointestinal tract [91-94]. Administering PPIs to twelve healthy volunteers for four weeks did not result in changes in diversity or changes in overall microbiota composition. However, abundance of specific taxa associated with *C. difficile* infection and gastrointestinal bacterial overgrowth increased, thereby potentially lowering colonization resistance against *C. difficile* [91].

Results of two mouse studies suggest that the reduced bactericidal effect, due to increased stomach pH, may be the most important factor for increased enteric infection risk. Mice received PPIs seven days prior to infection with the murine pathogen *Citrobacter rodentium*, which resulted in increased numbers of *C. rodentium* in the cecum one hour post inoculation as compared to control mice [96]. Similar results were observed in another study where treatment of mice with PPIs led to increased colonization of vancomycin-resistant enterococci and *Klebsiella pneumoniae* [97]. In spite of its general acceptance as a model for gut disturbances, it is important to note that mice were pre-treated with clindamycin, which may limit generalizability [97]. This is an important issue when studying effects of PPIs, as the combined use of medication in the human population complicates the study of the effects of PPIs on microbiota and CR. Even though large-scale studies have adjusted for cofounders to filter out the effect of PPIs on the gut microbiota, this does not represent a mechanistic study where only PPIs would be administered [92, 98].

Therefore, more mechanistic studies investigating how PPIs increase the risk for enteric infection are required. These studies should then exclusively administer PPIs to healthy human volunteers or animals.

Antidiabetics

Metformin is the primary prescribed drug for treatment of type II diabetes mellitus (T2DM) and mainly acts by reducing hepatic glucose production, thereby lowering blood glucose levels [99]. The current increase in the number of T2DM patients is unprecedented and it is therefore crucial to evaluate metformin's effect on gut microbiota and colonization resistance [100].

The microbiota of T2DM patients is, amongst other changes, characterized by a depletion in butyrate-producing bacteria [101, 102]. Metformin administration increased both the abundance of butyrate and other SCFA-producing bacteria, as well as fecal SCFA levels and may thus contribute to colonization resistance. The underlying mechanisms remain unknown [101, 103].

Another effect of metformin has been studied in an *in vitro* model, where it was found to reduce tight junction dysfunction of the gut barrier by preventing TNF-α induced damage to tight junctions [104]. Similar findings for improvement of tight junction dysfunction were demonstrated using two *in vivo* models, one using interleukin-10 deficient mice and one using a colitis mouse model [105, 106]. As tight junctions are a critical part of epithelial barrier integrity, alleviating their impaired functioning likely improves CR.
In conclusion, metformin may have beneficial effects on CR, as its ability to raise SCFA concentrations and improved tight junction function suggests. The effects of metformin on gut microbiota and CR in healthy organisms needs further evaluation.

Antipsychotics

The interest in whether antipsychotics affect gut microbiota composition and colonization resistance may surge after a recent publication demonstrating that antipsychotics target microbes based on their structural composition [107]. This led to the suggestion that antibacterial activity may not simply be a side effect of antipsychotics, but can be part of their mechanism of action [107]. Various antipsychotics have been investigated for their antibacterial effects, of which several will be highlighted here.

In an *in vitro* model, olanzapine has been demonstrated to completely inhibit growth of two potentially pathogenic bacteria, *E. coli* and *E. faecalis* [108]. Pimozide has been shown to inhibit internalization of several bacteria, including *L. monocytogenes* [109]. An *in vitro* screening test evaluated effects of fluphenazine on 482 bacterial strains, belonging to ten different genera. Growth inhibition was demonstrated in multiple

species, including five out of six *Bacillus spp.,* 95 out of 164 staphylococci, 138 out of 153 *V. cholerae* strains and *Salmonella* serovars Typhi and Typhimurium. Significant protection by administering fluphenazine was shown in a mouse model infected with *S*. Typhimurium, as viable cells in several organs was lower and overall survival was higher as compared to controls [110].

Antipsychotics can also be used in combination with antibiotics, to exert a synergistic antibacterial effect. Flupenthixol dihydrochloride (FD) was demonstrated to have antibacterial activity, both *in vitro* and *in vivo* [111]. Co-administration of FD and penicillin yielded extra protection against *S.* Typhimurium as compared to singular administration of either drug[111]. As antipsychotics have only recently been recognized for their potential antimicrobial effects, studies have only looked at the effects on pathogens. It is likely that gut commensals are also affected by these drugs, but future studies will have to confirm this hypothesis.

Apart from their potential antibacterial effects, several antipsychotics were shown to increase intestinal permeability in the distal ileum in rats, and therefore showing a possibly detrimental effect on CR [112]. Curiously enough, use of antidepressants was associated with increased risk of *C. difficile* infection development, although no underlying mechanism has been elucidated yet [113].

In conclusion, antipsychotics have definite antibacterial effects, but, to our knowledge, no studies have yet been performed regarding their effects on colonization resistance and bacterial enteric infection *in vivo*.

Colonization resistance towards specific bacterial enteric pathogens

Other than antibiotic resistance acquisition, enteric pathogens possess multiple virulence factors to overcome CR and cause infection. Some of these factors are common and apply to many bacterial species, others are organism-specific. Mechanisms implicated in antibiotic resistance development include horizontal gene transfer, mutational resistance and altering structure and thereby efficacy of the antibiotic molecule. Full reviews describing these mechanisms in depth can be found elsewhere [114, 115]. Here, the main focus will be on how several of the most prevalent and dangerous bacterial enteropathogens overcome the mechanisms providing CR as described herein, namely secretion of antimicrobial products, nutrient competition, mucus barrier integrity and bacteriophage deployment. As insufficient knowledge is available on how each specific enteropathogen overcomes CR by rendering bacteriophages ineffective, apart from the

well-known and conserved CRISPR-Cas, an overview of the currently known bacterial defense mechanisms will be given at the end of this review.

C. difficile

C. difficile-associated diarrhea is the most common hospital-acquired infection, causing more than 450.000 diarrheal cases per year in the United States alone [116]. Clinical symptoms can range from self-limiting diarrhea to bloody diarrhea, pseudomembranous colitis and ultimately death [117]. However, also in healthy individuals CR is not always successful against this opportunistic pathogen, resulting in asymptomatic colonization in 2-15% of the healthy population [118]. The reason why some asymptomatically colonized patients do not develop infection, while others do, may well be found in the gut microbiome, although no mechanisms have yet been elucidated. *C. difficile* contains a pathogenicity locus with the information to produce its two major toxins, TcdA and TcdB. The significance of a third toxin, called binary toxin, is less clear. Toxin production in the colon is facilitated by disruption of the native gut microbiota, for instance through antibiotic use [119].

Effects of SCFAs on *C. difficile* throughout its life cycle are currently unclear [120-122]. In an antibiotic-treated mouse model, decreased SCFA levels were associated with impaired CR against *C. difficile* [120]. CR was subsequently restored six weeks after ending antibiotic treatment with a concomitant increase in SCFAs, probably resulting from restoration of the fermentative activity of the microbiota [120]. Restoration of SCFA levels is also seen as an effect after fecal microbiota transplantations in humans [122]. However, SCFA supplementation could not induce a significant decrease in *C. difficile* shedding levels up to six weeks post infection [121]. No study has yet investigated whether *C. difficile* possesses any mechanisms by which it becomes resistant against the effects of SCFAs, which warrants further research.

Compared to the effects of SCFAs, there is more clarity on the effects of bile acids on *C. difficile*. Secondary bile acids are toxic to both *C. difficile* spores and vegetative cells, while primary bile acids generally stimulate growth and spore germination [123-125]. During antibiotic treatment, conversion of primary into secondary bile acids is suppressed and the reduction of secondary bile acids leads to a more favorable environment for *C. difficile* [120]. In addition, *C. difficile* isolates causing most severe disease in mice were also the isolates that showed highest resistance against lithocholic acid *in vitro* [126]. A relationship between disease score and deoxycholic acid could not be shown [126]. Secondary bile acid resistance may be strain-dependent, but further research is warranted to draw this conclusion with certainty.

Intrinsic anti-bacteriocin properties have been described for *C. difficile* [127, 128]. Nisin can

inhibit growth of vegetative cells and prevent spore germination of *C. difficile in vitro* [44]. However, this does not hold for all *C. difficile* strains, as the mutant strain MC119 had normal growth in sub-lethal concentrations. It was demonstrated that this resistance was at least partly due to export of nisin by an ABC-transporter [127]. Another identified mechanism was a net positive charge on the bacterial cell surface resulting in lower efficacy of nisin, since nisin is attracted to a low negative charge on the cell surface [128].

Using genome-scale metabolic models in antibiotic-treated mice, it was demonstrated that *C. difficile* does not necessarily compete for specific nutrients against specialized bacteria, but that it adapts to utilize a wide array of nutrients. This allows for colonization of diverse microbiomes, wherein *C. difficile* is not limited to a specific nutrient niche [129]. A follow-up study, also using a multi-omics approach, showed that *C. difficile* alters transcriptional activity of especially low abundant taxa. The main genes showing decreased transcription in these low abundant taxa during infection, as compared to mock infected mice, were carbohydrate-acquisition and utilization genes. A possible reason for this could be that *C. difficile* attempts to create its own nutrient niche to facilitate colonization [130].

However, others have found specific nutrients that may be important for *C. difficile* colonization and/or outgrowth. Three highly virulent ribotypes (RT), RT017, RT027 and RT078, have recently been demonstrated to utilize trehalose as a nutrient source [131, 132]. This was confirmed in a mouse model, where mice were challenged with spores of either RT027 or a non-trehalose metabolizing ribotype. After trehalose administration, RT027 mice showed higher mortality in a dose-dependent manner [131].

C. difficile post-antibiotic outgrowth depends partly on the production of succinate and sialic acid by commensals. *B. thetaiotaomicron* is capable of metabolizing polysaccharides and thereby produces sialic acid. Upon inoculation with *C. difficile,* monocolonized *B. thetaiotaomicron* mice had approximately a five times higher density of *C. difficile* in feces as compared to germ-free mice [133]. Expression levels of genes involved in sialic acid metabolism were increased in the *B. thetaiotaomicron* model, and, as expected, a sialidase-deficient *B. thetaiotaomicron* mutant led to highly reduced production of sialic acid and *C. difficile* density was lower [133]. Density of *C. difficile* was higher in *B. thetaiotaomicron* mice fed a polysaccharide-rich diet as compared to a chow diet [134]. The succinate to butyrate pathway was crucial for *C. difficile* expansion in *B. thetaiotaomicron* mice, as WT *C. difficile* was more effective in establishing infection than a succinate-transporter deficient *C. difficile* [134].

Micronutrient availability can affect virulence of *C. difficile*. High zinc levels have been demonstrated to exacerbate *C. difficile* infection in mouse models [135]. Mice fed a high-

zinc diet had higher toxin levels, higher pro-inflammatory cytokines levels and increased loss of barrier function. Furthermore, it was shown that calprotectin, a zinc-binding protein, was important for limiting zinc availability to *C. difficile* during infection [135].

Together, these studies demonstrate the importance of specific nutrients used by *C. difficile* to establish colonization and infection.

Efficient colonization of the epithelial barrier is made possible by flagella and pili [136, 137]. When mice were inoculated with flagellated or non-flagellated *C. difficile* strains, higher levels of flagellated C. *difficile* were found in mouse cecum [136]. The exact destination of non-flagellated *C. difficile* remained unknown, as levels were not measured in feces or in sections of the small intestine. Regarding pili, it has been shown that type IV pili were not playing a role in initial colonization, but were crucial for epithelial adherence and long-lasting infection [137].

_S._ Typhimurium

S. Typhimurium is a nontyphoidal S*almonella* and an important cause of gastroenteritis in humans. It was estimated that globally 3.4 million invasive nontyphoidal *Salmonella* infections occur each year, of which 65.2% are attributable to serovar Typhimurium [138]. It mostly causes self-limiting, non-bloody diarrhea in otherwise healthy individuals. However, it can lead to bloodstream infections and metastatic spread with eventually death in especially infants and immunocompromised individuals [138, 139]. *S.* Typhimurium contains two pathogenicity islands, SPI1 and SPI2. SPI1 mostly contains information for causing intestinal disease and cell invasion, while SPI2 is necessary for intracellular survival [140].

Effects of SCFAs on *S.* Typhimurium are not yet well defined. Butyrate and propionate have been demonstrated to reduce expression of invasion genes, while acetate increased their expression in *S.* Typhimurium [141, 142]. However, conflicting results exist. A *S.* Typhimurium knockout mutant, unable to metabolize butyrate, caused less inflammation than a WT *S.* Typhimurium, suggesting that butyrate is crucial for *S.* Typhimurium virulence [143]. Furthermore, this study demonstrated that butyrate was necessary for expression of invasion genes in mouse models. In contrast, propionate inhibited *S.* Typhimurium in a dose-dependent manner *in vitro*, probably due to disturbance of intracellular pH [144]. In an *in vivo* setting, it was demonstrated that a cocktail of propionate-producing *Bacteroides* species was sufficient to mediate CR against *S.* Typhimurium [144].

S. Typhimurium has developed mechanisms to overcome bile acids encountered in the gut. When exposed to individual bile acids at sub-lethal levels *in vitro*, it can become

resistant to originally lethal levels by changing gene and protein expression of several virulence regulators [145, 146]. In addition, it has been demonstrated that a mixture of cholate and deoxycholate confers a synergistic inhibition on invasion gene expression in *S.* Typhimurium [147].

Innate resistance of *S.* Typhimurium against bacteriocins produced by Gram-positive bacteria is naturally conferred through its Gram-negative outer membrane [148].

Usage of nutrients produced by gut microbiota is believed to facilitate *S.* Typhimurium outgrowth. By causing inflammation and thereby altering microbiota composition, *S.* Typhimurium provides itself with a competitive advantage [149, 150].

Metabolic profiling in mice showed increased luminal lactate levels in the inflamed gut during *S.* Typhimurium infection, which could result from a depletion in butyrate-producing bacteria [149]. When butyrate is scarce, enterocytes switch to glycolysis with lactate as end product. Lactate is an important nutrient for *S.* Typhimurium, as indicated by decreased colonization of cecal and colonic lumen by a *S.* Typhimurium mutant lacking two lactate dehydrogenases [149]. As explained in the introduction, an anaerobic milieu is maintained in the gut during homeostatic conditions. However, diffusion of oxygen from the tissue to the lumen is enabled by inflammation caused by *S.* Typhimurium, which alters enterocyte metabolism [151]. Oxygen can then be used by *S.* Typhimurium to ferment several carbohydrates through respiration [152-155]. In conclusion, these findings suggest that *S.* Typhimurium creates its own niche in the gut by causing inflammation, subsequently shifting microbiota composition and thereby nutrient availability, so that it can optimally colonize and expand.

An intact and well-functioning mucus layer is crucial for protection against *S.* Typhimurium infection. WT mice infected with the attenuated Δ*aroA* strain, which causes severe colitis, showed increased *muc2* gene expression and MUC2 production [156]. Mortality and morbidity was high in Δ*muc2* mice and higher numbers of the pathogen were found in their liver, ceca and close to the epithelial layer [156].

S. Typhimurium may profit from mucin-degrading commensal microbiota. In a gnotobiotic mouse model, complementation with mucin degrading *A. muciniphila* during *S.* Typhimurium infection allowed *S.* Typhimurium to dominate the bacterial community five days p.i [157]. This was not caused by an absolute increase in cell number, but by a decrease in other microbiota members. In addition, the complementation with *A. muciniphila* led to increased inflammation, as indicated by increased histopathology scores and protein and mRNA levels of pro-inflammatory cytokines. Although generally considered a beneficial bacterium, *A. muciniphilia* exacerbated *S.* Typhimurium

infection by thinning the mucus layer, thereby promoting translocation of the pathogen to the epithelial layer [157].

Enterohemorrhagic *E. coli*

Shiga-toxin producing *E. coli* (STEC) comprises a group of *E. coli* strains capable of producing Shiga-toxins. Enterohemorrhagic *E. coli* (EHEC) is a subgroup of STEC causing more severe disease, often with complications. Each year, approximately 100,000 people are infected by the most common EHEC serotype, O157:H7 [158]. Clinical presentation includes abdominal pain and bloody diarrhea which can progress into toxin-mediated hemolytic uremic syndrome [159]. Virulence of EHEC strains is mostly encoded by Shiga toxin genes, *stx1* and *stx2*, and by locus of enterocyte effacement (*lee*) genes, which are imperative for initial attachment to epithelial cells [160].

At present, outcomes regarding the effects of SCFAs on EHEC are mixed [161-165]. LEE protein and gene expression was already enhanced at 1.25mM of butyrate, while for acetate and propionate, only minor changes were detected at 20mM, with acetate giving a repressive effect. In a separate growth experiment, acetate was more efficient in inhibiting growth of EHEC as compared to butyrate and propionate [162]. Acetate was observed to have small repressive effects on EHEC in the study by Nakanishi *et al*., and this was also found by Fukuda *et al.* [162, 165]. Mice fed acetylated starch prior to infection showed higher fecal acetate levels and improved survival rate compared to starch-fed mice [165]. Acetate also prevented gut barrier dysfunction as measured by transepithelial electrical resistance and prevented translocation of the Shiga toxin to the basolateral side of the epithelial cells [165]. In Caco2 cells, EHEC epithelial adherence was 10-fold higher when grown on butyrate than on acetate or propionate [162]. These results indicate that butyrate may be less effective in inhibiting EHEC growth and potentially colonization as compared to acetate and propionate, for which the exact pathways and genes involved have been elucidated [162, 163]. In contrast, butyrate was found to be effective against EHEC in a pig model [161]. Piglets given sodium butyrate two days prior to being infected with EHEC showed no symptoms 24 hours p.i, while the control group developed multiple signs of disease, e.g. histopathological signs of kidney damage. The sodium butyrate group did not show any signs of inflammation and shed less viable cells compared to the control group within 48h [161]. *In vitro* assays demonstrated that butyrate enhanced bacterial clearance, ultimately making the authors suggest that butyrate can be developed as a new drug to treat EHEC [161].

EHEC has multiple traits to fight against the potentially deleterious effects of bile acids. Bile acid mixtures upregulated gene expression of the *A*crAB efflux pump and downregulated *ompF*, a gene encoding for an outer membrane porin [166]. In addition, other genes responsible for limiting penetration of bile acids through the membrane (*basR* and

basS), were upregulated, and this effect was concentration-dependent. Interestingly, the bile acid mixtures did slightly downregulate *stx2* subunit genes, encoding for Shiga toxin production [166].

EHEC possesses natural resistance against bacteriocins, especially nisin, through its Gram-negative outer membrane, as described in the chapter on bacteriocins. Three EHEC strains were screened for, amongst others, potential resistance against several colicinogenic *E. coli* strains [167]. *In vitro*, resistance against *E. coli* strains producing a single colicin was observed, but resistance was rarely observed against multiple colicins and could never be linked to acquiring a specific plasmid [167].

Nutrient competition for proline and several sugars between EHEC and commensal *E. coli* strains is described in the introductory section. In addition, ethanolamine (EA), a source of carbon, nitrogen and energy for EHEC, has been investigated. It was demonstrated that EA could diffuse across the bacterial membrane and that the *eut* genes were crucial for metabolizing EA. *Eut* sequences were absent in native bacterial genomes in the bovine gut, apart from commensal *E.coli*, indicating that EA provides a nutrient niche for *E. coli*. When the *eutB* gene was knocked out in EDL933, it was outcompeted by commensal *E. coli* due to its inability of utilizing EA, indicating its critical importance for colonization [168]. During further transcriptomic investigations of EA utilization, it was noticed that genes involved in gluconeogenesis were upregulated if no glucose was supplemented. A knockout of two genes within the gluconeogenesis pathway led to a growth defect in a coculture with the wildtype [169]. This is in line with a previous finding that optimal usage of gluconeogenic substrates by EDL933 is important for colonization [170]. Since this effect was seen in a medium consisting of bovine small intestinal contents, the relevance for the human gut remains unclear [169].

Co-culturing of EHEC with *B. thetaiotaomicron* led to an upregulation of genes involved in nutrient competition in EHEC as compared to culturing EHEC alone [171]. In addition, presence of *B. thetaiotaomicron* resulted in upregulation of multiple virulence genes including *lee*, likely due to regulation of a transcription factor involved in sensing carbon metabolite concentrations in the environment [171]. Using a combination of *in vitro* and *in vivo* methods, Pacheco *et al.* showed that fucose cleaved from mucins by *B. thetaiotaomicron* could be an important nutrient for upregulating virulence and intestinal colonization of EHEC [172]. Interestingly, fucose sensing and subsequent regulation of virulence genes was more important for successful colonization than utilization of fucose for energy. This example indicates that nutrients cannot only be utilized for energy, but that they can be important environmental signals for properly regulating timing of virulence [172].

Human colonoid monolayers were used to study initial colonization mechanisms of EHEC [173]. This study showed that EHEC disturbs the tight junctions, preferentially attaches to mucus producing cells and subsequently impairs the mucus layer [173]. In addition, by using various *in vitro* models, it was demonstrated that the metalloprotease StcE, produced by EHEC, enables degradation of MUC2 in the inner mucus layer which may pave the way to the epithelial surface [174].

S. flexneri

Shigella infections mostly occur in developing countries, with *S. flexneri* as the most frequently found species [175]. Annually, an estimated 164,000 people die of shigellosis worldwide [176]. Clinical presentation includes a wide variety of symptoms, including severe diarrhea, possibly containing blood and mucus, and abdominal pain [160]. *S. flexneri* contains a virulence plasmid (pINV) which is necessary for invasion of epithelial cells and intracellular survival [160].

No studies seem to have investigated resistance mechanisms of *S. flexneri* against SCFAs yet. Butyrate has been investigated as a potential therapeutic agent as it counteracts a putative virulence mechanism of *S. flexneri*, namely decreasing LL-37 expression in the gut [177, 178]. By suppressing LL-37 expression *S. flexneri* is able to colonize deeper into intestinal crypts [178]. Butyrate was able to increase rectal LL-37 expression in a subgroup of patients, which was associated with lower inflammation in rectal mucosa and lower levels of pro-inflammatory cytokines [177]. However, butyrate treatment did not seem to impact clinical recovery [177].

The type three secretion system (T3SS) which is able to directly inject bacterial protein into host cells and cause infection, is considered a key virulence factor. *S. flexneri* T3SS can sense and bind secondary bile acid deoxycholate, which leads to co-localization of protein translocators at the needle tip [179, 180]. In *S. flexneri* mutants lacking the needle structure, the deoxycholate-associated adhesion and invasion of *S. flexneri* to host epithelial cells was diminished [181]. At physiological levels of bile salts, *S. flexneri* is able to grow normally *in vitro*, but at increased concentrations growth is significantly reduced [182]. Transcriptomics showed that during exposure to physiological bile salt levels, genes involved in drug resistance and virulence were upregulated, which was subsequently confirmed using reverse transcription-quantitative PCR (RT-qPCR). Deletion of a multidrug efflux pump led to sensitivity to bile salts and growth inability, confirming the importance of this pump in bile salt resistance [182].

Bacteriocin resistance has not been well studied in *S. flexneri*, but downregulating antimicrobial peptide production in the gut is suggested to be an important virulence mechanism [183]. The downregulation of LL-37 early in infection was demonstrated both

in gut biopsies of patients and in cell lines [183]. Since protein and gene expression were not downregulated to the same degree, the authors speculated that there is an interference mechanism during active transcription of LL-37. Transcription of other antimicrobial peptides was also downregulated, especially in the human β-defensin hBD family [178, 183]. It was demonstrated that *S. flexneri* shows high sensitivity to LL-37 and hBD-3 peptides *in vitro* [178]. This suggests that by downregulating expression of antimicrobial peptides, *S. flexneri* creates an environment in which it can survive and ultimately cause severe disease.

It is unknown how *S. flexneri* competes and utilizes nutrients in the luminal side of the gut. Therefore, a short description will be given on how the bacterium rewires host cell metabolism for supporting its survival after entering the host cells. These findings might be translatable, and can at least provide insight in potential nutrient usage of *S. flexneri* in the lumen. Using a combination of metabolomics and proteomics it was demonstrated that *S. flexneri* does not alter host cell metabolism in HeLa cells, but that it captures the majority of the pyruvate output [184]. Pyruvate was demonstrated to be a crucial carbon source for *S. flexneri* cultured on a HeLa derivative, using metabolomics, transcriptomics and bacterial mutants [185]. *S. flexneri* converts pyruvate into acetate via a very quick, but energy-inefficient pathway, allowing for rapid expansion of the bacterium intracellularly without rapid destruction of the host cell [184].

S. flexneri possess special systems to alter mucus composition. Human colonoid monolayers infected with *S. flexneri* showed increased extracellular release of mucins [186]. The increased extracellular mucins were trapped at the cell surface which surprisingly favored access of *S. flexneri* to the apical surface, subsequently promoting cell invasion and cell-to-cell spread [186]. Furthermore, expression of several genes encoding for production of mucins and mucin glycosylation patterns were altered [186]. Together, these results suggest that *S. flexneri* can alter the mucus environment such that it can promote its own virulence.

C. jejuni

C. jejuni is associated with food-borne gastroenteritis and is estimated to cause more than 800,000 infections annually in the USA alone [187]. Major clinical symptoms include diarrhea (both with and without blood), fever and abdominal cramping [160]. In rare cases, it can give rise to the Guillain-Barré syndrome and reactive arthritis [187]. It is a commensal bacterium in avian species and it is not yet well understood why it causes disease in humans [188].

There is a distinct lack of research on the resistance mechanisms of *C. jejuni* against SCFAs, but one study found that SCFAs are important for colonization in chickens [189].

Acetinogenesis, the conversion of pyruvate to acetate, is a crucial metabolic pathway for optimal colonization of *C. jejuni*. Mutants unable to use this pathway show impaired colonization and decreased expression of acetinogenesis genes. Upon encountering a mixture of SCFAs at physiological levels, this mutant was surprisingly able to restore acetinogenesis gene expression to WT levels. Therefore, it was investigated whether expression of acetinogenic genes differs throughout the intestinal tract, as SCFAs are most abundant in distal parts of the intestine. It was observed that both gene expression and *C. jejuni* levels were highest in the cecum. The authors suggested that *C. jejuni* can monitor SCFA levels in the gut, so that in response it can express colonization factors [189]. As this is the only study suggesting this hypothesis, further research is required for validation.

Results regarding bile acid resistance in *C. jejuni* are mixed, which may stem from using different animal models or bile acids. A specific multidrug efflux pump, CmeABC, was important for bile resistance in chickens [190]. Δ*cmeABC* mutants showed impaired growth *in vitro* and unsuccessful colonization in chicken upon cholate administration, while cholate did not affect growth and colonization of the WT [190]. This suggests that the efflux pump is critical for proper colonization of *C. jejuni* by mediating bile-acid resistance. Another study elucidated the effects of secondary bile acids on *C. jejuni* [191]. Upon administration of deoxycholate prior to, and during, infection, mice showed decreased colitis. Unexpectedly, *C. jejuni* luminal colonization levels were not affected [191]. In conclusion, *C. jejuni* colonization seems not to be affected by bile acids, but may be important in limiting disease progression.

Bacteriocin resistance is not common in *C. jejuni*. Multiple *C. jejuni* (n=137) isolates were screened for resistance against two anti-*Campylobacter* bacteriocins, OR-7 and E-760, produced by the gut inhabitants *L. salivarius* and *Enterococcus faecium*. However, no isolates were found to harbor resistance [192]. In a follow-up study, chickens were successfully colonized with a *C. jejuni* strain prior to bacteriocin treatment, with the aim of studying bacteriocin resistance. Resistance developed in most chickens, but was lost upon ending bacteriocin administration, suggesting resistance instability *in vivo* [193].

In contrast to most other enteric pathogens, *C. jejuni* does not metabolize carbohydrates as its main energy source. It is unable to oxidize glucose, fructose, galactose and several disaccharides, including lactose, maltose and trehalose, resulting from the absence of 6-phosphofructokinase [194-197]. Fucose could be metabolized by some *C. jejuni* strains, due to the occurrence of an extra genomic island [197]. Main energy sources for *C. jejuni* are organic acids, including acetate, and a limited number of amino acids [198-200]. It is currently unclear what these metabolic adaptations mean for its colonization potential,

but it is possible that *C. jejuni* occupies a unique macronutrient niche.
Iron regulation systems are critical for colonization and persistence of *C. jejuni*. In presence of sufficient iron, transporter and acquisition genes are downregulated [201]. Mutants lacking genes involved in either iron acquisition or transport were severely impaired in colonizing the chick gut [201]. Free iron concentrations are extremely low in the gut, which forces *C. jejuni* to utilize other iron sources. It was demonstrated that lactoferrin and transferrin can also be used for this purpose and molecular pathways have been described [202]. In short, transferrin-bound iron can only be utilized if it is in close proximity to the bacterial cell surface. Thereafter, it is most likely that iron is freed from the bacterial cell surface proteins, transported across the outer membrane and subsequently internalized by an ABC-transporter [202]. Additionally, both in an *in vitro* setting and in a controlled human infection model with *C. jejuni* the most upregulated genes were involved in iron acquisition [188, 203]. These results suggest that iron regulation is maintained extremely well, and that *C. jejuni* can obtain sufficient iron even in a harsh environment as the gut.

C. jejuni resides in the mucus layer prior to invading the epithelial cell. It can cross and reside here because of its powerful flagellum, which can change in conformation or rotation upon being challenged by higher viscosity [204, 205]. *C. jejuni* can hereby cross the mucus layer at speeds which cannot be met by other enteric pathogens, and the flagellum can subsequently be used as an adhesin [205, 206].

Another important characteristic for *C. jejuni* 's success in crossing the mucus layer is its helix-shape. In a mouse model, a WT strain or either of two rod shaped *C. jejuni* bacteria, Δ*pgp1* or Δ*pgp2*, were administered to cause infection [207]. Rod-shaped mutants were demonstrated to be mostly non-pathogenic, whereas the WT strain caused severe inflammation. Mutants were to some extent able to colonize the mucus layer, but could not cross it, explaining their non-pathogenicity [207].

V. cholerae

V. cholerae is one of the first bacterial pathogens where the microbiota has been considered to play an important role against infection [208]. It is mainly prevalent in contaminated brackish or salt water and can cause outbreaks, particularly during wars and after natural disasters. In the first two years following the earthquake in Haiti, 2010, more than 600,000 people were infected with *V. cholerae* serogroup O1, biotype Ogawa, resulting in more than 7,000 deaths [209]. The clinical course is characterized by watery diarrhea, which can be so severe that it can result in dehydration, hypovolemic shock and death [210]. *V. cholerae* colonizes the small intestine by employing the toxin-coregulated pilus, after which it can cause severe infection and clinical symptoms through cholera enterotoxin production [210].

V. cholerae is able to utilize its acetate switch, the shift from elimination to assimilation of acetate, to increase its own virulence [211]. In a *Drosophila* model, it was demonstrated that *crbRS* controlled the acetate switch, while *acs1* was required for acetate assimilation [211]. When either of these genes were knocked-out, mortality decreased. Competition experiments demonstrated that WT *V. cholerae* had a growth advantage over strain when the Δ*crbS* strain and WT *V. cholerae* strains were administered together in a 9:1 ratio. This led the authors to suggest that acetate utilization may be important early in infection, when low levels of *V. cholerae* cells are present [211]. Furthermore, acetate consumption led to dysregulation of host insulin signaling pathways, ultimately leading to intestinal steatosis and increased mortality. Dysregulation of host insulin signaling was not observed in Δ*crbS* or Δ*acs1*, further confirming the role of acetate in *V. cholerae* virulence [211].

V. cholerae has a master regulator, *toxT*, which can directly activate several virulence factors including toxin production. Cholera toxin production was reduced by 97% when *V. cholerae* was grown in presence of bile, which could be reversed after growing the same cells in bile-free medium for a few hours [212]. C*tx* and *tcpA*, encoding for cholera toxin and the major structural unit of the toxin-coregulated pilus and regulated by *toxT*, were highly repressed during bile exposure [212]. Additionally, motility was increased approximately 1.6-fold in presence of bile [212]. To elucidate which exact components of bile acids were responsible for the repression of these virulence genes, bile was fractionated. It was found that several unsaturated fatty acids strongly repressed *ctx* and *tcpA* and that they upregulated expression of *flrA*, leading to increased motility [213]. The reason for upregulation of *flrA* and downregulation of *tcpA* could be that the flagellum increases the speed of passing through the mucus layer, while the pilus would only slow it down. When lower concentrations of bile at the epithelial surface are encountered, expression can be reversed [214].

Two outer membrane porins, OmpU and OmpT, are directly regulated by the master regulator *toxR*. Upon encountering bile acids, *ompU* and *ompT* are regulated in such a way that bile acid entrance is prevented [215, 216]. Furthermore, Δ*toxR* mutants are more sensitive to bile acids due to changed outer membrane composition [215]. Recently, it was shown that *toxR* also regulates *leuO* [217]. *LeuO* was demonstrated to confer bile resistance independent of the two porins, although its exact resistance mechanism is not yet elucidated [217].

Bacteriocin resistance in *V. cholerae* has, to our knowledge, not been studied and future studies will have to reveal whether any resistance is present.

An important nutrient through which *V. cholerae* gains a competitive advantage is

sialic acid, a component of the mucus layer. Using streptomycin pre-treated mice who were given a mutant strain defective in sialic acid transport (Δ*siaM)*, it was shown that sialic acid is not required for initial colonization, but that it is important for persistent colonization [218]. Competition assays of the two mutant strains in mouse intestine (small intestine, cecum and large intestine) showed that Δ*siaM* was less fit to compete in each environment, further indicating the necessity of sialic acid utilization for niche expansion of *V. cholerae* [218].

The El Tor strain may have a competitive advantage over 'classical' strains due to its differential carbohydrate metabolism [219]. When grown in a glucose-rich medium, classical strains display a growth defect as compared to El Tor. It was observed that this was due to production of organic acids through glucose metabolism, leading to acidification of the medium. El Tor biotypes were found to produce acetoin, a neutral compound, and decrease organic acid production. This prevented acidification of the medium, leading to better growth. El Tor strains were also more successful in colonizing mice, especially when extra glucose was administered. The classical types were shown to be able to produce acetoin, but glucose only led to a minor increase in transcription of genes necessary for acetoin production [219]. These studies have shown that specific metabolic pathways are used by *V. cholerae* to successfully colonize the gut.

One of the first studies on how the mucus layer can potentially be crossed by *V. cholerae* was reported almost 50 years ago [220]. Here, motile and non-motile strains were compared for pathogenicity after administration to mice. It was observed that motile strains were almost always deadly 36 hours p.i, while most non-motile strains had a mortality of under 35% [220]. One hypothesis offered by the authors was that together with mucinase, the flagellum could effectively pass the mucus barrier [220]. Specific mucin degradation mechanisms employed *by V. cholerae* have been identified since, with hemagglutinin/ protease (Hap), and TagA being the major ones [221-225]. Presence of mucins, limitation of carbon sources and bile acids maximized production of Hap, while glucose could partly reverse this effect [221]. This may indicate that during conditions as encountered in the gut, *V. cholerae* quickly aims to cross the mucus layer and be in close contact with the epithelial cells. TagA, which is similar to StcE as described for EHEC, is also capable of degrading mucin [222]. In conclusion, *V. cholerae* has developed a way of sensing environmental conditions, and in response to these, is able to upregulate virulence factors which can degrade mucins. A simplified overview of *V. cholerae* virulence factors opposing CR can be found in Fig. 2.

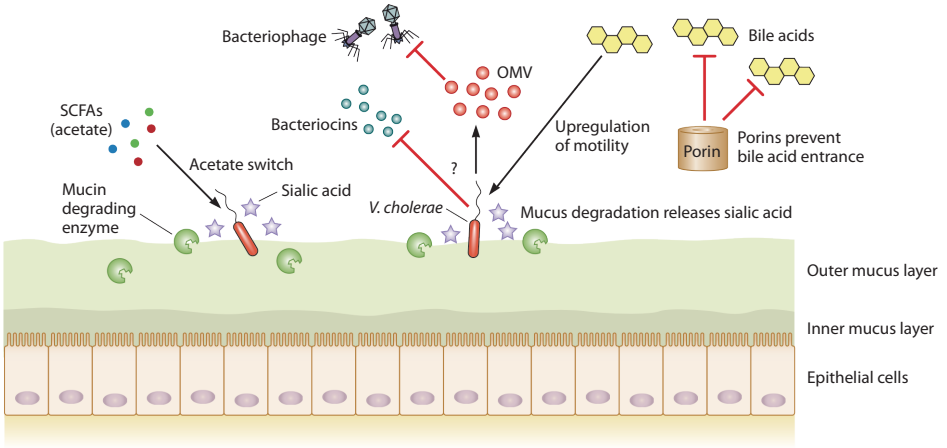


Figure 2: *Vibrio cholerae* uses a wide array of mechanisms to overcome CR. First, it employs its acetate switch to use acetate for upregulating its own virulence. Nothing about potential bacteriocin resistance is presently known, and this subject remains to be studied. To protect itself from bacteriophages, *V. cholerae* produces outer membrane vesicles (OMVs) which act as a decoy binding site for the attacking phages (see section: Bacterial defense mechanisms against bacteriophages). Regulation of outer membrane porins is such that they prevent entry of bile acids when they are encountered. By employing specific mucin-degrading enzymes, *V. cholerae* releases sialic acid and subsequently metabolizes it.

Y. enterocolitica

Yersiniosis is mostly contracted through contaminated food or water with *Y. enterocolitica*, and its prevalence is much higher in developing countries than in high-income nations [160, 226]. It is characterized by mild gastroenteritis, abdominal pain and is usually self-limiting, though pseudo-appendicitis illnesses can occur [160]. Virulence is mostly conferred through presence of a 64-75 kb plasmid on which several virulence genes are present, including *yadA*, which is crucial for epithelial adherence [227].

Resistance of *Y. enterocolitica* against antibacterial compounds has not been much studied. One study investigated effects of SCFAs, including acetic acid and propionic acid, on *Y. enterocolitica* at 4°C. *Y. enterocolitica* was less sensitive to acetic acid when cultured anaerobically than under aerobic culturing. Propionic acid was similarly effective in inhibiting growth with both culture methods [228]. Even though conditions like 4°C are not representative for the intestinal environment, this study might provide some initial clues on the effects of SCFAs on *Y. enterocolitica*. It is clear that more research is required to further elucidate potential resistance mechanisms.

ompR, a transcriptional regulator in *Y. enterocolitica*, is probably able to upregulate expression of the AcrAB-TolC efflux pump, which, in turn, is regulated by two components of the efflux pump, *acrR* and *acrAB* [229]. A mixture of bile acids, but not

the secondary bile acid deoxycholate, was found to be the strongest inducer of *acR* and *acrAB* [229]. Whether the upregulation of these efflux pump components contributes to bile acid resistance, remains to be elucidated.

Bacteriocin resistance is so far mostly unknown in *Y. enterocolitica*. WA-314 and 8081 are both 1B:O8 strains that are highly infective in murine models [230]. WA-314 possesses a putative colicin cluster for colicin production, but no expression was observed in a spot-on-lawn assay with 8081 and the colicin-sensitive *E. coli* K12 [230]. It is likely that no specific resistance against colicin is present, as colicin has been shown to effectively inhibit *Y. enterocolitica* infections *in vivo* [59].

Like most other enteric pathogens, *Y. enterocolitica* has sophisticated systems to acquire sufficient iron. Using these systems, *Y. enterocolitica* may be more efficient at scavenging iron than commensal members, thereby providing itself with a competitive advantage. *Y. enterocolitica* expresses yersiniabactin, *ybt*, a highly efficient siderophore and a crucial component for lethality in mouse models [231, 232]. The exact mechanisms for iron uptake and transport have been extensively reviewed elsewhere [233]. Proteomics analysis revealed that *Y. enterocolitica* serovar 1A, whose pathogenic role is unclear, uses different proteins to successfully scavenge iron, as it lacks the Ybt protein [234].

Y. enterocolitica is the only pathogenic Yersinia species which can metabolize sucrose, cellobiose, indole, sorbose and inositol [235]. Additionally, it can degrade EA and 1,2-PD by using tetrathionate as a terminal electron acceptor [235].

Mucus layer invasion and adherence of *Y. enterocolitica* have been elucidated in great detail several decades ago [236-240]. The YadA protein is used for initial attachment to the mucus [240]. The preferential binding side on mucins is their carbohydrate moiety, but binding to mucin proteins is also possible under specific conditions [238]. *Y. enterocolitica* uses a plasmid, pYV, with mucin-degradation enzymes to thin the mucus layer, facilitating crossing of the mucus layer [237, 240]. *Y. enterocolitica* containing the pYV plasmid is not only able to successfully invade and degrade the mucus layer, but is also highly efficient in multiplying in this environment [240]. After interacting with the mucus layer, its bacterial cell surface was altered so that *Y. enterocolitica* became less efficient in colonizing the brush border [240]. This may be a host response mechanism to prevent *Y. enterocolitica* invasion in deeper tissues. In a rabbit infection model, persistent goblet cell hyperplasia and increased mucin secretion was observed throughout the small intestine over 14 days [236]. The extent of hyperplasia was associated with severity of mucosal damage, indicating a compensatory mechanism. Mucin composition changed in infected rabbits, with a decrease in sialic acid and an increase in sulfate [236].

L. monocytogenes

L. monocytogenes causes listeriosis, a food-borne disease. Listeriosis is not highly prevalent, with an estimated 23,150 people infected in 2010 worldwide, but has a high mortality rate of 20-30% [241]. The most common syndrome is febrile gastroenteritis, but complications can develop, such as bacterial sepsis and meningitis [241]. This is especially relevant for vulnerable patient groups, such as immunocompromised individuals, neonates and fetuses [242]. Virulence genes are present on an 8.2-kb pathogenicity island, which includes internalin genes necessary for invading host cells [243].

Culturing *L. monocytogenes* in presence of high levels of butyrate leads to incorporation of more straight-chain fatty acids in the membrane [244, 245]. This is not a natural state for *L. monocytogenes*, as normally its membrane consists for a very high percentage of branched-chain fatty acids. When subsequently exposed to LL-37, it displays a survival defect as compared to bacteria not grown in presence of butyrate [244]. It was not elucidated whether this survival defect was due to increased stress, altered membrane composition or differentially regulated virulence factors. Effects of propionate on *L. monocytogenes* growth, metabolism and virulence factor expression are dependent on temperature, oxygen availability and pH [246]. Therefore, it is not possible to ascribe a general function to propionate in relation to *L. monocytogenes*.

L. monocytogenes possesses several bile acid resistance mechanisms, and *in vitro* transcriptome and proteome analyses have provided insight into these. Transcriptomics analysis revealed that in response to cholic acid, amongst others, two efflux pumps were upregulated, *mdrM* and *mdrT* [247]. BrtA was shown to regulate expression of the efflux pumps, and to be able to sense bile acid levels. Bacterial abundance was determined in multiple organs of mice infected with knockout strains of either efflux pump, but not in the intestine [247]. Proteomic analyses found many changes in response to bile salts and included proteins associated with efflux pumps, metabolism and DNA repair [248].

Bile salt hydrolases (BSH) are another way of combatting encountered bile acids. It was demonstrated that all *Listeria* species which infect mammals showed BSH enzyme activity. BSH was crucial during infection of guinea pigs, demonstrated by the decreased ability of Δ*bsh* to cause a persistent infection [249]. At decreased pH levels, e.g. in the duodenum, bile salts are more acidic and show higher toxicity [250]. However, this toxicity seems to be strain-dependent [251]. The strain responsible for a 2011 outbreak even displayed higher bile resistance at pH 5.5 than at 7.0, further indicating that bile susceptibility may be strain-dependent [251].

As discussed in the introductory section on bacteriocins, the Abp118 bacteriocin produced by *L. salivarius*, protected mice from *L. monocytogenes* infection [55].

However, several bacteriocins have been shown ineffective against *L. monocytogenes* and responsible mechanisms have been partly elucidated. Innate nisin resistance has been associated with multiple loci [252]. One crucial gene was *anrB*, encoding for a permease in an ABC transporter. Loss of this gene resulted in high sensitivity, not only to nisin, but also to several other bacteriocins [252]. The mannose phosphotransferase system (Man-PTS), encoded by *mptACD*, is a main sugar uptake system and two of its outer membrane proteins, IIC and IID, can serve as a class II bacteriocin receptor [253]. In natural resistant and spontaneous resistant strains, a reduced expression of *mptC* and *mptD* was observed, although this could not be linked to receptor mutations [254]. The *mpt* operon is partly regulated by *manR*, and a *manR* mutant did not show any activation of the *mpt* operon [255]. Development of bacteriocin resistance was to some extent dependent on available carbohydrates [256]. Several sugar sources impaired growth of *L. monocytogenes* when exposed to bacteriocin leucocin A. Increased sensitivity to leucocin A was hypothesized to relate to sugar uptake by Man-PTS. When specific sugars are present, cells may not downregulate this system even in presence of bacteriocins, which possibly allows leucocin A to use the Man-PTS as a docking molecule [256]. Not only does *L. monocytogenes* display bacteriocin resistance, it also produces a bacteriocin, Lysteriolysin S, which modifies the gut microbiota such that intestinal colonization is promoted [257]. *Allobaculum* and *Alloprevotella*, genera known to contain SCFA-producing strains, were significantly decreased in mice treated with Lysteriolysin S. *L. monocytogenes* strains unable to produce Lysteriolysin S were impaired in competing with native gut microbiota and colonized less efficiently [257].

Most reports about metabolic adaptations of *L. monocytogenes* have logically described intracytosolic adaptations, as *L. monocytogenes* replicates intracellularly [258]. Limited information is available on nutrient competition of *L. monocytogenes* inside the lumen. Comparison of genome sequences between colonizing *Listeria* and non-colonizing *Listeria* led to identification of, amongst others, a vitamin B12-dependent 1,2-propanediol (1,2-PD) degradation pathway in colonizing *Listeria*, dependent on the *pduD* gene [259]. Mice were co-infected with a Δ*pduD* strain and a WT strain. Within 3 hours after feeding, a large amount of the Δ*pduD* was shed in feces and 21 hours later the number of viable cells decreased significantly. At ten days p.i, the Δ*pduD* strain was completely cleared, while the WT strain shed for up to four more days. This indicates that the ability to degrade 1,2-PD offers *L. monocytogenes* a distinct competitive advantage [259].

Multiple adhesins and internalins have been characterized which facilitate *L. monocytogenes* retention in the mucus layer [260-263]. InlB, InlC, InlL and InlJ were demonstrated to bind to MUC2, but not to epithelial cell surface MUC1 [262, 263]. Histopathological analysis of a listeriosis rat model revealed that *L. monocytogenes* was present in the mucus layer after less than 3 hours p.i [261]. At this time point, very few *L. monocytogenes* were present on the epithelial cells [261].

Bacterial defense mechanisms against bacteriophages

As research investigating how each enteric pathogen overcomes CR by rendering bacteriophages ineffective is still in its infancy, this general section will describe the most employed resistance mechanisms. The bacteriophage infectious cycle involves a lytic and a lysogenic cycle. Phages have to bind to a receptor on the bacterial surface to be able to insert their genomic material, usually DNA, into the bacterial cytoplasm and subsequently circularize their DNA [264]. Here, lysogenic and lytic bacteriophages' mechanisms start to branch (Fig. 3). Lytic phages start DNA replication, assemble their proteins and pack their DNA into the typical bacteriophage shape with a capsid head and tail. After sufficient replication, phages use lytic enzymes to form holes in the bacterial cell membrane, eventually leading to lysis of the cell and phage spreading. Lysogenic phages integrate their DNA in the bacterial chromosome and become prophages. Reproduction is then ensured through vertical transmission, and upon induction, prophages can also enter the lytic cycle [265] (Fig. 3). In general, factors that induce the lytic phase are compounds or conditions with bactericidal effects, e.g. a DNA damaging-agent [266].

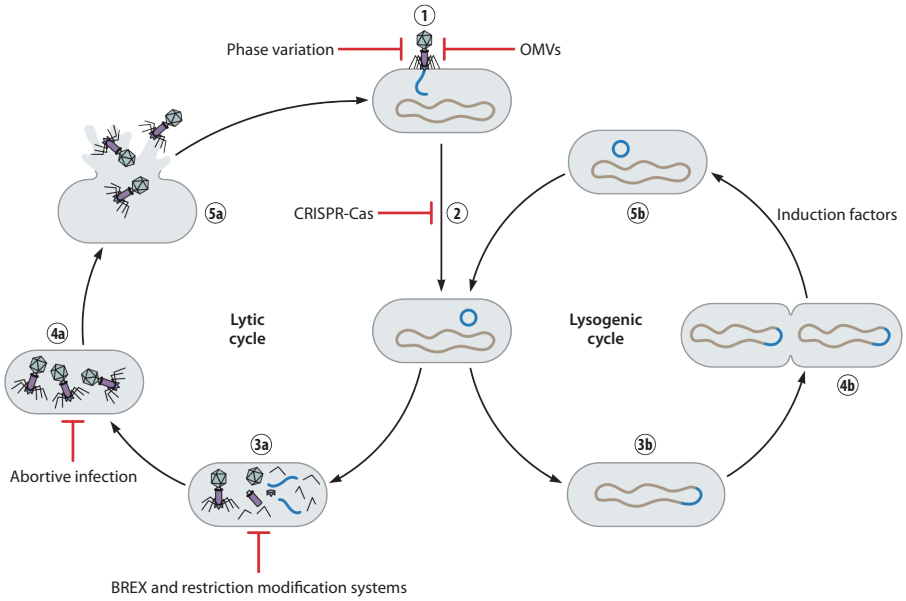


Figure 3: Lytic and lysogenic bacteriophage infection cycle with bacterial defense mechanisms. The first two steps (1 and 2) of infection are identical for the lytic and lysogenic cycle, namely phage binding followed by DNA insertion and DNA circularization. The lysogenic cycle then branches off by integrating its DNA into the bacterial chromosome and becoming prophage, thereby ensuring its replication (3b). Only upon encountering induction factors will the prophage leave the bacterial chromosome, after which it can enter the lytic cycle (4b and 5b). In the lytic cycle, phage DNA and protein is replicated and subsequently assembled into full phages (3a and 4a). The phages then lyse the bacterial cell, are released and can infect other bacteria (5a).

Bacteria possess multiple mechanisms to prevent killing by bacteriophages, starting with blocking attachment. This can be achieved through phase variation or production of OMVs. After phage DNA entry, CRISPR-Cas can recognize this foreign DNA and degrade it. Phage DNA and protein replication can be prevented by BREX and restriction modification systems, while full phage assembly can be prevented by abortive infection.

The first step for preventing bacteriophage infection is to prevent surface receptor recognition. Outer membrane vesicles are produced by Gram-negative bacteria and have several functions, including interbacterial communication [267]. They have highly similar surface composition as the bacterium and may thereby serve as decoys for attacking phages [268] (Fig. 3). Indeed, *V. cholerae* outer membrane vesicles were shown to neutralize a *V. cholerae* specific phage in a dose-dependent manner (Fig. 2) [268]. This effect was only seen when the O1 antigen, the bacteriophage target on *V. cholerae*, was included in the outer membrane vesicle structure [268].

V. cholerae possesses another mechanism to prevent O1 phage receptor recognition [269] (Fig. 3). Two genes necessary for O1 biosynthesis were shown to use phase variation to induce variation in the O1 antigen composition [269]. Mutants using phase variation were resistant to the O1 antigen phage, but displayed impaired colonization in a mouse model [269]. As the O1 antigen is an important virulence factor, e.g. for immune evasion, this demonstrates that enteric pathogens constantly have to deal with multiple CR mechanisms [269].

The second step in phage infection is injection of its DNA, and this can be prevented by superinfection exclusion systems which are mostly coded by prophages (Fig. 3). The *E. coli* prophage HK97 encodes for gp15, a probable inner transmembrane protein [270]. Remarkably, HK97 gp15 has putative homologues resembling the YebO protein family in many *Enterobacteriaceae* [270]. GP15 prevented DNA injection into the bacterial cytoplasm by preventing proper formation of a complex consisting of an inner membrane glucose transporter and part of the tape measure protein [270, 271]. This example illustrates how bacteria can incorporate phage DNA to prevent itself against future phage attacks.

DNA replication can be prevented by restriction-modification systems (Fig. 3). These systems consist of a methyltransferase and a restriction endonuclease. Exogenous DNA is not tagged by this methyltransferase, while ‘self’ DNA does get tagged [272, 273]. Subsequently, non-tagged DNA can be cleaved. This system is viewed as a primitive innate bacterial defense system. However, it was found that this system is not perfect, as these restriction-modification systems can also attack self-DNA [274].

Currently, many groups are actively investigating the adaptive bacterial immune system

CRISPR-Cas and this has been extensively reviewed elsewhere [275, 276]. CRISPR-Cas is present in about 45% of sequenced bacterial genomes, although it is unknown if its prevalence is similar in gut bacteria [277, 278]. In short, it consists of CRISPR arrays, sets of short repetitive DNA elements with variable DNA sequences (spacers) separating the repetitive DNA sets, and of an operon of CRISPR associated genes (Cas). Spacers are pieces of foreign DNA, derived from bacteriophage DNA or other mobile genetic elements such as plasmids. The defense mechanism consists of adaptation followed by expression and interference. During adaptation, Cas proteins can recognize foreign phage DNA and integrate a piece of this DNA as a new spacer into the CRISPR array. This allows the bacterium to build an immunological memory of all phages it previously encountered. The expression response entails transcription of the CRISPR array, followed by processing into smaller RNA pieces (crRNAs). CrRNAs consist of two outer parts of repeated DNA sequences, with a spacer in between. To form the eventual Cas-crRNA complex, crRNAs are combined with at least one Cas protein. This complex then travels through the bacterial cell and when it identifies a complementary DNA sequence, representative for the previously encountered bacteriophage, it cleaves and degrades this foreign DNA.

In 2015, a novel phage resistance system was discovered, called bacteriophage exclusion (BREX) [279]. BREX is able to block DNA replication, but does not prevent bacteriophage attachment to the bacterium (Fig. 3). It also uses methylation as guidance to identify self and exogenous DNA, but is different from restriction-modification systems as it does not cleave exogenous DNA [279]. Almost 10% of all bacterial genomes sequenced were found to have this BREX, suggesting that it is quite a conserved defense mechanism against bacteriophages [279]. In spite of this promising defense mechanism, no further papers have been released regarding BREX functioning in e.g. pathogenic bacteria.

Bacterial cells can perform an apoptosis-like action called abortive infection, resulting in death of the infected cell and hereby protecting surrounding bacterial cells [280] (Fig. 3). These systems have not been much elucidated for enteric pathogens at a molecular level, though, relevance of this system has been shown for the gut bacteria *S. dysenteriae* and *E. coli* [281, 282]. The abortive infection systems are best studied in *L. lactis*, a bacterium widely used in production of fermented foods [283].

Concluding remarks

Currently, bacterial enteric infections still cause a heavy disease burden worldwide. For many bacterial pathogens, the virulence factors involved in infection are understood, but less is known concerning the failure of gut microbiota to provide colonization resistance

against these enteropathogens. A more comprehensive understanding of why the microbiota fail to confer sufficient CR could lead to development of specific therapies aiming to restore CR. It is likely that not a single bacterium will be used as the 'holy grail' to restore CR, but that bacterial consortia with complementary functions will be used instead. This would be preferable over the currently often used FMT, where it is not well known what exact components are transferred to the patient. One could imagine that these consortia could not only be used to treat existing infections, but that they could also be administered prophylactically in susceptible patient groups. In addition, more attention has recently been given to several drugs that were previously not linked to gut health for their potentially disturbing effect on gut microbiota and perhaps CR. In conclusion, we reviewed many of the latest insights in the rapidly evolving fields of gut microbiota, colonization resistance and bacterial enteric infection. We are looking forward to the coming years, where undoubtedly more knowledge will be gained on gut microbiota and CR, ultimately leading to more microbiota-based therapies.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. B.V.H.H. and E.J.K. are supported by an unrestricted grant from Vedanta Biosciences Inc. W.v.S. is supported by a Royal Society Wolfson Research Merit Award (WM160092). V.B.Y. is supported by grant AI124255 from the National Institutes of Health (United States) and is a consultant for Vedanta Biosciences Inc.

Q.R.D., R.D.Z., B.V.H.H. and E.J.K. designed the structure and content of the review and performed literature research. Q.R.D. wrote the manuscript with guidance from R.D.Z and B.V.H.H. The manuscript was critically reviewed and revised by all authors (Q.R.D., R.D.Z., B.V.H.H., W.v.S., V.B.Y., E.J.K). All authors (Q.R.D., R.D.Z., B.V.H.H., W.v.S., V.B.Y., E.J.K) read and approved the final manuscript.

References

1. Sender R, Fuchs S, Milo R. 2016. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol 14:e1002533.
2. Sommer F, Backhed F. 2013. The gut microbiota--masters of host development and physiology. Nat Rev Microbiol 11:227-238.
3. Round JL, Palm NW. 2018. Causal effects of the microbiota on immune-mediated diseases. Sci Immunol 3:eaao1603.
4. Opazo MC, Ortega-Rocha EM, Coronado-Arrazola I, Bonifaz LC, Boudin H, Neunlist M, Bueno SM, Kalergis AM, Riedel CA. 2018. Intestinal microbiota influences non-intestinal related autoimmune diseases. Front Microbiol 9:432.
5. Li Q, Gao Z, Wang H, Wu H, Liu Y, Yang Y, Han L, Wang X, Zhao L, Tong X. 2018. Intestinal immunomodulatory cells (t lymphocytes): A bridge between gut microbiota and diabetes. Mediators Inflamm 2018:9830939.
6. Sadowsky MJ, Staley C, Heiner C, Hall R, Kelly CR, Brandt L, Khoruts A. 2017. Analysis of gut microbiota - An ever changing landscape. Gut Microbes 8:268-275.
7. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal flora. Science 308:1635-1638.
8. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD. 2019. A new genomic blueprint of the human gut microbiota. Nature.
9. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. 2009. A core gut microbiome in obese and lean twins. Nature 457:480-484.
10. Vollaard EJ, Clasener HA. 1994. Colonization resistance. Antimicrob Agents Chemother 38:409-414.
11. Becattini S, Littmann ER, Carter RA, Kim SG, Morjaria SM, Ling L, Gyaltshen Y, Fontana E, Taur Y, Leiner IM, Pamer EG. 2017. Commensal microbes provide first line defense against *Listeria monocytogenes* infection. J Exp Med 214:1973-1989.
12. Martens EC, Neumann M, Desai MS. 2018. Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. Nat Rev Microbiol 16:457-470.
13. Pickard JM, Zeng MY, Caruso R, Nunez G. 2017. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunol Rev 279:70-89.
14. Louis P, Flint HJ. 2017. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol 19:29-41.
15. Rios-Covian D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilan CG, Salazar N. 2016. Intestinal short chain fatty acids and their link with diet and human health. Front Microbiol 7:185.
16. Litvak Y, Byndloss MX, Baumler AJ. 2018. Colonocyte metabolism shapes the gut microbiota. Science 362.
17. Cummings JH, Pomare EW, Branch WJ, Naylor CPE, Macfarlane GT. 1987. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut 28:1221-1227.
18. Repaske DR, Adler J. 1981. Change in intracellular pH of *Escherichia coli* mediates the chemotactic response to certain attractants and repellents. J Bacteriol 145:1198-1208.
19. Roe AJ, O'Byrne C, McLaggan D, Booth IR. 2002. Inhibition of *Escherichia coli* growth by acetic acid a problem with methione byosynthesis and homocysteine toxicity. Microbiology 148:2215-2222.
20. Hofmann AF. 1999. The continuing importance of bile acids in liver and intestinal disease. Arch Intern Med 159:2647-2658.
21. Ridlon JM, Kang DJ, Hylemon PB. 2006. Bile salt biotransformations by human intestinal bacteria. J Lipid Res 47:241-259.
22. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MR, Jenq RR, Taur Y, Sander C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. Nature 517:205-208.
23. Hirano S, Nakam R, Tamaki M, Masuda N, Oda H. 1981. Isolation and characterization of thirteen intestinal microorganisms capable of 7 alpha-dehydroxylating bile acids. Appl Environ Microbiol 41:737-745.
24. Kang DJ, Ridlon JM, Moore DR, 2nd, Barnes S, Hylemon PB. 2008. *Clostridium scindens* baiCD and baiH genes encode stereo-specific 7alpha/7beta-hydroxy-3-oxo-delta4-cholenoic acid oxidoreductases. Biochim Biophys Acta 1781:16-25.
25. Sannasiddappa TH, Lund PA, Clarke SR. 2017. In vitro antibacterial activity of unconjugated and conjugated

bile salts on *Staphylococcus aureus*. Front Microbiol 8:1581.
26. Kurdi P, Kawanishi K, Mizutani K, Yokota A. 2006. Mechanism of growth inhibition by free bile acids in lactobacilli and bifidobacteria. J Bacteriol 188:1979-1986.
27. Watanabe M, Fukiya S, Yokota A. 2017. Comprehensive evaluation of the bactericidal activities of free bile acids in the large intestine of humans and rodents. J Lipid Res 58:1143-1152.
28. Thanissery R, Winston JA, Theriot CM. 2017. Inhibition of spore germination, growth, and toxin activity of clinically relevant *C. difficile* strains by gut microbiota derived secondary bile acids. Anaerobe 45:86-100.
29. Sayin SI, Wahlstrom A, Felin J, Jantti S, Marschall HU, Bamberg K, Angelin B, Hyotylainen T, Oresic M, Backhed F. 2013. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab 17:225-235.
30. Wells EJ, Berr F, Thomas LA, Hermon Dowling R, Dylemon PB. 2000. Isolation and characterization of cholic acid 7α-dehydroxylating fecal bacteria from cholesterol gallstone patients. J Hepatol 32:4-10.
31. Sorg JA, Sonenshein AL. 2010. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol 192:4983-4990.
32. Studer N, Desharnais L, Beutler M, Brugiroux S, Terrazos MA, Menin L, Schurch CM, McCoy KD, Kuehne SA, Minton NP, Stecher B, Bernier-Latmani R, Hapfelmeier S. 2016. Functional intestinal bile acid 7α-dehydroxylation by *Clostridium scindens* associated with protection from *Clostridium difficile* infection in a gnotobiotic mouse model. Front Cell Infect Microbiol 6:191.
33. Amrane S, Bachar D, Lagier JC, Raoult D. 2018. *Clostridium scindens* is present in the gut microbiota during *Clostridium difficile* infection: a metagenomic and culturomic analysis. J Clin Microbiol 56:e01663-17.
34. Termen S, Tollin M, Rodriguez E, Sveinsdottir SH, Johannesson B, Cederlund A, Sjovall J, Agerberth B, Gudmundsson GH. 2008. PU.1 and bacterial metabolites regulate the human gene CAMP encoding antimicrobial peptide LL-37 in colon epithelial cells. Mol Immunol 45:3947-3955.
35. Peric M, Koglin S, Dombrowski Y, Gross K, Bradac E, Ruzicka T, Schauber J. 2009. VDR and MEK-ERK dependent induction of the antimicrobial peptide cathelicidin in keratinocytes by lithocholic acid. Mol Immunol 46:3183-3187.
36. Rea MC, Sit CS, Clayton E, O'Connor PM, Whittal RM, Zheng J, Vederas JC, Ross RP, Hill C. 2010. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against *Clostridium difficile*. Proc Natl Acad Sci U S A 107:9352-9357.
37. Parks WM, Bottrill AR, Pierrat OA, Durrant MC, Maxwell A. 2007. The action of the bacterial toxin, microcin B17, on DNA gyrase. Biochimie 89:500-507.
38. Destoumieux-Garzon D, Peduzzi J, Thomas X, Djediat C, Rebuffat S. 2006. Parasitism of iron-siderophore receptors of *Escherichia coli* by the siderophore-peptide microcin E492m and its unmodified counterpart. Biometals 19:181-191.
39. Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH. 2004. Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. Mol Cell 14:739-751.
40. Cotter PD, Ross RP, Hill C. 2013. Bacteriocins - a viable alternative to antibiotics? Nat Rev Microbiol 11:95-105.
41. Rebuffat S. 2011. Bacteriocins from Gram-negative bacteria: A classification?, p 55-72. *In* Drider D, Rebuffat S (ed), Prokaryotic Antimicrobial Peptides: From Genes to Applications. Springer New York, New York, NY.
42. Rea MC, Ross RP, Cotter PD, Hill C. 2011. Classification of bacteriocins from Gram-positive bacteria, p 29-53. *In* Drider D, Rebuffat S (ed), Prokaryotic Antimicrobial Peptides: From Genes to Applications. Springer New York, New York, NY.
43. Severina E, Severin A, Tomasz A. 1998. Antibacterial efficacy of nisin against multidrug-resistant Gram-positive pathogens. J Antimicrob Chemother 41:341-347.
44. Le Lay C, Dridi L, Bergeron MG, Ouellette M, Fliss IL. 2016. Nisin is an effective inhibitor of *Clostridium difficile* vegetative cells and spore germination. J Med Microbiol 65:169-175.
45. Benkerroum N, Sandine WE. 1988. Inhibitory action of nisin against *Listeria monocytogenes*. J Dairy Sci 71:3237-3245.
46. Stevens KA, Sheldon BW, Klapes NA, Klaenhammer TR. 1991. Nisin Treatment for Inactivation of *Salmonella* Species and other gram-negative bacteria. Appl Environ Microbiol 57:3613-3615.
47. Prudencio CV, Mantovani HC, Cecon PR, Vanetti MC. 2015. Differences in the antibacterial activity of nisin and bovicin HC5 against *Salmonella* Typhimurium under different temperature and pH conditions. J Appl Microbiol 118:18-26.
48. Singh AP, Prabha V, Rishi P. 2013. Value addition in the efficacy of conventional antibiotics by nisin against *Salmonella*. PLoS One 8:e76844.

49. Singh AP, Preet S, Rishi P. 2014. Nisin/beta-lactam adjunct therapy against *Salmonella enterica* serovar Typhimurium: a mechanistic approach. J Antimicrob Chemother 69:1877-1887.
50. Li Q, Montalban-Lopez M, Kuipers OP. 2018. Increasing the antimicrobial activity of nisin-based lantibiotics against gram-negative pathogens. Appl Environ Microbiol 84:e00052-18.
51. Ritchie JM, Greenwich JL, Davis BM, Bronson RT, Gebhart D, Williams SR, Martin D, Scholl D, Waldor MK. 2011. An *Escherichia coli* O157-specific engineered pyocin prevents and ameliorates infection by *E. coli* O157:H7 in an animal model of diarrheal disease. Antimicrob Agents Chemother 55:5469-5474.
52. Gillor O, Nigro LM, Riley MA. 2005. Genetically engineered bacteriocins and their potential as the next generation of antimicrobials. Curr Pharm Des 11:1067-1075.
53. Hasper HE, Kramer NE, Smith JL, Hilman JD, Zachariah C, Kuipers OP, de Kruijff B, Breukink E. 2006. An alternative bactericidal mechanism of action for lantibiotic peptides that target Lipid II. Science 313:1636-1637.
54. Wiedemann I, Breukink E, van Kraaij C, Kuipers OP, Bierbaum G, de Kruijff B, Sahl HG. 2001. Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity. J Biol Chem 276:1772-1779.
55. Corr SC, Riedel CU, O'Toole PW, Hill C, Gahan CGM. 2007. Bacteriocin production as a mechanism for the antiinfective activity of *Lactobacillus salivarius* UCC118. Proc Natl Acad Sci U S A 104:7617-7621.
56. Rea MC, Dobson A, O'Sullivan O, Crispie F, Fouhy F, Cotter PD, Shanahan F, Kiely B, Hill C, Ross RP. 2011. Effect of broad- and narrow-spectrum antimicrobials on *Clostridium difficile* and microbial diversity in a model of the distal colon. Proc Natl Acad Sci U S A 108 Suppl 1:4639-4644.
57. Rea MC, Alemayehu D, Casey PG, O'Connor PM, Lawlor PG, Walsh M, Shanahan F, Kiely B, Ross RP, Hill C. 2014. Bioavailability of the anti-clostridial bacteriocin thuricin CD in gastrointestinal tract. Microbiology 160:439-445.
58. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, Pudlo NA, Kitamoto S, Terrapon N, Muller A, Young VB, Henrissat B, Wilmes P, Stappenbeck TS, Nunez G, Martens EC. 2016. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell 167:1339-1353 e21.
59. Bosak J, Micenkova L, Hrala M, Pomorska K, Kunova Bosakova M, Krejci P, Gopfert E, Faldyna M, Smajs D. 2018. Colicin FY inhibits pathogenic *Yersinia enterocolitica* in mice. Sci Rep 8:12242.
60. Gillor O, Kirkup BC, Riley MA. 2004. Colicins and microcins: the next generation antimicrobials. Adv Appl Microbiol 54:129-146.
61. Sassone-Corsi M, Nuccio SP, Liu H, Hernandez D, Vu CT, Takahashi AA, Edwards RA, Raffatellu M. 2016. Microcins mediate competition among *Enterobacteriaceae* in the inflamed gut. Nature 540:280-283.
62. Rebuffat S. 2012. Microcins in action: amazing defence strategies of Enterobacteria. Biochem Soc Trans 40:1456-1462.
63. Walsh CJ, Guinane CM, Hill C, Ross RP, O'Toole PW, Cotter PD. 2015. In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. BMC Microbiol 15:183.
64. Momose Y, Hirayama K, Itoh K. 2008. 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 94:165-171.
65. Fabich AJ, Jones SA, Chowdhury FZ, Cernosek A, Anderson A, Smalley D, McHargue JW, Hightower GA, Smith JT, Autieri SM, Leatham MP, Lins JJ, Allen RL, Laux DC, Cohen PS, Conway T. 2008. Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. Infect Immun 76:1143-1152.
66. Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. 2013. Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. PLoS One 8:e53957.
67. Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ, Contreras H, Libby SJ, Fang FC, Raffatellu M. 2013. Probiotic bacteria reduce *Salmonella* Typhimurium intestinal colonization by competing for iron. Cell Host Microbe 14:26-37.
68. Borenstein E, Kupiec M, Feldman MW, Ruppin E. 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci U S A 105:14482-14487.
69. Peterson LW, Artis D. 2014. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. Nat Rev Immunol 14:141-153.
70. Kashyap PC, Marcobal A, Ursell LK, Smits SA, Sonnenburg ED, Costello EK, Higginbottom SK, Domino

SE, Holmes SP, Relman DA, Knight R, Gordon JI, Sonnenburg JL. 2013. Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. Proc Natl Acad Sci U S A 110:17059-64.
71. Johansson ME, Larsson JM, Hansson GC. 2011. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci U S A 108 Suppl 1:4659-4665.
72. Atuma C, Strugala V, Allen A, Holm L. 2001. The adherent gastrointestinal mucus gel layer thickness and physical state in vivo. Am J Physiol Gastrointest Liver Physiol 280:G922-G929.
73. Schroeder BO, Birchenough GMH, Stahlman M, Arike L, Johansson MEV, Hansson GC, Backhed F. 2018. Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. Cell Host Microbe 23:27-40 e7.
74. Jakobsson HE, Rodriguez-Pineiro AM, Schutte A, Ermund A, Boysen P, Bemark M, Sommer F, Backhed F, Hansson GC, Johansson ME. 2015. The composition of the gut microbiota shapes the colon mucus barrier. EMBO Rep 16:164-177.
75. Keen EC. 2015. A century of phage research: bacteriophages and the shaping of modern biology. Bioessays 37:6-9.
76. Mirzaei MK, Maurice CF. 2017. Ménage à trois in the human gut: interactions between host, bacteria and phages. Nat Rev Microbiol 15:397-408.
77. Wittebole X, De Roock S, Opal SM. 2014. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence 5:226-235.
78. Yen M, Cairns LS, Camilli A. 2017. A cocktail of three virulent bacteriophages prevents *Vibrio cholerae* infection in animal models. Nat Commun 8:14187.
79. Fischer S, Kittler S, Klein G, Glunder G. 2013. Impact of a single phage and a phage cocktail application in broilers on reduction of *Campylobacter jejuni* and development of resistance. PLoS One 8:e78543.
80. Duerkop BA, Clements CV, Rollins D, Rodrigues JL, Hooper LV. 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proc Natl Acad Sci U S A 109:17621-17626.
81. Henein A. 2013. What are the limitations on the wider therapeutic use of phage? Bacteriophage 3:e24872.
82. Schooley RT, Biswas B, Gill JJ, Hernandez-Morales A, Lancaster J, Lessor L, Barr JJ, Reed SL, Rohwer F, Benler S, Segall AM, Taplitz R, Smith DM, Kerr K, Kumaraswamy M, Nizet V, Lin L, McCauley MD, Strathdee SA, Benson CA, Pope RK, Leroux BM, Picel AC, Mateczun AJ, Cilwa KE, Regeimbal JM, Estrella LA, Wolfe DM, Henry MS, Quinones J, Salka S, Bishop-Lilly KA, Young R, Hamilton T. 2017. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant *Acinetobacter baumannii* infection. Antimicrob Agents Chemother 61:e00954-17.
83. Chan BK, Turner PE, Kim S, Mojibian HR, Elefteriades JA, Narayan D. 2018. Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa*. Evol Med Public Health 2018:60-66.
84. Keeney KM, Yurist-Doutsch S, Arrieta MC, Finlay BB. 2014. Effects of antibiotics on human microbiota and subsequent disease. Annu Rev Microbiol 68:217-235.
85. Becattini S, Taur Y, Pamer EG. 2016. Antibiotic-induced changes in the intestinal microbiota and disease. Trends Mol Med 22:458-478.
86. Forgacs I, Loganayagam A. 2008. Overprescribing proton pump inhibitors. BMJ 336:2-3.
87. Leonard J, Marshall JK, Moayyedi P. 2007. Systematic review of the risk of enteric infection in patients taking acid suppression. Am J Gastroenterol 102:2047-2056.
88. Wei L, Ratnayake L, Phillips G, McGuigan CC, Morant SV, Flynn RW, Mackenzie IS, MacDonald TM. 2017. Acid-suppression medications and bacterial gastroenteritis: a population-based cohort study. Br J Clin Pharmacol 83:1298-1308.
89. Hafiz RA, Wong C, Paynter S, David M, Peeters G. 2018. The risk of community-acquired enteric infection in proton pump inhibitor therapy: systematic review and meta-analysis. Ann Pharmacother 52:613-622.
90. Bavishi C, Dupont HL. 2011. Systematic review: the use of proton pump inhibitors and increased susceptibility to enteric infection. Aliment Pharmacol Ther 34:1269-1281.
91. Freedberg DE, Toussaint NC, Chen SP, Ratner AJ, Whittier S, Wang TC, Wang HH, Abrams JA. 2015. Proton pump inhibitors alter specific taxa in the human gastrointestinal microbiome: a crossover trial. Gastroenterology 149:883-5 e9.
92. Imhann F, Bonder MJ, Vich Vila A, Fu J, Mujagic Z, Vork L, Tigchelaar EF, Jankipersadsing SA, Cenit MC, Harmsen HJ, Dijkstra G, Franke L, Xavier RJ, Jonkers D, Wijmenga C, Weersma RK, Zhernakova A. 2016. Proton pump inhibitors affect the gut microbiome. Gut 65:740-748.
93. Jackson MA, Goodrich JK, Maxan ME, Freedberg DE, Abrams JA, Poole AC, Sutter JL, Welter D, Ley RE,

Bell JT, Spector TD, Steves CJ. 2016. Proton pump inhibitors alter the composition of the gut microbiota. Gut 65:749-756.

94. Takagi T, Naito Y, Inoue R, Kashiwagi S, Uchiyama K, Mizushima K, Tsuchiya S, Okayaama T, Dohi O, Yoshida N, Kamada K, Ishikawa T, Handa O, Konishi H, Okuda K, Tsujimoto Y, Ohnogi H, Itoh Y. 2018. The influence of long-term use of proton pump inhibitors on the gut microbiota: an age-sex-matched case-control study. J Clin Biochem Nutr 62:100-105.
95. Sieczkowska A, Landowski P, Gibas A, Kaminska B, Lifschitz C. 2018. Long-term proton pump inhibitor therapy leads to small bowel bacterial overgrowth as determined by breath hydrogen and methane excretion. J Breath Res 12:036006.
96. Yasutomi E, Hoshi N, Adachi S, Otsuka T, Kong L, Ku Y, Yamairi H, Inoue J, Ishida T, Watanabe D, Ooi M, Yoshida M, Tsukimi T, Fukuda S, Azuma T. 2018. Proton pump inhibitors increase the susceptibility of mice to oral infection with enteropathogenic bacteria. Dig Dis Sci 63:881-889.
97. Stiefel U, Rao A, Pultz MJ, Jump RL, Aron DC, Donskey CJ. 2006. Suppression of gastric acid production by proton pump inhibitor treatment facilitates colonization of the large intestine by vancomycin-resistant *Enterococcus spp.* and *Klebsiella pneumoniae* in clindamycin-treated mice. Antimicrob Agents Chemother 50:3905-3907.
98. Imhann F, Vich Vila A, Bonder MJ, Lopez Manosalva AG, Koonen DPY, Fu J, Wijmenga C, Zhernakova A, Weersma RK. 2017. The influence of proton pump inhibitors and other commonly used medication on the gut microbiota. Gut Microbes 8:351-358.
99. Foretz M, Guigas B, Bertrand L, Pollak M, Viollet B. 2014. Metformin: from mechanisms of action to therapies. Cell Metab 20:953-966.
100. Smyth S, Heron A. 2005. Diabetes and obesity: the twin epidemics. Nature Medicine 12:75-80.
101. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Costea PI, Kultima JR, Li J, Jorgensen T, Levenez F, Dore J, Meta HITc, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528:262-266.
102. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55-60.
103. Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Manneras-Holm L, Stahlman M, Olsson LM, Serino M, Planas-Felix M, Xifra G, Mercader JM, Torrents D, Burcelin R, Ricart W, Perkins R, Fernandez-Real JM, Backhed F. 2017. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med 23:850-858.
104. Zhou HY, Zhu H, Yao XM, Qian JP, Yang J, Pan XD, Chen XD. 2017. Metformin regulates tight junction of intestinal epithelial cells via MLCK-MLC. Eur Rev Med Pharmacol Sci 21:5239-5246.
105. Deng J, Zeng L, Lai X, Li J, Liu L, Lin Q, Chen Y. 2018. Metformin protects against intestinal barrier dysfunction via AMPKalpha1-dependent inhibition of JNK signalling activation. J Cell Mol Med 22:546-557.
106. Xue Y, Zhang H, Sun X, Zhu MJ. 2016. Metformin improves ileal epithelial barrier function in interleukin-10 deficient mice. PLoS One 11:e0168670.
107. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, Brochado AR, Fernandez KC, Dose H, Mori H, Patil KR, Bork P, Typas A. 2018. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 555:623-628.
108. Morgan AP, Crowley JJ, Nonneman RJ, Quackenbush CR, Miller CN, Ryan AK, Bogue MA, Paredes SH, Yourstone S, Carroll IM, Kawula TH, Bower MA, Sartor RB, Sullivan PF. 2014. The antipsychotic olanzapine interacts with the gut microbiome to cause weight gain in mouse. PLoS One 9:e115225.
109. Lieberman LA, Higgins DE. 2009. A small-molecule screen identifies the antipsychotic drug pimozide as an inhibitor of *Listeria monocytogenes* infection. Antimicrob Agents Chemother 53:756-764.
110. Dastidar SG, Chaudhury A, Annadurai S, Roy S, Mookerjee M, Chakrabarty AN. 1995. In vitro and in vivo antimicrobial action of fluphenazine. J Chemother 7:201-206.
111. Jeyaseeli L, Dasgupta A, Dastidar SG, Molnar J, Amaral L. 2012. Evidence of significant synergism between antibiotics and the antipsychotic, antimicrobial drug flupenthixol. Eur J Clin Microbiol Infect Dis 31:1243-1250.
112. Cussotto S, Strain CR, Fouhy F, Strain RG, Peterson VL, Clarke G, Stanton C, Dinan TG, Cryan JF. 2018.

Differential effects of psychotropic drugs on microbiome composition and gastrointestinal function. Psychopharmacology (Berl).
113. Rogers MA, Greene MT, Young VB, Saint S, Langa KM, Kao JY, Aronoff DM. 2013. Depression, antidepressant medications, and risk of *Clostridium difficile* infection. BMC Med 11.
114. Munita JM, Arias CA. 2016. Mechanisms of antibiotic resistance. Microbiol Spectr 4:VMBF-0016-2015.
115. Davies J, Davies D. 2010. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74:417-433.
116. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC. 2015. Burden of *Clostridium difficile* infection in the United States. N Engl J Med 372:825-834.
117. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. 2016. *Clostridium difficile* infection. Nat Rev Dis Primers 2:16020.
118. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Pechine S, Wilcox MH, Kuijper EJ. 2018. Understanding *Clostridium difficile* colonization. Clin Microbiol Rev 31:e00021-17.
119. Keller JJ, Kuijper EJ. 2015. Treatment of recurrent and severe *Clostridium difficile* infection. Annu Rev Med 66:373-386.
120. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, Huffnagle GB, J ZL, Young VB. 2014. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat Commun 5:3114.
121. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R, Schreiber F, Brandt C, Deakin LJ, Pickard DJ, Duncan SH, Flint HJ, Clark TG, Parkhill J, Dougan G. 2012. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. PLoS Pathog 8:e1002995.
122. Seekatz AM, Theriot CM, Rao K, Chang YM, Freeman AE, Kao JY, Young VB. 2018. Restoration of short chain fatty acid and bile acid metabolism following fecal microbiota transplantation in patients with recurrent *Clostridium difficile* infection. Anaerobe 53:64-73.
123. Sorg JA, Sonenshein AL. 2008. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. J Bacteriol 190:2505-2512.
124. Sorg JA, Sonenshein AL. 2009. Chenodeoxycholate is an inhibitor of *Clostridium difficile* spore germination. J Bacteriol 191:1115-1117.
125. Wilson KH. 1983. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. J Clin Microbiol 18:1017-1019.
126. Lewis BB, Carter RA, Pamer EG. 2016. Bile acid sensitivity and in vivo virulence of clinical *Clostridium difficile* isolates. Anaerobe 41:32-36.
127. McBride SM, Sonenshein AL. 2011. Identification of a genetic locus responsible for antimicrobial peptide resistance in *Clostridium difficile*. Infect Immun 79:167-176.
128. McBride SM, Sonenshein AL. 2011. The dlt operon confers resistance to cationic antimicrobial peptides in *Clostridium difficile*. Microbiology 157:1457-1465.
129. Jenior ML, Leslie JL, Young VB, Schloss PD. 2017. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. mSystems 2:e00063-17.
130. Jenior ML, Leslie JL, Young VB, Schloss PD. 2018. *Clostridium difficile* alters the structure and metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization. mSphere 3:e00261-18.
131. Collins J, Robinson C, Danhof H, Knetsch CW, van Leeuwen HC, Lawley TD, Auchtung JM, Britton RA. 2018. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. Nature 553:291-294.
132. Collins J, Danhof H, Britton RA. 2018. The role of trehalose in the global spread of epidemic *Clostridium difficile*. Gut Microbes:1-6.
133. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM, Sonnenburg JL. 2013. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. Nature 502:96-99.
134. Ferreyra JA, Wu KJ, Hryckowian AJ, Bouley DM, Weimer BC, Sonnenburg JL. 2014. Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. Cell Host Microbe 16:770-777.
135. Zackular JP, Moore JL, Jordan AT, Juttukonda LJ, Noto MJ, Nicholson MR, Crews JD, Semler MW, Zhang Y, Ware LB, Washington MK, Chazin WJ, Caprioli RM, Skaar EP. 2016. Dietary zinc alters the microbiota and decreases resistance to *Clostridium difficile* infection. Nat Med 22:1330-1334.
136. Tasteyre A, Barc MC, Collignon A, Boureau H, Karjalainen T. 2001. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. Infect Immun 69:7937-7940.

137. McKee RW, Aleksanyan N, Garret EM, Tamayo R. 2018. Type IV pili promote *Clostridium difficile* adherence and persistence in a mouse model of infection. Infect Immun 86:e00943-17.
138. Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. 2015. Global burden of invasive nontyphoidal *Salmonella* disease, 2010. Emerg Infect Dis 21:941-949.
139. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. 2012. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. Lancet 379:2489-2499.
140. Coburn B, Li Y, Owen D, Vallance BA, Finlay BB. 2005. *Salmonella enterica* serovar Typhimurium pathogenicity island 2 is necessary for complete virulence in a mouse model of infectious enterocolitis. Infect Immun 73:3219-3227.
141. Kwon YM, Ricke SC. 1998. Induction of acid resistance of *Salmonella* Typhimurium by exposure to short-chain fatty acids. Appl Environ Microbiol 64:3458-3463.
142. Lawhon S, Maurer R, Suyemoto M, Altier C. 2002. Intestinal short□chain fatty acids alter *Salmonella* Typhimurium invasion gene expression and virulence through BarASirA. Mol Microbiol 46:1451-1464.
143. Bronner DN, Faber F, Olsan EE, Byndloss MX, Sayed NA, Xu G, Yoo W, Kim D, Ryu S, Lebrilla CB, Baumler AJ. 2018. Genetic ablation of butyrate utilization attenuates gastrointestinal *Salmonella* disease. Cell Host Microbe 23:266-273 e4.
144. Jacobson A, Lam L, Rajendram M, Tamburini F, Honeycutt J, Pham T, Van Treuren W, Pruss K, Stabler SR, Lugo K, Bouley DM, Vilches-Moure JG, Smith M, Sonnenburg JL, Bhatt AS, Huang KC, Monack D. 2018. A gut commensal-produced metabolite mediates colonization resistance to *Salmonella* infection. Cell Host Microbe 24:296-307 e7.
145. van Velkinburgh JC, Gunn JS. 1999. PhoP-PhoQ-regulated loci are required for enhanced bile resistance. Infect Immun 67:1614-1622.
146. Hernandez SB, Cota I, Ducret A, Aussel L, Casadesus J. 2012. Adaptation and preadaptation of *Salmonella enterica* to bile. PLoS Genet 8:e1002459.
147. Eade CR, Hung CC, Bullard B, Gonzalez-Escobedo G, Gunn JS, Altier C. 2016. Bile acids function synergistically to repress invasion gene expression in *Salmonella* by destabilizing the invasion regulator hilD. Infect Immun 84:2198-2208.
148. Gänzle MG, Hertel C, Hammes WP. 1999. Resistance of *Escherichia coli* and *Salmonella* against nisin and curvacin A. Int J Food Microbiol 48:37-50.
149. Gillis CC, Hughes ER, Spiga L, Winter MG, Zhu W, Furtado de Carvalho T, Chanin RB, Behrendt CL, Hooper LV, Santos RL, Winter SE. 2018. Dysbiosis-associated change in host metabolism generates lactate to support *Salmonella* growth. Cell Host Microbe 23:54-64 e6.
150. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, Chaffron S, Macpherson AJ, Buer J, Parkhill J, Dougan G, von Mering C, Hardt WD. 2007. *Salmonella enterica* serovar Typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol 5:2177-2189.
151. Rivera-Chavez F, Zhang LF, Faber F, Lopez CA, Byndloss MX, Olsan EE, Xu G, Velazquez EM, Lebrilla CB, Winter SE, Baumler AJ. 2016. Depletion of butyrate-producing *Clostridia* from the gut microbiota drives an aerobic luminal expansion of *Salmonella*. Cell Host Microbe 19:443-454.
152. Spiga L, Winter MG, Furtado de Carvalho T, Zhu W, Hughes ER, Gillis CC, Behrendt CL, Kim J, Chessa D, Andrews-Polymenis HL, Beiting DP, Santos RL, Hooper LV, Winter SE. 2017. An oxidative central metabolism enables *Salmonella* to utilize microbiota-derived succinate. Cell Host Microbe 22:291-301 e6.
153. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, Sterzenbach T, Tsolis RM, Roth JR, Baumler AJ. 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. Proc Natl Acad Sci U S A 108:17480-17485.
154. Faber F, Thiennimitr P, Spiga L, Byndloss MX, Litvak Y, Lawhon S, Andrews-Polymenis HL, Winter SE, Baumler AJ. 2017. Respiration of microbiota-derived 1,2-propanediol drives *Salmonella* expansion during colitis. PLoS Pathog 13:e1006129.
155. Faber F, Tran L, Byndloss MX, Lopez CA, Velazquez EM, Kerrinnes T, Nuccio SP, Wangdi T, Fiehn O, Tsolis RM, Baumler AJ. 2016. Host-mediated sugar oxidation promotes post-antibiotic pathogen expansion. Nature 534:697-699.
156. Zarepour M, Bhullar K, Montero M, Ma C, Huang T, Velcich A, Xia L, Vallance BA. 2013. The mucin Muc2 limits pathogen burdens and epithelial barrier dysfunction during *Salmonella enterica* serovar Typhimurium colitis. Infect Immun 81:3672-3683.
157. Ganesh BP, Klopfleisch R, Loh G, Blaut M. 2013. Commensal *Akkermansia muciniphila* exacerbates gut inflammation in *Salmonella* Typhimurium-infected gnotobiotic mice. PLoS One 8:e74963.
158. Eppinger M, Cebula TA. 2015. Future perspectives, applications and challenges of genomic epidemiology

studies for food-borne pathogens: A case study of enterohemorrhagic *Escherichia coli* (EHEC) of the O157:H7 serotype. Gut Microbes 6:194-201.
159. Pai CH, Ahmed N, Lior H, Johnson WM, Sims HV, Woods DE. 1988. Epidemiology of sporadic diarrhea due to verocytotoxin-producing *Escherichia coli*: a two-year prospective study. J Infect Dis 157:1054-1057.
160. Carroll KC, Pfaller MA, Landry ML, McAdam AJ, Patel R, Richter SS, Warnock DW. 2019. Manual of Clinical Microbiology, 12th ed, vol 1. ASM Press, Washington D.C.
161. Xiong H, Guo B, Gan Z, Song D, Lu Z, Yi H, Wu Y, Wang Y, Du H. 2016. Butyrate upregulates endogenous host defense peptides to enhance disease resistance in piglets via histone deacetylase inhibition. Sci Rep 6:27070.
162. Nakanishi N, Tashiro K, Kuhara S, Hayashi T, Sugimoto N, Tobe T. 2009. Regulation of virulence by butyrate sensing in enterohaemorrhagic *Escherichia coli*. Microbiology 155:521-530.
163. Takao M, Yen H, Tobe T. 2014. LeuO enhances butyrate-induced virulence expression through a positive regulatory loop in enterohaemorrhagic *Escherichia coli*. Mol Microbiol 93:1302-1313.
164. Zumbrun SD, Melton-Celsa AR, Smith MA, Gilbreath JJ, Merrell DS, O'Brien AD. 2013. Dietary choice affects Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 colonization and disease. Proc Natl Acad Sci U S A 110:E2126-2133.
165. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. Nature 469:543-547.
166. Kus JV, Gebremedhin A, Dang V, Tran SL, Serbanescu A, Barnett Foster D. 2011. Bile salts induce resistance to polymyxin in enterohemorrhagic *Escherichia coli* O157:H7. J Bacteriol 193:4509-4515.
167. Schamberger GP, Diez-Gonzalez F. 2005. Assessment of resistance to colicinogenic *Escherichia coli* by *E. coli* O157:H7 strains. J Appl Microbiol 98:245-252.
168. Bertin Y, Girardeau JP, Chaucheyras-Durand F, Lyan B, Pujos-Guillot E, Harel J, Martin C. 2011. Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. Environ Microbiol 13:365-377.
169. Bertin Y, Deval C, de la Foye A, Masson L, Gannon V, Harel J, Martin C, Desvaux M, Forano E. 2014. The gluconeogenesis pathway is involved in maintenance of enterohaemorrhagic *Escherichia coli* O157:H7 in bovine intestinal content. PLoS One 9:e98367.
170. Miranda RL, Conway T, Leatham MP, Chang DE, Norris WE, Allen JH, Stevenson SJ, Laux DC, Cohen PS. 2004. Glycolytic and Gluconeogenic Growth of Escherichia coli O157:H7 (EDL933) and E. coli K-12 (MG1655) in the Mouse Intestine. Infection and Immunity 72:1666-1676.
171. Curtis MM, Hu Z, Klimko C, Narayanan S, Deberardinis R, Sperandio V. 2014. The gut commensal *Bacteroides thetaiotaomicron* exacerbates enteric infection through modification of the metabolic landscape. Cell Host Microbe 16:759-769.
172. Pacheco AR, Curtis MM, Ritchie JM, Munera D, Waldor MK, Moreira CG, Sperandio V. 2012. Fucose sensing regulates bacterial intestinal colonization. Nature 492:113-7.
173. In J, Foulke-Abel J, Zachos NC, Hansen AM, Kaper JB, Bernstein HD, Halushka M, Blutt S, Estes MK, Donowitz M, Kovbasnjuk O. 2016. Enterohemorrhagic *Escherichia coli* reduce mucus and intermicrovillar bridges in human stem cell-derived colonoids. Cell Mol Gastroenterol Hepatol 2:48-62 e3.
174. Hews CL, Tran SL, Wegmann U, Brett B, Walsham ADS, Kavanaugh D, Ward NJ, Juge N, Schuller S. 2017. The StcE metalloprotease of enterohaemorrhagic *Escherichia coli* reduces the inner mucus layer and promotes adherence to human colonic epithelium ex vivo. Cell Microbiol 19:e12717.
175. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM. 1999. Global burden of *Shigella* infections implications for vaccine development. Bull World Health Organ 77:651-666.
176. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM. 2018. Shigellosis. The Lancet 391:801-812.
177. Raqib R, Sarker P, Mily A, Alam NH, Arifuzzaman ASM, Rekha RS, Andersson J, Gudmundsson GH, Cravioto A, Agerberth B. 2012. Efficacy of sodium butyrate adjunct therapy in shigellosis a randomized, double-blind, placebo-controlled clinical trial. BMC Infect Dis 12:111.
178. Sperandio B, Regnault B, Guo J, Zhang Z, Stanley SL, Jr., Sansonetti PJ, Pedron T. 2008. Virulent *Shigella flexneri* subverts the host innate immune response through manipulation of antimicrobial peptide gene expression. J Exp Med 205:1121-1132.
179. Olive AJ, Kenjale R, Espina M, Moore DS, Picking WL, Picking WD. 2007. Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. Infect Immun 75:2626-2629.
180. Stensrud KF, Adam PR, La Mar CD, Olive AJ, Lushington GH, Sudharsan R, Shelton NL, Givens RS, Picking

WL, Picking WD. 2008. Deoxycholate interacts with IpaD of *Shigella flexneri* in inducing the recruitment of IpaB to the type III secretion apparatus needle tip. J Biol Chem 283:18646-18654.
181. Brotcke Zumsteg A, Goosmann C, Brinkmann V, Morona R, Zychlinsky A. 2014. IcsA is a *Shigella flexneri* adhesin regulated by the type III secretion system and required for pathogenesis. Cell Host Microbe 15:435-445.
182. Nickerson KP, Chanin RB, Sistrunk JR, Rasko DA, Fink PJ, Barry EM, Nataro JP, Faherty CS. 2017. Analysis of *Shigella flexneri* resistance, biofilm formation, and transcriptional profile in response to bile salts. Infect Immun 85:e01067-16.
183. Islam D, Bandholtz L, Nilsson J, Wigzell H, Christensson B, Agerberth B, Gudmundsson GH. 2001. Downregulation of bactericidal peptides in enteric infections a novel immune escape mechanism. Nat Med 7:180-185.
184. Kentner D, Martano G, Callon M, Chiquet P, Brodmann M, Burton O, Wahlander A, Nanni P, Delmotte N, Grossmann J, Limenitakis J, Schlapbach R, Kiefer P, Vorholt JA, Hiller S, Bumann D. 2014. *Shigella* reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. Proc Natl Acad Sci U S A 111:9929-9934.
185. Waligora EA, Fisher CR, Hanovice NJ, Rodou A, Wyckoff EE, Payne SM. 2014. Role of intracellular carbon metabolism pathways in *Shigella flexneri* virulence. Infect Immun 82:2746-2755.
186. Sperandio B, Fischer N, Joncquel Chevalier-Curt M, Rossez Y, Roux P, Robbe Masselot C, Sansonetti PJ. 2013. Virulent *Shigella flexneri* affects secretion, expression, and glycosylation of gel-forming mucins in mucus-producing cells. Infect Immun 81:3632-3643.
187. Kaakoush NO, Castano-Rodriguez N, Mitchell HM, Man SM. 2015. Global epidemiology of *Campylobacter* infection. Clin Microbiol Rev 28:687-720.
188. Crofts AA, Poly FM, Ewing CP, Kuroiwa JM, Rimmer JE, Harro C, Sack D, Talaat KR, Porter CK, Gutierrez RL, DeNearing B, Brubaker J, Laird RM, Maue AC, Jaep K, Alcala A, Tribble DR, Riddle MS, Ramakrishnan A, McCoy AJ, Davies BW, Guerry P, Trent MS. 2018. *Campylobacter jejuni* transcriptional and genetic adaptation during human infection. Nat Microbiol 3:494-502.
189. Luethy PM, Huynh S, Ribardo DA, Winter SE, Parker CT, Hendrixson DR. 2017. Microbiota-derived short-chain fatty acids modulate expression of *Campylobacter jejuni* determinants required for commensalism and virulence. MBio 8:e00407-17.
190. Lin J, Sahin O, Michel LO, Zhang Q. 2003. Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of *Campylobacter jejuni*. Infect Immun 71:4250-4259.
191. Sun X, Winglee K, Gharaibeh RZ, Gauthier J, He Z, Tripathi P, Avram D, Bruner S, Fodor A, Jobin C. 2018. Microbiota-derived metabolic factors reduce campylobacteriosis in mice. Gastroenterology 154:1751-1763 e2.
192. Hoang KV, Stern NJ, Saxton AM, Xu F, Zeng X, Lin J. 2011. Prevalence, development, and molecular mechanisms of bacteriocin resistance in *Campylobacter*. Appl Environ Microbiol 77:2309-2316.
193. Hoang KV, Stern NJ, Lin J. 2011. Development and stability of bacteriocin resistance in *Campylobacter spp*. J Appl Microbiol 111:1544-1550.
194. Velayudhan J, Kelly DJ. 2002. Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni* an essential role for phosphoenolpyruvate carboxykinase. Microbiology 148:685-694.
195. Line JE, Hiett KL, Guard-Bouldin J, Seal BS. 2010. Differential carbon source utilization by *Campylobacter jejuni* 11168 in response to growth temperature variation. J Microbiol Methods 80:198-202.
196. Gripp E, Hlahla D, Didelot X, Kops F, Maurischat S, Tedin K, Alter T, Ellerbroek L, Schreiber K, Schomburg D, Janssen T, Bartholomäus P, Hofreuter D, Woltemate S, Uhr M, Brenneke B, Grüning P, Gerlach G, Wieler L, Suerbaum S, Josenhans C. 2011. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. BMC Genomics 12:584.
197. Muraoka WT, Zhang Q. 2011. Phenotypic and genotypic evidence for L-fucose utilization by *Campylobacter jejuni*. J Bacteriol 193:1065-1075.
198. Velayudhan J, Jones MA, Barrow PA, Kelly DJ. 2003. L-serine catabolism via an oxygen-labile l-serine Dehydratase is essential for colonization of the avian gut by *Campylobacter jejuni*. Infect Immun 72:260-268.
199. Wright JA, Grant AJ, Hurd D, Harrison M, Guccione EJ, Kelly DJ, Maskell DJ. 2009. Metabolite and transcriptome analysis of *Campylobacter jejuni* in vitro growth reveals a stationary-phase physiological switch. Microbiology 155:80-94.
200. Leach S, Harvey P, Wait R. 1997. Changes with growth rate in the membrane lipid composition of and amino acid utilization by continuous cultures of *Campylobacter jejuni*. J Appl Microbiol 82:631-640.
201. Palyada K, Threadgill D, Stintzi A. 2004. Iron acquisition and regulation in *Campylobacter jejuni*. J Bacteriol 186:4714-4729.

202. Miller CE, Rock JD, Ridley KA, Williams PH, Ketley JM. 2008. Utilization of lactoferrin-bound and transferrin-bound iron by *Campylobacter jejuni*. J Bacteriol 190:1900-1911.
203. Liu MM, Boinett CJ, Chan ACK, Parkhill J, Murphy MEP, Gaynor EC. 2018. Investigating the *Campylobacter jejuni* transcriptional response to host intestinal extracts reveals the involvement of a widely conserved iron uptake system. MBio 9:e01347-18.
204. Ferrero R, Lee A. 1988. Motility of *Campylobacter jejuni* in a viscous environment comparison with conventional rod-shaped bacteria. J Gen Microbiol 134:53-59.
205. Szymanski CM, King M, Haardt M, Armstrong GD. 1995. *Campylobacter jejuni* motility and invasion of Caco-2 cells. Infect Immun 63:4295-4300.
206. McSweegan E, Walker RI. 1986. Identification and characterization of two *Campylobacter jejuni* adhesins for cellular and mucous substrates. Infect Immun 53:141-148.
207. Stahl M, Frirdich E, Vermeulen J, Badayeva Y, Li X, Vallance BA, Gaynor EC. 2016. The helical shape of *Campylobacter jejuni p*romotes in vivo pathogenesis by aiding transit through intestinal mucus and colonization of crypts. Infect Immun 84:3399-3407.
208. Freter R. 1955. The fatal enteric cholera infection in the guinea pig, achieved by inhibition of normal enteric flora. J Infect Dis 97:57-65.
209. Barzilay EJ, Schaad N, Magloire R, Mung KS, Boncy J, Dahourou GA, Mintz ED, Steenland MW, Vertefeuille JF, Tappero JW. 2013. Cholera surveillance during the Haiti epidemic--the first 2 years. N Engl J Med 368:599-609.
210. Kaper JB, Morris JG, Jr., Levine MM. 1995. Cholera. Clin Microbiol Rev 8:48-86.
211. Hang S, Purdy AE, Robins WP, Wang Z, Mandal M, Chang S, Mekalanos JJ, Watnick PI. 2014. The acetate switch of an intestinal pathogen disrupts host insulin signaling and lipid metabolism. Cell Host Microbe 16:592-604.
212. Gupta S, Chowdhury R. 1997. Bile affects production of virulence factors and motility of *Vibrio cholerae*. Infect Immun 65:1131-1134.
213. Chatterjee A, Dutta PK, Chowdhury R. 2007. Effect of fatty acids and cholesterol present in bile on expression of virulence factors and motility of *Vibrio cholerae*. Infect Immun 75:1946-53.
214. Sengupta C, Ray S, Chowdhury R. 2014. Fine tuning of virulence regulatory pathways in enteric bacteria in response to varying bile and oxygen concentrations in the gastrointestinal tract. Gut Pathog 6:38.
215. Provenzano D, Schuhmacher DA, Barker JL, Klose KE. 2000. The virulence regulatory protein ToxR mediates enhanced bile resistance in *Vibrio cholerae* and other pathogenic *Vibrio* species. Infect Immun 68:1491-1497.
216. Duret G, Delcour AH. 2010. Size and dynamics of the *Vibrio cholerae* porins OmpU and OmpT probed by polymer exclusion. Biophys J 98:1820-9.
217. Ante VM, Bina XR, Howard MF, Sayeed S, Taylor DL, Bina JE. 2015. *Vibrio cholerae* leuO transcription is positively regulated by toxR and contributes to bile resistance. J Bacteriol 197:3499-3510.
218. McDonald ND, Lubin JB, Chowdhury N, Boyd EF. 2016. Host-derived sialic acids are an important nutrient source required for optimal bacterial fitness in vivo. MBio 7:e02237-15.
219. Yoon SS, Mekalanos JJ. 2006. 2,3-butanediol synthesis and the emergence of the *Vibrio cholerae* El Tor biotype. Infect Immun 74:6547-6556.
220. Guentzel MN, Berry LJ. 1975. Motility as a virulence factor for *Vibrio cholerae*. Infect Immun 11:890-897.
221. Silva AJ, Pham K, Benitez JA. 2003. Haemagglutinin/protease expression and mucin gel penetration in El Tor biotype *Vibrio cholerae*. Microbiology 149:1883-1891.
222. Szabady RL, Yanta JH, Halladin DK, Schofield MJ, Welch RA. 2011. TagA is a secreted protease of *Vibrio cholerae* that specifically cleaves mucin glycoproteins. Microbiology 157:516-525.
223. Benitez JA, Spelbrink RG, Silva A, Philips TE, Stanley CM, Boesman-Finkelstein M, Finkelstein RA. 1997. Adherence of *Vibrio cholerae* to cultured differentiated human intestinal cells an in vitro colonization model. Infect Immun 65:3474-3477.
224. Booth BA, Boesman-Finkelstein M, Finkelstein RA. 1984. *Vibrio cholerae* hemagglutinin protease nicks cholera enterotoxin. Infect Immun 45:558-560.
225. Finkelstein RA, Hanne LF. 1982. Purification and characterization of the soluble hemagglutinin (cholera lectin) (produced by *Vibrio cholerae*. Infect Immun 36:1199-1208.
226. Rahman A, Bonny TS, Stonsaovapak S, Ananchaipattana C. 2011. *Yersinia enterocolitica*: Epidemiological studies and outbreaks. J Pathog 2011:239391.
227. Bancerz-Kisiel A, Pieczywek M, Lada P, Szweda W. 2018. The most important virulence markers of *Yersinia enterocolitica* and their role during infection. Genes (Basel) 9:E235.
228. El-Ziney MG, De Meyer H, Debevere JM. 1997. Growth and survival kinetics of *Yersinia enterocolitica* IP 383

O:9 as affected by equimolar concentrations of undissociated short-chain organic acids. Int J Food Microbiol 34:233-247.
229. Raczkowska A, Trzos J, Lewandowska O, Nieckarz M, Brzostek K. 2015. Expression of the AcrAB components of the AcrAB-TolC multidrug efflux pump of *Yersinia enterocolitica* is subject to dual regulation by OmpR. PLoS One 10:e0124248.
230. Garzetti D, Bouabe H, Heesemann J, Rakin A. 2012. Tracing genomic variations in two highly virulent *Y. enterocolitica* strains with unequal ability. BMC Genomics 13:467.
231. Carniel E, Mazigh D, Mollaret HH. 1987. Expression of iron-regulated proteins in *Yersinia* species and their relation to virulence. Infect Immun 55:277-280.
232. Heesemann J, Hantke K, Vocke T, Saken E, Rakin A, Stojiljkovic I, Berner R. 1993. Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65 000 Da and pesiticin sensitivity. Mol Microbiol 8:397-408.
233. Perry RD, Fetherston JD. 2011. Yersiniabactin iron uptake: mechanisms and role in *Yersinia pestis* pathogenesis. Microbes Infect 13:808-817.
234. Kanaujia PK, Bajaj P, Kumar S, Singhal N, Virdi JS. 2015. Proteomic analysis of *Yersinia enterocolitica* biovar 1A under iron-rich and iron-poor conditions indicate existence of efficiently regulated mechanisms of iron homeostasis. J Proteomics 124:39-49.
235. Reuter S, Connor TR, Barquist L, Walker D, Feltwell T, Harris SR, Fookes M, Hall ME, Petty NK, Fuchs TM, Corander J, Dufour M, Ringwood T, Savin C, Bouchier C, Martin L, Miettinen M, Shubin M, Riehm JM, Laukkanen-Ninios R, Sihvonen LM, Siitonen A, Skurnik M, Falcao JP, Fukushima H, Scholz HC, Prentice MB, Wren BW, Parkhill J, Carniel E, Achtman M, McNally A, Thomson NR. 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. Proc Natl Acad Sci U S A 111:6768-6773.
236. Mantle M, Atkins E, Kelly J, Thakore E, Buret A, Gall DG. 1991. Effects of *Yersinia enterocolitica* infection on rabbit intestinal and colonic goblet cells and mucin morphometrics, histochemistry, and biochemistry. Gut 32:1131-1138.
237. Mantle M, Basaraba L, Peacock SC, Gall DG. 1989. Binding of *Yersinia enterocolitica* to rabbit intestinal brush border membranes, mucus, and mucin. Ìnfect Immun 57:3292-3299.
238. Mantle M, Husar SD. 1994. Binding of Yersinia enterocolitica to purified, native small intestinal mucins from rabbits and humans involves interactions with the mucin carbohydrate moiety. Infect Immun 62:1219-1227.
239. Mantle M, Rombough C. 1993. Growth in and breakdown of purified rabbit small intestinal mucin by *Yersinia enterocolitica*. Infect Immun 61:4131-4138.
240. Paerregaard A, Espersen F, Jensen OM, Skurnik M. 1991. Interactions between *Yersinia enterocolitica* and rabbit ileal mucus growth, adhesion, penetration, and subsequent changes in surface hydrophobicity. Infect Immun 59:253-260.
241. de Noordhout CM, Devleesschauwer B, Angulo FJ, Verbeke G, Haagsma J, Kirk M, Havelaar A, Speybroeck N. 2014. The global burden of listeriosis: a systematic review and meta-analysis. Lancet Infect Dis 14:1073-1082.
242. McLauchlin J. 1990. Human listeriosis in Britain, 1967-85, a summary of 722 cases. Epidemiol Infect 104:181-189.
243. Roberts AJ, Wiedmann M. 2003. Pathogen, host and environmental factors contributing to the pathogenesis of listeriosis. Cell Mol Life Sci 60:904-918.
244. Sun Y, Wilkinson BJ, Standiford TJ, Akinbi HT, O'Riordan MX. 2012. Fatty acids regulate stress resistance and virulence factor production for *Listeria monocytogenes*. J Bacteriol 194:5274-5284.
245. Julotok M, Singh AK, Gatto C, Wilkinson BJ. 2010. Influence of fatty acid precursors, including food preservatives, on the growth and fatty acid composition of *Listeria monocytogenes* at 37 and 10degreesC. Appl Environ Microbiol 76:1423-1432.
246. Rinehart E, Newton E, Marasco MA, Beemiller K, Zani A, Muratore MK, Weis J, Steinbicker N, Wallace N, Sun Y. 2018. *Listeria monocytogenes* response to propionate is differentially modulated by anaerobicity. Pathogens 7:E60.
247. Quillin SJ, Schwartz KT, Leber JH. 2011. The novel *Listeria monocytogenes* bile sensor BrtA controls expression of the cholic acid efflux pump MdrT. Mol Microbiol 81:129-142.
248. Payne A, Schmidt TB, Nanduri B, Pendarvis K, Pittman JR, Thornton JA, Grissett J, Donaldson JR. 2013. Proteomic analysis of the response of *Listeria monocytogenes* to bile salts under anaerobic conditions. J Med Microbiol 62:25-35.
249. Dussurget O, Cabanes D, Dehoux P, Lecuit M, Consortium TELG, Buchrieser C, Glaser P, Cossart P. 2002. *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. Mol Microbiol 45:1095-1106.

250. Begley M, Gahan CGM, Hill C. 2002. Bile stress response in *Listeria monocytogenes* LO28: Adaptation, cross-protection, and identification of genetic loci involved in bile resistance. Appl Environ Microbiol 68:6005-6012.
251. White SJ, McClung DM, Wilson JG, Roberts BN, Donaldson JR. 2015. Influence of pH on bile sensitivity amongst various strains of *Listeria monocytogenes* under aerobic and anaerobic conditions. J Med Microbiol 64:1287-1296.
252. Collins B, Curtis N, Cotter PD, Hill C, Ross RP. 2010. The ABC transporter AnrAB contributes to the innate resistance of *Listeria monocytogenes* to nisin, bacitracin, and various -lactam antibiotics. Antimicrob Agents Chemother 54:4416-4423.
253. Diep DB, Skaugen M, Salehian Z, Holo H, Nes IF. 2007. Common mechanisms of target cell recognition and immunity for class II bacteriocins. Proc Natl Acad Sci U S A 104:2384-2389.
254. Kjos M, Nes IF, Diep DB. 2011. Mechanisms of resistance to bacteriocins targeting the mannose phosphotransferase system. Appl Environ Microbiol 77:3335-3342.
255. Dalet K, Cenatiempo Y, Cossart P, Consortium TELG, Héchard Y. 2001. A sigma(54)-dependent PTS permease of the mannose family is responsible for sensitivity of *Listeria monocytogenes* to mesentericin Y105. Microbiology 147:3263-3269.
256. Balay DR, Ganzle MG, McMullen LM. 2018. The effect of carbohydrates and bacteriocins on the growth kinetics and resistance of *Listeria monocytogenes*. Front Microbiol 9:347.
257. Quereda JJ, Dussurget O, Nahori MA, Ghozlane A, Volant S, Dillies MA, Regnault B, Kennedy S, Mondot S, Villoing B, Cossart P, Pizarro-Cerda J. 2016. Bacteriocin from epidemic *Listeria* strains alters the host intestinal microbiota to favor infection. Proc Natl Acad Sci U S A 113:5706-5711.
258. Chen GY, Pensinger DA, Sauer JD. 2017. *Listeria monocytogenes* cytosolic metabolism promotes replication, survival, and evasion of innate immunity. Cell Microbiol 19:e12762.
259. Schardt J, Jones G, Muller-Herbst S, Schauer K, D'Orazio SEF, Fuchs TM. 2017. Comparison between *Listeria sensu stricto* and *Listeria sensu lato* strains identifies novel determinants involved in infection. Sci Rep 7:17821.
260. Mariscotti JF, Quereda JJ, Garcia-Del Portillo F, Pucciarelli MG. 2014. The *Listeria monocytogenes* LPXTG surface protein Lmo1413 is an invasin with capacity to bind mucin. Int J Med Microbiol 304:393-404.
261. Pron B, Boumaila C, Jaubert F, Sarnacki S, Monnet J, Berche P, Gaillard J. 1998. Comprehensive study of the intestinal stage of listeriosis in a rat ligated ileal loop system. Infect Immun 66:747-755.
262. Linden SK, Bierne H, Sabet C, Png CW, Florin TH, McGuckin MA, Cossart P. 2008. *Listeria monocytogenes* internalins bind to the human intestinal mucin MUC2. Arch Microbiol 190:101-104.
263. Popowska M, Krawczyk-Balska A, Ostrowski R, Desvaux M. 2017. InlL from *Listeria monocytogenes* is involved in biofilm formation and adhesion to mucin. Front Microbiol 8.
264. Segev N, Laub A, Cohen G. 1980. A circular form of bacteriophage P1 DNA made in lytically infected cells of *Escherichia coli*. Virology 101:261-271.
265. Davies EV, Winstanley C, Fothergill JL, James CE. 2016. The role of temperate bacteriophages in bacterial infection. FEMS Microbiol Lett 363:fnw015.
266. Refardt D, Rainey PB. 2010. Tuning a genetic switch: experimental evolution and natural variation of prophage induction. Evolution 64:1086-1097.
267. Schwechheimer C, Kuehn MJ. 2015. Outer-membrane vesicles from gram-negative bacteria: biogenesis and functions. Nat Rev Microbiol 13:605-619.
268. Reyes-Robles T, Dillard RS, Cairns LS, Silva-Valenzuela CA, Housman M, Ali A, Wright ER, Camilli A. 2018. *Vibrio cholerae* outer membrane vesicles inhibit bacteriophage infection. J Bacteriol 200:e00792-17.
269. Seed KD, Faruque SM, Mekalanos JJ, Calderwood SB, Qadri F, Camilli A. 2012. Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in *Vibrio cholerae* O1. PLoS Pathog 8:e1002917.
270. Cumby N, Edwards AM, Davidson AR, Maxwell KL. 2012. The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. J Bacteriol 194:5012-5019.
271. Cumby N, Reimer K, Mengin-Lecreulx D, Davidson AR, Maxwell KL. 2015. The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of E. coli phage HK97. Mol Microbiol 96:437-447.
272. Haberman A, Heywood J, Meselson M. 1972. DNA Modification Methylase Activity of *Escherichia coli* Restriction Endonucleases K and P. Proc Natl Acad Sci U S A 69:3138-3141.
273. Brockes JP. 1972. The deoxyribonucleic acid modification enzyme of bacteriophage P1. Biochem J 127:1-10.
274. Pleska M, Qian L, Okura R, Bergmiller T, Wakamoto Y, Kussell E, Guet CC. 2016. Bacterial autoimmunity due to a restriction-modification system. Curr Biol 26:404-409.
275. Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016. Diverse evolutionary roots

and mechanistic variations of the CRISPR-Cas systems. Science 353:aad5147.
276. Amitai G, Sorek R. 2016. CRISPR-Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol 14:67-76.
277. Touchon M, Bernheim A, Rocha EP. 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. ISME J 10:2744-2754.
278. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52-W57.
279. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 34:169-183.
280. Gross SR. 1953. Abortive Infection of a Strain of *Escherichia Coli* by Colipphage T2. J Bacteriol 68:36-42.
281. Fields KL. 1969. Comparison of the action of colicins El and K on *Escherichia coli* with the effects of abortive infection by virulent bacteriophages. J Bacteriol 97:78-82.
282. Smith HS, Pizer LI, Pylkas L, Lederberg S. 1969. Abortive Infection of *Shigella dysenteriae* P2 by T2 Bacteriophage. J Virol 4:162-168.
283. Ainsworth S, Stockdale S, Bottacini F, Mahony J, van Sinderen D. 2014. The *Lactococcus lactis* plasmidome: much learnt, yet still lots to discover. FEMS Microbiol Rev 38:1066-1088.

Chapter 3

Opportunities and Challenges in Development of Live Biotherapeutic Products To Fight Infections

Development of LBPs against infection
The Journal of Infectious Diseases, 2021

Q. R. Ducarmon[1,2], E. J. Kuijper[1,2,3], B. Olle[4]

[1]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands
[2]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands
[3]Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[4]Vedanta Biosciences Inc., Cambridge, United States

J Infect Dis. 2021 Jun 16;223
(Supplement_3):S283-S289.

Abstract

Treatment of bacterial infections with broad spectrum antibiotics is a strategy severely limited by the decreased ability of the perturbed resident microbiota to control expansion of antibiotic resistant pathogens. Live Biotherapeutic Products (LBPs) could provide an alternative to antibiotics in infection control by restoring gut colonization resistance and controlling expansion of resistant strains, an important therapeutic need not being addressed with existing anti-infective drug modalities. We review opportunities and challenges in developing LBPs for MDRO colonization and infection control, with a focus on commercial FMT-like products and defined bacterial consortia, and spanning considerations related to availability of models for rational drug candidate selection and dose regimen selection, good manufacturing practice, intellectual property, and commercial viability.

Focus and definitions

FDA defines LBPs as "a biological product that: 1) contains live organisms, such as bacteria; 2) is applicable to the prevention, treatment, or cure of a disease or condition of human beings; and 3) is not a vaccine" [1]. Within FDA, The Center for Biologics Evaluation and Research (CBER) is responsible for regulating LBPs, and their licensure is obtained by approval of a biologics license application (BLA) [2]. A number of drug modalities currently being advanced meet the definition of LBP, including procedures to transplant fecal microbiota or spore fractions from fecal microbiota, as well as products of defined composition, such as single bacterial strains, engineered bacterial strains, and defined bacterial consortia. Furthermore, LBPs may be administered orally, rectally, topically, or as injectables. This piece focuses on orally and rectally delivered LBPs consisting of natural, unmodified bacteria, which have drawn most interest to date in the context of antimicrobial resistance (AMR), and excludes injectables, topicals, and LBPs consisting of engineered bacterial strains. Development considerations pertinent to engineered LBPs have been reviewed elsewhere [3].

Opportunity for Live Biotherapeutic Products (LBPs) in the context of AMR

The gut is a reservoir for numerous multi-drug resistant organisms (MDRO), including *Enterobacteriaceae* such as *Escherichia coli*, *Klebsiella pneumonia*e or *Enterobacter aerogenes*, and Enterococci such as *Enterococcus faecium* and *E. faecalis*. Antibiotic use associated with a range of medical procedures results in collateral damage to the gut microbiota resulting in an increased risk for development of infections, including by *Clostridioides difficile*, and acquired colonization with MDRO [4-6].

The extensive use of antibiotics can also contribute to bacteria developing antibiotic resistance mechanisms. As physicians have become more aware of these threats and antibiotic stewardship programs have been put in place, antibiotic sales volumes have dropped in the US [7], which combined with severe pricing pressure, has led to an exodus of pharmaceutical companies from anti-infective drug development. LBPs could contribute to breaking this vicious cycle in several ways. First, the expansion of resident or acquired MDROs could be kept in check by helping restore the host microbiota after an antibiotic perturbation. This could be particularly useful in vulnerable populations such as patients undergoing hematopoietic stem cell transplantation, intestinal surgeries, organ transplants, chemotherapy, or dialysis. Second, LBPs should not contribute to selecting resistant strains from susceptible populations and therefore no LBP stewardship should be necessary. Supporting this promise, fecal microbiota transplantation (FMT)

has shown high efficacy in prevention of recurrent *Clostridioides difficile* infection (CDI) [8] and defined bacterial consortia have shown promise in rodents models of vancomycin-resistance *Enterococcus* (VRE) infection [9].

Recent work has started to shed light on the mechanisms of post-antibiotic gut microbiome recovery, paving the way for developing targeted prevention strategies [10]. Modes of action through which LBPs may achieve successful eradication or prevent colonization of MDROs include competition for nutrients, production of short-chain fatty acids (SCFA), conversion of primary to secondary bile acids, and production of bacteriocins, among others [11, 12]. There is some limited evidence that administration of an LBP consisting of a single microorganism can help prevent C. *difficile* infection [13], and significantly more evidence that administration of complex bacterial communities such as FMT and defined bacterial consortia [8, 14] can be therapeutically useful. Therefore, we focus here on challenges specific to LBPs consisting of FMT and defined bacterial consortia.

Undefined vs defined LBPs

The two main categories of LBPs that have received attention from drug developers to address AMR are commercial FMT and FMT-like procedures and defined LBPs consisting of single bacteria or consortia of bacteria. FMT and FMT-like procedures consist of full bacterial communities from fecal donor samples or spore fractions of such communities, administered rectally [15] or orally [16, 17]. The focus for these procedures is on standardizing the steps to identify and screen healthy fecal donors, and process, store, ship, and administer the stool formulations. Given the variation of microbiota composition across individuals and over time, the composition of the resulting products varies with each donation and is thus undefined in nature. Regulation of these products is primarily concerned with the process by which they are prepared for transplantation, rather than its undefined contents. Defined LBPs, in contrast, consist of a limited set of bacterial species produced by fermentation from clonal cell banks, resulting in a final product of defined, standardized composition. Regulation of defined LBPs is concerned with both the process as well as the specific components of the product, with an increased scrutiny of characteristics of the component strains such as their genetic identity and their potential for transferring virulence or antibiotic resistance genes to other members of the microbiota.

Challenges to development of LBPs

Biological complexity

The single most formidable challenge for development of LBPs is perhaps the sheer

complexity of the biology being uncovered, and yet to be uncovered, on the role of host microbial communities in human health and disease. The mechanisms by which bacteria influence host phenotypes are often highly pleiotropic, rendering reductionistic potency assays to be of limited value in the development of LBPs. The understanding of the role of host microbial communities in disease is only partial, making selection of optimal patient populations for clinical studies a complex endeavor. The knowledge of the fundamental rules that govern assembly of microbial communities is still in its infancy, making bottom-up approaches to rational construction of drug candidates consisting of bacterial communities rudimentary for now. Perhaps the most salient departure from traditional development of drugs based on small molecules, proteins, or oligonucleotides is that, while transdisciplinary approaches using chemistry, biology, and computational science have been successful to enable these modalities, development of LBPs needs to rely heavily on insights from microbial ecology, a discipline largely ignored by the pharmaceutical industry to date.

Determination of Pharmacokinetic - Pharmacodynamic relationships and dose regimen selection

Pharmacokinetics (PK) is the study of how the host affects the fate of an exogenously administered drug. In the context of LBPs, this should ideally include studying how abundantly and durably the product strains colonize the host, and what proportion of the product strains colonize a given host at a given time. Pharmacodynamics (PD) is the study of how a drug affects the organism, traditionally with a focus on the biochemical and physiologic effects of the drug on the host. This definition is still relevant to LBPs, but study of the PD of an LBP additionally requires understanding the ecological effects of the drug on the host resident microbial community.

It is impossible to fully understand the action of a drug unless the relationship between drug exposure and effect has been reasonably well described. The inherent batch-to-batch variability in the composition of FMT and FMT-like procedures makes it challenging to reasonably describe in a quantitative manner the relationship between PK and PD and thus rationalize clinical successes and clinical failures. In contrast, quantifying with precision the PK of defined LBPs is complex but feasible. It needs to address the non-trivial technical problem of discriminating exogenously administered LBP strains from closely-related resident strains in the host's bacterial community. This has been achieved by culturing the strains in a defined LBP and obtaining high quality, complete genome sequences from which unique genetic markers can be derived and used to track strain-level engraftment from metagenomic sequencing of DNA isolated from fecal material [18, 19]. Recent clinical work following this approach has started to illuminate some basic features of LBP PK that are likely to be generalizable, specifically showing that higher dose, more frequent administration, and pretreatment with short

courses of antibiotics to create a niche for the LBP to engraft can significantly improve the abundance and durability of LBP strain colonization, as well as the proportion of LBP strains that colonize [18].

Quantifying the PK of a defined LBP is thus already possible. *Predicting* the PK of an LBP however, remains a significant challenge. Successful colonization of LBPs will likely be a function of a combination of features of the LBP, the host resident microbiota, the host, and other environmental factors. Key features of the LBP that influence colonization include dose, dose frequency, and species traits that may help with engraftment in the gut (e.g. a shared evolutionary history with the host). Features of the resident bacterial community that influence the success of invasion by an exogenous LBP may include bacterial density [20], diversity, and community structure, among others. Features of the host that may influence colonization include disease status, age, the host immune system, and host genetics. Finally, other environmental factors including diet and previous or concomitant drug use may have particularly salient effects on PK. Among drug-LBP interactions, interactions with broad spectrum antibiotics represent a case of particular medical interest in the context of AMR. Antibiotic perturbation can significantly lower the bacterial density and diversity of the resident gut community [10], thus freeing up resources for invaders and creating a niche for LBP engraftment.

The factors outlined above, combined with the host specificity of bacterial communities, render use of rodent animal models of limited value in the selection of dose and dose regimens for human studies. Healthy volunteer studies, controlled human infection models (CHIM), or ultimately dose-ranging studies in patients provide a more representative, albeit expensive alternative to determining PK-PD relationships. Early clinical efforts in the microbiome field omitted dose ranging exploration altogether before advancing drug candidates to late stage efficacy studies, and this may have been a factor contributing to clinical failures [21].

Exploration of pharmacodynamic effects of LBPs on the resident host microbiota is complicated by the myriad community features revealed by metagenomics, metabolomics and proteomics analyses. This work could be significantly aided by the use of standardized indices of gut microbiota health or disease susceptibility. Such indices may rely on measures of community structure that correlate with clinical outcomes. For example, oligodomination by certain opportunistic pathogens has been strongly associated to risk of infection [22]. In the context of AMR, such indices may support the development of LBPs by quantifying in a simple, easy to comprehend manner, the risk that a given patient may become infected and/or dominated by a pathogen, and by serving as a surrogate measure of the contribution of an LBP towards outcomes such as lowering infection risk for that patient [23, 24].

Intellectual Property

Obtaining patent protection is an essential component of successful drug development. An important element in obtaining patent protection for LBPs in the US has been navigating requirements codified in the United States Code (USC) as 35 USC § 101 (utility). This requirement defines the boundaries of what is patent eligible, to the exclusion of "natural phenomena". The US Supreme Court in *Mayo* [25] and *Myriad* [26] limited the scope of patent eligibility for natural products, making it more difficult to obtain composition of matter claims covering such products. Nevertheless, several applicants have obtained composition of matter claims on defined LBPs [27] by arguing successfully that their claims combine additional elements that result in new functional properties that yield something that is "significantly more" than what exists in nature [28]. Obtaining broad patent coverage can be more challenging for FMT products, which have not received broad composition of matter claims due to a combination of factors including lack of differentiation from what exists in nature, lack of novelty over prior art, and inability to sufficiently describe the composition of the FMT preparation. Instead, applicants have resorted to pursuing narrower method of use claims highlighting unique modifications to the stool preparation process such as filtration, lyophilization, or encapsulation steps [28].

Challenging marketplace for anti-infectives

The last decade has seen a massive exodus of pharmaceutical companies from anti-infective drug development due to structural economic issues that ultimately result in the inability to make meaningful profits from selling antibiotics. A first, salient issue with the economic marketplace is that US hospitals are strongly incentivized to use cheaper antibiotics whenever possible, unless there is absolute clinical need for more expensive antibiotics, because US insurers pay for in-patient antibiotics as part of a lump sum to hospitals, and thus cheap antibiotics increase hospital profit margins [29]. For example, fidaxomicin has been proven superior to vancomycin, a cheap generic, in sustained cure of CDI, but its uptake has been limited due to pricing concerns [30]. Most drugs are not paid for like this. Oral LBPs which do not require administration in a hospital setting and can instead be taken at home may partially circumvent this issue, but ultimately only proposed reforms currently before Congress like the DISARM Act can fix this structural issue. A second issue with the marketplace is that stewardship programs aimed at limiting spread of AMR put downwards pressure on sales volumes of new antibiotics, the use of which is left as a last resort. This has led to calls for new regulation delinking antibiotic sales volume from return on investment through prizes or insurance-like models [31]. The current framework for AMR stewardship is based on small molecule antibiotics and focuses on limiting the selection of resistant strains as a result of antibiotic use. The mechanisms of action by which LBPs help restore the gut microbiota and its colonization resistance against pathogens are highly unlikely to elicit selection of resistant strains, and in fact could help limit the expansion of host-resident

resistant strains which could otherwise thrive in a perturbed microbiota. As a result, we predict that there should not be downwards pressure on sales volumes of a hypothetical successful LBP anti-infective.

Good Manufacturing Practice

Manufacturing of FMT-like products and defined LBPs intended for oral or rectal administration (and thus, likely based on anaerobic organisms) has a few shared challenges. These include minimizing exposure to oxygen, in particular in steps of the process where the organisms are metabolically active and preserving the viability of bacterial cells during processing and storage. A variety of factors influence the viability of bacteria during the manufacturing process and subsequent storage, including oxygen exposure, growth media, shearing, composition of the buffer solutions used to suspend the bacteria before freezing or freeze-drying, cooling rate, and freeze-thaw cycles, among others. The problem of maintaining cell viability during freezing and particularly during freeze-drying for long term storage deserves special attention, as it is one of the most technically challenging steps of manufacturing an LBP. During freezing and freeze-drying, the bacterial cell wall is exposed to mechanical forces due to formation of ice crystals inside and outside the cell, which can disrupt the membrane and kill the cell. During freeze-drying, furthermore, the process of removing water by sublimation generates osmotic pressures that can damage the cell membrane. Optimization of freeze-drying cooling cycles and development of buffer solutions containing cryoprotectants or lyoprotectants is therefore an important step to ensure the long-term preservation of LBPs. While preservation conditions for a number of aerobes and some facultative anaerobes such as E. *coli*, and Lactobacillus and Lactococcus species has been described in the literature, there is very little published on the topic of preservation of anaerobic gut commensals [32]. Further complicating the matter of long-term preservation of LBPs, the efficiency of cooling regimes and cryoprotectant and lyoprotectant substances can be highly bacterial species-specific.

There are certain manufacturing considerations that are unique to FMT-like products. Feces are a heterogenous substance composed of bacteria, viruses, fungi, food, and host secretions, which does not naturally yield itself to precise characterization. Consequently, manufacturing considerations emphasize rigorous donor screening and processing of stool donations, and relatively de-emphasize in-depth characterization of the composition, which would vary with every donation. FMT is performed using suspensions made of donor stool from carefully selected and screened healthy individuals. Donors undergo extensive health questionnaires and their blood and stool samples are analyzed for a list of known pathogenic viruses, bacteria, and parasites before being accepted. Recently, some amendments have been introduced to this process as a result of FDA's issuance of a series of safety alerts on the potential risks of life-threatening infections with the

use of FMT [33, 34] and on the risk of transmission of Sars-CoV-2 with FMT, leading to a halting of FMT studies in the US during 2020. Processing of stool donations varies depending on whether the final formulation is intended for oral or rectal administration. Stool samples may undergo a series of steps to filter the non-microbial components of stool, or the non-spore forming bacterial components of stool, depending on the product. FMT drug product may be released after meeting a specification of potency consisting of an estimate of the total aggregate of viable organisms present in the product, and the same assays may be used to demonstrate the FMT product stability for the planned duration of the clinical studies in which it is being used.

Defined LBP manufacturing considerations, by virtue of the composition being known and standardized, can put increased emphasis on the characterization of the component strains and less emphasis on an in-depth understanding of the donor from whom the strains were originally isolated. FDA expects a description of the drug substance including the biological name of each of the strains and strain designations, the original source of each of the strains, their passage history, and a description of the phenotype and genotype of the product strains [1]. Furthermore, sponsors are expected to characterize their LBPs using assays that assure the identity, purity, and potency of the drug substance and final drug product, and to apply these same assays over time as part of a stability program to ensure the product remains within specification for the duration of the proposed clinical studies. Identity tests are expected to detect each of the bacterial strains that compose the LPB, and to discriminate among LBP component strains. High quality genome sequences for each strain can provide an authoritative identification of each organism and enable comprehensive identification of potentially undesirable safety traits such as antibiotic resistance genes or virulence factors. A further assessment of the risk of transmission of such genes to relevant microbial flora (for example, based on proximity to mobile elements) is of particular interest. Sponsors are also expected to determine the antibiotic resistance phenotypes of the LBP strains, with a particular focus towards identifying clinically relevant antibiotics that can be used as rescue therapies in the event of an infection suspected to be caused by LBP strains. Purity tests are expected to show the absence of contaminating bacteria or yeast above acceptable limits. Potency tests commonly used for LBPs assess the product viability, for example in terms of viable CFUs per dose.

Defined LBPs are manufactured starting from clonal cell banks via fermentation, which may require optimization of growth media and physiological parameters like mixing, temperature, pH, retention time, and redox potential. After fermentation, bacteria are harvested by downstream steps such as filtration, which may require selection of appropriate filtration membranes and optimization of process variables such as transmembrane pressure and flow rates to minimize shear-induced damage to the

bacterial cell. A further challenge inherent to multi-strain defined LBPs manufactured as monocultures is that the number of banking campaigns, production runs, and characterization assays required scales linearly with the number of strains in the product. Taken together, these considerations impose a significant burden on drug developers but also create an opportunity to innovate: a non-trivial amount of the advances made by LBP developers will originate in their process development and GMP manufacturing activities.

Preclinical and clinical models to discover LBPs and study their pharmacology

While not strictly required by FDA, use of in vivo and in vitro models to test the efficacy and characterize the mechanism of action of LBP candidates prior to use in humans can be a sensible business decision. A challenge in use of animal models to study efficacy of microbiome drugs is that it is not always clear what microbiome endpoints are the most relevant surrogates of therapeutic efficacy. For example, pinpointing a specific microbiome endpoint most predictive of efficacy in treating immune or metabolic disease is not straightforward. An advantage of designing LBPs for use in AMR is the relative clarity of the microbiome endpoints used to quantify efficacy and their relation to the therapeutic goal: the microbiome endpoint of an animal model used for efficacy testing may be reduction or elimination of MDRO carriage in the gut (e.g., carbapenem-resistant *Enterobacteriaceae* [CRE], extended spectrum beta-lactamase [ESBL], or VRE), and the therapeutic goal may be to prevent infection outcomes with that same MDRO. Rodents, for example, have been colonized (at least temporarily) with pathogenic MDRO strains that infect humans, without resorting to surrogate mouse pathogens [14], and used to rationally select defined bacterial consortia that reduce intestinal colonization. Whether the surrogate endpoints of decolonization models truly predict clinical outcomes of LBPs will have to be demonstrated in future clinical studies.

A challenge in measuring efficacy of LBPs in AMR applications is the difficulty in anticipating which patients will be exposed to the pathogen, become colonized, and develop disease, which complicates execution of clinical studies powered on the basis of disease outcome endpoints. CHIM, where carefully selected human volunteers are deliberately infected with well-characterized infectious agents in a controlled setting can be an effective way of measuring the efficacy of a drug agent in these circumstances. CHIM have the advantage of decreasing the number of patients needed to detect efficacy in phase 2 and 3 trials, and have been used for testing vaccines in early in clinical development, dating back to 1900 [35]. CHIM offer the opportunity to study the physiological, immunological and metabolic changes that occur upon infection, including potentially assessing the role of the gut microbiome in transmission of antibiotic resistance and virulence genes.

Conclusion

Identification of commensal bacteria that can restore gut colonization resistance after antibiotics in high-risk patients is an important new strategy to prevent infection and transmission of MDROs. Use of LBPs as anti-infectives could circumvent a key limitation of antibiotics, namely the need for stewardship driven by selective pressure on resistant strains, while providing a potentially safe and convenient way of restoring the microbiota after antibiotic use in high risk patient populations.

Notes

Supplement sponsorship
This work is part of a supplement sponsored by the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health (NIH) and the Centers for Disease Control and Prevention (CDC).

Conflicts of interest
B.O. is an employee of Vedanta Biosciences Inc, E.J.K. is supported by an unrestricted grant from Vedanta Biosciences Inc and has performed research for Cubist, Novartis, and Qiagen and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur, and Seres Therapeutics. Q.R.D. reports no potential conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Financial support
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

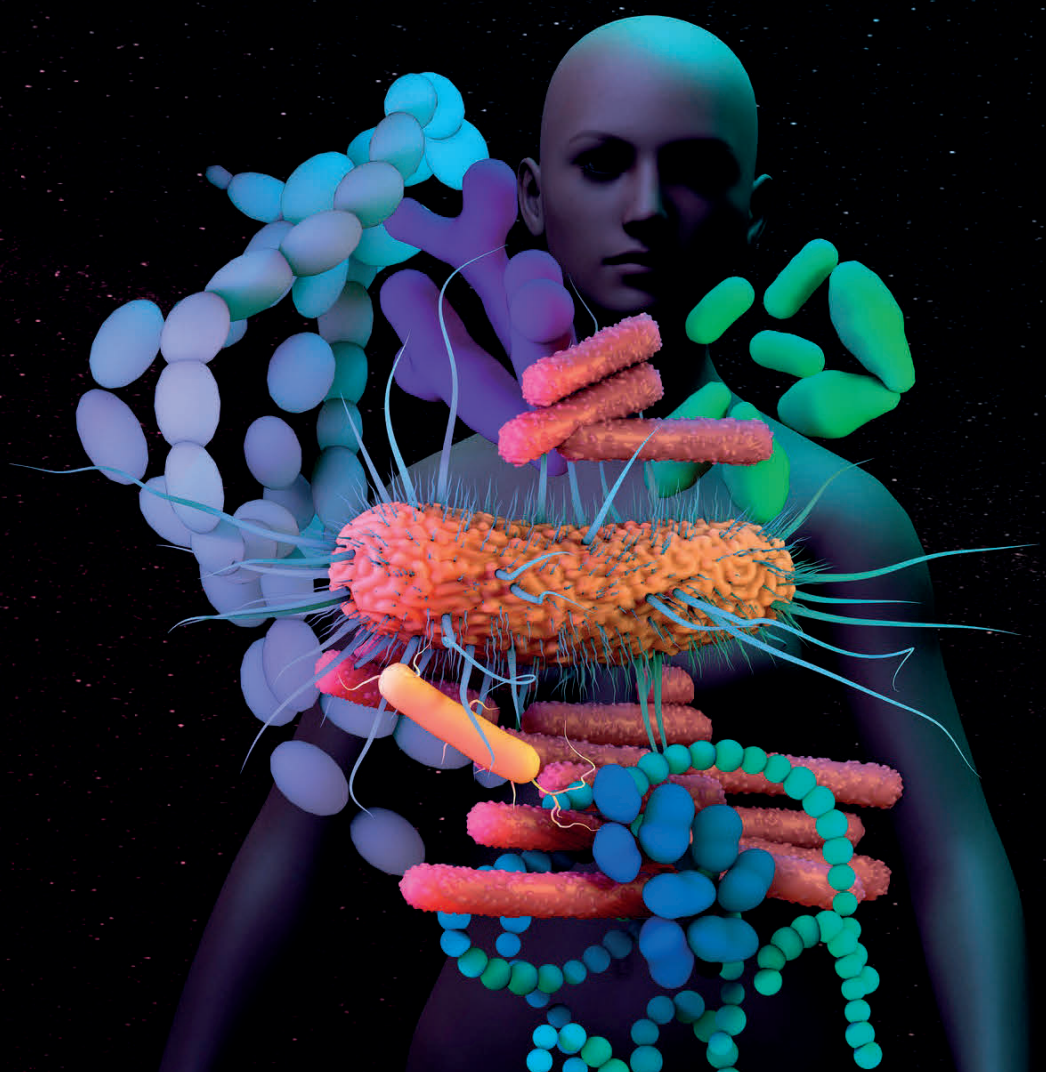
References

1. FDA. Early Clinical Trials with Live Biotherapeutic Products: Chemistry, Manufacturing, and Control Information (FDA, Washington, DC, 2016). https://www.fda.gov/files/vaccines,%20blood%20&%20biologics/published/Early-Clinical-Trials-With-Live-Biotherapeutic-Products--Chemistry--Manufacturing--and-Control-Information--Guidance-for-Industry.pdf, **2016**.
2. Dreher-Lesnick SM, Stibitz S, Carlson PE, Jr. U.S. Regulatory Considerations for Development of Live Biotherapeutic Products as Drugs. Microbiol Spectr **2017**; 5.
3. Charbonneau MR, Isabella VM, Li N, Kurtz CB. Developing a new class of engineered live bacterial therapeutics to treat human diseases. Nat Commun **2020**; 11:1738.
4. Thabit AK, Varugehese CA, Levine AR. Antibiotic use and duration in association with Clostridioides difficile infection in a tertiary academic medical center: A retrospective case-control study. Anaerobe **2019**; 59:126-30.
5. Weber D, Jenq RR, Peled JU, et al. Microbiota Disruption Induced by Early Use of Broad-Spectrum Antibiotics Is an Independent Risk Factor of Outcome after Allogeneic Stem Cell Transplantation. Biol Blood Marrow Transplant **2017**; 23:845-52.
6. Baggs J, Jernigan JA, Halpin AL, Epstein L, Hatfield KM, McDonald LC. Risk of Subsequent Sepsis Within 90 Days After a Hospital Stay by Type of Antibiotic Exposure. Clin Infect Dis **2018**; 66:1004-12.
7. Suda KJ, Hicks LA, Roberts RM, Hunkler RJ, Matusiak LM, Schumock GT. Antibiotic Expenditures by Medication, Class, and Healthcare Setting in the United States, 2010-2015. Clin Infect Dis **2018**; 66:185-90.
8. 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.
9. Kim SG, Becattini S, Moody TU, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. Nature **2019**; 572:665-9.
10. Chng KR, Ghosh TS, Tan YH, et al. Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. Nat Ecol Evol **2020**; 4:1256-67.
11. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev **2019**; 83:e00007-19.
12. Keith JW, Pamer EG. Enlisting commensal microbes to resist antibiotic-resistant pathogens. J Exp Med **2019**; 216:10-9.
13. Gerding DN, Sambol SP, Johnson S. Non-toxigenic Clostridioides (Formerly Clostridium) difficile for Prevention of C. difficile Infection: From Bench to Bedside Back to Bench and Back to Bedside. Front Microbiol **2018**; 9:1700.
14. Caballero S, Kim S, Carter RA, et al. Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. Cell Host Microbe **2017**; 21:592-602.e4.
15. Rebiotix. Rebiotix and Ferring announce world's first with positive preliminary pivotal Phase 3 data for investigational microbiome-based therapy RBX2660. Saint-Prex, Switzerland: https://www.rebiotix.com/news-media/press-releases/rebiotix-announces-worlds-first-positive-pivotal-phase-3-data-investigational-microbiome-based-therapy-rbx2660/, **2020**.
16. Finch. Finch Therapeutics Announces Positive Topline Results from Randomized Controlled Trial of CP101, an Oral Microbiome Drug, for the Prevention of Recurrent C. difficile Infection. Somerville, MA, USA: https://finchtherapeutics.com/blog/finch-therapeutics-announces-positive-topline-results-from-randomized-controlled-trial-of-cp101-an-oral-microbiome-drug-for-the-prevention-of-recurrent-cdiff, **2020**.
17. Businesswire. Seres Therapeutics Announces Positive Topline Results from SER-109 Phase 3 ECOSPOR III Study in Recurrent C. difficile Infection. Cambridge, MA, USA: https://www.businesswire.com/news/home/20200810005194/en/Seres-Therapeutics-Announces-Positive-Topline-Results-SER-109, **2020**.
18. Bobilev D, Bhattarai S, Menon R, et al. VE303, a Rationally Designed Bacterial Consortium for Prevention of Recurrent Clostridioides difficile (C. Difficile) infection (rCDI), Stably Restores the Gut Microbiota After Vancomycin (vanco)-Induced Dysbiosis in Adult Healthy Volunteers (HV). Open Forum Infect Dis **2019**; 6:S60.
19. Maldonado-Gómez MX, Martínez I, Bottacini F, et al. Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. Cell Host Microbe **2016**; 20:515-26.
20. Britton GJ, Contijoch EJ, Spindler MP, et al. Defined microbiota transplant restores Th17/RORγt(+) regulatory T cell balance in mice colonized with inflammatory bowel disease microbiotas. Proc Natl Acad Sci U S A **2020**.
21. Seres. Seres Therapeutics announces key findings from SER-109 Phase 2 study analyses. Cambridga, MA,

USA: https://ir.serestherapeutics.com/news-releases/news-release-details/seres-therapeutics-announces-key-findings-ser-109-phase-2-study, **2017**.
22. Taur Y, Xavier JB, Lipuma L, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis **2012**; 55:905-14.
23. Halpin AL, McDonald LC. Editorial Commentary: The Dawning of Microbiome Remediation for Addressing Antibiotic Resistance. Clin Infect Dis **2016**; 62:1487-8.
24. Blount K, Jones C, Deych E, Shannon B. Evaluating a Prototype Microbiome Health Index (MHI) as a Measure of Microbiome Restoration Using Data Derived From Published Studies of Fecal Microbiota Transplant to Treat recurrent Clostridium difficile Infections (rCDI). IDWeek 2018. San Francisco, California, **2018**.
25. Mayo Collaborative Services v. Prometheus Laboratories, Inc., 566 US 66, **2012**.
26. Association for Molecular Pathology v. Myriad Genetics, Inc., 569 US 576, **2013**.
27. Vedanta. Vedanta patents. https://www.vedantabio.com/pipeline/intellectual-property.
28. FitzGerald MJ, Spek EJ. Microbiome therapeutics and patent protection. Nat Biotechnol **2020**; 38:806-10.
29. Outterson K, Powers JH, Daniel GW, McClellan MB. Repairing the broken market for antibiotic innovation. Health Aff (Millwood) **2015**; 34:277-85.
30. Louie TJ, Miller MA, Mullane KM, et al. Fidaxomicin versus vancomycin for Clostridium difficile infection. N Engl J Med **2011**; 364:422-31.
31. Rex JH, Outterson K. Antibiotic reimbursement in a model delinked from sales: a benchmark-based worldwide approach. Lancet Infect Dis **2016**; 16:500-5.
32. Bircher L, Geirnaert A, Hammes F, Lacroix C, Schwab C. Effect of cryopreservation and lyophilization on viability and growth of strict anaerobic human gut microbes. Microb Biotechnol **2018**; 11:721-33.
33. DeFilipp Z, Bloom PP, Torres Soto M, et al. Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. N Engl J Med **2019**; 381:2043-50.
34. FDA. Fecal Microbiota for Transplantation: Safety Alert - Risk of Serious Adverse Events Likely Due to Transmission of Pathogenic Organisms. https://www.fda.gov/safety/medical-product-safety-information/fecal-microbiota-transplantation-safety-alert-risk-serious-adverse-events-likely-due-transmission, **2020**.
35. Roestenberg M, Hoogerwerf MA, Ferreira DM, Mordmüller B, Yazdanbakhsh M. Experimental infection of human volunteers. Lancet Infect Dis **2018**; 18:e312-e22.

Part II

Optimization and standardization of laboratory and computational procedures for microbiome research



Chapter 4

Toward standards in clinical microbiota studies: comparison of three DNA extraction methods and two bioinformatic pipelines

DNA isolation and informatics for microbiota studies
mSystems, 2020

Q.R. Ducarmon[a, b], B.V.H. Hornung[a,b], A.R. Geelen[a,b], E.J. Kuijper[a,b], R.D. Zwittink[a,b]

[a]Center for Microbiome Analyses and Therapeutics, Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands
[b]Experimental Bacteriology, Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands

mSystems. 2020 Feb 11;5(1):e00547-19.
Supplementary information available online.

Abstract

When studying the microbiome using next generation sequencing, DNA extraction method, sequencing procedures and bioinformatic processing are crucial to obtain reliable data. Method choice has been demonstrated to strongly affect the final biological interpretation. We assessed the performance of three DNA extraction methods and two bioinformatic pipelines for bacterial microbiota profiling through 16S rRNA gene amplicon sequencing, using positive and negative controls for DNA extraction and sequencing, and eight different types of high- or low-biomass samples. Performance was evaluated based on quality control passing, DNA yield, richness, diversity and compositional profiles. All DNA extraction methods retrieved the theoretical relative bacterial abundance with maximum three-fold change, although differences were seen between methods, and library preparation and sequencing induced little variation. Bioinformatic pipelines showed different results for observed richness, but diversity and compositional profiles were comparable. DNA extraction methods were successful for feces and oral swabs and variation induced by DNA extraction methods was lower than inter-subject (biological) variation. For low-biomass samples, a mixture of genera present in negative controls and sample-specific genera, possibly representing biological signal, were observed. We conclude that the tested bioinformatic pipelines perform equally with pipeline-specific advantages and disadvantages. Two out of three extraction methods performed equally well, while one method was less accurate regarding retrieval of compositional profiles. Lastly, we again demonstrate the importance of including negative controls when analyzing low bacterial biomass samples.

Importance

Method choice throughout the workflow of a microbiome study, from sample collection to DNA extraction and sequencing procedures, can greatly affect results. This study evaluated three different DNA extraction methods and two bioinformatic pipelines by including positive and negative controls, and various biological specimens. By identifying an optimal combination of DNA extraction method and bioinformatic pipeline use, we hope to contribute to increased methodological consistency in microbiota studies. Our methods were not only applied to commonly studied samples for microbiota analysis, e.g. feces, but also on more rarely studied, low-biomass samples. Microbiota composition profiles of low-biomass samples (e.g. urine and tumor biopsies) were not always distinguishable from negative controls, or showed partial overlap, confirming the importance of including negative controls in microbiota studies, especially when low bacterial biomass is expected.

Introduction

Humans constantly interact with microbes that are present in the environment and reside on or within the human body. Recently, the attention for microbes has shifted from an exclusive interest in the pathogenicity of specific microbes toward the potential beneficial role of the microbiota in human health [1]. The gastrointestinal tract contains the highest number of microbes and has been the most extensively studied body site of all human microbial communities [2]. However, many other body sites are inhabited by various microbes composing a specific microbiota, such as the oral region, skin and urogenital system. Microbial complexity varies between these niches, e.g. a healthy vaginal microbiota is often mainly composed of a few *Lactobacillus* strains, while gut and skin microbiota are usually more diverse [3].

A limiting factor in current microbiome research is that comparison of various study results is often difficult due to the application of different methodologies and lack of appropriate controls. These differences can affect data outcomes and lead to variation as large as biological differences [4]. Variation can be introduced throughout the entire workflow, from sample collection, storage and processing to data analysis [5-8]. Recently, more attention has been devoted to standardizing the workflow of microbiome research. For instance, it was observed that DNA extraction has a large impact on obtained data [4, 9] and consensus has been achieved regarding the application of bead-beating to increase efficiency of cell wall lysis and thereby improve the yield of Gram-positive bacterial DNA [10]. Nevertheless, various kits and in-house extraction methods are used across different laboratories. Recently, Costea *et al.* evaluated 21 DNA extraction methods

across three continents and suggested one protocol, named protocol Q, as 'golden standard' for human fecal samples [9]. They stated that it was unknown whether this method is optimal for other samples than fecal material, e.g. for low-biomass samples. To evaluate performance of DNA extraction for low-biomass samples, it is crucial to include multiple negative controls to allow for identification of bacterial DNA introduced during the entire workflow, from sample collection to sequencing [11].

As part of optimizing the procedures for 16S rRNA gene amplicon sequencing-based microbiota studies in our facility, we evaluated three DNA extraction methods and two bioinformatic pipelines using various positive controls and negative controls. In addition, we applied these DNA extraction methods to various biological specimens.

Results and discussion

Mock communities pass quality control

We evaluated three different DNA extraction methods and two bioinformatic pipelines for microbiota profiling through 16S rRNA gene amplicon sequencing (Fig 1) using several positive and negative controls. Included positive controls were two bacterial mock communities and one DNA standard. Included negative controls were DNA extraction controls and sequencing controls. Quality control (QC) passing (DNA concentration and intact genomic fragment) were evaluated to determine extraction method performance. It was expected that positive controls would pass QC, while negative controls would not. Regarding mock communities, all extractions using Zymo and Q passed QC, while for Magna one extraction did not pass QC for both the ATCC mock community and Zymo mock community (Table S3). This was not unexpected, as mock communities were diluted for extraction using Magna and, therefore, DNA concentrations were lower. Negative extraction controls did not pass QC for Q and Magna, but they did for Zymo. This likely represents a higher contamination load during the extraction process for Zymo, which was also reflected by higher DNA concentrations (Table S3). A full overview of all samples included in this study, their QC passing and DNA concentrations can be found in Table S4.

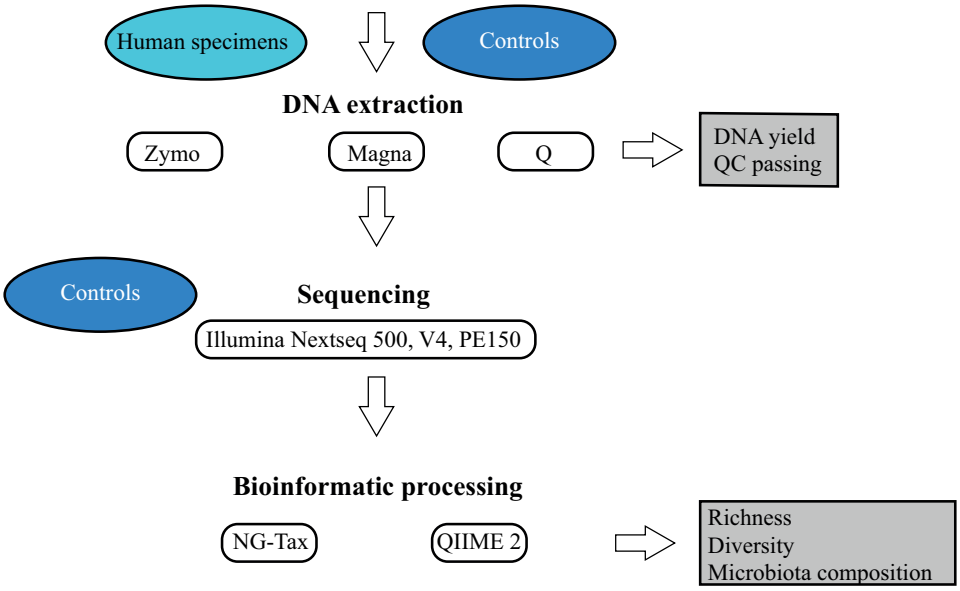


Figure 1: Study design workflow. DNA was extracted from human specimens and positive and negative controls using three different DNA extraction methods. DNA extraction performance was assessed on DNA yield and QC passing. Extracted DNA, and positive and negative sequencing controls were sequenced. Raw sequencing data was processed using two bioinformatic pipelines. Performance was assessed on microbiota composition, richness and diversity.

Positive controls: Classification, richness, diversity and relative species abundance

Primer choice in combination with bioinformatic pipeline choice may limit correct classification of all bacterial species in mock communities

Performance of the three extraction methods in combination with two bioinformatic pipelines, NG-Tax and QIIME 2, was evaluated on correctly identifying richness, diversity and relative abundances from bacterial mock communities and a DNA standard. Richness and diversity were computed at the OTU level and at genus level. Analysis of compositional profiles was performed at genus level. Both pipelines failed to classify one organism from either mock community; NG-Tax did not detect *Cutibacterium* from the ATCC mock, while QIIME 2 did not detect *Salmonella* from the Zymo mock. The inability to detect *Cutibacterium* is most likely a combination of different internal settings and filtering steps in the computational pipelines and a primer choice issue, since the universal 515F and 806R primers are known to poorly amplify *Cutibacterium acnes* [12]. Poor amplification of *C. acnes* results in limited read numbers, which may be filtered out during bioinformatic processing. These issues could likely be solved by choosing primers targeting different 16S rRNA gene regions, or by using adapted V4 region primers which do allow for accurate amplification of *Cutibacterium* [12, 13]. Regarding

QIIME 2 and the inability to detect *Salmonella*, there was an *Enterobacteriaceae* family with approximately expected relative abundance for *Salmonella*, and we were therefore confident this represented *Salmonella*. This *Enterobacteriaceae* family was subsequently included as *Salmonella*, and designated *Enterobacteriaceae (Salmonella)*. This classification error likely resulted from the fact that *Enterobacteriaceae* members cannot always be discriminated based on the 16S rRNA V4 region [14].

DNA standard and Zymo mock community data can be recovered independent of extraction protocol or pipeline

The Zymo mock and DNA standard consist of respectively cell material or DNA of eight bacterial species and two fungal species. As the 16S rRNA gene was targeted, fungi should not be detected. Therefore, theoretical richness is eight and theoretical Shannon diversity was calculated to be 2.01.

Regarding the DNA standard, NG-Tax overestimated OTU-based richness for both duplicates, DNA 1 and DNA 2 (Fig 2A, table S3). Richness was however accurately retrieved at genus level (Fig 2C). The same was observed regarding diversity, which was overestimated at the OTU level (Fig 2B), but accurate at genus level (Fig 2D). QIIME 2 approached theoretical richness and diversity values at the OTU level (Fig 2A+B, table S3). Richness slightly improved at genus level (Fig 2C), while diversity did not differ from OTU-based diversity (Fig 2D). Thus, QIIME 2 better estimated richness and diversity at the OTU level, while NG-Tax performed better at genus level (Table S3). This likely stems from NG-Tax finding an inflated richness due to assignment of multiple OTUs from a single organism (e.g. multiple *Enterococcus* OTUs). When OTUs are collapsed at genus level, this is no longer a problem, probably explaining why NG-Tax can perform better at genus level, while performing worse at the OTU level.

Compositional profiles of DNA 1 and DNA 2 are highly similar to theoretical abundance (Fig 3A+B). To quantify differences in compositional profiles, Bray-Curtis dissimilarity and Kullback-Leibler divergence (Fig 3C-F) [15] and fold errors for each taxon (Fig 4) were determined. For the dissimilarity and divergence values, a value of zero represents an identical microbiota composition to the theoretical expectation. NG-Tax obtained values closer to zero than QIIME 2 for both DNA 1 and DNA 2, although the difference is minimal (Fig 3 and Table S2) and the performance of both pipelines can therefore be regarded as equal. A similar conclusion can be drawn from the fold errors (Fig 4), since both pipelines accurately retrieved expected relative abundance, with all genera having a fold error between -1.5 and 1.5 (Table S3).

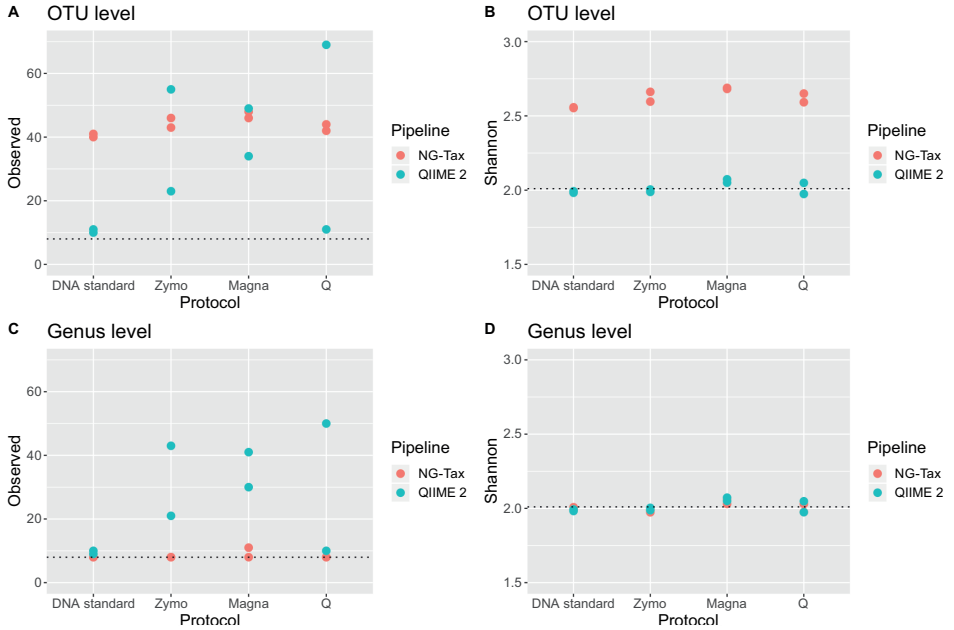


Figure 2: Richness (observed OTUs) and diversity (Shannon) computed for Zymo DNA and Zymo mock at OTU level (A+B) and at genus level (C+D) for each combination of bioinformatic pipeline and DNA extraction method. Dashed lines indicate theoretical values.

Similar analyses were performed for the Zymo mock to evaluate performance of DNA extraction methods in combination with the bioinformatic pipelines. All DNA extraction methods, independent of pipeline, resulted in OTU-based richness above 20 for most samples, far higher than theoretical expectance (Fig 2A). This is especially noteworthy for QIIME 2, as it was highly accurate in retrieving correct richness for the DNA standard, in contrast to NG-Tax. Zymo and Q protocols in combination with NG-Tax retrieved accurate genus level-based richness, while a slightly inflated richness was observed for Magna (Fig 2C). No extraction method was consistent in retrieving correct genus level-based richness in combination with QIIME 2. Regarding diversity, all DNA extractions, independent of pipeline, retrieved highly accurate values at genus level (Table S3). At the OTU level, however, the NG-Tax pipeline resulted in overestimation of diversity independent of DNA extraction method, and can therefore be considered a result of bioinformatic processing. Magna extraction resulted in Bray-Curtis and Kullback-Leibler values closer to zero than Zymo and Q, independent of pipeline (Fig 3C-F and Table S3). A similar conclusion can be drawn from the fold errors, which are lowest for Magna and pipeline-independent (Fig 4 and Table S3).

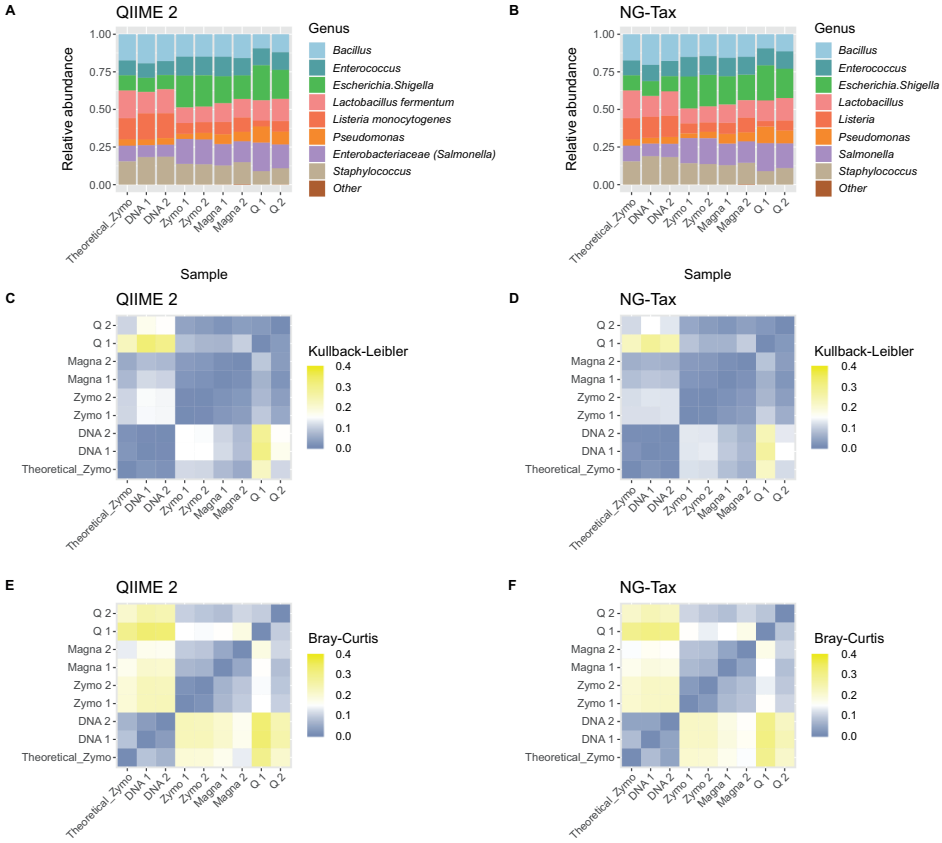


Figure 3: Compositional profiles at the genus level for QIIME 2 (A) and NG-Tax (B) for Zymo mock, theoretical composition is indicated in the first bar graph. Comparison of compositional profiles expressed by Kullback-Leibler divergence (C+D) and Bray-Curtis dissimilarity (E+F) per pipeline. QIIME 2 results are shown in figure C+E, NG-Tax results are shown in figure D+F. For both Kullback-Leibler and Bray-Curtis measures, 0 indicates an identical compositional profile, while higher numbers indicate more dissimilar profiles.

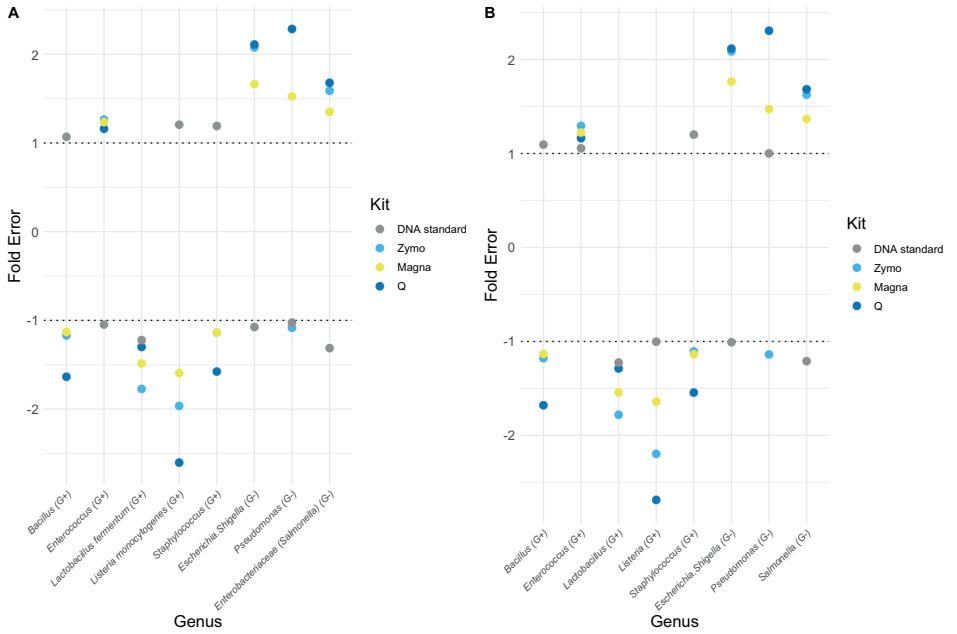


Figure 4: Fold error per bacterium as compared to theoretical values for QIIME 2 (A) and NG-Tax (B). Genera are ordered based on being Gram-positive or Gram-negative. A value above 1 represents overestimation, and a value below -1 represents underestimation.

Taken together, results obtained from the DNA standard indicate that QIIME 2 and NG-Tax perform equally well in general, except for overestimation of the OTU level richness and diversity when using NG-Tax. Results obtained from the Zymo mock, which is a better representation of the full procedure for a microbiota study, indicate that richness is most accurate at the genus level using Zymo or Q in combination with the NG-Tax pipeline. In addition, bacterial microbiota composition profiles are best retrieved using Magna, followed by Zymo, and are pipeline-independent.

In concordance with current literature [9] and independent of extraction method, a general underestimation of Gram-positive bacteria was observed, with *Enterococcus* being the sole exception (Fig 4). This is most likely due to incomplete cell wall lysis of Gram-positive bacteria. Based on the DNA standard and the Zymo mock, we conclude that Zymo and Magna in combination with either pipeline are the best performing combinations (Table S3). However, when high-throughput DNA extraction is required (e.g. for large cohort studies), Magna may be preferred from a practical point of view, although it overestimates richness independent of pipeline.

In general, overestimation of OTUs may stem from the 100% identity setting for clustering, combined with the natural divergence of the 16S rRNA gene [16, 17]. There is no current consensus on OTU identity setting, and cut-offs between 97% and 100% are most commonly used [18]. An advantage of the 100% cut-off is that unique taxa differing a single nucleotide are clustered into different OTUs. A disadvantage is that, as intragenomic diversity in the 16S rRNA gene is common within bacterial genomes, a 100% cut-off can lead to multiple OTUs stemming from a single bacterium and thereby inflate richness [17]. In addition, using a 100% cut-off can theoretically inflate richness due to sequencing errors and requires computational denoising. Apart from biological explanations, the different algorithms and internal filtering steps used in QIIME 2 and NG-Tax can affect the outcome for richness.

ATCC mock is recovered incorrectly, independent of extraction protocol or pipeline
The ATCC mock consists of 20 unique bacterial species, with four of them belonging to two genera (*Staphylococcus* and *Streptococcus*). Therefore, theoretical richness at OTU level would be 20, but eighteen at the genus level. In addition, these 20 unique bacterial species come from different environments, including gut, oral and skin microbiota.

No values close to the theoretical profiles for the ATCC mock for any extraction method/bioinformatic pipeline were observed, and one sample from Q consisted almost entirely of non-classifiable reads (Fig 5), indicating sample-related issues. *Bacillus* was highly overrepresented in all other samples, with a relative abundance over 30% in Zymo and Magna extracted samples, while 6.13% is expected. Curiously, after the first mechanical lysis step in Q, we could culture *Bacillus cereus* and *Cutibacterium acnes* (identification scores of 1.90 and 2.00, respectively), and *Bacillus cereus* (identification score 2.05) after mechanical lysis in Zymo. This is clinically important, as it means that infectious materials cannot be considered safe or non-infectious after mechanical lysis. As culturing of *B. cereus* indicates that cell wall lysis was incomplete, it would be expected that its relative abundance was underestimated, contrarily to what was observed. Another research group recently reported a similar overrepresentation of *Bacillus* in the ATCC community [19]. ATCC itself was also unable to retrieve abundances close to theoretical expectation, neither with 16S rRNA gene amplicon sequencing nor with shotgun sequencing [20]. Several reasons could explain this discrepancy between theoretical profiles and obtained profiles. For example, physical cell-to-cell interactions or presence of different metabolites may interfere with DNA extraction [16, 21]. Therefore, based on this synthetic community, no conclusions on the optimal extraction-pipeline combination could be made. This proposed positive control prompts the question whether mock communities are always reliable for assessing performance of DNA extraction methods. As can be observed from the Zymo mock, DNA extraction kits do not necessarily inflict observed deviations, but may rather be a result of mock community-specific properties.

Outcomes may depend on extraction kit / community type combination, indicating the potential necessity to use a positive control that strongly resembles the investigated microbiome.

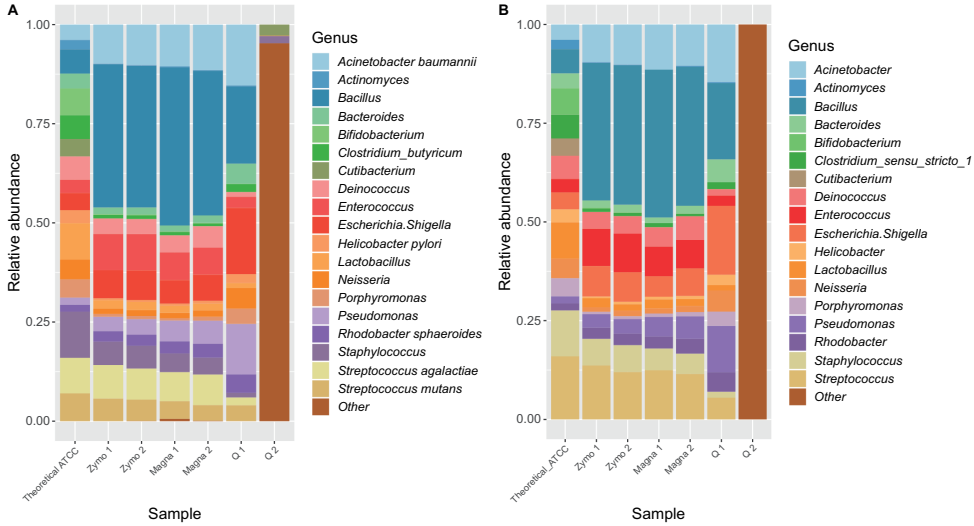


Figure 5: Compositional profiles at the genus level for QIIME 2 (A) and NG-Tax (B) for the ATCC mock. Genus '*Other*' is the sum of the relative abundance of all genera not listed in the legend.

Negative controls: inconsistently contaminated

Negative controls were taken along for each extraction method to check for kit-specific contaminants, which is especially relevant for deciding whether low-biomass samples contain real microbiota. Regarding Zymo, clear kit-contaminants were *Pseudomonas* and *Delftia* (Fig S2A+C), consistent across the different pipelines at the genus level, and with previous findings [11, 22]. For Magna and Q, specific contaminants were less obvious, although *Pseudomonas* was present. Generally, negative controls mostly consisted of genera commonly found in gut and oral microbiota, most of them also previously described as contaminants [11]. In addition, negative sequencing controls were taken along, and here no consistent contaminants could be observed (Fig S2B+D). Potential contamination sources are multifold, such as kit contamination, index hopping, or well-to-well contamination [23, 24]. Index-hopping is however not a likely source of contamination, as the negative control for Magna was sequenced in different lanes, and profiles look highly similar (Fig S2A+C). Additionally, we did not observe index-hopping in our positive controls.

One of the contaminants we identified has not been previously described as a contaminant,

namely *Clostridioides*. This likely represents *C. difficile*, and contamination by this bacterium can be explained by the fact that DNA extractions were performed in our National Reference Laboratory for *C. difficile*, which probably contains minor amounts of *C. difficile* spores during most time points. *C. difficile* contamination on laboratory surfaces has also recently been described in another clinical microbiology laboratory [25].

By incorporating this information with the Zymo positive controls, it can be concluded that Zymo and Magna are most optimal. Magna most accurately captured the expected community profile, while kit-specific contaminants are clear and easy to discriminate from biological signal using Zymo (Table S2).When investigating different biological sample types it would be ideal to use a kit for which kit contaminants do not overlap with the biological signal, e.g. *Pseudomonas* contamination when studying sputum samples from cystic fibrosis patients who are frequently colonized with *Pseudomonas* spp. However, this would require contaminants to be stable across batches, which has been shown to not be the case [22].

Automatic Magna extraction yields lowest DNA for biological samples

Twenty-seven biological samples were available per extraction protocol (Table S1) and Q was most successful in passing QC (22/27), followed by Zymo (20/27) and Magna (17/27) (Table S3), although differences were not statistically significant (Cochran's Q-test, p=0.178). QC passing was based on DNA concentration and intact genomic fragments. DNA concentrations were on average lowest for Magna, while yields were comparable between Q and Zymo (Figure S1). Processing of raw sequencing data from biological samples was performed using the NG-Tax pipeline at the genus level.

Fecal microbiota analysis is only slightly affected by the applied DNA extraction methods

DNA extracted from fecal samples using the three different protocols all passed QC. Magna, Zymo and Q achieved an average concentration of approximately 29 ng/µl, 111 ng/µl and 212 ng/µl, respectively (Fig. S1). While DNA yield varied between extraction methods, all were sufficient for sequencing. Microbiota profiles were comparable between extraction methods for each sample (Figure S3A). In addition, differences in compositional profiles were quantified using Kullback-Leibler divergence (Figure 6A). This heatmap shows that technical variation induced by DNA extraction method is much lower than biological variation between feces samples. Profiles of the feces donors contained many bacterial genera commonly present in fecal microbiomes [26, 27]. Healthy fecal microbiomes largely consist of Bacteroidetes and Firmicutes phyla (~90%), while Actinobacteria and Proteobacteria are present in smaller proportions. At the genus level, *Bacteroides*, *Prevotella* and *Faecalibacterium* are among the most prevalent genera [3], all of which were found in high abundance herein.

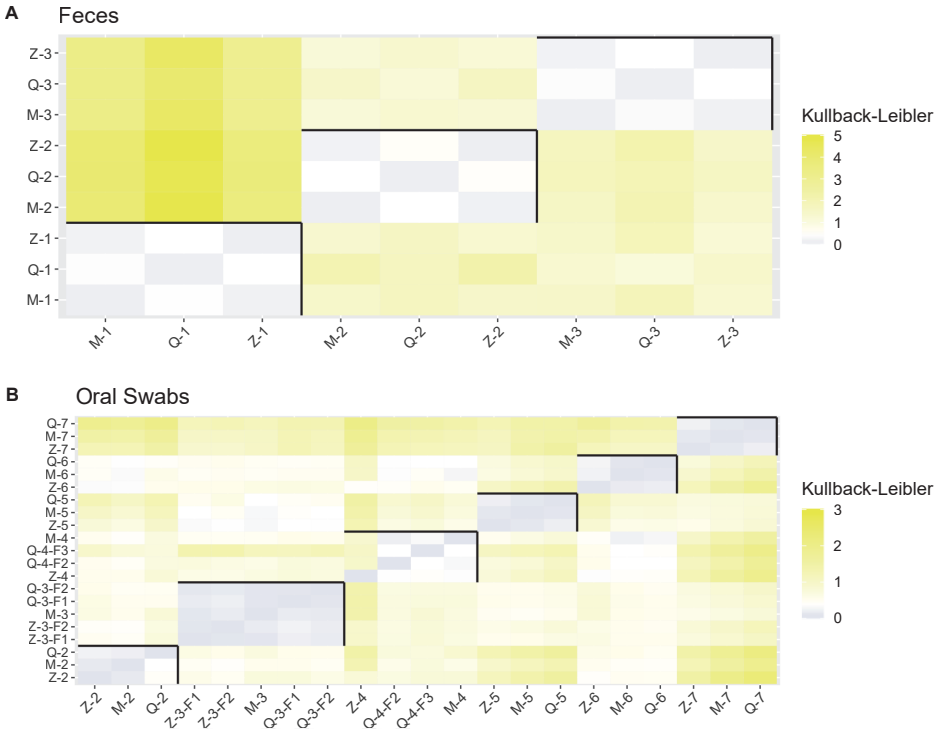


Figure 6: Kullback-Leibler divergence heatmap of feces (A) and oral swabs (B). Black lines group unique biological samples. Gray indicates highly similar composition, while yellow indicates divergence in composition. F1-F2-F3 represent samples which have been sequenced in duplicate, but on different flow cells.

Microbiota profiles of oral swabs are consistent, despite low DNA yields

Out of eighteen DNA extractions, fifteen extractions passed QC for oral swabs. Only for Zymo, all extractions passed QC. DNA yields were highly variable for all extraction methods, ranging from 0.12 to 6.34 ng/µl. Half of the extractions (nine/eighteen) yielded a concentration below one ng/µl. All compositional profiles were dominated by *Streptococcus*, *Prevotella* spp., *Haemophilus* and *Veillonella*, which was individual-independent. In addition, technical variation induced by DNA extraction and subsequent steps was lower than biological variation (Fig 6B). The oral microbiota, like the gut microbiota, is highly diverse. Nevertheless, a certain core of genera (e.g. *Streptococcus* spp. and *Prevotella* spp.*)* is present in most people, all of which were found in our study [3, 28, 29]. Together, the good QC passing rate, DNA concentrations and consistency of compositional profiles between extraction methods lead us to conclude that all three methods work well for oral swabs.

Applied methodology yields inconsistent results for the urine microbiota

During the last decade, microbiota studies showed that urine contains a bacterial microbiota [30, 31]. Despite using 30-40 ml of urine and centrifugation prior to extraction [32], we were not able to convincingly capture a urinary microbiota for all samples (Fig S3C). DNA concentrations were high for an infected sample (between thirteen and 42 ng/µl), but concentrations for the other samples were between 0.11 and 0.99 ng/µl. Six out of nine samples passed QC. For the infected sample with a high bacterial load, we were able to classify the cause of infection to *Enterobacteriaceae*, which is in agreement with the fact that most UTIs are caused by members of *Enterobacteriaceae*. One urine sample showed high similarity to negative controls for respective kits, with non-classifiable reads for Q and Magna, and high relative abundance of *Pseudomonas* for Zymo (Fig S3C). Another urine sample contained a high *Lactobacillus* relative abundance, which has previously been shown to be prevalent in urine samples [31]. *Lactobacillus* spp. could be cultured in 15% of urine samples collected by a transurethral catheter and was thereby the most prevalent genus cultured [31]. Another small-scale study found that in five out of six patients, *Lactobacillus* was detected in midstream urine samples and its relative abundance was between 22 and 80% [30]. In addition, presence of *Atopobium*, *Gardnerella*, *Prevotella* and *Anaerococcus* point towards an existing urinary microbiota [33]. However, *Pseudomonas,* a common Zymo kit contaminant, was still found in this urine sample, and for Magna more than 25% of reads could not be classified (Fig S3C). This could indicate that the biological signal is not much stronger than contamination, and therefore a mixed profile is observed. Further efforts and method optimization should be undertaken, although this can be difficult to implement in routine work [34]. In addition, culturing could be used as a follow-up method to confirm that contaminants are not viable bacteria, but rather bacterial DNA.

Saliva samples with long storage time and multiple freezing-thawing cycles seem unsuitable for microbiota research

DNA yield from included saliva samples was lower as compared to literature [35, 36] (Fig S1). Only a single DNA extraction had a concentration of slightly above one ng/µl (1.18; Table S4), while all other extractions had concentrations between 0.04 and 0.68 ng/µl. This is most likely associated with storage duration (~fifteen years) and the fact that samples were thawed and refrozen several times. This also explains why only three out of nine DNA extractions passed QC. The included saliva samples were chosen as investigators within our facility were interested to see if microbiota studies could be performed using these samples. Compositional profiles consisted of a mixture of genera present in the normal oral microbiota (*Oribacterium*, *Prevotella_7*, *Prevotella_9* and *Streptococcus*) [3], genera present in our negative controls (*Pseudomonas*, *Delftia*) and non-classifiable reads (Fig S3D). In combination with low DNA yields, it is likely that a mixture between biological signal and contamination signal is present. Therefore,

we consider the applied extraction methods unsuitable for saliva samples with a long duration of storage time and multiple freezing-thawing cycles.

The colorectal cancer microbiota present in biopsies was indistinguishable from negative controls or fecal microbiota

As colorectal cancer development has been associated with specific gut bacteria, we were interested to see if colorectal cancer tissue itself also contained bacteria [37, 38]. DNA concentrations were sufficient for all samples to pass QC, but extracted DNA was likely mostly human-derived. Two of three extraction methods were not successful, as samples extracted using Zymo and Magna showed high similarity to their respective negative controls (Fig S3E). Using Q, *Bacteroides*, *Fusobacterium* and *Gemella* were identified, all being previously associated with colorectal cancer development [37, 39]. Several gut commensals, including *Faecalibacterium* and *Escherichia-Shigella* were present in both the negative controls and these colorectal cancer samples. It is therefore difficult to discriminate whether these are contaminant bacteria, or whether they represent biological signal.

We hypothesized that by spinning down the material, the supernatant would contain more bacteria than the cancer tissue. DNA concentrations of supernatant were between 0.16 and 2.32 ng/µl, and seven out of nine DNA extractions passed QC (Table S4). For one sample, it was clear that across all methods many genera were observed which were present in negative controls (e.g. *Pseudomonas*), or reads could not be classified at all (Fig S3F). A second sample seemed to contain a real microbiota. Profiles were consistent across extraction methods, did not contain many contaminants and had specific bacteria previously linked to colorectal cancer (e.g. *Fusobacterium*) [37]. The third sample showed a profile reflecting a mix between biological signal and technical contamination. Profiles were consistent across methods and contained genera representative of a gut microbiota, but also contained non-classifiable reads and contamination. Therefore, profiles are likely a mixture of biological signal and technical contamination, and further optimization is necessary prior to using this sample type for experimental studies. We have the same recommendation for colorectal cancer sample types as for urine, as discussed above.

It remains unclear whether HPV-negative vulvar squamous cell carcinoma biopsies contain a bacterial microbiota

Vulvar squamous cell carcinoma (VSCC) has different etiological pathways, of which one is associated with human papilloma virus (HPV). The counterpart is non-virally related and is frequently associated to lichen sclerosis, a benign chronic inflammatory lesion and *TP53* mutations [40, 41]. We extracted DNA from HPV-negative VSCC tissue as a pilot study to determine if investigating the relationship between bacterial microbiota and HPV-negative VSCC would be potentially feasible. DNA concentrations were high

(Fig S1), only for three extractions below one ng/µl, and eight out of nine extractions passed QC. However, DNA was probably again largely human-derived. This was reflected in the obtained microbiota profiles, as most reads were not classified or the profiles showed high similarity to negative controls (e.g. high abundance of *Pseudomonas*) (Fig S3G). Therefore, it is unlikely that this cancer tissue contains bacteria, or bacteria are so lowly abundant that they are overshadowed by contamination load. In general, the vulvar microbiota has not been extensively studied. A recent study on vulvar microbiota observed that *Lactobacillus*, *Corynebacterium*, *Finegoldia*, *Staphylococcus and Anaerococcus* are most abundant on this body site, but the use of negative controls was not reported [42]. These genera are also part of the vaginal microbiota, and might be sampling contamination or reflect high similarity between vulvar and vaginal microbiota. A large amount of formalin-fixed VSCC materials are stored in a biobank at our facility. To investigate whether this sample collection could be used for microbiota profiling, DNA was extracted from three formalin-fixed VSCC samples. DNA concentrations were all below 0.3 ng/µl, and only two out of nine extractions passed QC (Table S4). One sample extracted with Q was excluded from further analysis, as no reads were present after sequencing. Extraction and sequencing of formalin-fixed material poses additional problems, as DNA molecules could be highly fragmented and too short for amplicon sequencing of the V4 region [43]. For Zymo, samples resembled negative controls, with *Delftia* and *Pseudomonas* being highly abundant (Fig S3H). The same samples had completely different microbiota profiles when using protocol Q or Magna. Both extraction methods showed genera commonly found in the lower urogenital tract, including *Streptococcus*, *Prevotella* and *Gordonia* [3, 27]. However, many of these genera were also detected in negative controls. In combination with low DNA yield and inconsistent profiles across extraction methods, we conclude that no reliable bacterial microbiota profile could be identified in these samples. For both VSCC types, we suggest the same way forward as for urine samples.

Sample groups with and without biological signal cluster apart

Lastly, we performed t-distributed stochastic neighbor embedding (t-SNE) clustering using Bray-Curtis measures on all samples used in the present study (Fig 7) [44]. Based on microbiota composition as measured by Bray-Curtis, t-SNE projects points in a two-dimensional space, while maintaining local structures present in high-dimensional space. Clear clusters could be identified for Zymo positive controls, feces, oral swabs and ATCC mock (all but one sample) (Fig 7). Other biological samples and negative controls were more dispersed throughout the plot, indicating that either more biological or technical variation was present. This is in agreement with our detailed analysis, showing that their microbiota cannot necessarily be distinguished from the negative controls. This highlights the importance of including negative controls in microbiota studies, which has previously been shown in two studies aiming to unravel the placental microbiota [45, 46],

and is increasingly recognized in the field. It is currently unclear whether a placental microbiota exists, but when comparing placental samples of healthy deliveries to included negative controls, microbiota compositions could not be distinguished [45, 46].

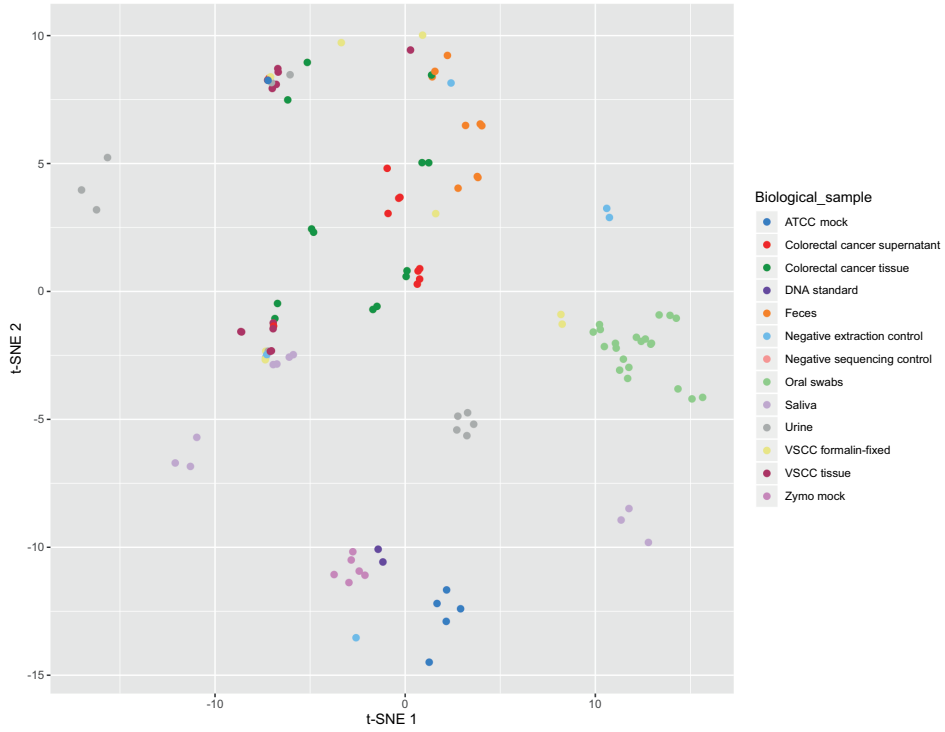


Figure 7: Bray-Curtis distance measures visualized by t-distributed stochastic neighbour embedding **(**t-SNE) for all samples. Each dot in the plot represents a single sample, and short distances between samples indicate high similarity.

Strengths and limitations

The current study had several strengths and limitations. By using a positive control of cell material with a corresponding DNA standard, we differentiated variation induced from sequencing procedures and DNA extraction. We demonstrate the importance of using positive and negative controls in microbiota studies, and show that negative controls are crucial for interpretation of low-biomass samples. Another strength of the study was that for several higher biomass biological samples (feces and oral swabs), we showed that technical variation was much smaller than biological variation. A shortcoming of the study is that we did not perform any other quantification next to 16S

rRNA gene sequencing (e.g. qPCR), which may be particularly useful for quality control of the ATCC mock. Furthermore, the current study used only three unique samples of most biological sample types. Especially for samples for which DNA extraction was challenging (urine samples, colorectal cancer supernatant), a higher number of unique samples would have allowed for a more thorough evaluation.

Conclusion

The current study evaluated three DNA extraction methods and two bioinformatic pipelines for bacterial microbiota profiling using several positive and negative controls, and a range of biological specimens. All three extraction methods quite accurately retrieved theoretical abundance of the Zymo mock, but not of the ATCC mock. For DNA extraction, we recommend using the Zymo and Magna protocol, since they showed good overall performance for all samples. Sequencing procedure only induced minor variation, as shown using a DNA standard. We furthermore showed that the NG-Tax and QIIME 2 pipelines perform equally well overall, each having their specific flaws.
By including negative controls and comparing these with low-biomass samples, we evaluated whether low-biomass samples consisted of technical noise, biological signal or a mixture. In most cases, identification of a unique microbiota was not achieved, highlighting the importance of negative controls and sufficiently sensitive methods. The results from this study can help other microbiome study groups to select an appropriate DNA extraction method and bioinformatic pipeline. Lastly, we hope this study contributes to further awareness of the usage of controls, especially when studying low-biomass samples.

Materials and Methods

Sample collection and pre-processing
Eight different biological specimens were included in this study, namely feces, urine, saliva, oral swabs, colorectal cancer tissue, colorectal cancer supernatant, vulvar squamous cell carcinoma tissue and formalin-fixed vulvar squamous cell carcinoma. Of each biological specimen, three unique samples were included. Only for oral swabs, six unique samples were included (Table S1). These samples were anonymized and treated according to the medical ethical guidelines described in the Code of Conduct for Proper Secondary Use of Human Tissue of the Dutch Federation of Biomedical Scientific Societies. A detailed overview of sample types, sample processing and storage conditions can be found in Table S1.

Mock communities and DNA standard

Two mock communities (ZymoBiomics Microbial Community Standard, Zymo Research, Irvine, California, USA and 20 Strain Even Mix Whole Cell Material ATCC® MSA2002™, ATCC, Wesel, Germany) were included as positive controls for DNA extraction. Exact composition and relative abundances of 16S rRNA gene copies was provided on the product sheet for ZymoBiomics Microbial Community standard (hereafter referred to as Zymo mock), while for ATCC® MSA2002™ (hereafter referred to as ATCC mock) we calculated expected 16S rRNA gene profiles based on genomic information (Table S2). ZymoBiomics Microbial Community DNA Standard (hereafter referred to as DNA standard) was taken along as a positive sequencing control.

DNA extraction

Procedures

Cancer samples were pre-processed for DNA extraction comparably to a recent study on pancreatic cancer microbiota [47], urine samples according to a recent publication on how to study urinary microbiota [32] and other samples according to in-house methods for sample processing (Table S1). For solid cancer samples, the beating steps during pre-processing were performed using a Qiagen TissueLyser LT (Qiagen Benelux, Venlo, the Netherlands) at 50Hz for one minute (Table S1). As single saliva samples did not contain sufficient volume for multiple extractions, several samples of the same individual were pooled to obtain the appropriate volume. DNA was extracted in duplicate from three unique samples for each biological material, only for oral swabs from six unique samples, and from the two mock communities. DNA was extracted using three different extraction protocols (see Protocols section), and for each protocol a negative extraction (no sample) was included in duplicate. The DNA standard was taken along in duplicate. DNA was quantified using a Qubit 3.0 Fluorometer (Invitrogen, Breda, the Netherlands) and the Qubit™ dsDNA HS Assay Kit (Thermo Fisher, Landsmeer, the Netherlands). A schematic overview of the study setup is shown in Figure 1.

DNA extraction protocols

Detailed protocols, including all minor adaptations, are present in Supplementary Methods. DNA extraction was performed using three methods: 1) the Quick-DNA Fecal/Soil Microbe kit (hereafter referred to as Zymo) (Zymo Research) according to manufacturer instructions with minor adaptations, 2) protocol Q (hereafter referred to as Q) [9] and 3) automated DNA extraction with MagNA Pure 96 ™ (hereafter referred to as Magna) (Roche Diagnostics, Almere, the Netherlands) using the MagNA Pure 96 DNA and viral NA small volume kit (Roche Diagnostics), according to standard operating procedures with minor adaptations. Mock communities were diluted to 10^4-10^5 cells per sample for extraction using Magna. For Q, several buffers and other materials were not provided in the kit and therefore purchased elsewhere, namely BeadBug™ prefilled

tubes with 2.0 mL capacity and 0.1 mm Zirconium beads (Sigma-Aldrich, Zwijndrecht, the Netherlands), RNase A, DNase and protease-free water (10 mg/mL) (Thermo Fisher, the Netherlands) and TE buffer (Thermo Fisher).

MALDI-TOF Mass Spectrometry (Biotyper)

To verify whether all bacteria of the ATCC mock were lysed after the first mechanical lysis step of both Zymo and Q, the lysate was plated on a tryptic soy agar plate containing 5% sheep (VWR International, Amsterdam, the Netherlands), and aerobically and anaerobically incubated at 37°C for five days. The MALDI Biotyper system was used (Bruker Daltonics, Germany) to identify the bacterial species. Samples were prepared in the following way: A bacterial colony was taken from the culturing plate and spread in duplicate on single spots on a Bruker polished steel targetplate. Subsequently, one µl of 70% formic acid was added on each single spot and when dried, one µl prepared Bruker Matrix HCCA according to clinical laboratory protocols was added per spot. The Bruker polished steel targetplate was then used for MALDI-TOF MS Biotyper analysis.

Library preparation and 16S rRNA gene amplicon sequencing

Of each duplicate DNA extraction from biological specimens, the duplicate with highest genomic DNA concentration was used for sequencing. Duplicate samples from controls were both sequenced. Quality control, library preparation and sequencing were performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEXTflex™ 16S V4 Amplicon-Seq Kit (BiooScientific, TX, USA) and Illumina NextSeq 500 (paired-end, 150bp) according to their standard operating procedures. QC passing was based on intact genomic DNA and DNA concentrations measured by GenomeScan B.V. Therefore, those DNA concentrations were used for downstream analysis. Several samples were sequenced on multiple lanes, which is indicated in all relevant figures and tables.

Sequencing data analysis

Read filtering, operational taxonomic unit (OTU)-picking and taxonomic assignment were performed using two different bioinformatic pipelines, QIIME 2 and NG-Tax 0.4 [48, 49], both using the Silva_132_SSU Ref database for taxonomic classification [50]. For both pipelines, a read length of 120 bp was chosen based on quality of reads. The following settings were applied for QIIME 2: forward and reverse read length of 120 bp, quality control using Deblur, identity level of 100% (default). The following settings were applied for NG-Tax: forward and reverse read length of 120 bp, ratio OTU abundance of 2.0 (default), classify ratio of 0.9 (default), minimum threshold of 0.1% (default), identity level of 100% (default), error correction of 98.5 (default). Prior to the NG-Tax run, potential left over primers were removed with cutadapt v. 1.9.1 [51], in paired-end mode, with additional setting -e 0.2 (increased error tolerance, 20%). This setting was required since NG-Tax first creates a smaller custom database, based on the

used primers. During further processing, data has to be primer sequence free, as the primer sequence is removed from the smaller database. Furthermore, all sequences with any deviating barcode in the fastq header were changed to the original barcode to allow inclusion into the NG-Tax pipeline.

The obtained OTU-tables were filtered for OTUs with a number of sequences less than 0.005% of the total number of sequences [52]. Downstream analysis was performed in R (v3.6.1), mainly using the phyloseq (v.1.28.0), microbiome (v.1.6.0) and ggplot2 (v.3.2.0) packages [53-55]. Alpha diversity was computed at both the OTU and genus levels, while analysis of compositional profiles was performed at the genus level. Kullback-Leibler divergence and Bray-Curtis dissimilarity measure heatmaps were computed by first deleting genera that had a relative abundance of zero in all investigated samples (positive controls, feces and oral swabs) and subsequent calculation of the respective measure. All R code is available upon request from the corresponding author.

Notes

Data accessibility

All raw sequencing data used in the current study are deposited in the European Nucleotide Archive with accession number PRJEB34118.

Acknowledgements

We thank all collaborating partners who provided us with clinical biospecimens, namely Liz Terveer, Eric Berssenbrugge, Erik Giltay, Noel de Miranda, Jitske van den Bulk, Natalja ter Haar and Kim Kortekaas. We also thank Eric Claas for support with the use of the MagNA Pure 96 ™ and the clinical diagnostic laboratory for identification of bacteria with MALDI-TOF.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. BH and EK are supported by an unrestricted grant from Vedanta Biosciences Inc. EK has performed research for Cubist, Novartis and Qiagen, and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur and Seres Therapeutics. The companies had no role in the study and writing of this manuscript.

Author contributions

QD, BH, AG, EK and RZ conceptualized and designed the study. QD and AG performed practical laboratory work. BH and RZ processed raw sequencing data. QD analyzed

data, prepared figures and wrote the manuscript under supervision of BH and RZ. All authors interpreted data, read and revised drafts of the manuscript, and approved the final version.

References

1. Guarner F, Malagelada JR. 2003. Gut flora in health and disease. Lancet 361:512-9.
2. Rajilic-Stojanovic M, Smidt H, de Vos WM. 2007. Diversity of the human gastrointestinal tract microbiota revisited. Environ Microbiol 9:2125-2136.
3. Human Microbiome Project C. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207-14.
4. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Microbiome Quality Control Project C, Abnet CC, Knight R, White O, Huttenhower C. 2017. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 35:1077-1086.
5. Fouhy F, Deane J, Rea MC, O'Sullivan O, Ross RP, O'Callaghan G, Plant BJ, Stanton C. 2015. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. PLoS One 10:e0119355.
6. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, Thomson JM, Consortium UIG, Satsangi J, Flint HJ, Parkhill J, Lees CW, Hold GL. 2014. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. PLoS One 9:e88982.
7. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. Microbiome 3:26.
8. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res 43:e37.
9. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, Kultima JR, Hayward MR, Coelho LP, Allen-Vercoe E, Bertrand L, Blaut M, Brown JRM, Carton T, Cools-Portier S, Daigneault M, Derrien M, Druesne A, de Vos WM, Finlay BB, Flint HJ, Guarner F, Hattori M, Heilig H, Luna RA, van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'Toole PW, Orvain C, Patil KR, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, Slezak K, Veiga P, Versalovic J, Zhao L, Zoetendal EG, Ehrlich SD, Dore J, Bork P. 2017. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 35:1069-1076.
10. Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, Dore J, Guarner F, Manichanh C. 2014. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiol 14:112.
11. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. 2019. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol 27:105-117.
12. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, Knight R. 2011. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet 13:47-58.
13. Zeeuwen P, Boekhorst J, Ederveen THA, Kleerebezem M, Schalkwijk J, van Hijum S, Timmerman HM. 2017. Reply to Meisel et al. J Invest Dermatol 137:961-962.
14. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, Wong GK. 2016. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Front Microbiol 7:459.
15. Kullback S, Leibler RA. 1951. On Information and Sufficiency. The Annals of Mathematical Statistics 22:79-86.
16. Hornung BVH, Zwittink RD, Kuijper EJ. 2019. Issues and current standards of controls in microbiome research. FEMS Microbiol Ecol 95.
17. Sun DL, Jiang X, Wu QL, Zhou NY. 2013. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. Appl Environ Microbiol 79:5962-9.
18. Patin NV, Kunin V, Lidstrom U, Ashby MN. 2013. Effects of OTU clustering and PCR artifacts on microbial diversity estimates. Microb Ecol 65:709-19.
19. Scherz V, Aeby S, Bertelli C, Greub G. 2019. Microbiota profiling of saliva: unexpected results of a preliminary study, abstr ECCMID, Amsterdam, 2019.
20. ATCC. 2018. Microbiome standard & research solutions. https://www.lgcstandards-atcc.org/~/media/A36E4E940E3F40B1A5E14D6851E9348F.ashx. Accessed April 10th 2019.

21. Angelakis E, Bachar D, Henrissat B, Armougom F, Audoly G, Lagier JC, Robert C, Raoult D. 2016. Glycans affect DNA extraction and induce substantial differences in gut metagenomic studies. Sci Rep 6:26276.
22. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87.
23. Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, Granger B, Green L, Howd T, Mason T, Vicente G, Dasilva M, Brodeur W, DeSmet T, Dodge S, Lennon NJ, Gabriel S. 2018. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics 19:332.
24. Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R. 2019. Quantifying and Understanding Well-to-Well Contamination in Microbiome Research. mSystems 4.
25. Reigadas E, Vazquez-Cuesta S, Onori R, Villar-Gomara L, Alcala L, Marin M, Martin A, Munoz P, Bouza E. 2019. Clostridioides difficile contamination in the environment of a clinical microbiology laboratory and laboratory workers. Clin Microbiol Infect doi:10.1016/j.cmi.2019.06.027.
26. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD, Wang J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464:59-65.
27. Lloyd-Price J, Abu-Ali G, Huttenhower C. 2016. The healthy human microbiome. Genome Med 8:51.
28. Lamont RJ, Koo H, Hajishengallis G. 2018. The oral microbiota: dynamic communities and host interactions. Nat Rev Microbiol 16:745-759.
29. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. 2005. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 43:5721-32.
30. Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K, Fitzgerald M, Mueller ER, Schreckenberger P, Dong Q, Nelson DE, Brubaker L. 2012. Evidence of uncultivated bacteria in the adult female bladder. J Clin Microbiol 50:1376-83.
31. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. J Clin Microbiol 52:871-6.
32. Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, Gregory WT, Nardos R, McWeeney SK. 2018. Community profiling of the urinary microbiota: considerations for low-biomass samples. Nat Rev Urol 15:735-749.
33. Bucevic Popovic V, Situm M, Chow CT, Chan LS, Roje B, Terzic J. 2018. The urinary microbiome associated with bladder cancer. Sci Rep 8:12157.
34. Boers SA, Hays JP, Jansen R. 2017. Novel micelle PCR-based method for accurate, sensitive and quantitative microbiota profiling. Sci Rep 7:45536.
35. Lim Y, Totsika M, Morrison M, Punyadeera C. 2017. The saliva microbiome profiles are minimally affected by collection method or DNA extraction protocols. Sci Rep 7:8523.
36. Vesty A, Biswas K, Taylor MW, Gear K, Douglas RG. 2017. Evaluating the Impact of DNA Extraction Method on the Representation of Human Oral Bacterial and Fungal Communities. PLoS One 12:e0169877.
37. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM, Brenner D, Fuchs CS, Meyerson M, Garrett WS. 2013. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe 14:207-15.
38. He Z, Gharaibeh RZ, Newsome RC, Pope JL, Dougherty MW, Tomkovich S, Pons B, Mirey G, Vignard J, Hendrixson DR, Jobin C. 2019. Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin. Gut 68:289-300.
39. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, Tsoi KKK, Wong MCS, Tse G, Chan MTV, Chan FKL, Ng SC, Wu JCY, Wu WKK, Yu J, Sung JJY, Wong SH. 2018. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. Gastroenterology 155:383-390.e8.
40. Abdulrahman Z, Kortekaas KE, De Vos Van Steenwijk PJ, Van Der Burg SH, Van Poelgeest MI. 2018. The immune microenvironment in vulvar (pre)cancer: review of literature and implications for immunotherapy. Expert Opin Biol Ther 18:1223-1233.
41. van de Nieuwenhof HP, van der Avoort IA, de Hullu JA. 2008. Review of squamous premalignant vulvar

lesions. Crit Rev Oncol Hematol 68:131-56.
42. Vongsa R, Hoffman D, Shepard K, Koenig D. 2019. Comparative study of vulva and abdominal skin microbiota of healthy females with high and average BMI. BMC Microbiol 19:16.
43. Fuks G, Elgart M, Amir A, Zeisel A, Turnbaugh PJ, Soen Y, Shental N. 2018. Combining 16S rRNA gene variable regions enables high-resolution microbial community profiling. Microbiome 6:17.
44. van der Maaten L, Hinton G. 2008. Visualizing Data using t-SNE. J Mach Learn Res 9:2579-2605.
45. Lauder AP, Roche AM, Sherrill-Mix S, Bailey A, Laughlin AL, Bittinger K, Leite R, Elovitz MA, Parry S, Bushman FD. 2016. Comparison of placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. Microbiome 4:29.
46. de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS, Smith GCS. 2019. Human placenta has no microbiome but can contain potential pathogens. Nature doi:10.1038/s41586-019-1451-5.
47. Pushalkar S, Hundeyin M, Daley D, Zambirinis CP, Kurz E, Mishra A, Mohan N, Aykut B, Usyk M, Torres LE, Werba G, Zhang K, Guo Y, Li Q, Akkad N, Lall S, Wadowski B, Gutierrez J, Kochen Rossi JA, Herzog JW, Diskin B, Torres-Hernandez A, Leinwand J, Wang W, Taunk PS, Savadkar S, Janal M, Saxena A, Li X, Cohen D, Sartor RB, Saxena D, Miller G. 2018. The Pancreatic Cancer Microbiome Promotes Oncogenesis by Induction of Innate and Adaptive Immune Suppression. Cancer Discov 8:403-416.
48. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ, Smidt H. 2016. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes[version 1; referees: 2 approved with reservations, 1 not approved]. F1000Research 5.
49. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson IIMS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints 6:e27295v2.
50. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.
51. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:3.
52. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods 10:57-9.
53. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217.
54. Lahti L, Shetty S. 2017. Tools for microbiome analysis in R. Microbiome package version 1.2.1,
55. Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media.

Chapter 5

Development of a novel computational tool for profiling of carbohydrate-active enzymes (CAZymes) in the human gut and its application in colorectal cancer cohorts

Unpublished data, manuscript in preparation

Q. R. Ducarmon[a,b,c,*], N. Karcher[a,*], H.L.P. Tytgat[d], C. Schudoma[a], G. Zeller[a]

[a]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany.
[b]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands.
[c]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands.
[d]Nestlé Institute of Health Sciences, Nestlé Research, Société des Produits Nestlé S.A., Lausanne, Switzerland
*these authors contributed equally

Abstract

Carbohydrate-active enzymes (CAZymes) are essential for the synthesis and breakdown of (complex) glycans and glycoconjugates. They are present in all living species, but are especially diverse in bacteria. In the gut microbiome, CAZymes are crucial for metabolizing complex carbohydrates of dietary and host origin, such as fiber and mucins, respectively. Currently, dbCAN2 is the most widely used computational tool for annotation of CAZymes in genomic data, but it cannot be directly applied to metagenomic data. dbCAN2 can identify protein sequences that are similar to those present in the CAZy database (the most comprehensive data and knowledge base on CAZymes) using Hidden Markov models (HMMs) specifically built for each CAZyme (sub-)family. However, detection accuracy (E-values cutoffs) has not been calibrated for these HMMs. A second challenge for wide application of this tool for metagenome analysis is the lack of systematic substrate annotations for CAZyme families, which are currently primarily grouped based on amino acid sequence similarity. A hierarchical annotation of CAZyme substrates would however be needed for functional interpretation of CAZyme profiles. To close this gap, the main aim of this study was to build the first tool for computing CAZyme profiles from shotgun metagenomic data, which can be interpreted in terms of substrate specificities. This entailed optimization of HMM E-values for precise detection of CAZymes and construction of a novel hierarchical substrate scheme to facilitate functional interpretation. Application of this tool using data from eight different colorectal cancer (CRC) cohorts revealed that CRC metagenomes were enriched in microbial CAZymes involved in glycosaminoglycan metabolism (p-value 7.44e^-04) and in peptidoglycan metabolism (3.52e^-02), and depleted in CAZymes involved in dietary fiber metabolism (p-value 3.68e^-04) as compared to control metagenomes, suggesting that known dietary risk factors, such as increased meat consumption/decreased fiber consumption in CRC, are reflected in the gut microbial CAZy repertoire.

Introduction

Carbohydrate-active enzymes (CAZymes) are a diverse group of enzymes which can build up and break down glycans and glycoconjugates, consisting of e.g. glycosyltransferases and glycosyl hydrolases. They are present in all domains of life, but an exceptionally diverse set is encoded in microbial genomes, thereby contributing to the extraordinary metabolic versatility of microorganisms[1]. Microbes present in the human gut possess a variety of CAZymes which, amongst others, aid in metabolizing (complex) carbohydrates consumed from the diet, one of the key functions of the gut microbiome[2]. As the human genome encodes for only 17 enzymes with limited capacity (involved in breakdown of sucrose, lactose and starch) for carbohydrate degradation[1], microbial CAZymes are imperative for metabolizing diet-derived complex carbohydrates.

Fermentation of these complex carbohydrates leads to production of metabolites including short-chain fatty acids (SCFAs), of which butyrate, propionate and acetate form 90-95% of the total SCFA pool in the human gut[3]. The contributions of SCFAs to maintaining host health are difficult to overstate, but amongst the most important functions are butyrate being the main energy source for enterocytes[4], regulating gut barrier integrity[5] and being ligands for a variety of receptors present on enteroendocrine and immune cells[5].

Apart from metabolizing complex carbohydrates that are stemming from the diet, CAZymes can also metabolize host glycans such as the mucus layer, with negative consequences in certain conditions. For example, upon exposing fiber-deprived mice to the pathogen *Citrobacter rodentium*, mice suffered from a far higher mortality than mice who were provided a normal diet. This was likely due to the microbiome starting to use host-derived glycans (mucus layer) as the main energy source due to the lack of fiber. This led to mucus layer erosion and thereby to closer proximity of *C. rodentium* to the gut epithelium in these fiber-deprived mice, which likely facilitated infection[6].

In the CAZy database, CAZyme families have been constructed and they are defined based on amino acid sequence similarity. For each family, at least one founding member is required to have been biochemically characterized[7]. These CAZyme families are rigorously curated and extensively evaluated before a new CAZyme family is accepted into the database[7]. This is one of the reasons why the CAZy database is regarded as the gold standard in the field. One drawback of the current classification scheme, based on amino acid sequence similarity, is that glycan substrate annotation is in many cases problematic. A CAZyme family can contain enzymes with similar amino acid sequences, while they can metabolize a variety of substrates[8]. This drawback can, however, not only be ascribed to the current classification scheme. For multiple CAZyme families

no substrate is known and to this day it remains unknown what determines substrate specificity. Nevertheless, the lack of a higher, systematic substrate system for CAZyme families for which substrates are known currently limits functional interpretability with regard to high-level substrate usage.

A main goal of researchers interested in CAZymes is to discover which CAZymes are encoded in their (meta)genomic data and whether CAZyme abundances differ meaningfully between groups of samples. One of the most popular tools to annotate CAZymes in genomic data is the dbCAN2 tool, which uses CAZyme (sub)family-specific Hidden Markov models (HMMs) to annotate gene sequences[9]. While this tool has proven valuable and remains widely used, optimization of E-value cutoffs (which aims to minimize conflicts between functional homology and sequence similarity) was performed on only six genomes[9]. This leaves considerable room for optimizing the detection accuracy of CAZyme annotation. In addition, dbCAN2 is an open-reading frame (ORF) annotation framework and not a profiler, which prevents estimating taxonomic or functional abundance profiles from a shotgun metagenome. From the ORF prediction it is not trivial to estimate (relative) abundances of CAZymes in a given metagenome, especially given that, due to their multimodular nature, multiple CAZyme families can occur within a single ORF. Lastly, to be able to annotate ORFs with CAZymes, researchers would first have to laboriously reconstruct these (via metagenome assembly, ORF prediction, and redundancy removal) from metagenomic data, posing a substantial challenge to many (non-)bioinformaticians.

Here, we aimed to develop the first easy-to-use tool to profile CAZymes from short-read metagenomic data and to optimize E-values for more accurate detection of CAZymes using HMMs. In addition, we suggest a novel substrate annotation scheme allowing for improved interpretation of the resulting CAZyme profiles in terms of broader substrate groups these act on. Lastly, we applied our tool on eight different colorectal cancer (CRC) cohorts to uncover novel associations between specific CAZymes, substrate metabolism and CRC.

Materials and methods

Generation of CAZyme (sub)family module sequence set

In order to obtain family-wise CAZy modules, we first downloaded all CAZy sequences (http://bcb.unl.edu/dbCAN2/download/Databases/V9/ dbCAN HMMdb release 9.0 and CAZyDB released on 07/30/2020). We then identified CAZy modules on these sequences using the dbCAN HMMs (from the same dbCAN2 release) using an E-value threshold of 1e-15 and a coverage threshold of 0.35 (default cut-offs on dbCAN server). We then

generated (sub)family-wise multiple sequence alignments of module sequences using mafft for the 29 largest families (containing most sequences) and mafft-linsi for the remaining sequences on representative sequences obtained from mmseqs2 clustered at 99% (clustered using mmseqs2 (arguments easy-cluster, –min-seq-id 0.99, --cov-mode 0, -c 0.5)[10, 11]. The clustering was done in order to avoid bias in building the HMMs due to overrepresentation of very similar sequences. Default parameters were used unless stated otherwise.

Optimization of HMM E-value cutoffs

To optimize HMM E-value cutoffs for each CAZy family we optimized HMM E-values using 5-fold blocked cross validation as well as an external negative sequence set: For each CAZy (sub)family and fold, we divided the module sequence set into two sets: The first set of sequences was used to build the module HMM and the second set was used as positive instances for evaluation, combined with the module sequences of all remaining CAZy (sub)families as well as non-CAZy sequences obtained from UniProt (see below) as negative instances. Division into training and test folds was done in a blocked fashion, where sequences in one block are always together in either training or test set. This was done to minimize information from the training set leaking into the test set. To define blocking groups, module sequences within each family were clustered at 60% sequence identity using mmseqs2 (arguments easy-cluster, –min-seq-id 0.60, --cov-mode 0, -c 0.5)[11]. Folds were designed in such a way that test sets never overlap with each other (Figure 1). This setup was applied to 330 CAZyme families with at least 150 sequences in the CAZy database. For the 207 families which had less than 150 sequences, but more than 30 sequences, cross validation was performed without blocking. Non-CAZy sequences were obtained from UniProt in the following manner. First, all manually curated sequences (Swiss-Prot) with an annotation score of five out of five (indicating experimental evidence at protein level) were downloaded (n=54,978 sequences, March 5th 2021). We subsequently filtered out all sequences with an annotated Enzyme Commission (EC) number present in CAZy, yielding 51,507 sequences. Since CBMs are non-catalytic (and thus have no EC number), we did not add UniProt sequences as negative instances when we optimized E-values for CBMs.

Finally, we determined family- and fold-wise optimal E-values by iterating over E-value thresholds (from 1e^0 to 1e^-200) and choosing the E-value that maximizes the F1-Score. The F1 score was calculated using the follow formula: 2*(recall*precision) / (recall + precision). Recall was calculated using the formula: TP / (TP + FN) and precision using TP / (TP + FP). A schematic overview of the workflow for building novel HMMs and optimization of their E-values can be found in Figure 1.

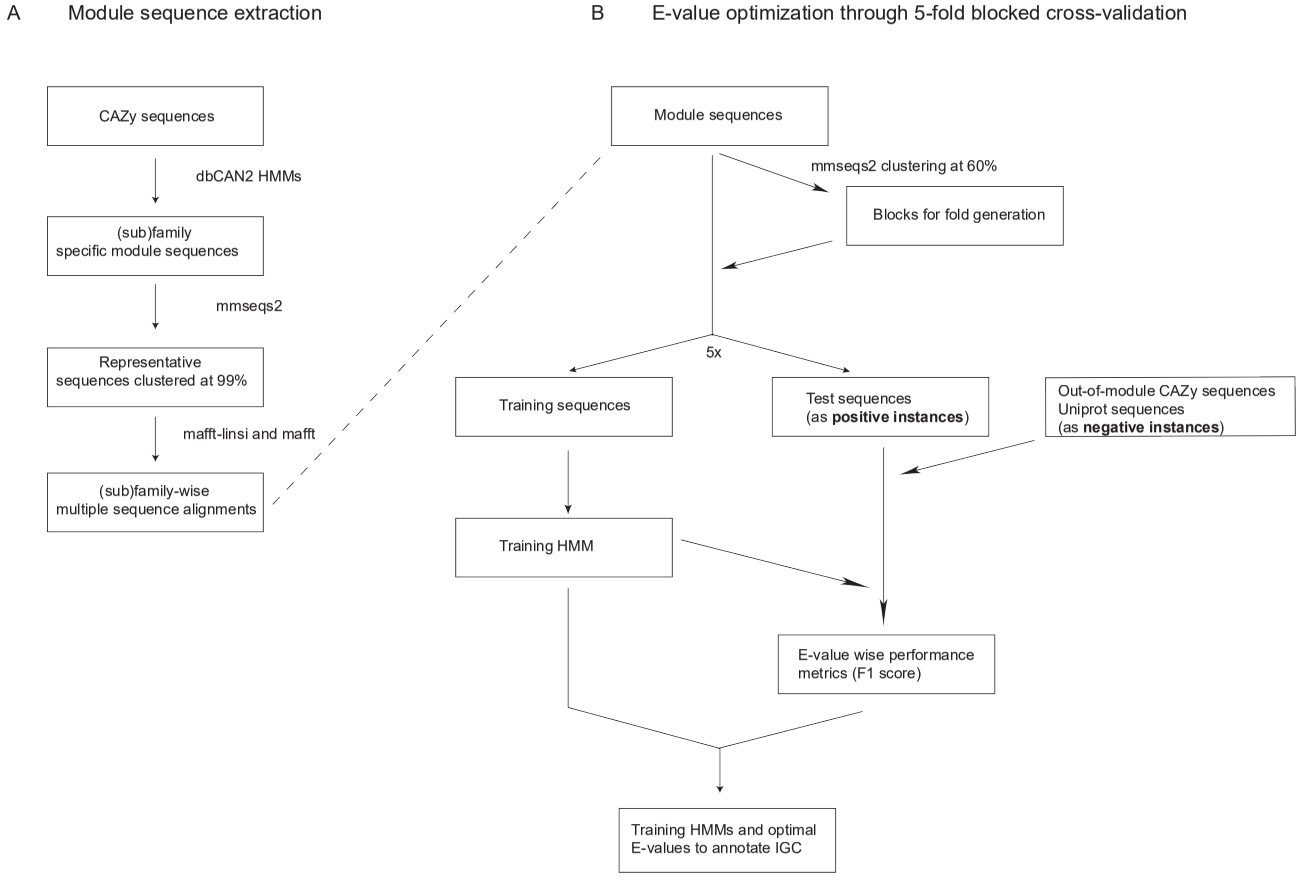


Figure 1: Schematic workflow of construction of the novel HMMs and optimization of their E-values. (A) Extraction of module subsequences from CAZy ORFs using the dbCAN2 HMMs. (B) E-value optimization for each of CAZy family using blocked cross-validation.

CAZyme annotation of the Integrated Gene Catalog (IGC)

To annotate the approximately 9.9 million genes present in the IGC[12] with CAZy module information, we ran the CV HMMs of all folds against the IGC and assigned module hits by determining the overlap of all fold-HMMs hits.

Community profiling of datasets

We investigated eight different colorectal cancer cohorts. Raw metagenomic sequencing data was processed using NGLess (v1.0.0) and accompanying tools[13, 14]. In short, raw sequence data was pre-processed by quality-based trimming and reads with quality value below 25 were discarded, followed by discarding reads shorter than 45 bp. Second, reads were aligned to the human genome (hg19 reference) and discarded if reads mapped with more than 90% sequence identity and an alignment length of at least 45 bp. Third, we mapped filtered reads against the Integrated Gene Catalog (IGC) using BWA-MEM and obtained BAM files were sorted using samtools sort[12, 14, 15]. Fourth, bedtools intersect (-bed -wo) was used to investigate overlap between the mapped reads and CAZy modules[16]. Lastly, we used isect_quant (https://github.com/cschu/gff_quantifier/blob/master/gffquant/isect_quant.py) to obtain CAZyme abundance profiles (which were corrected for the length of the respective CAZyme). Here, we focused on colorectal cancer (CRC) patients and controls (total n = 1225, CRC cases n = 632, controls n = 593) from eight different cohorts spanning seven countries and three continents[17-23].

Meta-analysis of differentially abundant CAZymes in CRC

We started our meta-analysis by investigating which CAZymes were differentially abundant between CRC and controls' metagenomes through a blocked Wilcoxon-rank sum test as available through the coin (v1.4-1) package, where study was the blocking factor[24]. Next, to investigate whether signatures were consistent across studies, univariate nonparametric Wilcoxon tests and false discovery rate (FDR) correction were performed as implemented by the SIAMCAT package (v1.10.0), on a per-study basis[25]. Generalized fold changes for both tests were calculated as implemented by the SIAMCAT package. P-values were adjusted using FDR correction (Benjamini-Hochberg method)[26] and adjusted p-values < 0.05 were considered significant.

Gene set enrichment analysis (GSEA)

Leveraging the extensive manual substrate annotations allowed us to perform GSEA to investigate whether there is differential metabolism of specific substrates between CRC patients and controls[27]. GSEA was performed using the R package fgsea(v1.16.0) and the fgseaMultilevel function with default parameters[28]. As input measure for fold change, generalized fold changes as obtained from SIAMCAT were used.

Results

Overview over the new CAZy profiler

We started building our profiler by annotating the IGC with our cross-validated HMMs and optimized E-value cutoffs, obtaining CAZy module annotations and locations on each ORF. Subsequently, we mapped quality-filtered and human-filtered reads from a given metagenomic sample to the entire IGC. We then computed overlaps of the aligned reads with the annotated CAZy modules using bedtools intersect[16]. We generated CAZy profiles by averaging individual CAZy module coverages (number of reads aligning to a module normalized by its length).

Overview of CAZy annotations on IGC

We started our analyses by obtaining a global overview of which CAZymes were annotated using our method in the IGC, a comprehensive gene catalogue constructed from metagenomes of the human gut microbiome[12]. We detected a total of 457 unique CAZymes in the IGC, which amounted to 225,946 unique ORFs (2.29% of IGC). The largest number of (sub)families was annotated in the glycoside hydrolases (GHs) category; 240 unique (sub)families, followed by 76 glycosyltransferases (GTs), 63 carbohydrate-binding modules (CBMs) 55 polysaccharide lyases (PLs), 14 carbohydrate esterases (CEs) and 9 auxiliary activities (AA)s.

Constructing a novel glycan substrate annotation scheme

To construct a more informative scheme for glycan substrate annotation, we grouped carbohydrate substrates into five biological categories which we systematically annotated for each commonly encountered glycan in the human gut (Supplementary Table 1). We hereby focused on CAZy categories GH, PL, CE and CBM for two main reasons. First, there is relatively little information about the precise substrates of many GT and AA families precluding their grouping into functional categories. Second, our main interest was in digestion rather than synthesis of (complex) carbohydrates and GTs are less involved in the former process.

Based on the CAZymes that we annotated in the IGC, a list of all potential substrates was generated. Based on literature[1, 6, 29-32], these substrates were further classified into broad biological categories: reflecting their origin (e.g. bacterial, plant), function in their original biological context (e.g. storage, structural) and function at destination/general biological role (e.g. dietary fiber, glycosaminoglycan) (Supplementary Table 1). We then annotated each CAZyme detected in the IGC with substrates and matched this to the annotations per glycan (Supplementary Table 2). This way of grouping and annotating CAZymes allows for a systematic investigation of substrate metabolism potential using metagenomic data.

Meta-analysis shows CAZymes to be differentially abundant between CRC patients and controls

Shotgun metagenomic data from eight different CRC study populations recruited in seven different countries (Austria, China, Germany, France, Italy, Japan and the US) were re-analyzed.

The first aim of this meta-analysis was to investigate whether CRC-specific CAZyme signatures could consistently be identified across the included studies. To this end, we pooled data from all studies and performed a blocked Wilcoxon test to account for study heterogeneity (as blocking factor). As a result, we identified 203 CAZymes to be significantly differentially abundant between CRC patients and controls at an FDR-adjusted p-value ≤ 0.05 (data not shown). Next, we aimed to identify which CAZyme signatures were consistent across the cohorts. Therefore, we explored the datasets by investigating differential abundance of CAZymes separately in each study by employing a Wilcoxon test and FDR correction. Significantly differentially abundant CAZymes were noted in all cohorts except from the US cohort, which previously showed weak associations between the microbiome and CRC, possibly related to the long-term sample storage (Figure 2)[22, 23].

Among the consistently more abundant CAZymes in controls across different cohorts were GH94, GH53, GH5_2, CE17, CBM48 and CBM2. Of these CAZymes, GH94, GH53, GH5_2 and CBM2 have the annotation of dietary fiber as function at destination, while CE17 has no known substrate and CBM48 is involved in glycogen metabolism. Collectively, this points towards increased dietary fiber metabolism in control metagenomes. In CRC metagenomes, CE11, GT19, GH20, GT107, GH123, CBM40 and GH33 were among the most consistently enriched CAZymes. We could not assign a function at destination to any of these CAZyme families, apart from GH20, which is involved in both dietary fiber and peptidoglycan metabolism.

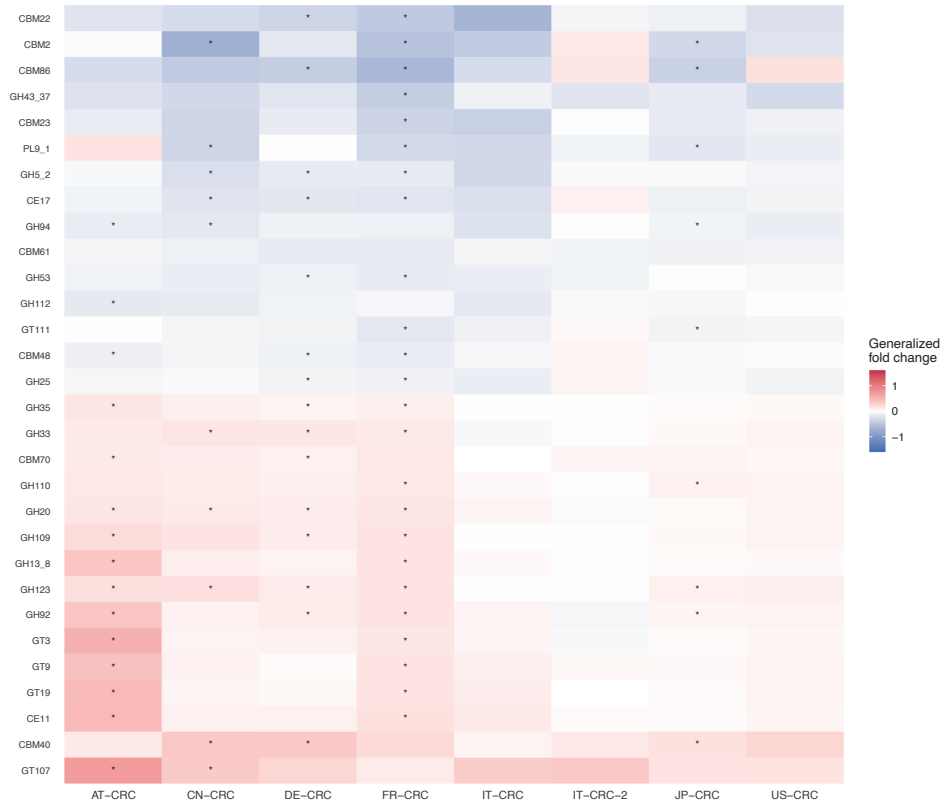


Figure 2: Meta-analysis on a per-study basis using Wilcoxon tests and FDR correction for p-values. We selected CAZymes to be displayed by taking the 15 most significantly differentially abundant CAZymes in both CRC metagenomes and in control metagenomes as obtained by the blocked Wilcoxon test using combined data from all eight cohorts. Note that both Italian cohorts contained significantly differentially abundant CAZymes, but these were not among the 15 most significant ones in either direction. The different studies are indicated in the columns of the heatmap and labeled underneath (AT; Austria, CN; China, DE; Germany, FR; France, IT; Italy, JP; Japan and US; United States). Red indicates CAZymes that are more abundant in CRC patients and blue indicates CAZymes more abundant in controls. CAZymes are ordered on the y-axis based on average generalized fold change values.

Enrichment analysis reveals decreased dietary fiber metabolism and increased glycosaminoglycan (GAG) metabolism in CRC patients

We next investigated whether dietary fiber, and other substrates, were enriched at the substrate level rather than the individual CAZyme level. To this end, we performed enrichment testing through GSEA. Leveraging our suggestion for novel manual substrate annotations in combination with GSEA, we could investigate enrichment of CAZymes at the substrate category level rather than at the individual CAZyme level. Enrichment analysis was performed at the level of the glycan's function at

destination, with testing performed for five large substrate categories (dietary fiber, glycosaminoglycans, peptidoglycan, glycogen and mucin). The difference in number of CAZymes annotated with these specific substrates is large, with 266 unique CAZymes being involved in metabolism of dietary fiber, 27 in glycosaminoglycan, 25 in glycogen, 16 in peptidoglycan and two in mucin. GSEA testing revealed an enrichment of dietary fiber metabolism in controls (adjusted p-value 7.44e^-04), while GAG and peptidoglycan were significantly enriched in CRC patients (adjusted p-value 3.68^-04 and 3.52e^-02, respectively) (Figure 3). GAG are important components of the extracellular matrix of animal tissues (for example in connective and skeletomuscular tissue) and, importantly, have never been detected in plants[33, 34].

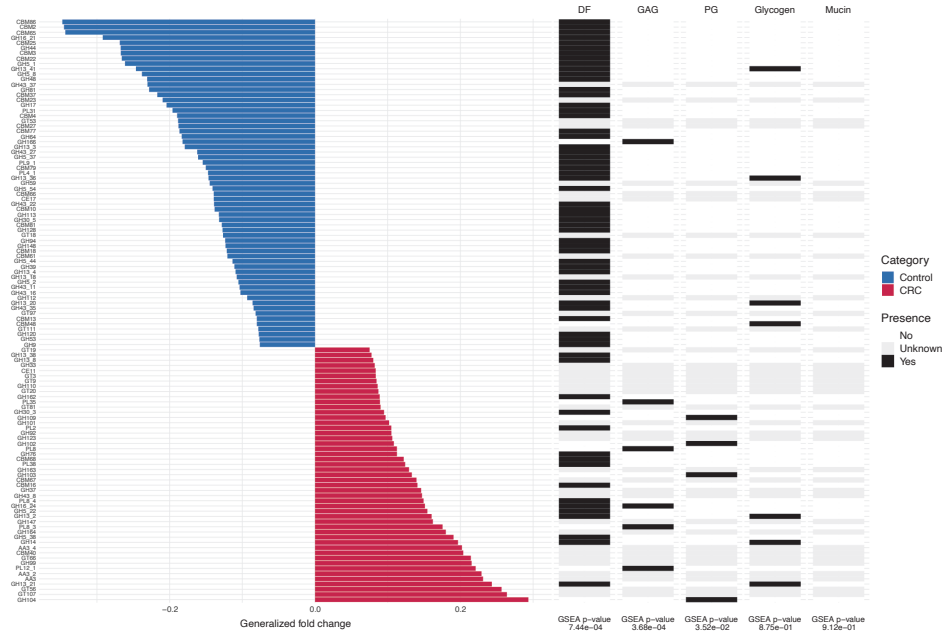


Figure 3: Differentially abundant CAZymes and substrate enrichment testing. Horizontal bars show significant abundance differences in CAZymes depicting the effect size by the generalized fold change, as identified by a blocked Wilcoxon test and FDR correction (left) and results of GSEA analysis per substrate category (right) with their respective indication whether a substrate was annotated by a specific CAZyme or not. GSEA was performed on all CAZymes which had a substrate annotation and not only the ones displayed in this plot. CAZymes with 'Unknown' substrate annotation were not included in GSEA. Dietary fiber (DF) was significantly enriched in controls, while glycosaminoglycans (GAG) and peptidoglycan (PG) were significantly enriched in CRC patients. On the right side of the plot, black indicates a CAZyme to be involved in specific substrate metabolism, white indicates it is not and grey indicates that no substrate could be assigned to the respective CAZyme. CAZymes on the left side were included if their adjusted p-value was <0.01 and had a generalized fold change of more than 0.075 or lower than -0.075. Red indicates CAZymes that are more abundant in CRC patients and blue indicates CAZymes more abundant in controls.

Discussion

We present a new tool for profiling CAZymes from shotgun metagenomic data generated from human fecal samples. To interpret these profiles with respect to broader substrate categories, we additionally derived a hierarchical substrate classification scheme and curated substrate information for all CAZyme families found in the human gut based on the scientific literature. To maximize the detection accuracy of CAZymes in uncharacterized (meta-)genomic sequences based on sequence homology, we systematically optimized and validated family-specific HMMs (and their E-value cutoffs) on a large set of manually annotated CAZyme sequences. By leveraging our novel substrate annotation scheme, we were for the first time able to perform enrichment testing with respect to substrate metabolism for CAZymes in metagenomic case-control studies. In an application to eight studies of CRC, these novel tools revealed a clear CRC-associated CAZyme signature that is largely consistent across the geographically diverse studies included in this meta-analysis.

Advantages of our newly developed tool over dbCAN2

The most widely used tool for detecting CAZymes in genomic data are dbCAN2 and its predecessor dbCAN[9, 35]. This tool is largely based on building HMMs from CAZy (sub)families from the CAZy database. While it has proven to be extremely valuable for annotating genomic data, the E-value cutoffs have not been optimized for detection accuracy. Here we used blocked cross-validation on the protein sequences from the CAZy database to optimize HMM E-value cutoffs to more accurately identify CAZymes. Going a step further than annotation of genomic data, we devised a profiling framework to enable users to directly profile CAZymes from human gut metagenomic data. This is a major advance, as it allows for straightforward comparative analyses of CAZyme profiles between (groups of) metagenomic samples using statistical tests suitable for metagenomic data. We will ultimately make the workflow presented in this manuscript (from raw read processing to obtaining matrices with CAZyme abundances) publicly available as open-source software to enable other researchers to easily profile the CAZyme repertoire in their metagenome. While we purely focused on the human gut microbiome, we plan for future updates to extend the CAZyme substrate annotations to gene sequences found in metagenomic gene catalogs available for other mammalian gut microbiomes and environmental communities, such as those in the environment (e.g. soil, ocean)[36-41].

Manual substrate annotations allow for informative grouping at higher functional levels and statistical enrichment testing

By detecting broader shifts in carbohydrate degradation capabilities of microbial communities, key insights into community function have been obtained, in some cases

with important implications for host physiology[42, 43]. However, this type of analysis is still not very commonly performed on human gut metagenomic data due to the following technical challenges. The combination of a lack of knowledge on substrate usage of specific CAZyme families and grouping CAZymes into families based on protein sequence similarity does not allow for easy interpretation of CAZyme profiles with regard to substrate utilization[7]. Therefore, it would be highly valuable to have a consistent higher-level substrate annotation that can be used to detect trends across families. A common workflow that researchers currently take when trying to assign functional information to CAZymes is to first investigate, through differential abundance testing, which CAZymes are differentially abundant between two groups of interest and then attempt to annotate these CAZymes only[17, 44]. However, this does not leverage a potential 'substrate-level' effect, as while individual CAZymes may not be significantly differentially abundant, their combined effect may be significant at the substrate level[27]. Such enrichment patterns can, however, be detected using our new substrate annotation scheme in combination with Kolmogorov-Smirnov type statistical tests that are also commonly applied in GSEA[27].

CAZyme and substrate enrichments correspond to and extend on previous nutritional epidemiological studies investigating CRC

Increased dietary fiber intake has long been linked to a reduced risk of developing CRC, although there is some conflicting evidence from nutritional epidemiological studies[45-48]. However, it has not been systematically investigated whether the gut microbiome may be involved in the physiological metabolic processes underlying this association[17]. Previous research showed specific CAZymes involved in fiber metabolism to be more abundant in controls and several CAZymes involved in host glycan metabolism (e.g. mucin) to be more abundant in CRC patients[17]. We investigated if this could be confirmed by applying our tool on data from eight different studies with patients recruited in seven different countries on three continents. Additionally, we moved away from exclusively testing individual CAZymes, but instead assessed enrichment of specific substrate categories through a GSEA. Importantly, we found dietary fiber metabolism to be increased in controls and an increase in GAG metabolism in CRC metagenomes (Figure 3). The latter possibly reflects increased meat consumption in CRC patients, which has been linked to an increased risk for developing CRC[49-51].

Unfortunately, no food frequency questionnaire or other detailed dietary information was collected from the CRC patients or controls, which prevents us from correlating reported dietary intake to abundance of specific CAZymes or to the observed enrichment in GAG metabolism. Importantly, while we see this enrichment in GAG in CRC metagenomes, it should be noted that this remains an association and cannot be interpreted as a causal relationship. Lastly, we observed an enrichment in peptidoglycan metabolism in CRC

metagenomes, which is in line with a previous observation of an increase in bacterial cell wall components in CRC metagenomes[17], but this result remains difficult to interpret.
In conclusion, our novel tool allows for accurate profiling of CAZymes from metagenomic shotgun sequence data and the annotation scheme enables substrate-based analysis and interpretation. Together, the tool and substrate annotation scheme pave the way for more informative functional analyses from metagenomic data with respect to the CAZyme repertoire. We applied our tool and annotation scheme to investigate the role of CAZymes and substrate metabolism in CRC metagenomes and discovered specific CAZymes and substrates to be enriched and depleted in CRC metagenomes. While both the dietary fiber and GAG signatures are consistent with known dietary risk factors of CRC, it remains unknown whether the gut microbiome mediates the potentially beneficial and harmful effects of these dietary components in CRC development, or whether the gut microbiome simply adapts to the diet and plays no mechanistic role in the diet-CRC relationship. Therefore, future studies should aim to delineate whether the effects of dietary factors on CRC development are microbiome-mediated or not.

Notes

Funding
QD was supported by a FEMS Research and Training Grant. NK and GZ are supported by EMBL core funding, GZ is partially funded by the German Federal Ministry of Education and Research (BMBF; the de. NBI network no. 031A537B and grant no. 031L0181A).

Author contributions
QD, NK and GZ conceptualized and designed the study. QD, NK and CS performed all bioinformatic analyses. QD and HT performed substrate annotations and QD wrote the manuscript with contributions from NK and GZ. GZ supervised the study. All authors discussed and approved the manuscript.

Supplementary tables

Table S1: Table with annotations at several functional categories per glycan. Dietary fiber (DF), non-starch polysaccharides (NSP), peptidoglycan (PG), glycosaminoglycan (GAG).

Glycan	Origin	Function_in_ origin	Function_ destination_1	Function_d estination_2	Function_ destination_3
arabinogalactan	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin
arabinoxylan	Plant	Structural	DF	NSP	Hemicellulose
beta-glucan	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose
beta-mannan	Plant	Structural	DF	NSP	Mannans_and_ Heteromannans
cellobiose	Plant	Structural	DF	NSP	Cellulose
cellulose	Plant	Structural	DF	NSP	Cellulose
galactomannan	Plant,Fungal	Structural	DF	NSP	Mannans_and_ Heteromannans
glucomannan	Plant,Fungal,Bacteria	Structural	DF	NSP	Mannans_and_ Heteromannans, Hemicellulose
xylan	Plant	Structural	DF	NSP	Hemicellulose
xyloglucan	Plant	Structural	DF	NSP	Hemicellulose
amylopectin	Plant	Storage	DF	Resistant_starch	Unknown
amylose	Plant	Storage	DF	Resistant_starch	Unknown
arabinan	Plant	Structural	DF	NSP	Pectin
glycogen	Animal	Storage	Glycogen	Unknown	Unknown
pullulan	Fungal	Structural	DF	Unknown	Unknown
starch	Plant	Storage	DF	Resistant_starch	Unknown
alginate	Algae,Bacteria	Structural	DF	NSP	Gum
carrageenan	Algae	Structural	DF	NSP	Gum
galactan	Algae,Plant	Structural	DF	NSP	Pectin
laminarin	Algae	Storage	DF	NSP	Unknown
porphyran	Algae	Structural	DF	NSP	Unknown
fucoidan	Algae	Structural	DF	NSP	Unknown
ulvan	Algae	Structural	DF	NSP	Unknown
alpha-mannan	Fungal	Structural	DF	NSP	Unknown
chitin	Animal,Fungal	Structural	DF	Unknown	Unknown
chitosan	Animal,Fungal	Structural	DF	Unknown	Unknown
chitodextrin	Animal,Fungal	Structural	DF	Unknown	Unknown
dextrin	Plant	Storage	DF	Resistant_ oligosaccharides	Unknown
chondroitin	Animal	GAG	GAG	Unknown	Unknown
chondroitin_sulfate	Animal	GAG	GAG	Unknown	Unknown
heparin	Animal	GAG	GAG	Unknown	Unknown
heparin_sulfate	Animal	GAG	GAG	Unknown	Unknown
hyaluronan	Animal,Bacteria	GAG	GAG	Unknown	Unknown
mucin	Animal	Unknown	Mucin	Unknown	Unknown
fructan	Plant	Storage	DF	NSP	Unknown
inulin	Plant	Storage	DF	Resistant_ oligosaccharides	Unknown
homogalacturonan	Plant	Structural	DF	NSP	Pectin
rhamnogalacturonan	Plant	Structural	DF	NSP	Pectin
melibiose	Fungal	Unknown	Unknown	Unknown	Unknown

Glycan	Origin	Function_in_ origin	Function_ destination_1	Function_d estination_2	Function_ destination_3
raffinose	Plant	Unknown	DF	Resistant_ oligosaccharides	Unknown
agarose	Algae	Structural	DF	NSP	Unknown
peptidoglycan	Microbial	Structural	PG	PG	PG
lichenin	Plant	Storage	DF	NSP	Unknown
GlcNAc	Animal,Bacteria	Structural	Unknown	Unknown	Unknown
GalNAc	Animal,Bacteria	Structural	Unknown	Unknown	Unknown
sialic_acid	Animal,Bacteria, Archaea,Algae	Unknown	Unknown	Unknown	Unknown

Table S2: Table with annotations for all CAZymes annotated in the IGC (when substrates could be assigned).

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_1	All_domains_of_life	Unknown	Unknown	Unknown	Unknown	broad
GH_2	All_domains_of_life	Unknown	Unknown	Unknown	Unknown	broad
GH_3	All_domains_of_life	Unknown	Unknown	Unknown	Unknown	broad
GH_4	All_domains_of_life	Unknown	Unknown	Unknown	Unknown	broad
GH_5	All_domains_of_life	Structural, Storage	DF,Glycogen	NSP	Hemicellulose,Cellulose, Unknown	cellulose
GH_5_1	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin
GH_5_2	Plant,Bacteria,Fungal, Animal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin,chitosan
GH_5_4	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin,xylan,xyloglucan
GH_5_5	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin
GH_5_7	Plant,Fungal,Bacteria	Structural	DF	NSP	Mannans_and_ Heteromannans	beta-mannan,mannan,galactomannan,glucomannan
GH_5_8	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Mannans_ and_ Heteromannans,Unknown	beta-glucan,cellulose,lichenin,mannan,galactomannan, glucomannan
GH_5_13	Unknown	Unknown	Unknown	Unknown	Unknown	alpha-arabinoside,beta-galactofuranoside
GH_5_18	Unknown	Unknown	Unknown	Unknown	Unknown	beta-mannoside
GH_5_21	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_5_22	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin,xylan
GH_5_23						glucopyranoside
GH_5_25	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Mannans_ and_ Heteromannans,Unknown	beta-glucan,cellulose,lichenin,mannan,galactomannan, glucomannan
GH_5_26	Plant,Microbial, Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin
GH_5_35	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_5_36	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose,Mannans_ and_ Heteromannans	beta-glucan,mannan,galactomannan,glucomannan
GH_5_37	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_5_38	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin
GH_5_41	Plant,Fungal,Bacterial	Structural	DF	NSP	Mannans_and_ Heteromannans	mannan,galactomannan,glucomannan
GH_5_44	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
GH_5_45	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	beta-glucoside,alpha-arabinofuranoside,arabins,arabinoxylan,arabinogalactan
GH_5_46	Plant,Bacteria, Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	beta-glucan,cellulose,lichenin
GH_5_52	Plant,Bacteria, Fungal	Structural	DF	NSP	Hemicellulose,Cellulose	beta-glucan,cellulose
GH_5_54	Plant	Structural	DF	NSP	Cellulose	cellulose
GH_8	Plant,Bacteria, Fungal, Animal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	cellulose,beta-glucan,xylan,lichenin,chitosan
GH_9	Plant,Bacterial, Fungal	Structural	DF	NSP	Hemicellulose,Cellulose, Mannans_and_ Heteromannans	cellulose,xylan,beta-glucan,xyloglucan,glucomannan
GH_10	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_11	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_12	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	cellulose,lichenin,beta-glucan,xyloglucan
GH_13	All_domains_of_life	Storage, Structural	DF,Glycogen	Unknown	Unknown	broad
GH_13_2	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_polysaccharides
GH_13_3	Plant	Storage	DF	Resistant_starch	Unknown	starch
GH_13_4	Plant		DF	Unknown,Resistant starch		amylosucrose
GH_13_5	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_polysaccharides
GH_13_8	Plant	Storage	DF	Resistant_starch	Unknown	amylose
GH_13_9	Plant	Storage	DF	Resistant_starch	Unknown	amylose
GH_13_10	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_ polysaccharides,trehalose
GH_13_11	Plant,Animal	Storage	DF,Glycogen	Resistant_starch, Unknown	Unknown	glycogen,amylopectin

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_13_12	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	starch,glycogen,related_ polysaccharides,pullulan, amylopectin
GH_13_13	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen
GH_13_14	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen,starch, related_polysaccharides
GH_13_16						maltose
GH_13_18						sucrose,glucosylglycerate
GH_13_19	Plant	Storage	DF	Resistant_starch	Unknown	amylose,amylaceous_polysaccharides
GH_13_20	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen,starch, related_polysaccharides,maltodextrin
GH_13_21	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_polysaccharides
GH_13_26						alpha-glucan
GH_13_27	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_polysaccharides
GH_13_28	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen,starch, related_polysaccharides
GH_13_29	Plant	Storage	DF	Resistant_starch	Unknown	trehalose,starch
GH_13_30	Plant	Storage	DF	Resistant_starch	Unknown	starch
GH_13_31	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_ polysaccharides,sucrose
GH_13_32	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	starch,glycogen,related_ polysaccharides,pullulan, amylopectin,amylaceous_polysaccharides
GH_13_33						maltose
GH_13_36	Animal,Fungal,Plant	Storage, Structural	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_ polysaccharides,pullulan, maltodextrin
GH_13_37	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_polysaccharides
GH_13_38	Plant	Storage	DF	Resistant_starch	Unknown	starch
GH_13_39	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen,starch, related_polysaccharides

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_13_41	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	pullulan,amylopectin,glycogen,starch, related_polysaccharides
GH_13_42	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_ polysaccharides, amlyaceous_polysaccharides
GH_14	Animal,Plant	Storage	Glycogen,DF	Unknown,Resistant_ starch	Unknown	starch,glycogen,related_ polysaccharides
GH_15	Plant	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	trehalose,alpha-glucan,dextran,isomaltos accharide
GH_16_3	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Unknown	lichenin,beta-glucan,lamirin,paramylon,pa chyman,keratan_sulfate
GH_16_6	Algae,Plant	Structural	DF	NSP	Pectin	transglycosylases,glucan,galactan
GH_16_8	Algae,Plant	Structural	DF	NSP	Pectin	transglycosylases,glucan,galactan
GH_16_12	Algae	Structural	DF	NSP	Unknown	porphyran
GH_16_14	Algae	Structural	DF	NSP	Unknown	agarose
GH_16_15	Algae	Structural	DF	NSP	Unknown	agarose
GH_16_16	Algae	Structural	DF	NSP	Unknown	agarose,porphyran
GH_16_17	Algae	Structural	DF	NSP	Gum	carrageenan
GH_16_21	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Unknown	lichenin,beta-glucan,lamirin,xylan
GH_16_24	Plant,Bacteria, Fungal, Algae,Animal,Bacteria	Structural, GAG	DF,GAG	NSP,Unknown	Hemicellulose,Gum, Unknown	beta-glucan,karatan,hyaluronan,carrage enan
GH_16_25	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Unknown	lichenin,beta-glucan
GH_17	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Unknown	beta-glucan,lichenin,abscisate
GH_18	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	chitin,chitodextrin,peptidoglycan,high-mannose_glycoproteins
GH_19	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	chitin,chitodextrin,peptidoglycan
GH_20	Bacterial,Animal	Structural	DF,PG	PG	Unknown,PG	gangliosides,chitobiose
GH_23	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	chitin,chitodextrin,peptidoglycan
GH_24	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	chitodextrin,peptidoglycan

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_25	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	chitodextrin,peptidoglycan
GH_26	Plant,Fungal,Bacteria	Structural	DF	NSP	Mannans_and_ Heteromannans, Hemicellulose	mannan,galactomannan,glucomannan,beta-mannan,xylan
GH_27	Animal,Plant,Unknown	Structural	DF	NSP	Gum,Pectin,Hemicellulose,Unknown	alpha-galactoside,N-acetylgalactosamine,alpha-glucan,arabinopyranoside
GH_28	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan,pectin, homogalacturonan, polygalacturote
GH_29	Algae,Animal	Structural	DF	NSP	Unknown, HMO	fucoidan
GH_30_1	Plant,Algae	Structural	DF	NSP	Hemicellulose,Pectin	xylan,glucan,galactan,glucuronide
GH_30_2	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_30_3	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
GH_30_4	Unknown	Unknown	Unknown	Unknown	Unknown	beta-fucoside
GH_30_5	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	arabinogalactan,beta-galactan
GH_30_6	Unknown	Unknown	Unknown	Unknown	Unknown	beta-glucoside
GH_30_8	Plant	Structural	DF	NSP	Hemicellulose	xylan,glucuronoarabinoxylans
GH_30_9	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	beta-glucuronoside
GH_31	Plant	Storage	DF	Resistant_starch	Unknown	starch,xylose,sucrose,isomaltose
GH_32	Plant	Storage	DF	Resistant_ oligosaccharides	Unknown	sucrose,inulin,levan
GH_33	Animal,Microbial, Algae	Unknown	Unknown	Unknown	HMO,Unknown	sialic_acid
GH_34	Animal,Microbial, Algae	Unknown	Unknown	Unknown	Unknown	sialic_acid
GH_35	Animal,Fungal	Structural	DF		Pectin, Unknown	beta-galactoside,alpha-arabinoside,chitosan
GH_36	Plant	Unknown	DF	NSP, Resistant_ oligosaccharides	Gum, Unknown	alpha-galactoside,N-acetylgalactoseamine,raffinose
GH_37	Animal,Bacteria,Plant, Fungal	Storage	Unknown	Unknown	Unknown	trehalose
GH_38	Fungal	Structural	DF	NSP	Unknown	alpha-mannan
GH_39	Plant	Structural	DF	NSP	Hemicellulose	xylan,dermatan_sulfate

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_42	Plant,Animal	Unknown	Unknown	Unknown	Unknown	beta-galactoside,alpha-arabinoside
GH_43	Plant,Bacteria	Structural	DF	NSP	Hemicellulose,Pectin,Gum	broad
GH_43_1	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside,arabin, arabinoxylan,arabinogalactan
GH_43_2	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_3	Unknown	Unknown	Unknown	Unknown	Unknown	galactofuranoside
GH_43_4	Unknown	Unknown	Unknown	Unknown	Unknown	arabin
GH_43_5	Unknown	Unknown	Unknown	Unknown	Unknown	arabin
GH_43_7	Plant	Structural	DF	NSP	Hemicellulose	xylan,arabinofuranoside
GH_43_8						galactofuranoside
GH_43_9	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_10	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside,arabin, arabinoxylan,arabinogalactan
GH_43_11	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside,arabin, arabinoxylan,arabinogalactan
GH_43_12	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside,arabin, arabinoxylan,arabinogalactan
GH_43_16	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan, xylobiose,arabinofuranoside, arabin, arabinoxylan,arabinogalactan
GH_43_17	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_18	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_19	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_22	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside, arabin,arabinoxylan,arabinogalactan
GH_43_23	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_24	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	arabinogalactan

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_43_26	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_27	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside, arabin,arabinoxylan,arabinogalactan
GH_43_28	Plant	Structural	DF	NSP	Hemicellulose	xylan,arabinofuranoside
GH_43_29	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside, arabin,arabinoxylan,arabinogalactan
GH_43_30	Unknown	Unknown	Unknown	Unknown	Unknown	galactofuranoside
GH_43_31	Unknown	Unknown	Unknown	Unknown	Unknown	galactofuranoside
GH_43_33	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_43_34	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinofuranoside,arabin,arabinoxylan, arabinogalactan,galactofuranoside
GH_43_35	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xylan,xylobiose,arabinofuranoside, arabin,arabinoxylan,arabinogalactan
GH_43_37	Unknown	Unknown	Unknown	Unknown	Unknown	galactofuranoside,arabin
GH_44	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	cellulose,lichenin,beta-glucan, xyloglucan
GH_45	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Mannans_ and_ Heteromannans,Unknown	cellulose,lichenin,beta-glucan, xyloglucan,mannan,galactomannan, glucomannan
GH_46	Animal,Fungal	Structural	DF	Unknown	Unknown	chitosan
GH_47	Fungal	Structural	DF	NSP	Unknown	alpha-mannan
GH_48	Plant,Bacteria, Fungal, Animal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Unknown	cellulose,lichenin,beta-glucan,chitin, chitodextrin
GH_49	Fungal	Structural	DF	Unknown	Unknown	dextran,pullulan
GH_50	Algae	Structural	DF	NSP	Unknown	agarose
GH_51	Plant,Bacteria,F ungal	Structural, Storage	DF	NSP	Gum,Pectin,Hemicellulose,Cellulose, Unknown	cellulose,lichenin,beta-glucan,xylan, arabinofuranoside,arabin,arabinoxylan, arabinogalactan
GH_52	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_53	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	arabinogalactan
GH_54	Plant	Structural	DF	NSP	Hemicellulose	xylan,alpha-arabinosides

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_55	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan,lamirin,paramylon, pachyman
GH_57	Plant	Storage	DF	Resistant_starch	Unknown	amylose,galactoside
GH_58	Animal,Bacteria, Archaea,Algae	Unknown	Unknown	Unknown	Unknown	sialic_acid
GH_59	Unknown	Unknown	Unknown	Unknown	Unknown	beta-galactosides,ceramide_derivatives
GH_63	Plant	Storage	DF	Resistant_starch	Unknown	starch,alpha-glucan,mannosylglycerate, glucosylglycerate
GH_64	Unknown	Unknown	DF	NSP	Hemicellulose	lamirin,paramylon,pachyman
GH_65	Unknown	Unknown	Unknown	Unknown	Unknown	matlose,trehalose,kojibiose
GH_66	Bacteria	Unknown	DF	Unknown	Unknown	alpha-glucan,dextran
GH_67	Plant	Structural	DF	NSP	Hemicellulose	alpha-glucuronoside,xylan
GH_68	Unknown	Unknown	DF	Unknown	Unknown	sucrose,beta-fructofuranoside
GH_70	Unknown	Unknown	DF	Unknown	Unknown	sucrose,alpha-glucan
GH_71	Unknown	Unknown	Unknown	Unknown	Unknown	isolichenin,pseudonigeran,nigeran
GH_73	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	peptidoglycan,chitodextrin,high-mannose_ glycopeptides
GH_74	Plant,Bacteria,Fungal	Structural, Storage	DF	NSP	Hemicellulose,Cellulose, Mannans_ and_ Heteromannans,Unknown	cellulose,lichenin,beta-glucan, xyloglucan,mannan,galactomannan, glucomannan
GH_76	Plant,Fungal	Storage Structural	DF	Resistant_starch,NSP	Unknown	alpha-mannan,starch
GH_77	Unknown	Unknown	Unknown	Unknown	Unknown	alpha-glucan
GH_78	Plant	Structural	DF	NSP	Pectin	alpha-rhamnoside,rhamnogalacturonan
GH_79	Animal,Bacteria	GAG	GAG	Unknown	Unknown	beta-glucuronoside,heparin_sulfate, hyaluronate
GH_80	Animal,Fungal	Structural	DF		Unknown	chitosan
GH_81	Unknown	Unknown	DF	NSP	Hemicellulose	lamarin,paramylon,pachyman
GH_82	Algae	Structural	DF	NSP	Gum	carrageenan
GH_83	Animal,Bacteria, Archaea,Algae	Unknown	Unknown	Unknown	Unknown	sialic_acid
GH_84	Animal,Bacteria	Unknown	Unknown	Unknown	Unknown	glycoproteins
GH_85	Plant,Unknown	Unknown	Unknown	Unknown	Unknown	high-mannose_glycopeptides

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_86	Algae	Structural	DF	NSP	Unknown	agarose,porphyran
GH_87	Unknown	Unknown	Unknown	Unknown	Unknown	alpha-glucan,isolichenin, pseudonigeran,nigeran
GH_88	Unknown	GAG	GAG	Unknown	Unknown	glycosaminoglycans
GH_89	Animal	Unknown	Mucin	Unknown	Unknown	mucin
GH_90	Unknown	Unknown	Unknown	Unknown	Unknown	rhamnose
GH_91	Plant	Storage	DF	Resistant_ oligosaccharides	Unknown	inulin
GH_92	Unknown	Unknown	Unknown	Unknown	Unknown	glycoproteins
GH_93	Unknown	Unknown	Unknown	Unknown	Unknown	alpha-arabin
GH_94	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose,Cellulose	cellobiose,cellodextrin,chitobiose, beta-glucan
GH_95	Plant,Algae,Animal	Structural	DF	NSP	Hemicellulose,Unknown, HMO	fucoidan,fucoside,xyloglucan,HMO, blood_group_glycoconjugates
GH_96	Algae	Structural	DF	NSP	Unknown	agarose
GH_97	Unknown	Unknown	Unknown	Unknown	Unknown	maltose
GH_98	Plant	Structural	DF	NSP	Hemicellulose	beta-galactoside,blood_group_antigen, xylan
GH_99	Unknown	Unknown	Unknown	Unknown	Unknown	glucose-substituted_mannose,fungal_ cell_wall
GH_101	Unknown	Unknown	Unknown	Unknown	Unknown	N-acetylgalactoseamine
GH_102	Microbial	Structural	PG	PG	PG	peptidoglycan
GH_103	Microbial	Structural	PG	PG	PG	peptidoglycan
GH_104	Microbial	Structural	PG	PG	PG	peptidoglycan
GH_105	Plant,Bacteria,Algae	Structural	DF	NSP	Gum,Pectin,Unknown, Pectin	rhamnogalacturonan,ulvan, arabinogalactan
GH_106	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_108	Animal,Fungal, Microbial	Structural	DF,PG	,PG	Unknown,PG	peptidoglycan,chitodextrin
GH_109	Microbial	Structural	PG	PG	PG	peptidoglycan
GH_110	Unknown	Unknown	Unknown	Unknown	Unknown	blood_group_antigen
GH_111	Unknown	Unknown	Unknown	Unknown	Unknown	N-acetylglucoseamine

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_112	Unknown	Unknown	Unknown	Unknown	Unknown	galacto-N-biose,lacto-N-biose
GH_113	Plant,Fungal,Bacteria	Structural	DF	NSP	Mannans_and_ Heteromannans	mannan,galactomannan, glucomannan
GH_115	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_116	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
GH_117	Algae	Structural	DF	NSP	Gum,Unknown	beta-galactofuranoside,agarose, carrageenan
GH_120	Plant	Structural	DF	NSP	Hemicellulose	xylan
GH_121	Plant		DF			hydroxproline-rich_glycoproteins
GH_122	Plant	Storage	DF	Resistant_starch	Unknown	starch
GH_123	Unknown	Unknown	Unknown	Unknown	Unknown	N-acetylgalactoseamine
GH_125	Fungal	Structural	DF	NSP	Unknown	alpha-mannan
GH_126	Plant	Storage	DF	Resistant_starch	Unknown	starch
GH_127	Plant	Unknown	DF	NSP	Pectin, Hemicellulose	Unknown
GH_128	Unknown	Unknown	DF	NSP	Hemicellulose	lamirin,paramylon,pachyman
GH_129	Animal,Microbial	Unknown, Structural	Mucin,PG	Unknown,PG	Unknown,PG	mucin,peptidoglycan
GH_130	Unknown	Unknown	Unknown	Unknown	Unknown	beta-mannosides
GH_133	Animal	Storage	Glycogen	Unknown	Unknown	glycogen
GH_136	Animal,Unknown	Unknown	Unknown	Unknown	HMO,Unknown	lacto-N-tetraose
GH_137	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_138	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_139	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_140	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_141	Plant,Algae	Structural	DF	NSP	Hemicellulose,Unknown	xylan,fucoidan
GH_142	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
GH_143	Unknown	Unknown	Unknown	Unknown	Unknown	rhamnogalacturonan
GH_144	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
GH_145	Unknown	Unknown	DF	NSP	Gum,Unknown	gum-arabic
GH_146	Unknown	Unknown	Unknown	Unknown	Unknown	sugar_beet_arabin
GH_147	Unknown	Unknown	Unknown	Unknown	Unknown	beta-galactoside
GH_148	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
GH_149	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan,lamirin
GH_150	Algae	Structural	DF	NSP	Gum	carrageenan
GH_151	Unknown	Unknown	Unknown	Unknown	HMO,Unknown	blood_group_antigen
GH_153	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
GH_154	Plant,bacteria	structural	DF	NSP	Gum, pectin	beta-glucuronoside
GH_156	Animal,Bacteria, Archaea,Algae	Unknown	Unknown	Unknown	Unknown	sialic_acid
GH_157	Unknown	Unknown	DF	NSP	Hemicellulose	lamirin,paramylon,pachyman
GH_158	Unknown	Unknown	DF	NSP	Hemicellulose	lamirin,paramylon,pachyman
GH_159	Unknown	Unknown	Unknown	Unknown	Unknown	beta-galactofuranosides
GH_160	Bacteria	Unknown	Unknown	Unknown	Unknown	bacterial_polysaccharides
GH_161	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan,lamirin
GH_162	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
GH_163	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
GH_164	Unknown	Unknown	Unknown	Unknown	Unknown	beta-mannosides
GH_165	Unknown	Unknown	Unknown	Unknown	Unknown	beta-galactoside
GH_166	Bacteria	GAG	GAG	Unknown	Unknown	bacterial_polysaccharides
GH_167	Algae	Structural	DF	NSP	Gum	carrageenan
PL_1_2	Plant	Structural	DF	NSP	Pectin	pectate
PL_1_6	Plant	Structural	DF	NSP	Pectin	pectate
PL_2	Plant	Structural	DF	NSP	Pectin	pectate
PL_2_1	Plant	Structural	DF	NSP	Pectin	pectate
PL_2_2	Plant	Structural	DF	NSP	Pectin	pectin
PL_4	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_4_1	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_4_4	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_6_1	Algae,Bacteria	Structural	DF	NSP	Gum	dermatan_sulfate,alginate
PL_6_2	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
PL_8	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,chondroitin,xanthan, chondroitin_sulfate,dermatan_ sulfate,heparin,heparin_sulfate

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
PL_8_1	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,chrondoitin
PL_8_2	Animal	GAG	GAG	Unknown	Unknown	heparin,chrondoitin,dermatan
PL_8_3	Animal	GAG	GAG	Unknown	Unknown	chondroitin_sulfate,dermatan_sulfate
PL_8_4	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_9_1	Plant	Structural	DF	NSP	Pectin	pectate,rhamnogalacturonan
PL_9_2	Plant	Structural	DF	NSP	Pectin	Unknown
PL_9_4	Plant	Structural	DF	NSP	Pectin	Unknown
PL_10	Plant	Structural	DF	NSP	Pectin	pectate
PL_10_1	Plant	Structural	DF	NSP	Pectin	pectate
PL_10_2	Plant	Structural	DF	NSP	Pectin	pectate
PL_11	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_11_1	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_11_2	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan
PL_12	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_12_1	Animal	GAG	GAG	Unknown	Unknown	heparin
PL_12_2	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_13	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_15	Algae,Bacteria,Animal	Structural,GAG	DF,GAG	NSP,Unknown	Gum,Unknown	heparin,heparin_sulfate,alginate
PL_15_1	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_15_2	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_16	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,chondroitin
PL_17	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_17_1	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_17_2	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_18	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_21	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_21_1	Animal	GAG	GAG	Unknown	Unknown	heparin,heparin_sulfate
PL_22	Plant	Structural	DF	NSP	Pectin	pectin
PL_22_1	Plant	Structural	DF	NSP	Pectin	pectin
PL_23	Animal	GAG	GAG	Unknown	Unknown	chondroitin
PL_26	Plant	Structural	DF	NSP	Pectin	rhamnogalacturonan

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
PL_27	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	arabinogalactan
PL_28	Algae	Structural	DF	NSP	Unknown	ulvan
PL_29	Animal,Bacteria	GAG	GAG	Unknown	Unknown	chondroitin_sulfate,dermatan_sulfate, hyaluronic_acid
PL_30	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,chondroitin
PL_31	Algae,Bacteria	Structural	DF	NSP	Gum	alginate,beta-glucuron
PL_32	Algae,Bacteria	Structural	DF	NSP	Gum	alginate
PL_33	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,gellan,chondroitin_sulfate
PL_33_1	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron,chondroitin
PL_33_2	Animal,Bacteria	GAG	GAG	Unknown	Unknown	chondroitin_sulfate,dermatan_sulfate, hyaluronan
PL_35	Animal	GAG	GAG	Unknown	Unknown	chondroitin
PL_37	Algae,Animal	Structural,GAG	DF,GAG	NSP,Unknown	Unknown	heparan_sulfate,dermatan_sulfate, chondroitin_sulfate, ulvan
PL_38	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin	glucuronan
PL_40	Algae	Structural	DF	NSP	Unknown	ulvan
CE_1	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
CE_2	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides
CE_3	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides
CE_4	Plant,Animal,Fungal, Microbial	Structural	DF,PG	NSP,PG	Hemicellulose,Unknown,PG	peptidoglycan,chitin,xylan
CE_5	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides,cutin
CE_6	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides
CE_7	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides, cephalosporin
CE_8	Plant	Structural	DF	NSP	Pectin	pectin
CE_9	Microbial	Structural	PG	PG	PG	peptidoglycan
CE_11	Bacterial	Unknown	Unknown	Unknown	Unknown	LPS
CE_12	Plant	Structural	DF	NSP	Hemicellulose,Pectin	pectin,rhamnogalacturonan,xylan, xylooligosaccharides
CE_14	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
CE_15	Plant,Bacteria	structural	DF	NSP	Gum, pectin	lignin,lignocellulose
CE_17	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
CBM_2	Plant,Animal,Fungal	Structural	DF	NSP	Cellulose,Hemicellulose, Unknown	cellulose,chitin,xylan
CBM_3	Plant	Structural	DF	NSP	Cellulose	cellulose
CBM_4	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	xylan,beta-glucan
CBM_5	Animal,Fungal	Structural	DF	Unknown	Unknown	chitin
CBM_6	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	xylan,beta-glucan
CBM_9	Plant	Structural	DF	NSP	Hemicellulose	xylan
CBM_10	Plant	Structural	DF	NSP	Cellulose	cellulose
CBM_12	Animal,Fungal	Structural	DF	Unknown	Unknown	chitin
CBM_13	Plant	Structural	DF	NSP	Hemicellulose	mannose,xylan
CBM_15	Plant	Structural	DF	NSP	Hemicellulose	xylan
CBM_16	Plant,Fungal,Bacteria	Structural	DF	NSP	Cellulose,Mannans_and_ Heteromannans	cellulose,glucomannan
CBM_18	Animal,Fungal	Structural	DF	Unknown	Unknown	chitin
CBM_20	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_21	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_22	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	xylan,beta-glucan
CBM_23	Unknown	Unknown	Unknown	Unknown	Unknown	mannan
CBM_25	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_26	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_27	Unknown	Unknown	Unknown	Unknown	Unknown	mannan
CBM_30	Plant	Structural	DF	NSP	Cellulose	cellulose
CBM_32	Plant	Structural	DF	NSP	Mannans_and_ Heteromannans	galactose,lactose,lamirin,pustulan, beta-mannan,mannooligosaccharides
CBM_34	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_35	Plant	Structural	DF	NSP	Hemicellulose	xylan,manns,mannooligosaccharides, beta-galactan
CBM_36	Plant	Structural	DF	NSP	Hemicellulose	xylan,xylooligosaccharides
CBM_37	Plant,Animal,Fungal	Structural	DF	NSP,	Cellulose,Hemicellulose, Unknown	xylan,chitin,cellulose,alfalfa_cell_walls, ba_stems,wheat_straw
CBM_38	Plant	Storage	DF	Resistant_ oligosaccharides	Unknown	inulin
CBM_40	Animal,Bacteria, Archaea,Algae	Unknown	Unknown	Unknown	Unknown	sialic_acid

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
CBM_41	Plant,Animal,Fungal	Storage, Structural	DF,Glycogen	Resistant_starch, Unknown	Unknown	alpha-glucan,starch,glycogen, amylose,pullulan
CBM_42	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	arabinoxylan,arabin,arabinogalactan
CBM_47	Algae	Structural	DF	NSP	Unknown	fucoidan
CBM_48	Animal	Storage	Glycogen	Unknown	Unknown	glycogen
CBM_50	Animal,Fungal, Microbial	Structural	DF,PG	PG	Unknown,PG	peptidoglycan,chitin
CBM_51	Unknown	Unknown	Unknown	Unknown	Unknown	galactose,blood_group_antigen
CBM_52	Plant,Bacteria ,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_54	Plant,Animal, Fungal	Structural	DF	NSP,	Hemicellulose,Unknown	xylan,chitin,yeast_cell_wall
CBM_55	Animal,Fungal	Structural	DF	Unknown	Unknown	chitin
CBM_56	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_57	Unknown	Unknown	Unknown	Unknown	Unknown	beta-glucuronide
CBM_58	Unknown	Unknown	Unknown	Unknown	Unknown	maltoheptaose,acarbose
CBM_61	Unknown	Unknown	Unknown	Unknown	Unknown	beta-galactan
CBM_62	Plant,Bacteria	Structural	DF	NSP	Gum,Pectin,Hemicellulose	xyloglucan,arabinogalactan,galactomann
CBM_63	Plant	Structural	DF	NSP	Hemicellulose,Cellulose	cellulose,arabinoxylan
CBM_64	Plant	Structural	DF	NSP	Cellulose	cellulose
CBM_65	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	xyloglucan,beta-glucan
CBM_66	Unknown	Unknown	Unknown	Unknown	Unknown	inulan,levan
CBM_67	Unknown	Unknown	Unknown	Unknown	Unknown	rhamnose,mannose
CBM_68	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_69	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_70	Animal,Bacteria	GAG	GAG	Unknown	Unknown	hyaluron
CBM_71	Unknown	Unknown	Unknown	Unknown	Unknown	lactose,LacNAc
CBM_72	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose,Cellulose	cellulose,beta-glucan,xylan,beta-mannan
CBM_73	Animal,Fungal	Structural	DF	Unknown	Unknown	chitin
CBM_75	Plant	Structural	DF	NSP	Hemicellulose	xyloglucan
CBM_76	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_77	Plant	Structural	DF	NSP	Pectin	pectin
CBM_79	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan

(Sub) family	Origin_Glycan	Function_in_ origin	Function_ destination_1	Function_ destination_2	Function_destination_3	Glycan_annotation
CBM_80	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_81	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_82	Plant	Storage	DF	Resistant_starch	Unknown	starch
CBM_84	Unknown	Unknown	Unknown	Unknown	Unknown	xanthan
CBM_85	Plant,Bacteria,Fungal	Structural	DF	NSP	Hemicellulose	beta-glucan
CBM_86	Plant	Structural	DF	NSP	Hemicellulose	xylan
CBM_87	Animal,Bacteria	GAG	GAG	Unknown	Unknown	galactoseaminogalactan

References

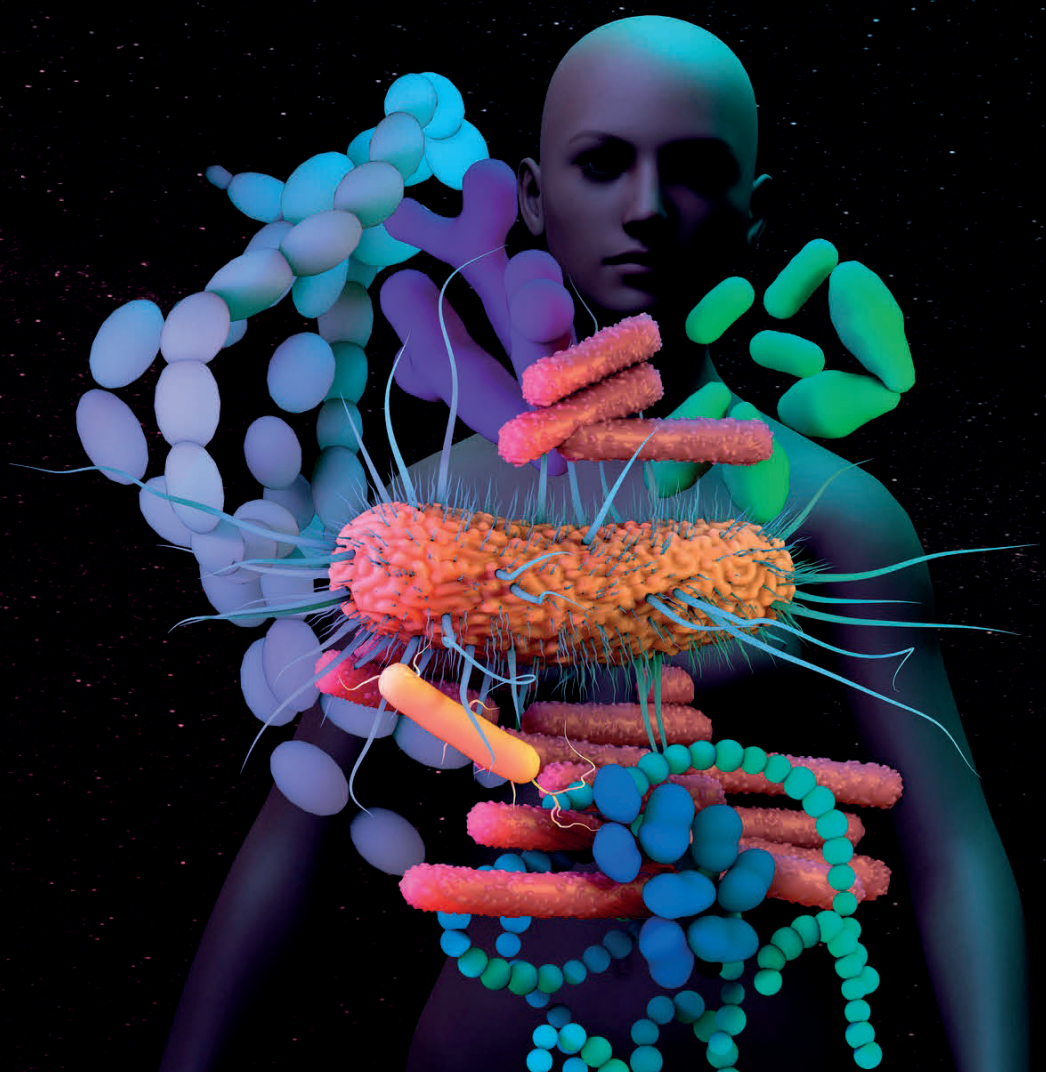
1. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol. 2013;11(7):497-504.
2. Makki K, Deehan EC, Walter J, Bäckhed F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. Cell Host Microbe. 2018;23(6):705-15.
3. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilán CG, Salazar N. Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. Front Microbiol. 2016;7:185.
4. Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science. 2018;362(6418).
5. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes. 2016;7(3):189-200.
6. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016;167(5):1339-53.e21.
7. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490-5.
8. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280 (Pt 2)(Pt 2):309-16.
9. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46(W1):W95-w101.
10. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80.
11. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026-8.
12. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32(8):834-41.
13. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7(1):84.
14. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;preprint arXiv:1303.3997.
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
16. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.
17. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014;10(11):766.
18. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, et al. Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing. PLoS One. 2016;11(5):e0155362.
19. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun. 2015;6:6528.
20. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017;66(1):70-8.
21. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat Med. 2019;25(6):968-76.
22. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 2019;25(4):667-78.
23. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25(4):679-89.
24. Hothorn T, Hornik K, van de Wiel MA, Zeileis A. A Lego System for Conditional Inference. The American Statistician. 2006;60(3):257-63.
25. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease

comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol. 2021;22(1):93.
26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B (Methodological). 1995;57(1):289-300.
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
28. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. BioRxiv. 2021:060012.
29. Ausland C, Zheng J, Yi H, Yang B, Li T, Feng X, et al. dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. Nucleic Acids Res. 2021;49(D1):D523-d8.
30. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol. 2012;10(5):323-35.
31. Patova OA, Golovchenko VV, Ovodov YS. Pectic polysaccharides: structure and properties. Russ Chem Bull. 2015;63(9):1901-24.
32. Stephen AM, Champ MM, Cloran SJ, Fleith M, van Lieshout L, Mejborn H, et al. Dietary fibre in Europe: current state of knowledge on definitions, sources, recommendations, intakes and relationships to health. Nutr Res Rev. 2017;30(2):149-90.
33. Yamada S, Sugahara K, Ozbek S. Evolution of glycosaminoglycans: Comparative biochemical study. Commun Integr Biol. 2011;4(2):150-8.
34. Namburi RB, Berteau O, Spillmann D, Rossi M. Chondroitinase AC: A host-associated genetic feature of Helicobacter bizzozeronii. Vet Microbiol. 2016;186:21-7.
35. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40(Web Server issue):W445-51.
36. Lesker TR, Durairaj AC, Gálvez EJC, Lagkouvardos I, Baines JF, Clavel T, et al. An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. Cell Rep. 2020;30(9):2909-22.e6.
37. Li J, Zhong H, Ramayo-Caldas Y, Terrapon N, Lombard V, Potocki-Veronese G, et al. A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. Gigascience. 2020;9(6).
38. Pan H, Guo R, Zhu J, Wang Q, Ju Y, Xie Y, et al. A gene catalogue of the Sprague-Dawley rat gut metagenome. Gigascience. 2018;7(5).
39. Youngblut ND, de la Cuesta-Zuluaga J, Reischer GH, Dauser S, Schuster N, Walzer C, et al. Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. mSystems. 2020;5(6).
40. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Ocean plankton. Structure and function of the global ocean microbiome. Science. 2015;348(6237):1261359.
41. Acinas SG, Sánchez P, Salazar G, Cornejo-Castillo FM, Sebastián M, Logares R, et al. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. Commun Biol. 2021;4(1):604.
42. von Schwartzenberg RJ, Bisanz JE, Lyalina S, Spanogiannopoulos P, Ang QY, Cai J, et al. Caloric restriction disrupts the microbiota and colonization resistance. Nature. 2021;595(7866):272-7.
43. Wastyk HC, Fragiadakis GK, Perelman D, Dahan D, Merrill BD, Yu FB, et al. Gut-microbiota-targeted diets modulate human immune status. Cell. 2021.
44. Bhattacharya T, Ghosh TS, Mande SS. Global Profiling of Carbohydrate Active Enzymes in Human Gut Microbiome. PLoS One. 2015;10(11):e0142038.
45. Kunzmann AT, Coleman HG, Huang WY, Kitahara CM, Cantwell MM, Berndt SI. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Am J Clin Nutr. 2015;102(4):881-90.
46. Aune D, Chan DS, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011;343:d6617.
47. Ben Q, Sun Y, Chai R, Qian A, Xu B, Yuan Y. Dietary fiber intake reduces risk for colorectal adenoma: a meta-analysis. Gastroenterology. 2014;146(3):689-99.e6.
48. Goodlad RA. Dietary fibre and the risk of colorectal cancer. Gut. 2001;48(5):587-9.
49. Chao A, Thun MJ, Connell CJ, McCullough ML, Jacobs EJ, Flanders WD, et al. Meat consumption and risk of colorectal cancer. JAMA. 2005;293(2):172-82.
50. Sandhu MS, White IR, McPherson K. Systematic review of the prospective cohort studies on meat consumption

and colorectal cancer risk: a meta-analytical approach. Cancer Epidemiol Biomarkers Prev. 2001;10(5):439-46.
51. Gurjao C, Zhong R, Haruki K, Li YY, Spurr LF, Lee-Six H, et al. Discovery and features of an alkylating signature in colorectal cancer. Cancer Discov. 2021.

Part III

Identifying microorganisms associated with colonization resistance in clinical studies



Chapter 6

The bacterial gut microbiota of adult patients infected, colonized or non-colonized by *Clostridioides difficile*

Microorganisms, 2020

M.J.T. Crobach*,[1], Q.R. Ducarmon*,[1,2], E.M. Terveer[1,2,3], C. Harmanus[1], I.M.J.G. Sanders[1], C.M. Verduin[4], E.J. Kuijper[1,2,3], R.D. Zwittink[1,2]

[1]Experimental Bacteriology, Department of Medical Microbiology, Center for Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands
[2]Center for Microbiome Analyses and Therapeutics, Department of Medical Microbiology, Center for Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands
[3]Netherlands Donor Feces Bank, Leiden, The Netherlands
[4]Department of Microbiology and Infection Prevention, Amphia Hospital, Breda, The Netherlands
*Equal contribution

Microorganisms. 2020 May 6;8(5):677.
Supplementary information available online.

Abstract

Gut microbiota composition in patients with *Clostridioides difficile* colonization is not well investigated. We aimed to identify bacterial signatures associated with resistance and susceptibility to *C. difficile* colonization (CDC) and infection (CDI). Therefore, gut microbiota composition from patients with CDC (n=41), with CDI (n=41) and without CDC (controls, n=43) was determined through 16S rRNA gene amplicon sequencing. Bacterial diversity was decreased in CDC and CDI patients ($p<0.01$). Overall microbiota composition was significantly different between control, CDC and CDI patients ($p=0.001$). Relative abundance of *Clostridioides* (most likely *C. difficile*) increased stepwise from controls to CDC and CDI patients. In addition, differential abundance analysis revealed that CDI patients' gut microbiota was characterised by significantly higher relative abundance of *Bacteroides* and *Veillonella* as compared to CDC patients and controls. Control patients had significantly higher *Eubacterium hallii* and *Fusicatenibacter* abundance than colonized patients. Network analysis indicated that *Fusicatenibacter* was negatively associated with *Clostridioides* in CDI patients, while *Veillonella* was positively associated with *Clostridioides* in CDC patients. Bacterial microbiota diversity is decreased in both CDC and CDI patients, but harbour a distinct microbiota.. *Eubacterium hallii* and *Fusicatenibacter* may indicate resistance against *C. difficile* colonization and subsequent infection, while *Veillonella* may indicate susceptibility to colonization and infection by *C. difficile*.

Introduction

Clostridioides difficile, formerly named *Clostridium difficile*, is an anaerobic, Gram-positive, spore-forming bacillus. It is the main causative agent of nosocomial diarrhea, with antibiotic use as most important risk factor. Nowadays, also community-associated diarrhea due to *C. difficile* is increasingly reported. Clinical symptoms arise when *C. difficile* spores germinate within the intestine and the viable bacteria start to produce toxins. The secretion of Toxin A (TcdA) and Toxin B (TcdB) leads to inflammation of the large intestine [1]. The clinical presentation may range from mild diarrhea to a life-threatening toxic megacolon [1]. However, the ingestion of *C. difficile* spores does not always lead to the development of symptomatic disease. *C. difficile* can also be silently present in the gut, without causing any symptoms. This condition is called asymptomatic *C. difficile* colonization [2]. Patients colonized with *C. difficile* play an important role in disease epidemiology, as they act as a reservoir for onward transmissions[3, 4] and they may also progress to infection themselves, especially in the presence of an underlying illness [5-7].

It is believed that the bacterial gut microbiota plays an important role in determining the susceptibility to colonization and subsequent infection with *C. difficile.* In patients with *C. difficile* infection (CDI), a lower richness and diversity, and decreased relative abundances of Bacteroidetes, *Ruminococcaceae* and *Lachnospiraceae* members have been described [8, 9]. The gut microbiota in *C. difficile* colonized patients is less well characterised[8, 10], but may give more insight into mechanisms that allow for colonization whilst protecting against infection. A previous study identified specific gut metabolites associated with colonization and infection by *C. difficile*, but did not determine microbiota composition [11]. In order to identify which bacterial signatures are associated with resistance and susceptibility to *C. difficile* colonization and CDI, we compared the gut microbiota of CDI patients, patients with *C. difficile* colonization (CDC) at hospital admission and patients without CDI or CDC at admission.

Materials and Methods

Subjects and sample collection

This study was designed to compare the gut microbiota between three groups: patients with *C. difficile* colonization (CDC) on hospital admission, patients without *C. difficile* colonization on hospital admission (controls) and hospitalised patients with CDI. For the first two groups, fecal samples were obtained from CDC and control patients admitted to Leiden University Medical Center (LUMC) or Amphia hospital as part of the CDD ("*Clostridium difficile* dragerschap" [carriership]) study, a study designed to determine

the prevalence of CDC at hospital admission, conducted between January 2015 and March 2016. Adult patients admitted to predefined medical and surgical wards were eligible for enrolment. Stool samples were requested within 72 hours of hospital admission. If patients were discharged home within 72hrs, a stool sample was collected at home and returned to the hospital by mail or in person. Colonized patients were defined as patients who tested positive for *C. difficile* by stool culture and were not clinically suspected of CDI within the first 72hrs of admission. For each colonized patient, the first consecutive patient with a negative stool culture for *C. difficile* was included to form the control group. For the third group, fecal samples were obtained from adult patients hospitalised in the LUMC and diagnosed with CDI between July 2015 and May 2017. All CDI cases had to comply with the definitions valid in the Dutch surveillance protocol [12], and CDI diagnosis was based on CDI symptoms in combination with laboratory CDI testing in agreement with the recommendations of the European Centre for Disease Control and Prevention [13]. *C. difficile* culturing and molecular diagnostics were performed as described below in the 'microbiological analysis' section. Patients initially participating in the CDD study but diagnosed with CDI within 72hrs of admission were included in the CDI group.

The LUMC institutional review board served as the central institutional review board and had no objection to the performance of the study. At the Amphia hospital, the directing board had no objection to the performance of the study. Stool samples from CDC and control patients were collected under verbal consent and written informed consent from these patients was obtained for collection of additional data (see below). A waiver for informed consent from CDI patients was obtained.

Microbiological analyses

Microbiological analyses were performed at the National Reference Laboratory for Clostridium difficile (LUMC, The Netherlands). Fecal samples were initially stored at 2-6°C and tested on the day of receipt, or the following working day in case of weekends or holidays.

Fecal samples from CDC and control patients were cultured on CLO plates (containing cefoxitin, amphotericin B and cycloserin, BioMérieux, The Netherlands) and after ethanol shock on CLO plates and CNA plates (containing colistin and nalidixinic acid, BioMérieux, The Netherlands). Suspicious colonies were tested by GDH PCR to confirm the presence of *C. difficile* [14]. In addition, a multiplex PCR for TcdA, TcdB and binary toxin genes was performed on the isolates to determine if CDC patients were colonized by a toxigenic or non-toxigenic strain [15].

Fecal samples from patients with suspected CDI were tested according to standard operating procedures which included an assay to detect free *C. difficile* toxins [16]. In addition, positive

tested samples were cultured for presence of *C. difficile* as described above.

C. difficile isolates from CDC patients and CDI patients were PCR ribotyped as previously described [17].

Patient data collection

Demographical data and data about medication use during the last three months (until admission for CDC patients and controls or until sample submission for CDI patients), previous hospitalisation in the last year and previous CDI episodes (ever and within the last eight weeks) were collected by questionnaires and electronic chart review (CDC and control patients) or chart review only (CDI patients). Recurrent CDI was defined as a new diarrheal episode within two to eight weeks after a previous diarrheal episode due to *C. difficile* and *C. difficile* re-infection as a new diarrheal episode more than 8 weeks after the previous diarrheal episode due to *C. difficile*.

Epidemiological analyses were performed to compare characteristics between control, CDC and CDI patients by one-way ANOVA or chi-squared test using STATA SE version 15.1 (StataCorp, College Station, TX, US).

Microbiota analysis

Samples

A total of 125 fecal samples were included: 43 samples from control patients, 41 samples from CDC patients and 41 samples from CDI patients. Samples from control and CDC patients were in 74/84 patients (88%) obtained within 72hrs after admission. Fecal samples were submitted from home by 15/84 patients (17.9%), while from the other 69/84 patients (82.1%) fecal samples were collected in the hospital.

DNA extraction, library preparation and sequencing

DNA was extracted from 0.1 gram faeces using the Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (ZymoResearch, CA, USA). Quality control, library preparation and sequencing were performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEXTflex™ 16S V4 Amplicon-Seq Kit (BiooScientific, TX, USA) and the Illumina HiSeq4000 platform (paired-end, 150bp). An average of 2.117.322 (707.362 – 5.742.717) reads per sample was obtained. Raw sequencing data is available in the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under study accession PRJEB30586.

Sequencing data analysis

Read filtering, operational taxonomic unit (OTU)-picking and taxonomic assignment were performed using the NGTax 0.4 pipeline with following settings: forward read length of 150, reverse read length of 120, ratio OTU abundance of 2.0, classify ratio of

0.9, minimum threshold of $1*10^{-7}$, identity level of 97%, error correction of 98.5, using the Silva_132_SSU Ref database [18-20]. The obtained OTU-table was filtered for OTUs with a number of sequences less than 0.005% of the total number of sequences [21]. A couple of technical duplicates were included for DNA extraction (n=3 samples) and sequencing (n=6 samples) procedures, indicating high replicability of results (Figure S1). Three negative controls were included from DNA extraction onwards and contained less than 1% of the number of reads obtained from fecal samples.

All microbiota analyses and data visualisation were performed in R (v3.5.1), using the packages phyloseq (v1.24.2), vegan (v2.5-2), ggplot2 (v3.0.0), DESeq2 (v1.20.0), microbiome (v1.2.1) and SpiecEasi (v0.1.4) [22-27]. Visualisation of network analysis was performed in Cytoscape (v3.7.0) [28]. Results were considered significant if $p \leq 0.05$, or Benjamini-Hochberg corrected $p \leq 0.05$ for differential abundance analysis. Prior to differential abundance testing (DESeq2) and network analysis (SpiecEasi), the OTU-table was filtered for OTUs present in less than 25% of samples. Nucleotide sequences belonging to the *Clostridioides* genus were blasted using the NCBI standard nucleotide blast, with 16S ribosomal sequences (Bacteria and Archaea) selected as the reference database, to determine if the sequence had a better hit to *C. difficile* or *C. mangenotii*. Kruskal-Wallis followed by post-hoc Dunn's testing was performed to compare Shannon diversity indices between the patient groups. Permutational multivariate analysis of variance (PERMANOVA) was performed using the 'adonis' function with 999 permutations and Bray-Curtis distances to separately investigate associations between microbiota composition and various clinical variables. Each clinical variable was tested separately using PERMANOVA. SPIEC-EASI, using the Meinshausen-Buhlman method for graph estimation of the network, was performed for network analysis with lambda.min.ratio=0.01, nlambda=20 and rep.num=99. This method is robust to many characteristics of 16S amplicon data, such as compositionality and dimensionality [22]. OTUs without a direct edge connection to another OTU were removed for visualisation purposes.

Results

Epidemiology

Thirty CDI episodes were primary episodes, seven were recurrent episodes and four were *C. difficile* re-infections. From four patients, two different episodes were included in this study. This mixture of primary episodes, recurrences and re-infections reflects the true CDI population, as recurrence and re-infection are common. Previous CDI, both within and beyond the last eight weeks, was common among CDI patients (17.1% for both) whereas it was uncommon in CDC patients (2.4% and 7.3%) and no previous CDI was recorded in controls. Antibiotics were used in the last three months in 97.6%

of CDI patients, 73.2% of CDC patients and 59.5% of control patients (p<0.001). The *C. difficile* PCR ribotype distribution in CDC and CDI patients is shown in Figure S2. Patient characteristics are shown in Table 1.

Table 1. Subject characteristics.

	CDI patients (n=41)	CDC patients (n=41)	Control patients (n=43)	p-value
Age in years, mean (SD)	57.5 (17.6)	55.3 (18.7)	57.8 (13.5)	0.76
Sex				0.47
Male	22/41 (53.7%)	22/41 (53.7%)	28/43 (65.1%)	
Female	19/41 (46.3%)	19/41 (46.3%)	15/43 (34.9%)	
Previous CDI				
Last 8 weeks	7/41 (17.1%)	1/41 (2.4%)	0/42 (0%)	0.003
>8 weeks earlier	7/41 (17.1%)	3/41 (7.3%)	0/42 (0%)	0.02
Current CDI episode				
primary episode	30/41 (73.2%)			
persistent primary episode	2/41 (4.9%)			
1st recurrence of primary episode	3/41 (7.3%)			
2nd recurrence of primary episode	1/41 (2.4%)			
5th recurrence of primary episode	1/41 (2.4%)			
1st reinfection	1/41 (2.4%)			
2nd reinfection	2/41 (4.9%)			
2nd recurrence of first reinfection	1/41 (2.4%)			
Previous hospitalisation (last year)	29/41 (70.7%)	30/41 (73.2%)	19/42 (45.2%)	0.01
Comorbidities				
IBD	2/41 (4.9%)	7/41 (17.1%)	2/42 (4.8%)	0.08
Solid organ transplant	17/41 (41.5%)	9/41 (22.0%)	2/42 (4.8%)	<0.001
Solid malignancy	5/41 (12.2%)	6/41 (14.6%)	11/42 (26.2%)	0.2
Hematological malignancy	9/41 (22.0%)	0/41 (0%)	2/42 (4.8%)	0.001
Previous medication use (last 3 months)				
Antibiotics	40/41 (97.6%)	30/41 (73.2%)	25/42 (59.5%)	<0.001
Immunosuppressants	30/41 (73.2%)	17/41 (41.5%)	13/42 (31.0%)	<0.001
Chemotherapy	10/41 (24.4%)	2/41 (4.9%)	5/42 (11.9%)	0.03
PPI or antacids	31/41 (75.6%)	30/41 (73.2%)	19/42 (45.2%)	0.006

Bacterial community structure

To elucidate characteristics of the bacterial community structure, several tests were performed. We determined bacterial diversity using the Shannon index, performed PERMANOVA to relate microbiota composition to clinical factors, and clustered samples based on both weighted and unweighted UniFrac distance metrices for between-sample comparisons. Bacterial diversity was significantly higher in controls as compared to CDC and CDI patients (p<0.01), but did not differ between CDC and CDI patients (Figure 1).

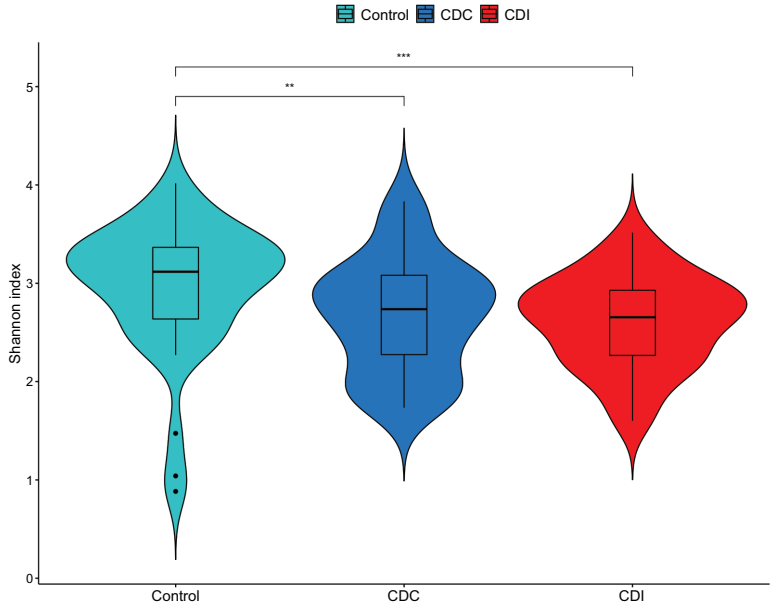


Figure 1: Violin plot of alpha diversity, as measured by the Shannon index, in control, CDC and CDI patients. The box plot shows the median, 25th and 75th percentile and whiskers indicate 1.5* interquartile range. *p<0.05, **p<0.01, ***p<0.001.

The most important clinical factor associated with microbiota composition was the patient group (PERMANOVA, p=0.001, R^2=0.075). Additional pairwise comparisons revealed that microbiota composition of control, CDC and CDI patients all differed from each other (PERMANOVA, p<0.01). The difference in microbiota composition between groups could also be observed via sample clustering based on unweighted UniFrac distance, but was less apparent, although still visible, using weighted UniFrac distance (Figure 2), reflecting differences in presence/absence of bacterial taxa rather than in their relative abundance. Here, microbiota composition of CDC patients are scattered, with some samples being more similar to CDI patients and others to controls. Clustering analysis solely on the CDC group showed no differentiation in microbiota composition by toxinogenic or non-toxinogenic *C. difficile* carriership, or by any other variable (data not shown).

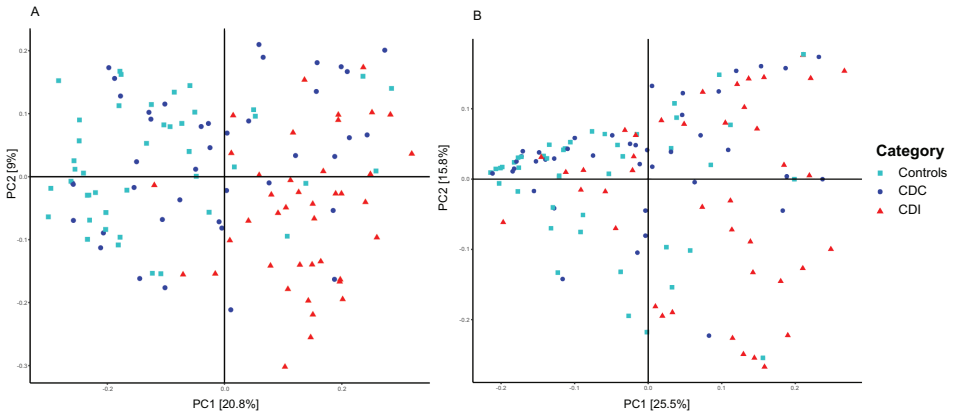


Figure 2. Principal Coordinates Analysis (PCoA) based on unweighted (A) and weighted (B) UniFrac distances. Each sample is represented by a shape and color according to its category. The percentage of variation explained by the two first PCoA dimensions is indicated on the respective axes.

In addition to patient group, overall microbiota composition was significantly affected by solid organ transplantation, previous CDI, PPIs/antacids, immunosuppressants and specific antibiotics, including vancomycin and metronidazole, which are commonly prescribed antibiotics for CDI treatment (Table S2). However, effect sizes were smaller for these clinical variables than for segregation by patient group. Since antibiotics are known to alter gut microbiota composition, we explored whether antibiotic use in the previous three months affected microbiota composition within the control and CDC group. This was indeed the case for control patients (PERMANOVA, p=0.035, R^2=0.044), but not for CDC patients (PERMANOVA, p=0.409, R^2=0.031). Within the control group, antibiotic use also impacted bacterial diversity, with a trend for increased diversity in the non-antibiotic group (p=0.0518). For these reasons, the control group was separated in controls with (C+AB) and without (C-AB) previous antibiotic use for differential abundance analysis.

Relative abundance of individual bacterial taxa

In order to study the differential abundance of bacterial taxa between the patient groups, DESeq2 analysis was performed. Relative abundance of *Clostridioides* showed a significant stepwise increase from C-AB (<0.01 ± <0.01%), C+AB (0.05 ± 0.2%) to CDC (0.7 ± 2.2%) and CDI patients (2.5 ± 2.9%) (Table S1A and S1B). It is, however, important to take prevalence into consideration, as *Clostridioides* reads were detected in only 26/41 CDC patients (63.4%) and in 38/41 CDI patients (92.7%). The nucleotide sequence belonging to this *Clostridioides* OTU resulted in a 100% sequence identity with two *C. difficile* strains, but only a 94% sequence identity with *C. mangenotii*.

Compared to CDC patients, C+AB and C-AB had an increased relative abundance of *Eubacterium hallii* (Figure 3, Table S1A,B). As expected, more and larger differences were observed between C-AB and CDC patients than between C+AB and CDC patients. In addition to an increase in *E. hallii,* the relative abundance of *Fusicatenibacter* was significantly higher in C-AB compared to CDC patients, while the relative abundance of several *Enterococci*, *Ruminococcus gnavus* and *Lachnoclostridium* were significantly lower (Figure 3, Table S1A,B).

Compared to C+AB and CDC patients, microbiota of CDI patients was characterised by a higher relative abundance of *Clostridioides*, *Bacteroides* and *Veillonella*, and by a lower abundance of genera belonging to the *Ruminococcaceae* family and Actinobacteria phylum (Figure 3, Table S1A,B). Many of these lower abundant genera are known short-chain fatty acid (SCFA)-producers and carbohydrate degraders. Additionally, CDI patients had increased relative abundance of *R. gnavus* and *Lachnoclostridium* compared to C+AB patients. To avoid antibiotic use bias, CDI patients were not compared to C-AB patients.

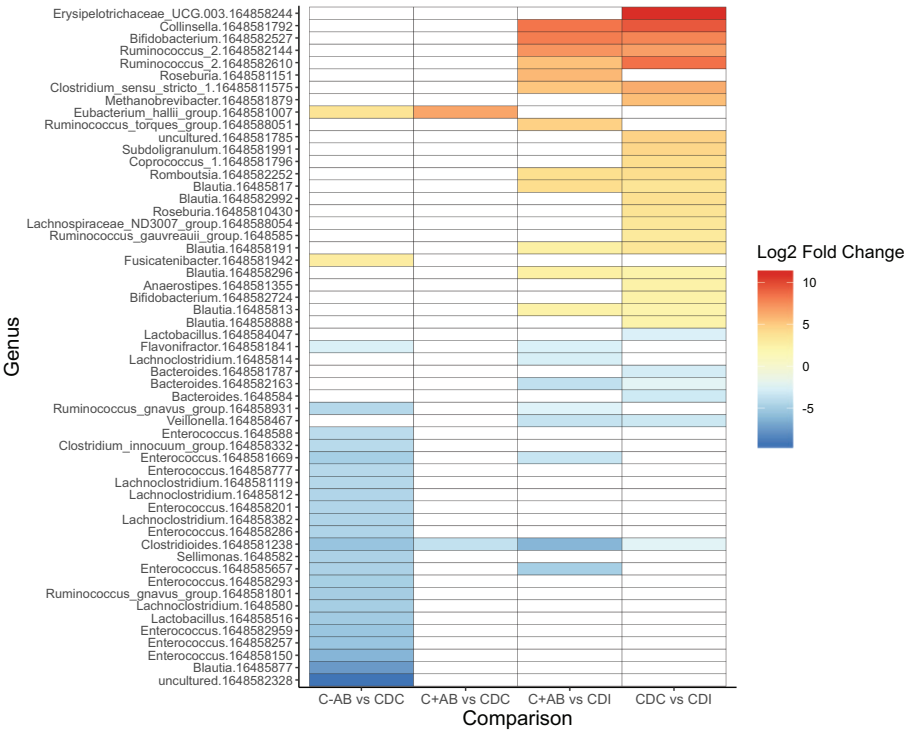


Figure 3: Heatmap showing differentially abundant bacterial taxa between C-AB, C+AB, CDC and CDI patients. Bacterial taxa with a Log2 fold change of at least (-)2.25 and a Benjamini-Hochberg corrected p-value ≤ 0.05 are shown on OTU-level. OTU numbers are indicated as 164858xxxxxx. A full overview can be found in Table S1A.

Bacterial networks

To investigate connectivity of the differentially abundant *Clostridioides* genus with other bacterial genera, network analysis was performed on microbiota composition profiles of CDC and CDI patients. In CDI patients, *Fusicatenibacter* was negatively associated with *Clostridioides* (Figure 4A). In CDC patients, a positive association between *Clostridioides* and *Veillonella* was observed (Figure 4B).

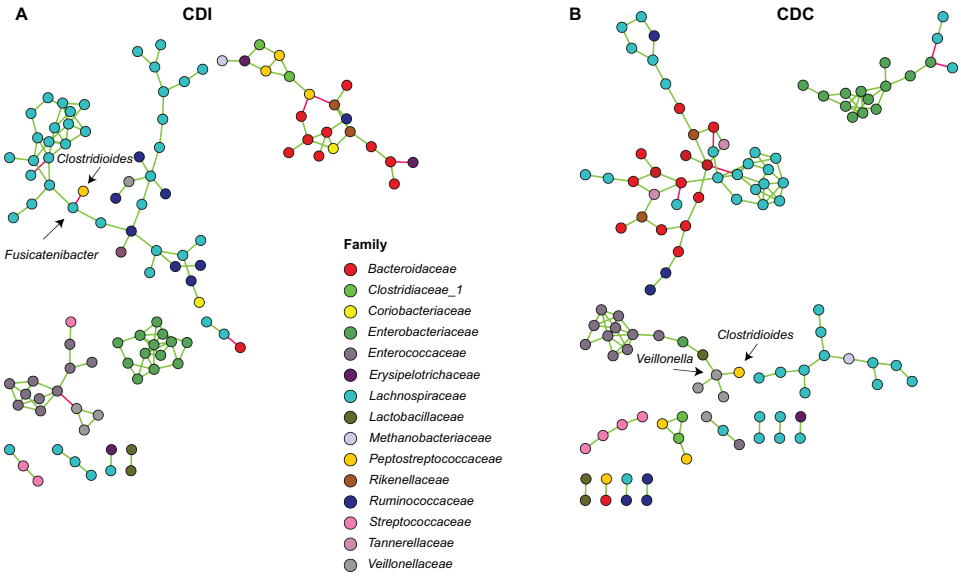


Figure 4: Association network analysis using SPIEC-EASI in CDI (A) and CDC (B) patients. Each node represents a single OTU and is colored according to family-level taxonomy. Green edges indicates a positive association between nodes, red edges indicate a negative association between nodes.

Discussion

It is generally accepted that CDI can develop due to a disturbed gut microbiota. In contrast, not much is known about the role of the gut microbiota in *C. difficile* colonization. In this study, the gut microbiota of patients with asymptomatic *C. difficile* colonization was characterised. CDC patients had unique gut microbiota signatures and bacterial taxa could be identified that may be of relevance for further mechanistic studies. While 16S rRNA gene amplicon sequencing did not allow for identification of *Clostridioides* in all colonized and CDI patients, its relative abundance increased in a step-wise manner from controls to colonized patients and CDI patients.

Bacterial diversity was decreased in CDC and CDI patients, and microbiota composition was mostly patient group-specific. Interestingly, microbiota composition was associated with previous antibiotic use within the control group, but not within the CDC group. This may suggest that CDC patients already have a disturbed bacterial community prior to colonization, independent of antibiotic treatment, although the underlying reason remains unclear. Another explanation could be that, as 73.2% of CDC patients had previous antibiotic use, too few CDC patients without antibiotic use were included to effectively identify an antibiotic treatment effect within this group.

Multiple differentially abundant genera were found between control, CDC and CDI patients, and included *Eubacterium hallii*, *Fusicatenibacter* and *Veillonella*. Bacterial network analysis showed that *C. difficile* was directly negatively associated with *Fusicatenibacter* in CDI patients, and directly positively associated with *Veillonella* in CDC patients. This may indicate that *Fusicatenibacter* may play a role in preventing CDI development, while *Veillonella* may play a role in *C. difficile* colonization, respectively. It has previously been hypothesized that *Eubacterium* species are protective against CDI development in asymptomatic carriers [10]. In our study, *E. hallii* was more abundant in controls (with and without antibiotic use) than in CDC patients. *E. hallii* is known to produce the three main SCFAs, propionate, acetate and butyrate [29, 30], and is increasingly investigated for its potential benefit in metabolic disease [31]. This bacterium may contribute to colonization resistance against *C. difficile* through SCFAs production, although the role of SCFAs against *C. difficile* remains debated [32, 33]. Possibly, *E. hallii* contributes to colonization resistance through secondary bile-acid production. Secondary bile acids are known to inhibit *C. difficile* growth and a secondary bile acid-producing bacterium, *Clostridium scindens*, enhances colonization resistance against *C. difficile* [34, 35]. *E. hallii* possesses *bsh* genes, necessary for deconjugation of conjugated bile acids, which is a crucial step prior to converting deconjugated bile acids into secondary bile acids [31]. However, although the most important enzyme for secondary bile acids conversion, 7α-dehydroxylase, was demonstrated to be present in *Eubacterium* species, no homologue has been detected in *E. hallii's* genome [31, 36-38].

Veillonella was more abundant in CDI patients compared to CDC patients and controls with prior antibiotic use, and was positively associated with *Clostridioides* in colonized patients in our study. *Veillonella* is normally found in the oral cavity where it can form dental plaques with *Streptococcus*, but is also found in atherosclerotic plaques and fecal samples from patients with atherosclerosis [39, 40]. *Veillonella and* streptococci may be metabolically linked through lactic acid, which also holds for other lactic-acid producing bacteria, like lactobacilli [41, 42]. *Lactobacillus* and *Veillonella* were indeed directly positively linked in our network analysis. While increased relative abundance of *Veillonella* may be a result of intrinsic resistance to multiple antibiotics, a recent *in vitro*

study showed that *Veillonella* increases when a dysbiotic microbiota is co-cultivated with *C. difficile* [43]. In addition, increased *Veillonella* abundance has been reported prior to CDI onset [44]. These studies, combined with our data, suggest that *Veillonella* is associated with *C. difficile* colonization and infection. It remains unclear whether *Veillonella* has a role in CDI development (e.g. via biofilm formation), or whether it simply outgrows as a result of altered metabolic pathways or unoccupied niches in the gut due to antibiotic use or *C. difficile* expansion.

Fusicatenibacter was differentially abundant between C-AB and CDC patients, and was negatively associated with *Clostridioides* in CDI patients in our study. This bacterium has only been cultured recently (2013) and we are the first to describe an association between *Fusicatenibacter* and *C. difficile* colonization or infection [45]. Previously, *Fusicatenibacter sacchivorans*, the only known species within the *Fusicatenibacter* genus, was shown to be increased in inactive ulcerative colitis (UC) patients and decreased in active UC, related to its positive association with IL-10 production [46].

Our study had some limitations. Almost all diagnosed CDI patients (39 of 41) came from the LUMC while CDC and controls were derived from both Amphia hospital and LUMC. As the LUMC is a university affiliated hospital instead of a general hospital, patient characteristics in these groups may not have been completely comparable. As such, solid organ transplants, previous hospitalisation, immunosuppressant use and chemotherapy were more frequent in LUMC. Several of these clinical variables significantly affected overall microbiota composition, which challenges studying the sole effect of CDI on microbiota composition. Another limitation is that a single stool sample was available. Therefore, it is impossible to determine if patients were transiently or persistently colonized by *C. difficile.* Patients classified as CDC might have included patients with only transient passage of spores [2]. Lastly, we have not performed functional characterisation of the microbiota, e.g. by metabolomics or transcriptomics.

However, our study had multiple important strengths. Firstly, this is the first study which investigates microbiota composition of *C. difficile* colonized patients, as compared to controls and CDI patients, with more than ten patients included per group. This allowed for more robust statistical analysis, and for detecting smaller and subtle changes within the composition. Secondly, controls in this study were not healthy controls. Instead, controls and CDC patients were selected from the same cohort of newly admitted patients and all three groups were hospitalised on the same wards to make the comparisons more fair. Thirdly, CDI was well defined. Although molecular testing is nowadays often used as a stand-alone test to diagnose CDI, these assays cannot discern colonization from infection [47]. In our study, all samples suspected of CDI were (also) tested with an assay detecting free toxins. Laboratory results were interpreted in conjunction with

clinical symptoms. According to the Dutch sentinel surveillance program and the ECDC criteria, patients had to have diarrhea for at least 2 days and/or pseudomembranous colitis at endoscopy and no other apparent cause of diarrhea. Although milder cases may have been missed by using these strict criteria, we are quite confident that our CDI group consisted of clinical relevant CDI cases requiring CDI treatment. Fourthly, duplicates for DNA extraction and sequencing were included to detect potential bias. All these duplicates showed very high similarity in composition profiles, demonstrating the reproducibility of DNA extraction and sequencing procedures (Figure S1).

Conclusion

We demonstrated that colonization and infection by *C. difficile* is associated with decreased bacterial diversity in the gut and differences in relative abundance of specific bacterial taxa including *Veillonella*, *Fusicatenibacter*, *Eubacterium hallii, Bacteroides* and members of the *Lachnospiraceae* and *Ruminococcaceae* families. Future studies could focus on functional characterisation of the microbiota, e.g. by metabolomics or transcriptomics, and on co-cultivation of specific bacteria, e.g. *Fusicatenibacter* with *C. difficile*, in light of *C. difficile* colonization and infection. In addition, it is relevant to determine if the observed gut microbiota changes are present before acquiring colonization and/or CDI, or merely a consequence.

Notes

Acknowledgments
We are indebted to all patients who participated in this study and we would like to thank Melanie Srodzinski, Inge van Duijn, Michelle de Raaf and René Vermaire for their help in obtaining fecal samples and patients' questionnaires. We thank Bastian Hornung for processing of the raw sequencing data.

Author contributions
Conceptualization, M.J.T.C., Q.R.D., E.J.K. and R.D.Z.; Data curation, M.J.T.C. and K.M.V.; Formal analysis, M.J.T.C., Q.R.D. and R.D.Z.; Funding acquisition, M.J.T.C. and E.J.K.; Investigation, M.J.T.C., C.H., I.M.J.G.S. and R.D.Z.; Methodology, M.J.T.C., Q.R.D. and R.D.Z.; Project administration, M.J.T.C. and K.M.V.; Resources, M.J.T.C. and K.M.V.; Software, Q.R.D. and R.D.Z.; Supervision, E.J.K. and R.D.Z.; Validation, M.J.T.C., Q.R.D. and E.M.T.; Visualization, M.J.T.C. and Q.R.D.; Writing – original draft, M.J.T.C. and Q.R.D.; Writing – review & editing, M.J.T.C., Q.R.D., E.M.T., C.H., I.M.J.G.S., K.M.V., E.J.K. and R.D.Z.. All authors contributed substantially, approved

the submitted version and agree to be personally accountable for their own contributions and for ensuring that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and documented in the literature.

Conflict of Interest

ET and EK are supported by an unrestricted grant from Vedanta Biosciences Inc. The sponsor had no role in the design, execution, interpretation, or writing of the study.

Ethics statement

The LUMC institutional review board served as the central institutional review board and had no objection to the performance of the study. At the Amphia hospital, the directing board had no objection to the performance of the study. Stool samples from CDC and control patients were collected under verbal consent and written informed consent from these patients was obtained for collection of additional data. A waiver for informed consent was obtained from CDI patients.

Funding

This work was supported by the Netherlands Organisation for Health Research and Development, ZonMW (grant 522001005). The funding source had no role in collection, analysis and interpretation of the data, writing of the manuscript or decision to submit the publication.

References

1. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. Clostridium difficile infection. Nat Rev Dis Primers **2016**; 2: 16020.
2. Crobach MJT, Vernon JJ, Loo VG, Kong LY, Pechine S, Wilcox MH, Kuijper EJ. Understanding Clostridium difficile colonization. Clin Microbiol Rev **2018**; 31(2).
3. Eyre DW, Griffiths D, Vaughan A, Golubchik T, Acharya M, O'Conor L, Crook DW, Walker AS, Peto TEA. Asymptomatic Clostridium difficile colonization and onward transmission. PLoS One **2013**; 8(11): e78445.
4. Kong LY, Eyre DW, Corbeil J, Raymond F, Walker AS, Wilcox MH, Crook DW, Michaud S, Toye B, Frost E et al. Clostridium difficile: Investigating transmission patterns between infected and colonized patients using whole genome sequencing. Clin Infect Dis **2019**; 68(2): 204-9.
5. Zacharioudakis IM, Zervou FN, Pliakos EE, Ziakas PD, Mylonakis E. Colonization with toxinogenic C. difficile upon hospital admission, and risk of infection: a systematic review and meta-analysis. Am J Gastroenterol **2015**; 110(3): 381-90; quiz 91.
6. Blixt T, Gradel KO, Homann C, Seidelin JB, Schonning K, Lester A, Houlind J, Stangerup M, Gottlieb M, Knudsen JD. Asymptomatic carriers contribute to nosocomial Clostridium difficile infection: A cohort study of 4508 patients. Gastroenterology **2017**; 152(5): 1031-41.e2.
7. Rodriguez C, Romero E, Garrido-Sanchez L, Alcain-Martinez G, Andrade RJ, Taminiau B, Daube G, Garcia-Fuentes E. Microbiota insights in Clostridium difficile infection and inflammatory bowel disease. Gut Microbes **2020**.
8. Zhang L, Dong D, Jiang C, Li Z, Wang X, Peng Y. Insight into alteration of gut microbiota in Clostridium difficile infection and asymptomatic C. difficile colonization. Anaerobe **2015**; 34: 1-7.
9. Antharam VC, Li EC, Ishmael A, Sharma A, Mai V, Rand KH, Wang GP. Intestinal dysbiosis and depletion of butyrogenic bacteria in Clostridium difficile infection and nosocomial diarrhea. J Clin Microbiol **2013**; 51(9): 2884-92.
10. Vincent C, Miller MA, Edens TJ, Mehrotra S, Dewar K, Manges AR. Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. Microbiome **2016**; 4: 12.
11. Robinson JI, Weir WH, Crowley JR, Hink T, Reske KA, Kwon JH, Burnham CD, Dubberke ER, Mucha PJ, Henderson JP. Metabolomic networks connect host-microbiome processes to human Clostridioides difficile infections. J Clin Invest **2019**; 130: 3792-806.
12. Vendrik KEW, Crobach MJT, Terveer EM, Harmanus C, Sanders IMJG, Kuijper EJ, Notermans DW, De Greeff SC, Alblas J, Van Dissel JT. Twelfth annual report of the National Reference Laboratory for Clostridium difficile and results of the sentinel surveillance May 2017 - May 2018. Bilthoven, Netherlands: RIVM, **2018**.
13. European Centre for Disease Prevention and Control. European Surveillance of Clostridium difficile infections. Surveillance protocol version 2.3. Stockholm, **2017**.
14. Paltansing S, van den Berg RJ, Guseinova RA, Visser CE, van der Vorm ER, Kuijper EJ. Characteristics and incidence of Clostridium difficile-associated disease in The Netherlands, 2005. Clin Microbiol Infect **2007**; 13(11): 1058-64.
15. Persson S, Torpdahl M, Olsen KE. New multiplex PCR method for the detection of Clostridium difficile toxin A (tcdA) and toxin B (tcdB) and the binary toxin (cdtA/cdtB) genes applied to a Danish strain collection. Clin Microbiol Infect **2008**; 14(11): 1057-64.
16. Terveer EM, Crobach MJ, Sanders IM, Vos MC, Verduin CM, Kuijper EJ. Detection of Clostridium difficile in feces of asymptomatic patients admitted to the hospital. J Clin Microbiol **2017**; 55(2): 403-11.
17. Fawley WN, Knetsch CW, MacCannell DR, Harmanus C, Du T, Mulvey MR, Paulick A, Anderson L, Kuijper EJ, Wilcox MH. Development and validation of an internationally-standardized, high-resolution capillary gel-based electrophoresis PCR-ribotyping protocol for Clostridium difficile. PLoS One **2015**; 10(2): e0118150.
18. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ, Smidt H. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes[version 1; referees: 2 approved with reservations, 1 not approved]. F1000Research **2016**; 5.
19. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **2013**; 41(Database issue): D590-6.
20. Ducarmon QR, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD. Toward standards in clinical microbiome

studies: comparison of three DNA extraction methods and two bioinformatic pipelines. bioRxiv 751123 **2019**.
21. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordin JI, Knight R, Mills DA, Caporaso JG. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods **2013**; 10(1): 57-9.
22. Kurtz ZD, Muller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol **2015**; 11(5): e1004226.
23. Lahti L, Shetty S. Tools for microbiome analysis in R. Microbiome package version 1.2.1. **2017**.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **2014**; 15(12): 550.
25. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One **2013**; 8(4): e61217.
26. Oksanen JF, Blanchet GF, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P et al. Vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan, **2018**.
27. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer Science & Business Media, **2009**.
28. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **2003**; 13(11): 2498-504.
29. Engels C, Ruscheweyh HJ, Beerenwinkel N, Lacroix C, Schwab C. The common gut microbe Eubacterium hallii also contributes to intestinal propionate formation. Front Microbiol **2016**; 7: 713.
30. Duncan SH, Louis P, Flint HJ. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl Environ Microbiol **2004**; 70(10): 5810-7.
31. Udayappan S, Manneras-Holm L, Chaplin-Scott A, Belzer C, Herrema H, Dallinga-Thie GM, Duncan SH, Stroes ESG, Groen AK, Flint HJ et al. Oral treatment with Eubacterium hallii improves insulin sensitivity in db/db mice. NPJ Biofilms Microbiomes **2016**; 2: 16009.
32. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, Huffnagle GB, Li JZ, Young VB. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. Nat Commun **2014**; 5: 3114.
33. Lawley TD, Clare S, Walker AW, Stares MD, Connor TR, Raisen C, Goulding D, Rad R, Schreiber F, Brandt C et al. Targeted restoration of the intestinal microbiota with a simple, defined bacteriotherapy resolves relapsing *Clostridium difficile* disease in mice. PLoS Pathog **2012**; 8(10): e1002995.
34. Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A et al. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature **2015**; 517(7533): 205-8.
35. Sorg JA, Sonenshein AL. Inhibiting the initiation of *Clostridium difficile* spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol **2010**; 192(19): 4983-90.
36. Doerner KC, Takamine F, LaVoie CP, Mallonee DH, Hylemon PB. Assessment of fecal bacteria with bile acid 7 alpha-dehydroxylating activity for the presence of bai-like genes. Appl Environ Microbiol **1997**; 63(3): 1185-8.
37. Shetty SA, Ritari J, Paulin L, Smidt H, De Vos WM. Complete Genome Sequence of Eubacterium hallii Strain L2-7. Genome Announc **2017**; 5(43).
38. Takamine F, Imamura T. Isolation and characterization of bile acid 7-dehydroxylating bacteria from human feces. Microbiol Immunol **1995**; 39(1): 11-8.
39. Mikx FH, van der Hoeven JS, Konig KG, Plasschaert AJ, Guggenheim B. Establishment of defined microbial ecosystems in germ-free rats. I. The effect of the interactions of streptococcus mutans or Streptococcus sanguis with Veillonella alcalescens on plaque formation and caries activity. Caries Res **1972**; 6(3): 211-23.
40. Koren O, Spor A, Felin J, Fak F, Stombaugh J, Tremaroli V, Behre CJ, Knight R, Fagerberg B, Ley RE et al. Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A **2011**; 108 Suppl 1: 4592-8.
41. Mashima I, Nakazawa F. The influence of oral Veillonella species on biofilms formed by Streptococcus species. Anaerobe **2014**; 28: 54-61.
42. van den Bogert B, Erkus O, Boekhorst J, de Goffau M, Smid EJ, Zoetendal EG, Kleerebezem M. Diversity of human small intestinal Streptococcus and Veillonella populations. Fems Microbiology Ecology **2013**; 85(2): 376-88.
43. Horvat S, Rupnik M. Interactions Between Clostridioides difficile and fecal Microbiota in in vitro batch model: Growth, sporulation, and microbiota changes. Front Microbiol **2018**; 9: 1633.
44. Khanna S, Montassier E, Schmidt B, Patel R, Knights D, Pardi DS, Kashyap P. Gut microbiome predictors of

treatment response and recurrence in primary Clostridium difficile infection. Aliment Pharmacol Ther **2016**; 44(7): 715-27.
45. Takada T, Kurakawa T, Tsuji H, Nomoto K. Fusicatenibacter saccharivorans gen. nov., sp. nov., isolated from human faeces. Int J Syst Evol Microbiol **2013**; 63(Pt 10): 3691-6.
46. Takeshita K, Mizuno S, Mikami Y, Sujino T, Saigusa K, Matsuoka K, Naganuma M, Sato T, Takada T, Tsuji H. A single species of Clostridium subcluster XIVa decreased in ulcerative colitis patients. Inflamm Bowel Dis **2016**; 22(12): 2802-10.
47. Crobach MJ, Planche T, Eckert C, Barbur F, Terveer EM, Dekkers OM, Wilcox MH, Kuijper EJ. European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for Clostridium difficile infection. Clin Microbiol Infect **2016**; 22 Suppl 4: S63-81.

Chapter 7

Dynamics of the bacterial gut microbiota during controlled human infection with *Necator americanus* larvae

Gut microbiota during *Necator americanus* infection
Gut Microbes, 2020

Q.R. Ducarmon[a,b], M.A. Hoogerwerf[c], J.J. Janse[c], A.R. Geelen[a,b], J.P.R. Koopman[c], R.D. Zwittink[a,b], J.J. Goeman[d], E.J. Kuijper[a,b], M. Roestenberg[c,e]

[a]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands
[b]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands
[c]Department of Parasitology, Leiden University Medical Center, Leiden, The Netherlands
[d]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands
[e]Department of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

Gut Microbes. 2020 Nov 9;12(1):1-15.
Supplementary information available online.

Abstract

Hookworms are soil-transmitted helminths that use immune-evasive strategies to persist in the human duodenum where they are responsible for anemia and protein loss. Given their location and immune regulatory effects, hookworms likely impact the bacterial microbiota. However, microbiota studies struggle to deconvolute the effect of hookworms from confounders such as coinfections and malnutrition. We thus used an experimental human hookworm infection model to explore temporal changes in the gut microbiota before and during hookworm infection. Volunteers were dermally exposed to cumulative dosages of 50, 100 or 150 L3 *Necator americanus* larvae. Fecal samples were collected for microbiota profiling through 16S rRNA gene amplicon sequencing at weeks zero, four, eight, fourteen and twenty. During the acute infection phase (trial week zero to eight) no changes in bacterial diversity were detected. During the established infection phase (trial week eight to twenty), bacterial richness (Chao1, p=0.0174) increased significantly over all volunteers. No relation was found between larval dosage and diversity, stability or relative abundance of individual bacterial taxa. GI symptoms were associated with an unstable microbiota during the first eight weeks and rapid recovery at week twenty. *Barnesiella*, amongst other taxa, was more abundant in volunteers with more GI symptoms throughout the study. In conclusion, this study showed that clinical GI symptoms following *N. americanus* infection are associated with temporary microbiota instability and relative abundance of specific bacterial taxa. These results suggest a possible role of hookworm-induced enteritis on microbiota stability.

Introduction

Helminths such as hookworms can have beneficial effects on auto-immune diseases[1, 2] such as celiac disease,[3, 4] but also cause eosinophilic gastroenteritis, anemia and protein loss and are therefore responsible for a high burden of disease in low- and middle-income countries.[5] As a part of the human gut microbiome in developing countries with a high rate of hookworm infections, hookworms can exert evolutionary pressure on the bacterial gut ecosystem through intestinal motility, mucin glycosylation, mucus secretion, epithelial damage and worm products.[6] For example, several helminths and their products have been shown to increase permeability of monolayers in cell culture.[7, 8] In addition, worm products can have direct antibacterial activity, thereby having the potential to directly alter the bacterial gut microbiota.[9, 10] However, the complex interplay between hookworms such as *Necator americanus* and the bacterial microbiota is largely unknown.

In real-world settings, most studies have focused on characterizing the gut microbiota of infected individuals in highly endemic regions with limited follow-up on individuals.[11, 12] However, effects of confounding factors cannot always be uncoupled from the bacteria-helminth relationship, as mixed helminth infections, other intestinal diseases and malnutrition are also common in endemic regions.[13] These factors may explain a large part of inconsistent findings between studies.[13] In addition, due to the high inter-individual variability of the microbiome, cross-sectional studies only yield limited information.

In the current study, we studied the effect of hookworm infection on the gut microbiota using a longitudinal model for human hookworm infection in healthy volunteers (controlled human hookworm infection model, CHHIM). Here, samples can be obtained at baseline, where the gut microbiome is unperturbed, and longitudinally in order to model the ecosystem's dynamics and perturbation after exposure to *N. americanus*. This model allows for studying the changes in the bacterial microbiota in the different stages of infection; skin penetration, (pulmonary) migration and gut establishment, a process which takes roughly four weeks.[14] In addition, potential confounding factors which could affect the outcome of studies investigating bacterial-helminth interactions are minimized.[13] The power of CHHIM to investigate changes in the human microbiota has been demonstrated in a small study where patients with celiac disease were experimentally infected.[15-17] Although a very small study (n=8), a minor increase in richness was seen after infection, while no changes in community, diversity or abundance of individual taxa occurred.[15] This study was however limited by the use of a low infectious inoculum of twenty larvae which resulted in egg output much lower than commonly found in endemic areas and by only including patients with celiac disease.[18] In this study we infected individuals with 50-150 L3 larvae, after which we found mean egg counts of around 1500 eggs per gram feces at plateau level, which is more in line with the

endemic situation where mild infection is defined by WHO as <2000 eggs per gram feces. [18, 19] Still, infection levels in CHHIMs are not fully comparable to areas with a high infectious burden, defined as >4000 eggs per gram by WHO. The current study had two main aims. First, to investigate temporal changes in the gut microbiota in response to different dosages (ranging from 50 to 150L3) *N. americanus* larvae in healthy young volunteers. Second, to investigate temporal differences in the gut microbiota between healthy volunteers experiencing different amounts of clinical symptoms.

Results

Results of the clinical trial have been published elsewhere.[19] Briefly, of the 24 randomized volunteers, twenty completed follow-up and were included in the microbiota analysis, providing a total of 100 fecal samples. The primary aim of the clinical trial was to investigate the effect of repeated infectious dosages on hookworm egg excretion and variability. From our 20 volunteers, eight (40%) were male and twelve (60%) were female and the mean age was 25.7 years (standard deviation 6.1 years). No volunteers had used antibiotics in the six weeks prior to enrolment. All volunteers developed patent hookworm infection as shown by positive microscopy for hookworm eggs at a median of eight weeks (range five-nine) after first skin infection with L3 larvae.[19] Abdominal adverse events in many volunteers starting at three to four weeks after infection were paralleled by eosinophil increases which likely marked the timepoint of arrival and establishment of the hookworm in the duodenum. Abdominal adverse events consisted of bloating, nausea, vomiting, diarrhea or abdominal cramping. Volunteers exposed to higher larval dosages (n=6 volunteers with 50L3 larvae, n=7 with 100L3 larvae and n=7 with 150L3 larvae) generally had higher egg loads in feces, but there was no relation between cumulative larval dosage and number and severity of adverse events.[19] Based on the severity, number and duration of adverse events, nine volunteers were classified into the "hi" GI symptoms group, whereas eleven were categorized into the "lo" GI symptoms group by two independent physicians. All volunteers with severe adverse events were placed in the "hi" category, together with two volunteers who did not have severe adverse events but moderate adverse events of long duration (Table S1). Median number of related abdominal adverse events was 4 in the whole cohort (range 0-10), split per dosage group this was 4.5 in the 50L3 group, 4 in the 100L3 group and 3 in the 150L3 group. This difference was not statistically significant. Originally, twelve volunteers were classified in the "hi" category, however, due to severe abdominal adverse events three participants from the "hi" group were treated early and could not be included in the microbiota analysis.

On average 28,600 reads (range=6,524-49,476 reads, median 29,244 reads) were generated per volunteer sample (total n=100), resulting in a total of 1,258 unique OTUs

(after filtering on 0.005% abundance). Both positive controls were highly similar to theoretical expectations, with the DNA standard (n=2) being more similar to theoretical expectation than DNA extraction controls (n=3) based on Bray-Curtis distances (Figure S1A + B). Two out of three negative extraction controls did not contain any reads post-filtering and one negative control contained only five reads in total.

High individual-specific clustering despite *N. americanus* infection

To explore data and understand potential shifts in microbiota composition, we performed t-Distributed Stochastic Neighbor Embedding (t-SNE), using Bray-Curtis dissimilarity of all samples, which revealed individual-specific clustering (Figure 1A), but no clear clustering according to GI symptoms group (Figure 1B) or larval dosage group (Figure 1C). Two individuals clustered separately, one of which had taken a course of amoxicillin (volunteer 18), while the other was strictly vegetarian (volunteer 11). It needs to be taken into account that t-SNE preserves the local structure rather than the global structure of the data (like in PCA), so large distances in the 2D plot do not necessarily reflect large distances in the high-dimensional space. Other people taking antibiotics during the study course, all for reasons unrelated to the study, did not show large compositional changes (Figure 1 + Figure S2 + Table 1).

Table 1: Volunteer characteristics. Included information is larval dosage group, GI symptom group and whether individuals took antibiotics during the study.

VolunteerID	Dosage_Group	GI_symptoms	Gender	Antibiotic use
1	C	Lo	Male	Amoxicilin, three times/day 500mg, for five days, between trial week zero and four.
2	C	Hi	Female	
3	C	Lo	Male	
4	B	Hi	Female	
5	A	Lo	Female	
6	C	Hi	Female	
7	B	Hi	Female	
8	A	Lo	Female	
9	B	Hi	Male	Single cefazolin administration, between trial week fourteen and twenty.
10	B	Lo	Male	
11	C	Lo	Male	
12	A	Hi	Male	
13	C	Lo	Male	
14	A	Hi	Female	Single azithromycin (1000mg), between trial week four and eight.
15	B	Lo	Female	
16	A	Lo	Female	Amoxicilin, three times/day 500mg, for five days, between trial week zero and four.
17	B	Lo	Male	
18	B	Hi	Female	Amoxicilin, three times/day 500mg, for five days, between trial week zero and four.
19	C	Hi	Female	
20	A	Lo	Female	

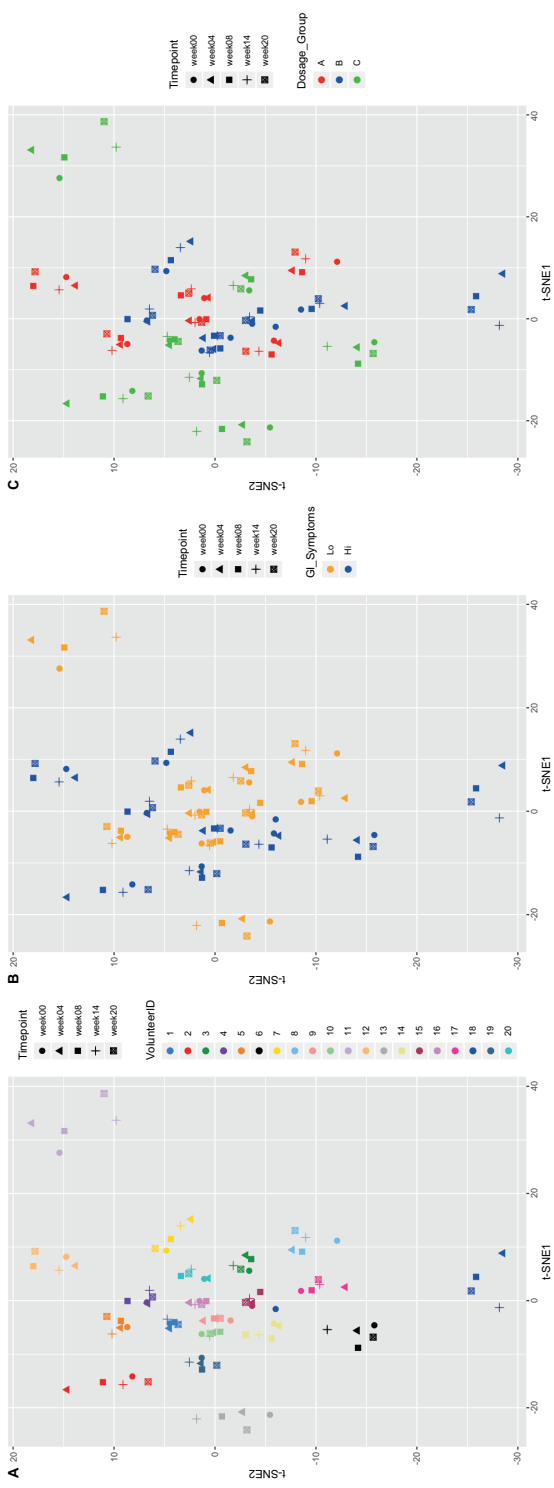


Figure 1: t-Distributed Stochastic Neighbor Embedding (t-SNE), using Bray-Curtis dissimilarity. Volunteer-specific clustering is observed, with no obvious shift according to timepoint. Volunteers (n=20) are colored by their volunteer number (A), the GI symptoms group (B) or the larval dosage group (C), while each shape corresponds to a timepoint.

Larval dosage does not differentially impact alpha diversity or stability in the acute phase of infection

To investigate whether larval dosages induce differential effects on the gut microbiota, we compared alpha diversity and stability measures between dosage groups. Group A (n=6 volunteers) received 50L3 larvae, group B (n=7 volunteers) 100L3 larvae and group C (n=7 volunteers) 150L3 larvae (Figure 2). First, we investigated potential changes in alpha diversity and stability during the acute phase of infection (which includes trial week zero, four and eight). To test for differences in these parameters, normality was tested using Shapiro-Wilk test and equal variance using an F-test. Subsequently, depending on outcome of these tests, appropriate tests were performed.

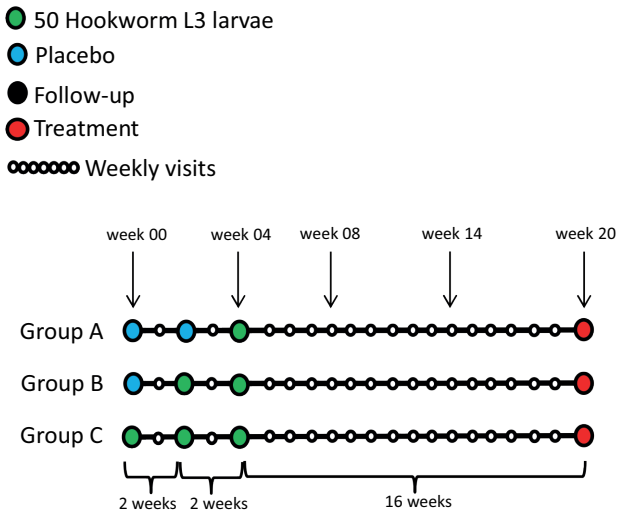


Figure 2: Study design. At indicated trial weeks (week zero, four, eight, fourteen and twenty) feces were collected for microbiota analysis.

We started to compare the effect of acute infection compared to an uninfected state. As at trial week four group A was not yet exposed, and group C twice (Figure 2), we compared their deltas at trial week four (Chao1/Shannon at week four minus Chao1/Shannon at week zero). Group B was not included in this analysis, since this group was infected at week two of the study and thus at the time samples were taken (week four), patent infection was not yet established in the gut. No differences in deltas were found at OTU level (independent t-test, p=.76) or genus level (Mann-Whitney test, p=.61) between A and C. No difference within group C between trial week zero and four at OTU level (paired t-test, p=.49) or genus level (paired t-test, p=.41) was observed either (Figure 3A

+ B). The same tests were performed for Shannon diversity, stability measures (1-Bray-Curtis and 1-Jaccard, Welch t-test p=.742 and independent t-test p=.219 respectively), but no differences were observed (Figure 3C-F).

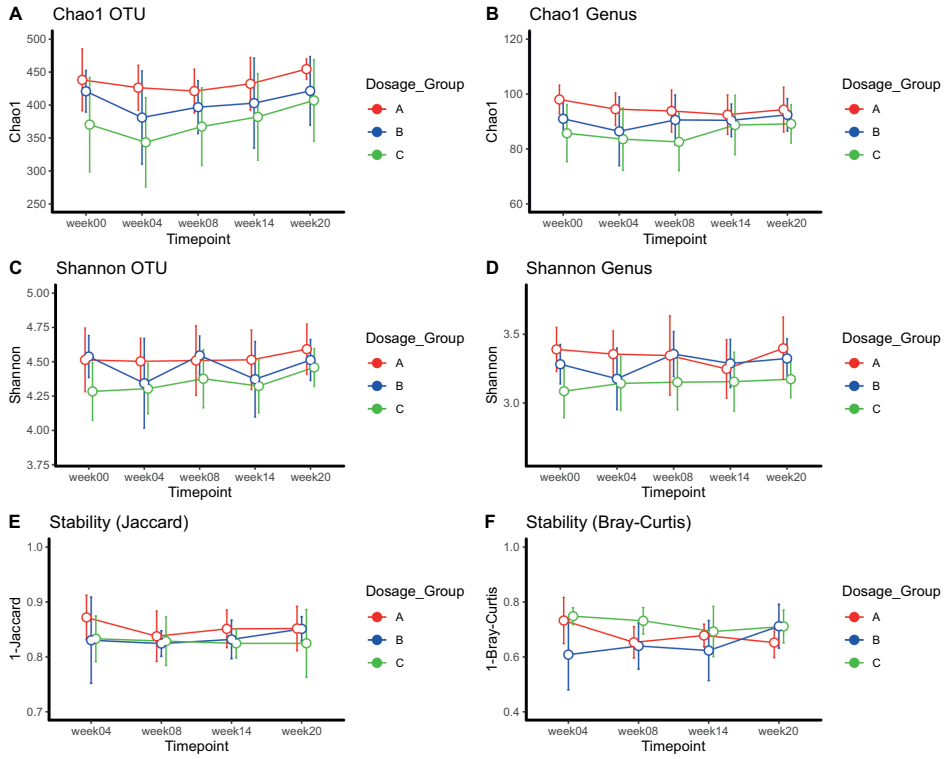


Figure 3: Richness (Chao1) and diversity plots at OTU and genus level (A-D) and stability measures (Jaccard and Bray-Curtis) for larval dosage groups (E-F). Total infectious dosages for group A (red): 50L3 larvae, group B (blue): 100L3 larvae and C (green): 150L3 larvae. Means and the 95% CI of the standard error of the mean are displayed. *p<.05, **p<.01, ***p<.001

At trial week eight all volunteers likely had established intestinal hookworm infection. Therefore, trial week zero and eight were compared to all individuals. No differences were observed in Chao1 at OTU level (Wilcoxon signed rank test, p=.391) nor at genus level (Wilcoxon signed rank test, p=.152) (Figure 3A + B). No differences were observed in Shannon diversity either (Figure 3C + D).

In conclusion, we did not observe any changes in diversity or stability of the microbiota during the acute phase of infection between or within dosage groups.

Microbiota richness increases in all volunteers during the established infection phase

Subsequently, we investigated the effect of established infection (trial week eight to twenty) on the gut microbiota using a linear mixed model (LMM). Chao1 at OTU level increased from trial week eight to twenty (p=.0174), and less clearly so at genus level (p=.0905) over all volunteers, but no differential effect between larval dosage groups was observed (Figure 3). No differences in Shannon diversity or stability were seen between or within larval dosage groups or over time across all individuals. In conclusion, we found an increased richness over all volunteers during established infection, but Shannon diversity and stability remained unchanged. It is however unclear whether this increased richness is a direct result of the infection, as no non-infected group was available at this time point.

Individual bacterial taxa do not display major differential changes between larval dosage groups

Lastly, we performed differential abundance analysis between larval dosage groups A and C over time using the MetaLonDA package and investigated an overall hookworm effect over all volunteers using DESeq2. Group B was not included in this analysis, as data was only available at timepoint two weeks after infection. At this timepoint no patent infection is established in the gut, but systemic effects or effects of early larval migration cannot be excluded. In addition, the antibiotic-induced effect on the gut microbiota of volunteer 18 (who was in group B) could affect the analysis, especially considering the small number of volunteers in each dosage group. MetaLonDA at genus revealed that *Dorea* was significantly increased between trial week four and week eight in group A (p=0.04) (Figure S3A). However, as this is the only differentially abundant taxa at a single time interval, this is unlikely to represent biological relevance. This analysis was also performed at OTU level (Table S2 and Figure S4A). No differences in relative abundance were observed across all volunteers from trial week zero to twenty either at both genus and OTU level (adj. p-value > 0.05). We subsequently continued analyzing the relationship between the gut microbiota and severity of GI symptoms.

Hi GI symptoms were associated with an unstable microbiota at trial week eight

Our next goal was to investigate whether baseline differences in microbiota composition could be associated with severity of GI symptoms, so we compared the “lo” (n=11 volunteers) and “hi” (n=9 volunteers) GI symptoms groups. No difference in Chao1 was observed at OTU level (independent t-test, p=0.244) or genus level (Mann-Whitney test, p=0.446) (Figure 4A + B) at week zero. Comparing week zero with week eight did not show differences at OTU level (Wilcoxon signed rank test for the “lo” group, p=0.391 and paired t-test for “hi” group, p=0.382, Figure 4A) or genus level (Wilcoxon-signed rank test for “lo”, p=0.152 and paired t-test for “hi”, p=0.132) (Figure 4B). No

differences were seen at trial week eight between symptom groups at OTU level (Mann-Whitney test, p=0.412) or genus level (independent t-test, p=0.674) (Figure 4A + B). The same tests were performed for Shannon diversity, but no differences were observed either (Figure 4C + D).

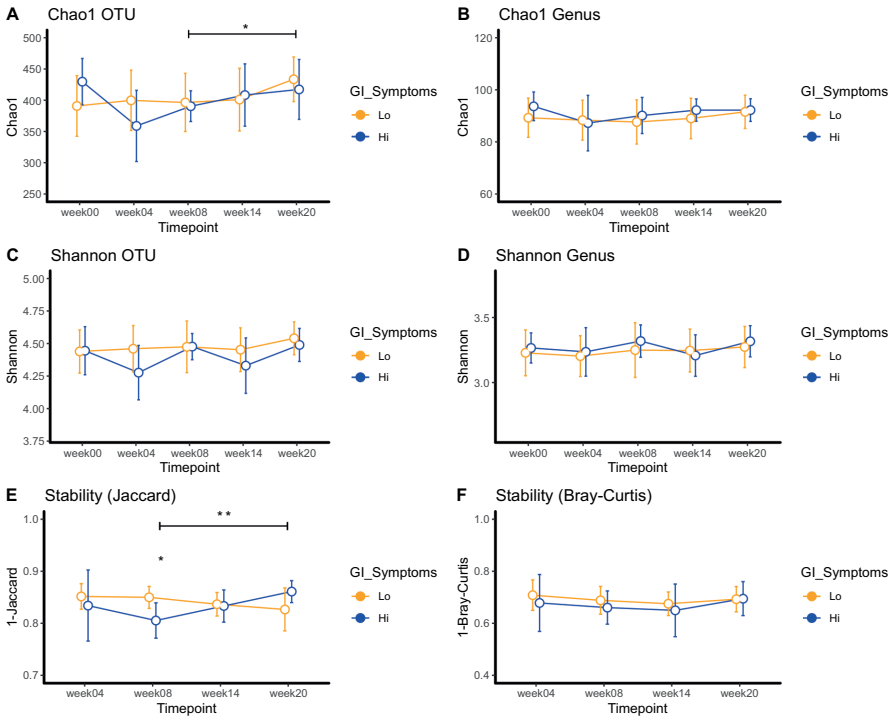


Figure 4: Richness (Chao1) and diversity plots at OTU and genus level (A-D) and stability measures (Jaccard and Bray-Curtis) for GI symptoms groups (E-F). For Figure A, significance between trial week eight and week twenty is for the "lo" GI symptoms group (orange). For Figure E, significance between trial week eight and week twenty is for the "hi" GI symptoms group (blue). Means and the 95% CI of the standard error of the mean are displayed. *p<.05, **p<.01, ***p<.001.

Microbiota stability of the "hi" GI symptoms group was significantly decreased at trial week eight as compared to the "lo" GI symptoms group (Jaccard, independent t-test, p=.036) (Figure 4E). No difference was found using Bray-Curtis dissimilarity (Figure 4F).

In conclusion, we did not find any changes in alpha diversity between GI symptoms groups, but microbiota stability (here Jaccard) was significantly reduced in the "hi" GI symptoms group at trial week eight compared to the "lo" GI symptoms group.

Microbiota stability recovers over time

LMM was applied to investigate changes in diversity during the established phase in the GI symptoms groups. We found a significant increase in Chao1 from trial week eight to twenty in the "lo" GI complaints group at OTU level (LMM, p=.045) (Figure 4A), but not at genus level (LMM, p=.120) (Figure 4B). No differences were found for Shannon diversity.

As previously mentioned, stability in the "hi" GI symptom group was reduced at trial week eight. This instability quickly recovered from trial week eight to twenty (Jaccard, paired t-test, p=.004) (Figure 4E). In addition, the slopes between the symptom groups were significantly different from trial week eight to twenty, confirming the recovery within the "hi" symptoms group (LMM interaction, p=.002). This also means that there was increased dissimilarity in the "lo" symptoms group in this time period. We further hypothesized that this stability may perhaps be related to eosinophil count, but we did not find a significant correlation between eosinophil count and microbiota stability when stratifying by trial week (Figure S5A). No differences in Bray-Curtis dissimilarity were observed during established infection, nor was a correlation found with eosinophil counts at any trial week (Figure S5B). Eosinophil counts are visualized per time point in Figure S6.

In summary, the "hi" GI symptom group was characterized by transient microbiota instability and subsequent recovery.

Specific bacterial taxa differ between symptoms groups during the entire study course

To investigate whether changes in individual bacterial taxa over the entire study course could be linked to symptom groups ('hi' n=9, 'lo' n=11), we employed the MetaLonDa package (Figure 5, Figure S3B-F and Table S2) and found several bacterial taxa in the "hi" GI symptoms group significantly increased at genus level. *Barnesiella* was found to be significantly increased in this group at all intervals from trial week zero to week twenty (p<.05), *Lachnospiraceae_ND_3007* was significantly higher from trial week zero to fourteen (p<.05), *Bilophila* was more abundant between trial week zero and four, and between trial week eight and twenty (p<.05) and *Escherichia-Shigella* was significantly more abundant between trial week zero and four (p<.05). In the "lo" GI symptoms group, *Allisonella* was more abundant between trial week fourteen and twenty, at which time a chronic infection had been established (p<.05). Lastly, relative abundance over time of these significantly different genera was visualized, to investigate whether significance was driven at the group level or by a single individual (Figure S3B-F). This showed that the difference in *Escherichia-Shigella* was driven by a single person, namely volunteer 18, while all other differences were largely group-driven.

When analysis was repeated without volunteer 18, *Escherichia-Shigella* was indeed non-significant (p=.292). In addition, the association with *Barnesiella* and *Bilophila* persisted throughout the study (p=.074 and p=.072), *Allisonella* became more abundant in the "lo" group throughout the entire study (p=.017) and *Oscillibacter* was more abundant in the "lo" group (p=.012) from week zero to fourteen (Table S2). All analyses were also performed at OTU level (Figure S4B+C and Table S2). These results confirm that differences in relative abundance of taxa between the symptom groups were largely group-driven, apart from *Escherichia-Shigella*.

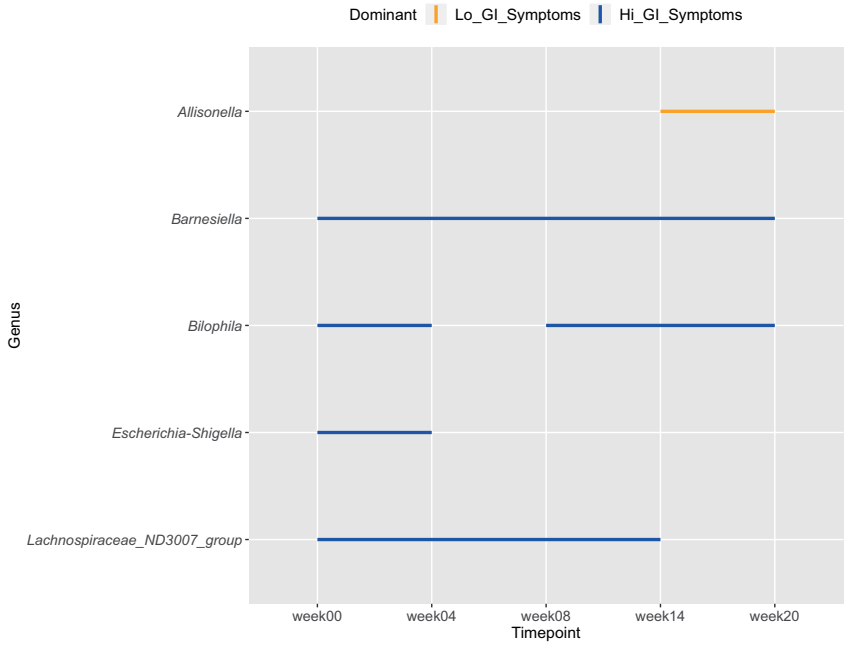


Figure 5: Time intervals of significantly different bacterial genera between GI symptoms groups. Each line interval represents a significant time interval, with significance being considered p<.05. Orange lines indicate higher abundance in the "lo" GI symptoms group, while blue indicates higher abundance in the "hi" GI symptoms group.

Discussion

Herewith we present the first longitudinal assessment of microbiota changes over the course of an experimental *N. americanus* infection in healthy individuals. Although no convincing relationship between microbiota and larval dosage was observed, stability of the bacterial microbiota was linked to severity of clinical symptoms. In addition, we found several statistically significant changes in relative abundance of individual

bacterial taxa over time between symptom groups.

We found a very strong volunteer-specific clustering, despite a patent hookworm infection. This corroborates previous findings showing that the gut microbiota is stable over time in healthy adults at a compositional level[20-22] and the previous assessment of experimental hookworm infections in patients with celiac disease where minor changes were detected over time.[16]

We detected an interesting link between microbiota stability from trial week eight to twenty and clinical GI symptoms. Recovery of stability in the "hi" symptoms groups leads us to believe that either volunteers with a more unstable microbiota in early weeks post-infection are more likely to experience GI symptoms during the infection, or the GI symptoms are caused by a more severe enteritis that also affects microbiota stability. The latter hypothesis seems most likely whereby symptoms are caused by an eosinophilic enteritis, with eosinophils having been described to correlate with severity of enteritis, [19, 23] and the enteritis may in turn affect the gut microbiota. Although cause and effect cannot be determined, it does suggest an important bacterial-helminth-host interplay, which deserves further investigation.

We observed increased richness over all volunteers during the established infection phase. This is in line with previous studies which have mostly shown that individuals with a parasitic infection show either equal or increased microbial richness and diversity.[11, 15, 24] However, we cannot be fully certain this is an infection-induced effect, as no "no-infection" control group was included in our study. It is unclear why alpha diversity may increase during hookworm infection, although several hypotheses can be formulated. Potentially, the immune regulatory effects of helminths might facilitate increased bacterial microbiota richness and diversity.[25, 26] On the other hand, expansion of the current bacterial community could also be promoted. Another hypothesis is that *N. americanus* affects the gut metabolome. While the effect of *N. americanus* infection on the full gut metabolome has not yet been investigated in humans to our knowledge, short chain fatty acid (SCFA) levels were measured in eight volunteers undergoing *N. americanus* infection.[27] Out of these eight volunteers, six showed an increase in total fecal SCFA, while two showed a reduction.[27] Even though this suggests an effect of *N. americanus* on the gut metabolome, this should be confirmed with a larger sample size. We observed several changes in individual bacteria taxa between the GI symptoms groups, although the biological relevance of these changes remains unclear. The increased abundance of *Barnesiella* and the decreased abundance of *Allisonella*, a histidine-consuming and histamine-producing taxon, in the "hi" symptoms group are puzzling. While *Barnesiella* is associated with a healthy microbiota and beneficial intestinal effects,[28-31] *Allisonella* and its metabolic product histamine are associated with

increased GI symptoms.[32-34] This would counterintuitively suggest that individuals with a microbiota generally regarded as more beneficial respond more heavily to hookworm invasion. The opposite holds true however for *Bilophila*, a taxon which thrives under high-fat and animal-based diets.[35, 36] It is associated with increased inflammation, impaired intestinal barrier function and production of hydrogen sulfide.[35-37] Being more abundant in the "hi" symptom group, this contradicts with the hypothesis that a more beneficial microbiota responds more vigorously to hookworm invasion. All in all, the relevance of these findings should be tested in larger groups and with more functional techniques than 16S rRNA gene amplicon sequencing. In addition, answering this hypothesis would require a clear definition of a 'healthy or beneficial' microbiota, a phenomenon which is currently incompletely understood.

The current study had several strengths and limitations. During the study period, five volunteers were prescribed antibiotics. While a clear effect of antibiotic usage was only observed in volunteer 18, we cannot exclude that an effect occurred in other volunteers and might have confounded our results. However, given the small sample size we were unable to partial out any confounders (e.g. antibiotic use and diet) which is a limitation of our study. It should be noted that investigating fecal material is probably not reflective for local microbiota changes in the duodenum and ideally duodenal biopsies would be taken, as was done previously by Giacomin *et al.*[17] This would however pose a sharply increased burden to volunteers. By using healthy volunteers, the effects of *N. americanus* infection on the bacterial gut microbiota could be studied without many external confounders. In addition, the longitudinal study setup allowed us to investigate the dynamics of the bacterial gut ecosystem. Implementation of well-controlled longitudinal studies were also recently described to be crucial for advancing the microbiome field.[38, 39] Future studies should include functional approaches (e.g. coupling metagenomics with metabolomics) to obtain insight into potential changes in microbial metabolism which could be a result of *N. americanus* infection. By using positive and negative controls, we confirmed that we should investigate richness and diversity both at OTU and genus level, and that our DNA extractions were well performed, although minimal contamination may have occurred. The original clinical trial reported the highest egg counts described in CHHI experiments in literature yet, reaching egg counts similar to those seen in mild infection in endemic areas, allowing for better comparison with natural infection.[18] Although the groups in this controlled study were small, the combination of infecting healthy volunteers with the highest infectious as of yet and describing their bacterial gut microbiota longitudinally is unique.

In conclusion, this is the first study to investigate longitudinal changes in gut microbiota during *N. americanus* infection in healthy individuals. We observed high stability of the gut microbiota despite this infection over the twenty-week study period, although

transient instability was observed in individuals with "hi" GI symptoms. These data open new avenues for exploring helminth-bacterial interaction in the human intestine.

Materials and methods

Twenty-four healthy male and female volunteers aged 18-45 years were included in a randomized controlled clinical trial investigating the effect of repeated infectious dosages on hookworm egg excretion and variability as previously described.[19] Volunteers were dermally exposed with two week intervals to either one, two or three dosages of 50 infectious larvae, resulting in cumulative dosages of 50, 100 and 150 larvae respectively. Study set-up was such that the 150 larvae group (C) received first infection at trial week zero, the 100 larvae group (B) at trial week two and the 50 larvae group (A) at trial week four. Volunteers were allocated equally to one of three groups at random according to an independently prepared randomization list. Group allocation was defined by the randomization number which was linked to the volunteer identification code at the first infection. Investigators and participants were blinded to group allocation. A schematic overview of study setup can be found in Figure 2.

Culture of larvae and procedure of infection was performed according to a previously described method.[40] In short, infective L3 larvae were cultured from feces from a chronic donor, were suspended in water and applied on upper arms and calves using gauzes. Volunteers were followed for twenty weeks after first exposure, after which treatment with albendazole was given to eradicate the infection. Volunteers visited the trial center weekly for collection of adverse events, safety laboratory evaluation and collection of fecal samples for egg count. Adverse events were collected at weekly visits. For every adverse event, time and date of onset, end, severity and causality was recorded. Adverse events were characterised using ICD-10 as unrelated, unlikely, possibly, probably or definitely related to hookworm infection, and mild (no interference with daily life), moderate (discomfort interfering with daily life) or severe (causing inability to perform usual daily activity). Adverse events were then assessed by two independent physicians who divided the participants in two groups with 'hi' and 'lo' adverse events. Originally, twelve volunteers were classified in the "hi" category. All volunteers with severe adverse events were placed in the "hi" category, together with two volunteers who did not have severe adverse events but moderate adverse events of long duration (Table S1). Consensus was reached for every participant. Unfortunately three participants in the "hi" group withdrew early from the trial due to severe abdominal adverse events and insufficient follow-up fecal samples were collected to include these participants in the microbiome analysis. For analysis, adverse events scored as possibly, probably and definitively related were considered related and were included in the symptom grouping.

Duration of adverse events was recorded. Detailed information on adverse events for each volunteer can be found in Table S1. Samples for analysis of fecal microbiota were collected at baseline and trial weeks four, eight, fourteen and twenty of the trial. For analysis of the relation between clinical symptoms and fecal microbiota, gastrointestinal symptoms were categorized as either "hi" or "lo". [19]Fecal egg counts were measured by microscopy using Kato-Katz. Eosinophils were measured weekly and egg counts were measured weekly from trial week five using Kato Katz microscopy. The trial was approved by the LUMC IRB (P17.224) and was registered at clinicaltrials.gov under NCT03257072.

Microbiota analysis

Fecal samples were aliquoted and immediately stored at -80°C. DNA was extracted from 0.1 gram feces using the Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (ZymoResearch, CA, USA) according to manufacturer instructions with minor adaptations, as described previously.[41] Quality control, library preparation and sequencing were performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEXTflex™ 16S V4 Amplicon-Seq Kit (BiooScientific, TX, USA) and the Illumina NovaSeq6000 platform (paired-end, 150bp). Raw read processing was performed using the NG-Tax 0.4 pipeline with following settings: forward read length of 120, reverse read length of 120, ratio OTU abundance of 2.0, classify ratio of 0.9, minimum threshold of $1*10^{-7}$, identity level of 100% and error correction of 98.5, using the Silva_132_SSU Ref database.[41-43] The obtained OTU table was filtered for OTUs with less than 0.005% relative abundance.[44] As quality controls for both DNA extraction and sequencing, we included ZymoBiomics Microbial Community Standard, (Zymo Research, Irvine, California, USA) ZymoBiomics Microbial Community DNA Standard (Zymo Research) and three negative DNA extraction controls. Raw sequencing data is available at ENA (https://www.ebi.ac.uk/ena) under accession number PRJEB36316. All analytical R code will be uploaded to GitHub upon acceptance of this manuscript.

Statistical analysis

All analyses were performed in R (v3.6.1) using the packages phyloseq (v1.28.0), microbiome (v1.6.0), Metalonda (v1.1.5), DESeq2 (v1.24.0), lme4 (1.1-21), lmerTest (v3.1-0).[45-50] Richness and diversity were computed at OTU and genus level, as richness was found to be overestimated based at OTU level in the positive controls. Genus level was obtained by agglomerating OTUs at genus level. Bray-Curtis and Jaccard indices were computed at genus level. Bray-Curtis and Jaccard indices were computed intra-individually, using trial week zero as the baseline measurement. As Bray-Curtis and Jaccard indices are dissimilarity indices, we computed 1- respective index to obtain similarity, where a value of 1 represents 100% similarity. For alpha diversity and stability measures, data was split into an "acute infection phase" (trial week zero to

eight) when most symptoms occurred and an "established infection phase"(trial week eight to twenty) when symptoms subsided. To test for differences in these parameters, normality was tested using Shapiro-Wilk test and variance was tested using an F-test. Subsequently, depending on outcome of the normality and variance test, independent t-tests, Welch t-tests, paired t-tests, Mann-Whitney U tests and Wilcoxon signed-rank tests were performed.. Clustering using t-Distributed Stochastic Neighbor Embedding (t-SNE) method was performed using the tsne_phyloseq function with default parameters.[51] t-SNE aims to preserve the local structure of the original high-dimensional space while projecting the data points in a low dimensional (2D) space. All Figure were created in R and only minimally formatted in Adobe Illustrator when necessary.

Correlation analysis

We used Spearman's Rank correlation to examine the relationship between eosinophil count and microbiota stability. Microbiota stability was defined in the same manner as previously, with Bray-Curtis and Jaccard indices computed intra-individually, using trial week zero as the baseline measurement. As both indices are dissimilarity indices, 1- respective index was computed to obtain similarity. Eosinophil count was measured weekly, and therefore each individual at each time point had a measured eosinophil count. Timepoints were stratified to account for the repeated measurements design. In order to avoid skewing of the correlation by baseline data, at which point eosinophils were low and microbiota was 100% similar due to baseline to baseline comparison, this timepoint was excluded.

Linear Mixed Models

We performed linear mixed modelling (LMM) using the lmer function from the lme4 package[49] for alpha diversity and both stability indices from trial week eight until week twenty, as all groups had established infection in the gut from trial week eight onwards. Volunteer ID was included as a random intercept to control for inter-individual baseline differences and repeated measurements design. Included fixed effects were dosage group/symptom group and timepoints. In case an interaction effect was suspected by visually inspecting plots, an additional interaction model was also performed with dosage group/symptom group*timepoint. Models were checked by inspecting whether residuals were normally distributed using qq-plots. P-values were obtained using the lmerTest package and considered significant when < 0.05.[50]

Time series modelling of individual taxa

Differential abundance testing was performed at genus and OTU level. The metagenomic longitudinal differential abundance method (MetaLonDA) package was used to identify differentially abundant taxa between groups over time.[47] It is a flexible method capable of handling inconsistencies often observed in human microbiome studies and relies

on two main modelling components, the negative binomial distribution for modelling read counts and smoothing spline ANOVA for modelling longitudinal profiles. The function metalondaAll was used with the following settings: n.perm=1000, fit.method="nbinomial", num.intervals=4, pvalue.treshold=0.05, adjust.method="BH", norm.method="median_ratio". These settings indicate that the function was run with 1000 permutations using the median ratio method to normalize count data and fitting a negative binomial distribution. Four intervals were included (between each included trial week) and p-values were corrected using the Benjamini-Hochberg procedure. DESeq2 was used to establish an overall time effect across all volunteers using the likelihood-ratio-test (full model included volunteer ID and timepoint, reduced model included volunteer ID) and for identifying differentially abundant taxa in pair-wise comparisons.[48] Prior to the use of both MetaLonDA and DESeq2, genera and OTUs were filtered for presence in at least 25% of all samples. Relevant tests performed are indicated in all Figures and in the text.

Notes

Acknowledgements

We would like to thank all working members from the Center for Microbiome Analyses and Therapeutics (CMAT) for their helpful input during work discussions and Ahmed Metwally for his help in implementing MetaLonDA. Lastly, we thank FEMS for awarding a Meeting Attendance Grant to QD to present results of this manuscript at Cell Symposia: Infection Biology in the Age of the Microbiome.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. EK is supported by an unrestricted grant from Vedanta Biosciences Inc. The original clinical trial was funded by a grant from Dioraphte Foundation. The funder had no influence on study design, data interpretation or preparation of the manuscript.

Author contributions

QD and MH wrote drafts of the manuscript. QD performed microbiota data analysis and created Figures. MH, JJ and JK performed the clinical study. AG performed DNA extraction and lab-related procedures. RZ processed raw sequence data. JG provided statistical advice. MR and EK designed the study. MR was the study principal investigator. All authors interpreted data, reviewed manuscript versions and approved the final manuscript.

References

1. Yazdanbakhsh M, Kremsner PG, van Ree R. Allergy, parasites, and the hygiene hypothesis. Science. 2002;296(5567):490-4.
2. Croese J, O'Neil J, Masson J, Cooke S, Melrose W, Pritchard D, et al. A proof of concept study establishing Necator americanus in Crohn's patients and reservoir donors. Gut. 2006;55(1):136-7.
3. Croese J, Giacomin P, Navarro S, Clouston A, McCann L, Dougall A, et al. Experimental hookworm infection and gluten microchallenge promote tolerance in celiac disease. J Allergy Clin Immunol. 2015;135(2):508-16.
4. McSorley HJ, Gaze S, Daveson J, Jones D, Anderson RP, Clouston A, et al. Suppression of inflammatory immune responses in celiac disease by experimental hookworm infection. PLoS One. 2011;6(9):e24092.
5. Collaborators GMaCoD. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1459-544.
6. Rapin A, Harris NL. Helminth–Bacterial Interactions: Cause and Consequence. Trends Immunol. 2018;39(9):724-33.
7. Rehman ZU, Deng Q, Umair S, Savoian MS, Knight JS, Pernthaner A, et al. Excretory/secretory products of adult Haemonchus contortus and Teladorsagia circumcincta which increase the permeability of Caco-2 cell monolayers are neutralised by antibodies from immune hosts. Vet Parasitol. 2016;221:104-10.
8. Hiemstra IH, Klaver EJ, Vrijland K, Kringel H, Andreasen A, Bouma G, et al. Excreted/secreted Trichuris suis products reduce barrier function and suppress inflammatory cytokine production of intestinal epithelial cells. Mol Immunol. 2014;60(1):1-7.
9. Midha A, Janek K, Niewienda A, Henklein P, Guenther S, Serra DO, et al. The Intestinal Roundworm Ascaris suum Releases Antimicrobial Factors Which Interfere With Bacterial Growth and Biofilm Formation. Front Cell Infect Microbiol. 2018;8:271.
10. Abner SR, Parthasarathy G, Hill DE, Mansfield LS. Trichuris suis: detection of antibacterial activity in excretory-secretory products from adults. Exp Parasitol. 2001;99(1):26-36.
11. Easton AV, Quiñones M, Vujkovic-Cvijin I, Oliveira RG, Kepha S, Odiere MR, et al. The Impact of Anthelmintic Treatment on Human Gut Microbiota Based on Cross-Sectional and Pre- and Postdeworming Comparisons in Western Kenya. mBio. 2019;10(2).
12. Rosa BA, Supali T, Gankpala L, Djuardi Y, Sartono E, Zhou Y, et al. Differential human gut microbiome assemblages during soil-transmitted helminth infections in Indonesia and Liberia. Microbiome. 2018;6(1):33.
13. Cortes A, Peachey LE, Jenkins TP, Scotti R, Cantacessi C. Helminths and microbes within the vertebrate gut - not all studies are created equal. Parasitology. 2019;146(11):1371-8.
14. Loukas A, Hotez PJ, Diemert D, Yazdanbakhsh M, McCarthy JS, Correa-Oliveira R, et al. Hookworm infection. Nat Rev Dis Primers. 2016;2(1).
15. Cantacessi C, Giacomin P, Croese J, Zakrzewski M, Sotillo J, McCann L, et al. Impact of experimental hookworm infection on the human gut microbiota. J Infect Dis. 2014;210(9):1431-4.
16. Giacomin P, Zakrzewski M, Croese J, Su X, Sotillo J, McCann L, et al. Experimental hookworm infection and escalating gluten challenges are associated with increased microbial richness in celiac subjects. Sci Rep. 2015;5(1).
17. Giacomin P, Zakrzewski M, Jenkins TP, Su X, Al-Hallaf R, Croese J, et al. Changes in duodenal tissue-associated microbiota following hookworm infection and consecutive gluten challenges in humans with coeliac disease. Sci Rep. 2016;6:36797.
18. WHO Expert Committee. Prevention and control of schistosomiasis and soil-transmitted helminthiasis. World Health Organ Tech Rep Ser. 2002;912:i-vi, 1-57, back cover.
19. Hoogerwerf MA, Koopman JPR, Janse JJ, Langenberg MCC, van Schuijlenburg R, Kruize YCM, et al. A randomized controlled trial to investigate safety and variability of egg excretion after repeated controlled human hookworm infection. J Infect Dis. 2020.
20. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et al. The long-term stability of the human gut microbiota. Science. 2013;341(6141):1237439.
21. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol. 2018;3(3):356-66.
22. Costea PI, Coelho LP, Sunagawa S, Munch R, Huerta-Cepas J, Forslund K, et al. Subspecies in the global human gut microbiome. Mol Syst Biol. 2017;13(12):960.

23. Croese J, Wood MJ, Melrose W, Speare R. Allergy controls the population density of Necator americanus in the small intestine. Gastroenterology. 2006;131(2):402-9.
24. Lee SC, Tang MS, Lim YA, Choy SH, Kurtz ZD, Cox LM, et al. Helminth colonization is associated with increased diversity of the gut microbiota. PLoS Negl Trop Dis. 2014;8(5):e2880.
25. Everts B, Smits HH, Hokke CH, Yazdanbakhsh M. Helminths and dendritic cells: sensing and regulating via pattern recognition receptors, Th2 and Treg responses. Eur J Immunol. 2010;40(6):1525-37.
26. Gaze S, McSorley HJ, Daveson J, Jones D, Bethony JM, Oliveira LM, et al. Characterising the mucosal and systemic immune responses to experimental human hookworm infection. PLoS Pathog. 2012;8(2):e1002520.
27. Zaiss MM, Rapin A, Lebon L, Dubey LK, Mosconi I, Sarter K, et al. The Intestinal Microbiota Contributes to the Ability of Helminths to Modulate Allergic Inflammation. Immunity. 2015;43(5):998-1010.
28. Mancabelli L, Milani C, Lugli GA, Turroni F, Cocconi D, van Sinderen D, et al. Identification of universal gut microbial biomarkers of common human intestinal diseases by meta-analysis. FEMS Microbiol Ecol. 2017;93(12).
29. Ubeda C, Bucci V, Caballero S, Djukovic A, Toussaint NC, Equinda M, et al. Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. Infect Immun. 2013;81(3):965-73.
30. Steinway SN, Biggs MB, Loughran TP, Jr., Papin JA, Albert R. Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. PLoS Comput Biol. 2015;11(5):e1004338.
31. Weiss GA, Chassard C, Hennet T. Selective proliferation of intestinal Barnesiella under fucosyllactose supplementation in mice. Br J Nutr. 2014;111(9):1602-10.
32. Garner MR, Flint JF, Russell JB. Allisonella histaminiformans gen. nov., sp. nov. A novel bacterium that produces histamine, utilizes histidine as its sole energy source, and could play a role in bovine and equine laminitis. Syst Appl Microbiol. 2002;25(4):498-506.
33. Bohn L, Storsrud S, Tornblom H, Bengtsson U, Simren M. Self-reported food-related gastrointestinal symptoms in IBS are common and associated with more severe symptoms and reduced quality of life. Am J Gastroenterol. 2013;108(5):634-41.
34. Wouters MM, Balemans D, Van Wanrooy S, Dooley J, Cibert-Goton V, Alpizar YA, et al. Histamine Receptor H1-Mediated Sensitization of TRPV1 Mediates Visceral Hypersensitivity and Symptoms in Patients With Irritable Bowel Syndrome. Gastroenterology. 2016;150(4):875-87.e9.
35. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505(7484):559-63.
36. Natividad JM, Lamas B, Pham HP, Michel ML, Rainteau D, Bridonneau C, et al. Bilophila wadsworthia aggravates high fat diet induced metabolic dysfunctions in mice. Nat Commun. 2018;9(1):2802.
37. Peck SC, Denger K, Burrichter A, Irwin SM, Balskus EP, Schleheck D. A glycyl radical enzyme enables hydrogen sulfide production by the human intestinal bacterium Bilophila wadsworthia. Proc Natl Acad Sci U S A. 2019;116(8):3171-6.
38. Schmidt TSB, Raes J, Bork P. The Human Gut Microbiome: From Association to Modulation. Cell. 2018;172(6):1198-215.
39. Spencer SP, Fragiadakis GK, Sonnenburg JL. Pursuing Human-Relevant Gut Microbiota-Immune Interactions. Immunity. 2019;51(2):225-39.
40. Hoogerwerf MA, Coffeng LE, Brienen EAT, Janse JJ, Langenberg MCC, Kruize YCM, et al. New Insights Into the Kinetics and Variability of Egg Excretion in Controlled Human Hookworm Infections. J Infect Dis. 2019;220(6):1044-8.
41. 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):e00547-19.
42. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ, et al. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes[version 1; referees: 2 approved with reservations, 1 not approved]. F1000Research. 2016;5.
43. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
44. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10(1):57-9.
45. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.
46. Lahti L, Shetty S. Tools for microbiome analysis in R. Microbiome package version 1.6.0. 2017.

47. Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome. 2018;6(1):32.
48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
49. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67:1-48.
50. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw. 2017.
51. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9:2579-605.

Chapter 8

Microbiota-associated risk factors for asymptomatic gut colonisation with multi-drug-resistant organisms in a Dutch nursing home

Genome Medicine, 2021

Quinten R. Ducarmon[a,b,*], Elisabeth M. Terveer[a,b,*], Sam Nooij[a,b], Michelle N. Bloem[a,b], Karuna E.W. Vendrik[a,c], Monique A.A. Caljouw[d], Ingrid M.J.G. Sanders[a], Sofie M. van Dorp[a,e], Man C. Wong[a], Romy D. Zwittink[a,b,*], Ed J. Kuijper[a,b,c,*]

*these authors contributed equally
[a]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands
[b]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands
[c]Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[d]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands
[e]Department of Internal Medicine and Geriatrics, Onze Lieve Vrouwe Gasthuis (OLVG Hospital), Amsterdam, The Netherlands

Genome Medicine. 2021 Apr 7;13(1):54.
Supplementary information available online.

Abstract

Background

Nursing homes residents have increased rates of intestinal colonisation with multidrug-resistant organisms (MDROs). We assessed the colonisation and spread of MDROs among this population, determined clinical risk factors for MDRO colonisation and investigated the role of the gut microbiota in providing colonisation resistance against MDROs.

Methods

We conducted a prospective cohort study in a Dutch nursing home. Demographical, epidemiological and clinical data were collected at four time points with two-month intervals (October 2016 - April 2017). To obtain longitudinal data, residents (n=27) were selected if they provided faeces at two or more time points. Ultimately, twenty-seven residents were included in the study and 93 faecal samples were analysed, of which 27 (29.0%) were MDRO-positive. Twelve residents (44.4%) were colonised with an MDRO at at least one time point throughout the six-month study.

Results

Univariable generalised estimating equation logistic regression indicated that antibiotic use in the previous two months and hospital admittance in the previous year were associated with MDRO colonisation. Characterisation of MDRO isolates through whole genome sequencing revealed *Escherichia coli* sequence type (ST)131 to be the most prevalent MDRO and ward-specific clusters of *E. coli* ST131 were identified. Microbiota analysis by 16S rRNA gene amplicon sequencing revealed no differences in alpha or beta-diversity between MDRO-positive and negative samples, nor between residents who were ever or never colonised. Three bacterial taxa (*Dorea, Atopobiaceae* and *Lachnospiraceae* ND3007 group) were more abundant in residents never colonised with an MDRO throughout the six-month study. An unexpectedly high abundance of *Bifidobacterium* was observed in several residents. Further investigation of a subset of samples with metagenomics showed that various *Bifidobacterium* species were highly abundant, of which *B. longum* strains remained identical within residents over time, but were different between residents.

Conclusions

Our study provides new evidence for the role of the gut microbiota in colonisation resistance against MDROs in the elderly living in a nursing home setting. *Dorea, Atopobiaceae* and *Lachnospiraceae* ND3007 group may be associated with protection against MDRO colonisation. Furthermore, we report a uniquely high abundance of several *Bifidobacterium* species in multiple residents and excluded the possibility that this was due to probiotic supplementation.

Background

Infections caused by multidrug-resistant organisms (MDROs) are a rising threat to global health and caused ~33,000 attributable deaths in Europe in 2015 [1]. Infections with MDROs are usually preceded by asymptomatic gut colonisation, and asymptomatically colonised individuals represent a potential transmission reservoir [2]. Nursing home residents are at increased risk for MDRO colonisation due to comorbidities resulting in increased healthcare contact and antibiotic use [3]. In addition, MDRO spread within a nursing home can be facilitated due to communal living, confined living space and incontinence of residents [4, 5]. This is similar to the transmission dynamics of *Clostridioides difficile*. The prevalence of MDROs and *C. difficile* varies between nursing homes from different countries, but large differences in prevalence can also be observed between different institutions in one country. For example, MDRO prevalence ranges from 0 to 47% in various nursing homes in the Netherlands [6-8] and from 0 to 75% in Ireland [5]. *C. difficile* colonisation prevalence ranges from 0 to 17% in Dutch nursing homes [9, 10], and from 0 to 10% in Germany [11]. These differences may reflect variation in individual nursing home infection prevention and control practices, antimicrobial stewardship, infrastructure, care load and presence of MDRO risk factors such as incontinence, recent hospitalisation and current antibiotic use. Colonisation resistance provided by the gut microbiome could contribute to preventing MDRO colonisation in the gut. The gut microbiome can provide colonisation resistance through secretion of antimicrobial products, nutrient competition, support of epithelial barrier integrity, bacteriophage deployment, and immune activation. However, current knowledge on the link between the microbiome and MDRO colonisation is limited [12, 13]. In travellers, an increase of antimicrobial resistance genes and *Escherichia coli* relative abundance in the microbiome were observed after acquisition and asymptomatic carriage of Extended-spectrum beta-lactamase (ESBL)-producing *E. coli*, but without clear differences in microbial community structure [14]. An exception to the understudied role of the microbiome in MDRO colonisation is vancomycin-resistant *Enteroccocus* (VRE). For example, it has recently been demonstrated that a lantibiotic-producer, in this case *Blautia producta*, could restore colonisation resistance against VRE [15].

To determine the prevalence and spread of MDROs in a Dutch nursing home, and to elucidate the role of the gut microbiota and clinical risk factors herein, we conducted a four-point-prevalence study and analysed clinical data of residents and whole-genome sequencing (WGS) data of MDRO isolates, in combination with gut microbiota analysis through 16S rRNA gene amplicon sequencing. In addition, we conducted more in-depth microbiota analysis in a selection of samples through metagenomics in order to further investigate findings from 16S rRNA gene amplicon analysis.

Methods

Study design

We conducted a prospective cohort study in which residents of a nursing home in the Netherlands were invited to participate. The prevalence, dynamics and risk factors of MDRO colonisation were studied in a non-outbreak situation. Demographical, epidemiological and clinical data of four time points with a two-month interval (October 2016 until April 2017) were collected. Microbiota analysis was performed on stool samples collected at the same four time points. Written informed consent was obtained from the resident or corresponding proxy. Ethical approval was granted by the medical ethics committee of the Leiden University Medical Center (No.P16.039). Sixty-four of 131 residents (49%) consented to participate. Data and corresponding faeces was collected from 60 residents (94%). To make optimal use of the longitudinal data from this study, residents were selected that provided faeces at at least two time points (n=47). For this study, we included residents who gave consent for additional analyses, from whom faeces was cultured for MDROs at at least two time points, and of which sufficient material was left for microbiota profiling at at least two time points (n=27 residents). The prevalence of MDRO was not statistically significant between the residents selected for microbiota analysis (12/27 residents and 27/93 time points) and those not selected (10/30 residents and 12/61 timepoints) (Chi-squared test, p=0.26).

Data and faeces collection

The nursing home consisted of 131 beds divided over eight wards of various sizes (12-35 beds). The wards had single en-suite rooms, except for three double rooms for couples. All wards had a separate dining area where freshly prepared meals were served daily and residents did not receive a specific diet or probiotics. In addition, the nursing home had a large communal recreation and shared physiotherapy area. Nursing staff was dedicated to specific wards, but occasionally staff cross-covered wards. For each consenting resident, socio-demographic and the following MDRO risk factor data were collected at each of the four time point using standardised ECDC definitions: care load indicators (disorientation, mobility, incontinence), hospitalisation in the previous six months, antibiotics (concomitant and in the previous six months), comorbidities, presence of an indwelling urinary catheter or wounds, history of past MDRO colonisation .

In addition, instructed caring staff collected fresh faeces on the four time points and subsequently stored the samples at 4^{o}C. Samples were transported within 72 hours to the laboratory (Leiden University Medical Center).

MDRO detection

Faecal samples were examined for multi-drug resistant bacteria by culturing within 8

hours after arrival at the laboratory and the faeces and cultured MDROs were subsequently stored at -20°C [9]. Based on national recommendations [17], the following micro-organisms were considered to be an MDRO: ESBL-producing *Enterobacterales; Enterobacterales* and *Acinetobacter spp.* resistant to both fluoroquinolones and aminoglycosides or carbapenemase-producing; carbapenemase-producing *Pseudomonas aeruginosa; P. aeruginosa* resistant to at least three of the following antibiotic classes*:* fluoroquinolones, aminoglycosides, ceftazidime and/or piperacillin; trimethoprim/sulfamethoxazole-resistant *Stenotrophomonas maltophilia;* or vancomycin resistant enterococci (VRE). Faecal samples were enriched in 15ml of Tryptic Soy Broth (TSB) and incubated for 18 hours at 35°C prior to plating on ChromID ESBL, ChromID VRE and MacConkey tobramycin agars (BioMérieux, Marcy l'Etiole, France) for 48 hours at 35°C [9]. The twenty samples of the first time-point were re-cultured two years after sampling, as these samples were initially enriched with TSB containing 8mg/L vancomycin and 0.25 mg/L cefotaxime. The samples were stored in -20°C with glycerol. All morphological different aerobic Gram-negative bacteria and enterococci were identified by the BD Bruker matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) Biotyper (Microflex, Bruker Daltonics, Bremen, Germany). Phenotypic antibiotic susceptibility testing was performed with the VITEK2 system (card N199, BioMérieux) using the European Committee of Antimicrobial Susceptibility Testing (EUCAST) breakpoints [18]. ESBL production was confirmed by a double disk method [19]. In addition, the faecal samples were screened for the presence of carbapenemase-producing Gram negative bacteria [19]. The minimum inhibitory concentration (MIC) of *Enterobacterales* with a meropenem MIC > 0.25 mg/L was confirmed with an antibiotic gradient strip method (Etest, BioMérieux). Strains with an MIC > 0.25 mg/L were further investigated by an in-house multiplex PCR to detect the most frequently found carbapenemase genes (KPC, VIM, NDM, OXA-48 and IMP). Additionally, *Clostridioides difficile* was cultured and characterised as previously described [20].

Risk factor analysis

Data from 27 nursing home residents (93 samples in total) were included for risk factor analysis. All analyses compared all MDRO-positive samples with all MDRO-negative samples, as extensive metadata was collected at each time point for each individual resident. To account for the repeated measurements design, generalised estimating equations (GEE) logistic regressions (using the geeglm() function in the geepack package) were performed with Resident number as cluster [21]. To identify clinical factors associated with MDRO colonisation, univariable GEE logistic regression was performed using variables for which ten or more 'events' were recorded, as previously recommended for logistic regression [22]. Factors with a p-value < 0.05 were included in multivariable GEE logistic regression analysis, as well as non-significant factors that were considered likely to influence MDRO colonisation risk based on expert opinion and

literature review. These factors were sex and current use of a urinary catheter. Lastly, we inspected possible multicollinearity between the variables included in the multivariable GEE logistic regression by computing variance inflation factors. While opinions differ on when a variance inflation factor can be considered considerable, we used the stringent variance inflation factor value of 2.5 here, as previously recommended, to obtain insight in possible multicollinearity [23].

Whole-genome sequencing of bacterial isolates and data processing

WGS analysis to characterise MDRO isolates was done at GenomeScan B.V. (Leiden, the Netherlands). Genome sequences were determined using the Illumina HiSeq 4000 platform (Illumina, San Diego, CA, USA) from DNA prepared by the QIAsymphony DSP Virus/Pathogen Midi Kit (Qiagen, Hilden, Germany) at Leiden University Medical Center following manufacturer's recommendations. Sequence libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for 150 bp paired-end sequencing.

Sequencing quality was evaluated with FastQC (version 0.11.8) [24] and MultiQC (version 1.7) [25]. Reads were assembled using a hybrid assembly strategy, starting with SKESA (version 2.3.0) [26] using default parameters for paired-end reads, followed by SPAdes (version 3.13.1) [27] using default parameters while providing SKESA's contigs with the '--untrusted-contigs' parameter. Assembly quality and length were checked after each step using QUAST (version 5.0.2) [28]. The scaffolds produced by SPAdes were used for subsequent analyses.

To evaluate assembly quality, all scaffolds were blasted (megablast version 2.9.0, parameters '-evalue 1e-10' and '-num_alignments 50') [29, 30] against the NCBI BLAST nt database (from July 13 2017) and taxonomically classified using the Lowest Common Ancestor algorithm implemented in Krona ktClassifyBLAST (version 2.7.1) [31]. Scaffolds classified as eukaryote were removed from further analysis. The remaining non-eukaryotic scaffolds were screened for the presence of antibiotic resistance genes using staramr (version 0.5.1, https://github.com/phac-nml/staramr) and ABRicate (version 0.8.13, https://github.com/tseemann/abricate) against the ResFinder database (from May 21 2019) [32]. The same scaffolds were also subjected to *in silico* multi-locus sequence typing (MLST) and core-genome MLST using SeqSphere (version 6.0.2, Ridom GmbH, Münster, Germany) [33] to determine Warwick sequence types (ST) and pairwise allele distances using the built-in *E. coli* scheme. Next, a pangenome analysis was conducted on the scaffolds using Roary (version 3.12.0) (34), for which the scaffolds were annotated using Prokka (version 1.13.4) [35]. Finally, a maximum-likelihood phylogenetic analysis was generated with IQTree (version 1.6.10, parameters '-b 500' and '-m MFP' for 500 bootstrap replicates and automatic model selection) [36] on the multiple sequence alignment of the core genomes generated by Roary. The selected phylogenetic model based on the

best Bayesian Information Criterion score was GTR+F+R2.
All tools were run with default parameters unless stated otherwise.

DNA extraction for gut microbiota analyses
DNA was extracted from 0.1 gram faeces (n = 93 samples) using the Quick-DNA™ Fecal/Soil Microbe Miniprep Kit (ZymoResearch, CA, USA) according to manufacturer instructions with minor adaptations, as described previously [37]. Beads were a mix of 0.1 and 0.5 mm size, and bead-beating was performed using a Precellys 24 tissue homogeniser (Bertin Technologies, France) at 5.5m/s for three times one minute with short intervals.

16S rRNA gene amplicon sequencing
Quality control, library preparation and sequencing were performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEXTflex™ 16S V4 Amplicon-Seq Kit (BiooScientific, TX, USA) and the Illumina NovaSeq6000 platform (paired-end, 150bp). Raw reads were processed using the NG-Tax 0.4 pipeline with following settings: forward read length of 120, reverse read length of 120, ratio OTU abundance of 2.0, classify ratio of 0.9, minimum threshold of $1*10^{-7}$, identity level of 100% and error correction of 98.5, using the Silva_132_SSU Ref database [38, 39]. Since a 100% identity level was used, amplicon sequence variants (ASVs) were obtained. The obtained ASV table was filtered for ASVs with less than 0.005% relative abundance [40]. Three ZymoBiomics Microbial Community Standards (Zymo Research, Irvine, California, USA), two ZymoBiomics Microbial Community DNA Standards (Zymo Research) and three negative DNA extraction controls were included as positive and negative controls for DNA extraction and sequencing procedures.

Metagenomic sequencing
Ten faecal samples (two samples from five residents) and two positive controls were selected for metagenomic shotgun sequencing. Quality control, library preparation and sequencing were performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEBNext® Ultra™ II FS DNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA) and the Illumina NovaSeq6000 platform (paired-end, 150bp). Raw shotgun sequencing reads were processed using the NGLess (v1.0.1) language and accompanying tools [41-45]. NGLess is a domain specific language especially designed for processing raw sequence data and designed for enabling user-friendly computational reproducibility. Pre-processing of raw data was performed as previously described [41]. In short, raw sequence data was first pre-processed by performing quality-based trimming and reads with quality value below 25 were discarded, followed by discarding reads shorter than 45 bp. Second, reads were aligned to the human genome (hg19 reference) and discarded if reads mapped with more than 90% sequence identity and an alignment

length of at least 45 bp. Third, taxonomic profiling was performed using the mOTUs2 (v2.5.1) tool using default parameters as previously described [44]. This profiler is based on ten household, universal, single-copy marker gene families and profiles bacterial species both with (ref-mOTUs) and without (meta-mOTUs) a sequenced reference genome. A relative abundance table was obtained as output.

Next to the read-based analysis described above, we used an assembly-based analysis pipeline, Jovian (version v0.9.6.1) [46]. In short, the pipeline checks read quality, trims low-quality reads, removes reads derived from the host organism (human) and de novo assembles reads into scaffolds which are then taxonomically classified and quantified. These classifications were used to support the read-based results and scaffolds of selected species were compared to one another using pyANI (version 0.2.10) to calculate pairwise average nucleotide identities [47].

Positive and negative controls for gut microbiota profiling

Included controls indicate good DNA extraction and sequencing performance

An average of 24,095 reads (range 4,841-68,057, median 22,775 reads) was generated per sample for 16S rRNA gene amplicon sequencing (total n=93), resulting in 1042 ASVs after filtering on 0.005% abundance. Both positive DNA sequencing controls (n=2) were highly similar to theoretical expectations (average fold change 1.11), while DNA extraction controls (n=3) were somewhat less similar to theoretical expectation (average fold change 1.81). One DNA extraction control showed a lower than expected abundance (~12 fold) of *Staphylococcus* for unknown reasons (Additional file 1: Fig S1A). Of the three included negative extraction controls, two did not contain any reads post-filtering and one negative control contained 21 reads, mostly from known contaminants such as *Delftia* and *Streptococcus*, as previously observed [37].

For metagenomic sequencing, the DNA extraction control and sequencing control closely matched theoretical profiles and eight mOTUs were identified, apart from a small fraction of unassigned reads (Additional file 1: Fig S1B).

Statistical analysis and visualisations

Analyses and visualisations were performed in R (v3.6.1), using the following packages: phyloseq (v1.28.0), microbiome (v1.6.0), Metalonda (v1.1.5), DESeq2 (v1.24.0), tidyverse packages (v1.2.1), pheatmap (v1.0.12) and ggplot2 (v3.2.0) [48-54].

Community composition analysis

Permutational multivariate analysis of variance (PERMANOVA) using Bray-Curtis dissimilarity was performed to test for differences in overall community composition. Prior to employing PERMANOVA testing, it was tested whether groups had homogenous

dispersions (homoscedasticity) using the betadisper function, as violation of this statistical assumption can lead to erroneous conclusions regarding PERMANOVA results. No heteroscedasticity was observed between groups. To account for the repeated measurements design, we used 'strata=Resident number'. Principal coordinates analysis (PCoA) based on Bray-Curtis dissimilarity were made and 95% confidence intervals were computed using the stat_ellipse function. Alpha diversity indices (observed ASVs/ observed genera and Shannon index) were compared using independent t-tests or Wilcoxon rank sum tests. For calculating intraindividual stability, Bray-Curtis dissimilarities between all samples of a resident were calculated, and this was averaged to obtain a mean stability per resident.

Differential abundance analysis

Differential abundance analysis between groups (MDRO-positive samples versus MDRO-negative samples was performed at genus level using DESeq2 and stratified per time point. Genera had to be present in at least 25% of samples to be included in the analysis. To correct for false discovery rate, p-values were corrected using the Benjamini-Hochberg procedure. Considering the low number of MDRO-positive samples per time point, adjusted p-values < 0.1 were included in visualisation of results.

Time series modelling of alpha diversity

Linear mixed models were applied to investigate the changes in alpha diversity over time between the ever colonised versus never colonised groups using the lme4 and lmerTest packages [55, 56]. Ever colonised was defined as having an MDRO-positive sample at at least one time point during the study, while never colonised was defined as having no MDRO-positive sample during the study. Resident number was included as a random intercept to control for inter-individual baseline differences and repeated measurements design. The included fixed effect was the interaction between 'ever colonised' and timepoint ('ever colonised'*timepoint). Models were inspected for normally distributed residuals using qq-plots and p-values <0.05 were considered significant.

Time series modelling of individual taxa

To identify temporal trends in differential abundance of bacterial genera, the metagenomic longitudinal differential abundance method (MetaLonDA) package was used [50]. Only residents with at least three available gut microbiota samples were included in this analysis (n=24 residents). Genera had to be present in at least 25% of samples to be included in the analysis. MetaLonDA is capable of handling inconsistencies often observed in human microbiome studies (e.g. missing samples) and relies on two main modelling components, the negative binomial distribution for modelling read counts and smoothing spline ANOVA for modelling longitudinal profiles. The function metalondaAll was used with the following settings: n.perm=1000,

fit.method=”nbinomial”, num.intervals=3, pvalue.treshold=0.05, adjust.method=”BH”, norm.method=”median_ratio”. These settings indicate that the function was run with 1000 permutations using the median ratio method to normalise count data and fitting a negative binomial distribution. P-values were corrected using the Benjamini-Hochberg procedure.

Results

Clinical risk factor analysis for MDRO colonisation

MDRO colonisation among nursing home residents is highly prevalent and dynamic over time

Of the 27 included residents, twelve (44.4%) were colonised by an MDRO at at least one time point; four (33.3%) were colonised at one time point and eight residents (66.7%) at more than one time point during the six-month study (Fig 1). Of the 93 faecal samples, 27 (29.0%) contained an MDRO. Fourteen samples (15.1% of all samples) from six different residents (22.2% of all residents) were positive for ESBL-producing bacteria, of which ten were *E. coli*, three *Enterobacter cloacae* and one *Citrobacter* non-*koseri*. The remaining thirteen MDRO isolates (14.0% of all samples) were both fluoroquinolone and aminoglycoside resistant *E. coli.* No carbapenemase-producing Gram negative bacteria, VRE and *Clostridioides difficile* were cultured. As MDROs in the current study are exclusively MDR *Enterobacterales*, we refer to MDR *Enterobacterales* as MDROs from here onwards.

Clinical risk factors are only associated with MDRO colonisation in univariable analysis

Analysis of MDRO-status of faecal samples and clinical data using univariable GEE logistic regression showed several factors related to an increased risk of MDRO colonisation, including bone fracture in medical history (p=0.031, odds ratio (OR) 4.39, 95% confidence interval (CI) 1.14-16.95), antibiotic use in the past two months (p=0.039, OR 3.06, 95% CI 1.06-8.85) and hospital admittance in the last year (p=0.043, OR 4.95, 95% CI 1.05-23.34). Based on expert opinion, we further included sex and present use of urinary catheter as variables in multivariable GEE logistic regression. After including all variables in a multivariable GEE logistic regression only antibiotic use in the past two months displayed a trend (p=0.088, OR 2.84, 95% CI 0.85-9.49), while hospital admittance in the past year (p=0.13, OR 3.78, 95% CI 0.69-20.70) and bone fracture in medical history (p=0.35, OR 1.95, 95% CI 0.48-8.00) became non-significant. Lastly, multicollinearity between the included variables was assessed by computing variance inflation factors, but no considerable collinearity was observed (variance inflation factors for all variables < 2).

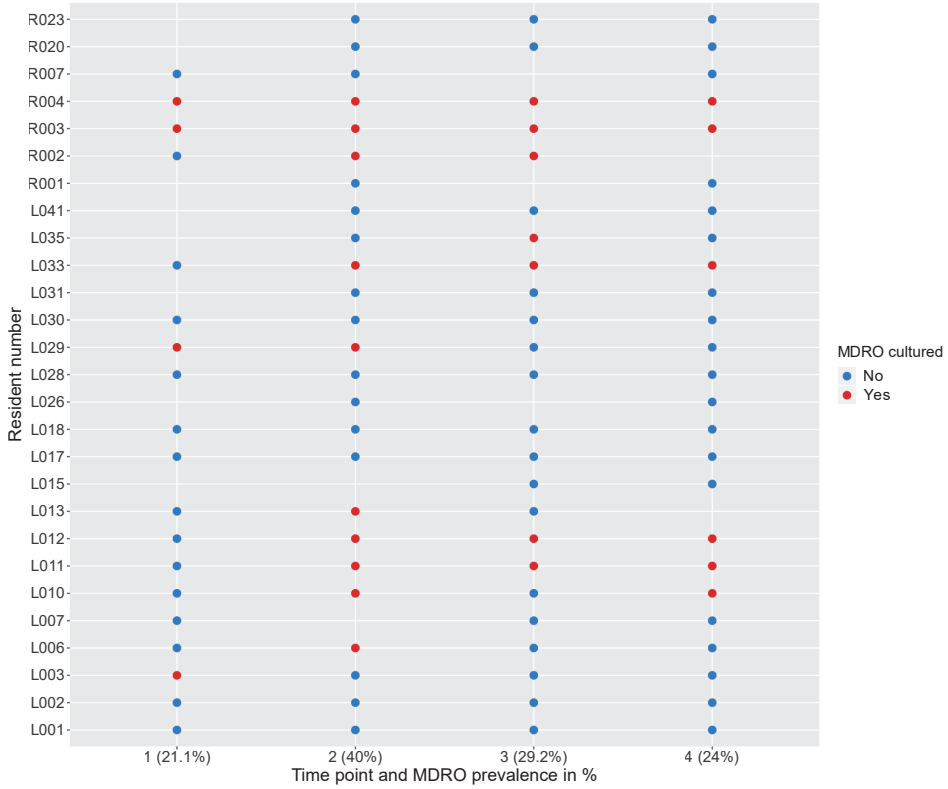


Figure 1: Overview of MDRO status for all samples of each resident over time. Blue colour indicates a negative MDRO culture, while red indicates a positive MDRO culture. Prevalence per time point is shown in percentage. Resident numbers are preceded by either 'R' or 'L', these letters indicate two physically separated buildings.

WGS of bacterial isolates

As most isolated MDRO strains were *E. coli* strains (22/27, 81.5%), we focused our analyses on this species. The 22 isolates were derived from 11 residents and were analysed by whole-genome analysis, including maximum likelihood phylogeny of core genes, accessory genome clustering, core-genome MLST and profiling of antibiotic resistance genes.

Genome-based clustering reveals a ward-specific E. coli *ST131 strain*

Based on pangenome analysis we identified core and accessory (non-core) genes, of which the accessory genes (5,057) were selected for clustering. Clustering based on presence/absence of these accessory genes showed a clear cluster of ST131 strains (Fig 2). Within the ST131 cluster, two separate clusters could be observed, one closely related cluster of twelve isolates belonging to three residents on ward A, and one cluster of four less related

isolates from four residents of four different wards. The isolates of three residents on ward A (R002, R003 and R004) have nearly identical accessory genes, suggesting that they were colonised with the same strain. In addition, these isolates have a nearly identical accessory genome over time, suggesting persistent colonisation of the same strain. Clustering based on the maximum likelihood phylogeny of core genes also resulted in a clear clustering of ST131 strains (data not shown). In addition, while the differences are smaller than in the accessory genome, ST131 strains from ward A still cluster apart from ST131 strains from other wards. Lastly, a core-genome MLST confirms clustering of ST131 strains on ward A (with up to two alleles difference) and shows that ST131 isolates from other wards are different (with more than 30 alleles difference) (Additional file 1: Fig S2). These results support the hypothesis that an ST131 strain was spread across ward A.

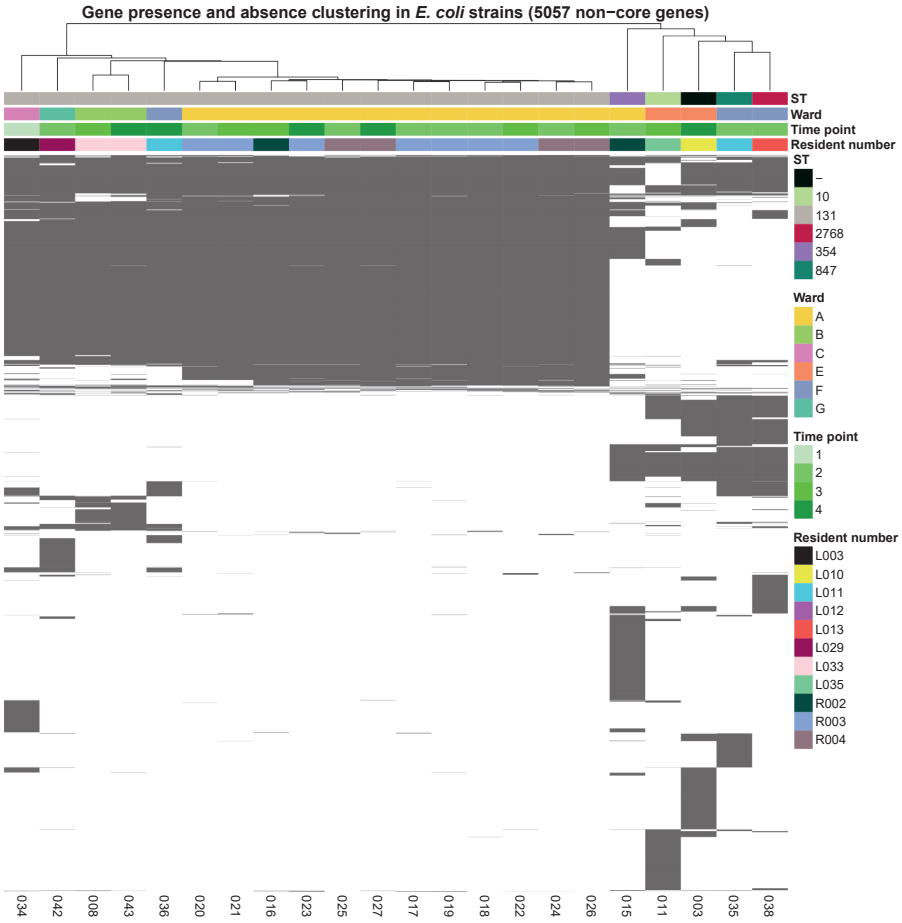


Figure 2: Overview of the accessory genome (non-core genes) of the 22 *E. coli* strains from eleven residents at different time points. Accessory genes are clustered based on the average linkage method using Euclidean distances. All (n=17) ST131 isolates cluster together, while

the other STs form a separate cluster. In addition, ST131 from ward A cluster together and are different from ST131 from other wards. The y-axis displays accessory genes and the x-axis isolate numbers. Black bars indicate presence and white bars absence of a gene.

Specific resistance genes are exclusive to certain wards

Next, the prevalence of antibiotic resistance genes was determined. Based on resistance gene absence/presence in the genome, ST131 largely clustered together (Fig 3), and again a cluster of ST131 belonging to residents of one ward (ward A) was observed. These strains were characterised by presence of nine resistance genes (*aac(6')-Ib-cr*, *aadA5*, *bla-CTX-M-15*, *blaOXA-1*, *catB3*, *dfrA17*, *mph(A)*, *sul1* and *tet(a)*). Three isolates belonging to ST131, 847 and 2786 from ward F clustered together, and these three strains (from two residents) contained the rifampicin resistance gene *arr-3*, which was not detected in other strains.

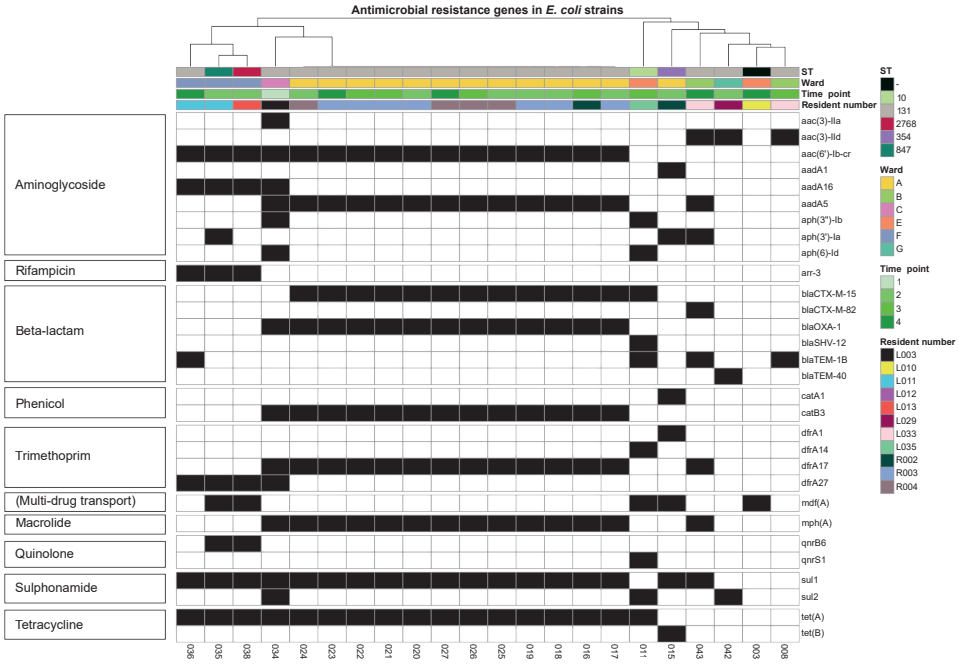


Figure 3: Heatmap of antibiotic resistance genes in the 22 *E. coli* isolates from eleven residents at different time points. Black boxes indicate presence of resistance gene, while white indicates absence of the resistance gene. Antibiotic resistance gene profiles are clustered by hierarchical clustering using Euclidian distances. Resident number, time, ward and time point are given as coloured annotations.

Gut microbiota analysis using 16S rRNA gene amplicon sequencing

A distinct gut microbiota between MDRO-positive and negative samples

First, alpha diversity (using observed ASVs/genera and Shannon index) was computed at both ASV and genus level to compare MDRO-positive with MDRO-negative samples. To account for repeated measures, we stratified these alpha diversity analyses by time point. No significant differences in alpha diversities at either level at any time point were observed (Additional file 1: Fig S3). Beta diversity was also not significantly different between these samples ($p=0.12$ and $R^2=0.049$) (Fig 4A). To identify individual bacterial taxa associated with MDRO status, differential abundance analysis was performed using DESeq2 at each time point. Several taxa were more abundant in MDRO-negative samples on multiple timepoints, namely *Atopobiaceae*, *Coprococcus_3*, *Dorea*, *Enorma*, *Holdemanella*, *Lachnospiraceae*, *Lachnospiraceae_ND3007_group*, *Phascolarctobacterium* and *Ruminococceae_UCG-014* (Additional file 1: Fig S4, Additional file 2: Table S1). Only three taxa (*Erysipelatoclostridium, uncultured_Coriobacteriales* and *uncultured_Ruminococcaceae* were more abundant in MDRO-positive samples at any time point.

MDRO colonisation is associated with consistent differences in relative abundance of specific bacterial taxa

Residents and their samples were further classified on having been MDRO-colonised at at least one time point during the study (ever, n=45 samples) or not (never, n=48 samples). There were no difference in alpha diversities over time between the groups (Additional file 1: Fig S5), nor in beta diversity (intra-individual stability) between the ever and never colonised group (independent t-test, $p = 0.2$) (Fig 4B).

Longitudinal differential abundance analysis between samples from ‘ever’ versus ‘never’ MDRO-colonised residents was performed to investigate whether differences in relative abundance were consistent over time. From each resident, at least three out of four samples should have been available to be included in this analysis, resulting in 45 samples from ever colonised residents and 42 samples from never colonised residents. Three taxa (*Atopobiaceae*, *Dorea* and *Lachnospiraceae_ND3007_group*) were consistently more abundant in ‘never’ colonised residents throughout the six months study period (Fig 5, Additional file 1: Fig S6). These taxa were also identified to be more abundant in MDRO-negative samples compared to MDRO-positive samples at two time points (Additional file 1: Fig S4).

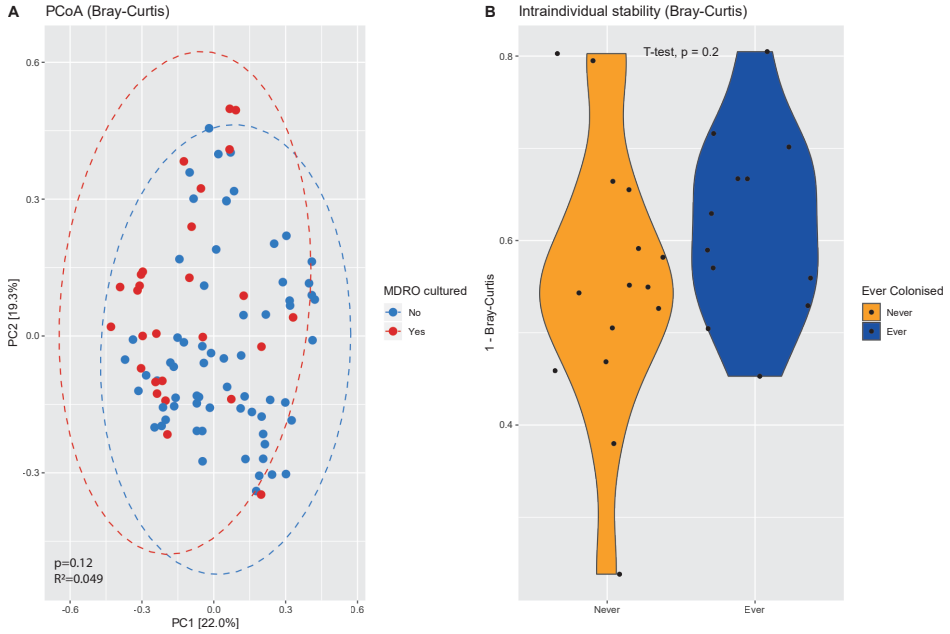


Figure 4: Bray-Curtis distance measures visualised by principle coordinates analysis (PCoA) for all (n=93) faecal samples based on whether an MDRO was cultured (A) and by mean intraindividual stability (1 - Bray-Curtis dissimilarity) between 'ever' and 'never' colonised residents (B). Each dot in the plot represents a single sample, and ellipses indicate 95% confidence intervals.

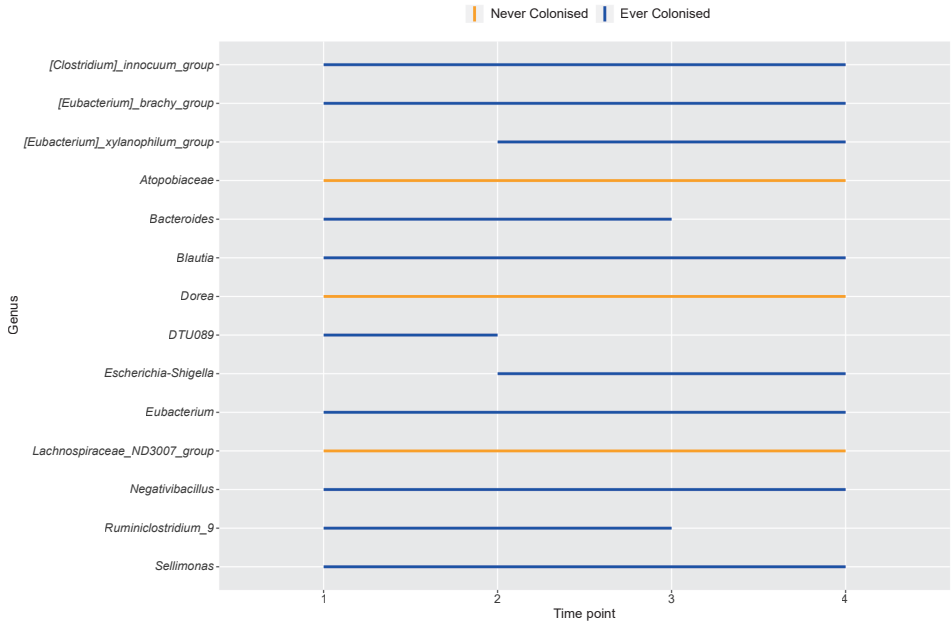


Figure 5: Time intervals of significantly different bacterial genera between ever (n=12) and never (n=15) MDRO colonised residents. Each line interval represents a significant time interval, with

significance being considered $p<0.05$. Orange lines indicate higher abundance in the never colonised group, while blue indicates higher abundance in the ever colonised group. If no coloured line is observed, the respective genus is not significantly differentially abundant between specific time points.

Lastly, we looked for intra-individual changes in pairs of samples of residents who either became MDRO colonised or were MDRO decolonised during the study period. For this, samples were analysed of an MDRO negative sample prior to an MDRO positive sample (n=8 residents), and vice versa; an MDRO positive sample followed by an MDRO negative sample (n=6 residents). Resident L10 could be included twice in the first comparison, but to avoid excessive impact of this resident on statistical analysis, it was included once. We then performed paired analyses for each of the two groups. However, no differences in alpha or beta diversity were observed, nor were any genera differentially abundant in any of the comparisons (data not shown).

Compositional profiles show very high abundance of Actinobacteria *members* Bifidobacterium *and* Collinsella

Next, we investigated the global microbiota profiles across all residents without a focus on MDRO colonisation. Compositional profiles at phylum and family level showed that the most abundant phylum in multiple residents was *Actinobacteria* (Fig 6A), which is in contrast to what is considered a 'normal' gut microbiota that generally consists of ~90% *Firmicutes* and *Bacteroidetes*. *Bifidobacterium* and *Collinsella* were the *Actinobacteria* members with highest relative abundance (Fig 6B).

Metagenome analysis using shotgun sequencing data of ten faecal samples

Not a single species, but several Bifidobacterium *species are highly abundant in residents*

The nursing home did not provide probiotics to their residents. However, the high abundance of *Bifidobacterium* in the residents' stools suggested otherwise. Ten stool samples from five residents with high *Bifidobacterium* and/or *Collinsella* relative abundance were further investigated by shotgun metagenomic sequencing, and two positive controls were included. The high relative abundance of *Bifidobacterium* and *Collinsella* could be confirmed and residents were colonised by seven highly abundant *Bifidobacterium* species, namely *B. adolescentis*, *B. angulatum*, *B. bifidium*, *B. breve*, *B. longum*, *B. pseudocatenulatum* and *B. ruminantium* (Fig 7A). From these species, *B. adolescentis*, *B. bifidum*, *B. breve* and *B. longum* are the most commonly used species in probiotics, although the others have been studied for probiotic properties as well [57].

Assembly-based method reveals that Bifidobacterium longum *strains are (almost) identical within residents, but not between residents*

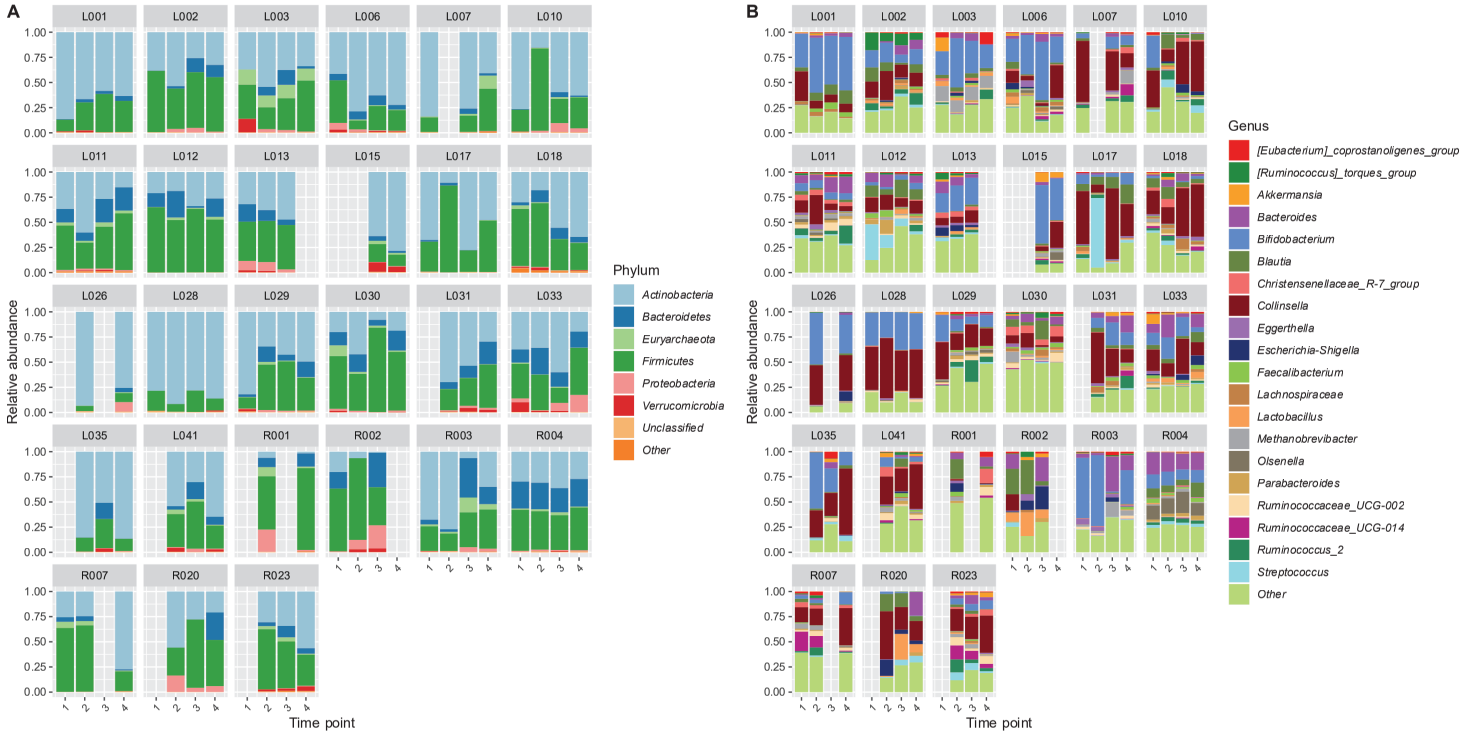


Figure 6: Compositional profiles at phylum level (A) and genus level (B) from 16S rRNA gene amplicon data of 27 residents at four time points. Other indicates the sum of all bacterial genera not specifically indicated in the legend. The y-axis displays relative abundance and the x-axis the study time point.

To investigate whether *Bifidobacterium longum* strains were identical between and within residents, we analysed the strains using *de novo* assemblies. *B. longum* was selected because of its high relative abundance in multiple samples, increasing the chance of recovering a full genome from the respective metagenomes and because it is commonly present in probiotics. Its genome size is about 2.5 Mb and contains a high GC content of ~60%. From samples of residents L001, L006 and L028, *B. longum* genomes larger than 2 Mb could be recovered, indicating that (nearly) full genomes were successfully obtained from the metagenome, but this was not the case for L031 and R003 (Additional file 3: Table S2). While average nucleotide identities were high between samples, strains from the same individual were more identical to themselves than to strains from other residents (Fig 7B). This indicates that residents do not carry the same *B. longum* strains. It should be noted that a full *B. longum* genome could not be retrieved for all residents. Lastly, *B. longum* genomes were compared to the NCBI reference genome (accession number NC_011593), the representative genome (NC_004307) and its plasmid (NC_004943) and several other *B. longum* strains (Fig 7B) to provide insight in what levels of divergence are to be expected between strains. Comparing these *B. longum* genomes from the NCBI database shows that unrelated *B. longum* strains have an average nucleotide identity (ANI) of between 0.956 and 0.988. This further confirms that *B. longum* strains between the nursing home residents were different (maximum ANI between strains from different residents 0.99) and that within residents strains were almost identical (ANI > 0.994), in case a nearly full genome could be retrieved.

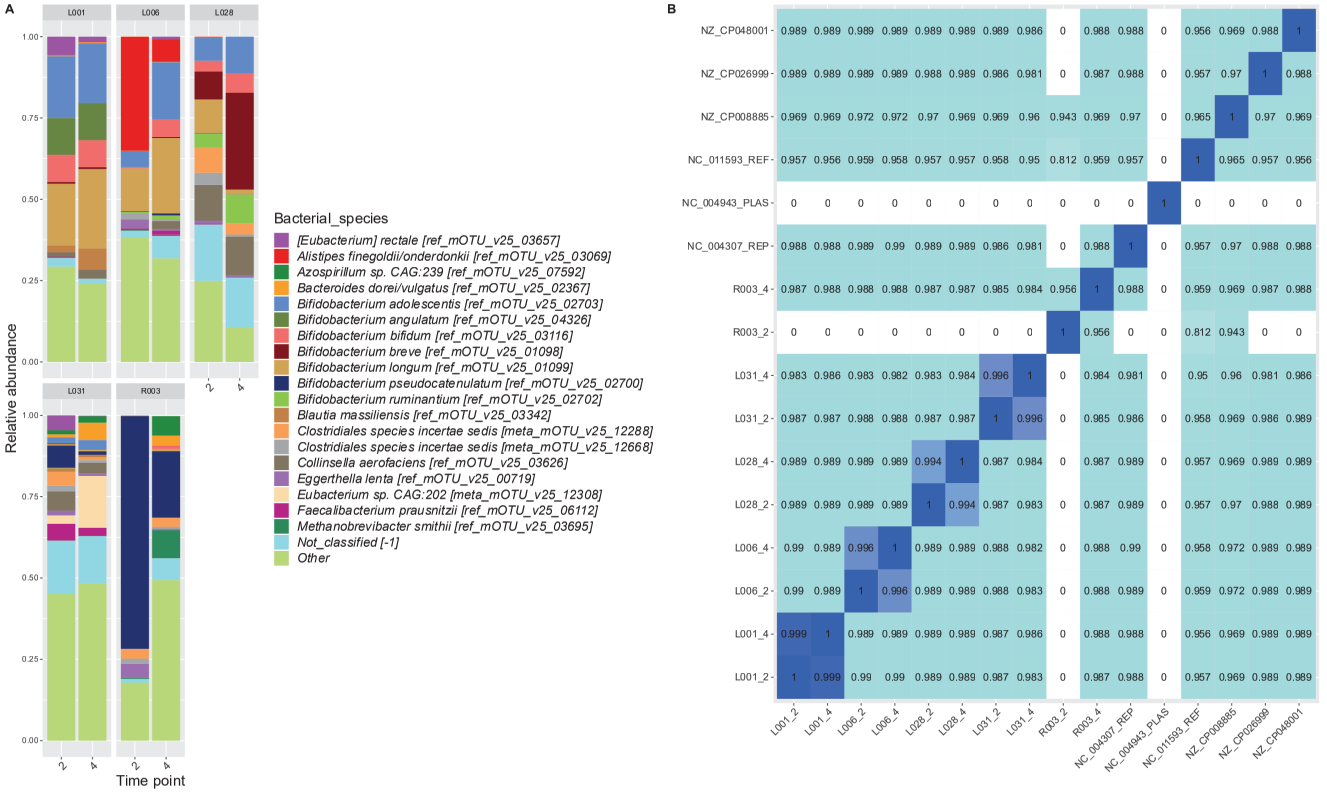


Figure 7: Compositional plot based on metagenomes of ten faecal samples from five residents using mOTUs (A) and average nucleotide identity between assembled *B. longum* strains and reference sequences (B). Relative abundance is shown of the twenty most abundant bacterial species in all samples and different bacterial species are indicated by colours. "*Other*" is the sum of the relative abundance of all species not listed in the colour key. Numbers on the x-axis indicate the resident number and study time point. Average nucleotide identity of *B. longum* strains as computed by pyANI. The sequence labelled "NC_004307_REP" in B is the representative genome on GenBank, the sequence with "NC_004943_PLAS" is its plasmid. The sequence with "NC_011593_REF" is the *B. longum* reference genome.

Discussion

We present a unique study on asymptomatic gut MDRO (in this study MDR *Enterobacterales*) colonisation in nursing home residents and performed a wide variety of analyses, namely clinical risk factor analysis, WGS of MDRO isolates and 16S rRNA gene amplicon sequencing and metagenomic sequencing of the gut microbiota. We identify possible risk factors for MDRO colonisation, potential spread of MDROs within a ward and microbial signatures associated with MDRO colonisation using 16S rRNA gene amplicon sequencing. Many of the MDRO-associated microbial signatures are consistent over the six-month time course of this study as shown by longitudinal modelling. Additionally, the unexpectedly high abundance of *Bifidobacterium* abundance in multiple residents was further investigated using metagenomic sequencing. We show that this high abundance is very unlikely to be stemming from probiotic supplementation, as *Bifidobacterium* species and *B. longum* strains differed between residents.

We observed a spread of *E. coli* ST131 within a ward, but not between wards, as the ST131 seemed ward-specific. *E. coli* ST131 was the most commonly found ST in our study, which is in line with previous results showing that this ST is major driver of the current worldwide spread of ESBL-producing *E. coli* [58, 59]. This sequence type is associated with community-acquired infections and older age, and is frequently observed in nursing homes in countries throughout Europe and the USA [7, 60-62]. While ST131 outbreaks are generally seen amongst and between various nursing homes, we concluded that spread of specific ST131 strains was restricted within wards. However, previous studies may have been limited by methods to characterise ST131, as they characterise strains only with regular MLST (of a limited number of housekeeping genes). By using pangenome analysis, we investigated the genetic differences in detail, allowing for discrimination of the ST131 strains between the wards. We conclude that MDRO transmission within nursing home wards seems to reflect that of household contacts [63]. This small scale MDRO spread was observed in the samples of 27 residents, one could hypothesize higher absolute numbers of related strains if all nursing home residents would have been screened. Not only strains can spread, plasmids are also able to move between different bacterial strains. For instance, three different *E. coli* ST types found at ward F contained *arr-3*, *aadA16* and *dfrA27*. Considering that these three genes are usually encoded on a plasmid [64, 65], it is possible that they spread between ST131 strains on ward F. However, definite conclusions cannot be made based on these results, as only three MDRO strains were detected in ward F.

Novel microbial signatures of MDRO colonisation were identified which could contribute to colonisation resistance against MDROs. Three taxa were consistently more abundant throughout the study in residents never colonised with an MDRO, namely

Dorea, *Lachnospiraceae_ND3007_group* and *Atopobiaceae*, and these taxa were also found to be more abundant in MDRO-negative samples at two time points. Increased relative abundance of *Dorea* and the *Lachnospiraceae* family has been shown to be associated with colonisation resistance against *Campylobacter* infection [66]. The relative abundance of *Dorea formicigenerans* was identified as a potential pre-liver transplant marker for subsequent MDRO colonisation [67] but another report did not mention *Dorea* as either a protective taxon or a risk factor [13]. While these results are conflicting, there is a possibility that different studies observed effects of different *Dorea* species or strains, which could theoretically have different or opposing effects on MDRO colonisation. Lastly, as clinical variables were not evenly distributed between compared groups, there is a possibility that observed differences in relative abundance of bacterial taxa can partially be attributed to these confounding factors.

We did not observe differences in alpha diversities between the different groups based on MDRO status. This contrasts several reports where MDRO colonisation was associated with a reduced alpha diversity, although conflicting evidence exists [13, 67, 68]. In addition, no difference in beta diversity was observed between the ever and never MDRO-colonised groups, nor between MDRO-positive and MDRO-negative samples. This contradicts findings in liver transplant patients and MDRO colonisation [67]. Conflicting results regarding diversities and microbial signatures could have multiple reasons. First, technical variation induced from the entire workflow starting with sample collection and ending with use of different statistical tools. Second, different MDRO types were studied between the various reports. In the current study, we mainly observed multi-drug resistant *E. coli*, while two other major studies investigating MDROs and gut microbiota found a larger variety of MDRO types [13, 67]. Considering that microbiome-mediated colonisation resistance is likely to be specific for individual bacterial species and most likely even bacterial strains, further studies should ideally focus on investigating single MRDOs in relation to the gut microbiota. Third, geographical locations of the studied cohorts were different, likely reflecting differences in gut microbiota composition due to varying dietary patterns and other cultural habits.

An unexpectedly high relative abundance of *Bifidobacterium* was observed in several residents in different wards. Such consistently high relative abundances have, to the best of our knowledge, not previously been described in adults or elderly. Incidental reports of an outgrowth of *Bifidobacterium* species in elderly in a long-term care facility have been described [69]. Rowan *et al*. observed a high relative abundance of *Bifidobacterium* species in two out of eleven elderly subjects (>15% relative abundance at at least one time point; mainly *B. longum*, *B. breve* and *B. adolescentis*), although potential explanations were not discussed.

It is known that in infancy the gut microbiota is largely dominated by *Bifidobacterium*, but that this high abundance declines with aging [70]. In addition, elderly mostly harbour *B. longum*, *B.nucleatum, B. pseudonucleatum* and *B. adolescentis*. While we found that these species were indeed among the most abundant, high relative abundances of *B. angulatum*, *B. bifidus*, *B. breve* and *B. ruminantium* were also observed. At first, we hypothesised that high *Bifidobacterium* relative abundance could be stemming from probiotic supplementation used on a voluntary basis by the nursing home residents, despite knowing that probiotics generally do not colonise very successfully [71, 72]. By performing metagenomic sequencing on a subset of samples, we showed this was unlikely to be the case, as different *Bifidobacterium* species were observed between residents. In addition, using strain-resolved metagenomics we show that *B. longum* strains were different between residents, but likely the same within residents. Our second hypothesis related to dietary patterns of residents, that perhaps a very monotonous diet could stimulate outgrowth of *Bifidobacterium*. However, residents consumed fresh, daily prepared meals according to a normal Dutch diet. It is unclear what the reasons and consequences of this high relative abundance of *Bifidobacterium* are in our residents. In combination with the observation that a high relative abundance of *Bifidobacterium* is not associated with protection against MDRO colonisation, this suggests that probiotics based on the *Bifidobacterium* species in our study may not effectively protect against MDRO colonisation.

This study has several limitations and strengths. First, our sample size and number of MDRO-positive samples was limited, preventing the application of a more extensive epidemiological risk factor analysis. Sample size was also a limiting factor in differential abundance testing between MDRO-positive and MDRO-negative samples per time point. Second, this study focused on a single nursing home and we can therefore not be certain that microbiota profiles are representative for residents of other (Dutch) nursing homes. Especially in light of our unique findings of high relative abundance of *Bifidobacterium* species, profiling the gut microbiota across other nursing homes would be important. Third, some wards had a very limited number of MDRO isolates, which hampered making definite conclusions about MDRO spread in those wards. Lastly, not all residents provided faecal samples on all four time points.

However, this study uses a unique combination of analyses for in-depth understanding of MDRO spread in a nursing home and the relation of MDRO colonisation with residents' microbiota. The longitudinal nature of our study setup allowed for 1) detection of robust associations between MDRO colonisation and specific microbial taxa and 2) identifying whether colonising MDRO strains were identical over time and 3) comparing *B. longum* strains within and between residents using strain-resolved metagenomics. In addition, the use of various statistical methods for identifying microbial taxa associated with

MDRO colonisation further strengthens our findings. Lastly, our finding of high relative abundance of *Bifidobacterium* in multiple residents warrants further investigation and confirmation by other studies.

Conclusions

Our study provides new evidence regarding the gut microbiota's potential in providing colonisation resistance against MDRO colonisation in a nursing home. Several specific taxa were identified which were consistently more abundant in residents never colonised with an MDRO throughout the six-month study. Considering that most of the detected MDROs were *E. coli* strains belonging to ST131, it may be especially interesting to test the potentially protective effect of these taxa against *E. coli* ST131. In addition, we report a uniquely high abundance of several *Bifidobacterium* species in multiple residents and excluded the possibility that this was due to probiotic supplementation. While the reasons for, and consequences of this high relative abundance remain unclear, it does suggest that probiotics based on *Bifidobacterium* species observed in our study are highly unlikely to prevent or eradicate MDRO colonisation in the gut of nursing home residents.

Declarations

Ethics approval and consent to participate

Written informed consent was obtained from the resident or his/her proxy. Ethical approval was granted by the 'Medisch Ethische Toetsings Commissie' of Leiden University Medical Centre (No.P16.039). The study was conducted in accordance with the guidelines of the Helsinki declaration. Sixty-four (49%) residents consented to participate. For this study, we included residents who gave consent for additional analyses, from whom faeces was cultured for MDROs at at least two time points, and of which sufficient material was left for microbiota profiling at at least two time points (n=27 residents).

Consent for publication

Not applicable.

Availability of data and materials

The sequencing data in this study are available at the European Nucleotide Archive (ENA) under accession number PRJEB37898 (https://www.ebi.ac.uk/ena/browser/view/PRJEB37898) (73). All data and R code necessary to reproduce analyses and

figures from this manuscript can be found at https://github.com/qducarmon/nursing_home_MDRO (74).

Competing interests

ET and EK are supported by an unrestricted grant from Vedanta Biosciences Inc. EK has performed research for Cubist, Novartis, and Qiagen and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur, and Seres Therapeutics. The companies had no role in the study or writing of the manuscript. The remaining authors declare that they have no competing interests.

Funding

This study was financially supported by RIVM with " Bijzonder resistente micro-organismen en *Clostridium difficile* dragerschap bij verpleeghuisbewoners:" (2016-2017).

Authors' contributions

QD analysed epidemiological data, 16S rRNA gene amplicon data, processed and analysed metagenomic data, created figures and wrote drafts of the manuscript. ET was the principal investigator, supervised the culturing experiments and wrote drafts of the manuscript. SN processed and analysed WGS data of MDRO isolates, processed and analysed metagenomic data, created figures and wrote drafts of the manuscript. MB performed practical work related to MDRO culturing and 16S rRNA gene amplicon sequencing. KV aided in the epidemiological analysis. MC designed the clinical study. IS performed practical work related to MDRO culturing. SD and MW designed the clinical study. RZ processed 16S rRNA gene amplicon data and supervised microbiota analyses. EK designed and supervised the clinical study**.** All authors read and approved the final manuscript.

Acknowledgements

We would like to thank all participating nursing home residents, their families and the staff of the nursing home: Woonzorgcentra Haaglanden, the Hague in the Netherlands
We would like to thank Renato Alves for his help and feedback regarding NGLess, Bastian Hornung for his guidance in analysis of MDRO isolates and Jacco Wallinga for his advice on risk factor analysis. In addition, we would like to thank all working members of CMAT for their valuable feedback during work discussions.

Abbreviations

MDROs: Multi-drug resistant organisms
ESBL: Extended-spectrum beta-lactamase
VRE: Vancomycin-resistant *Enteroccocus*
WGS: Whole-genome sequencing
TSB: Tryptic soy broth
MALDI-TOF: Matrix-assisted laser desorption ionisation-time of flight
EUCAST: European committee of antimicrobial susceptibility testing
MIC: Minimum inhibitory concentration
GEE: Generalised estimating equations
ST: Sequence type
ASV: Amplicon sequence variant
PERMANOVA: Permutational multivariate analysis of variance
MetaLonDA: Metagenomic longitudinal differential abundance method
MLST: Multi-locus sequence typing
OR: Odds ratio
CI: Confidence interval
ANI: Average nucleotide identity

References

1. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis. 2019;19(1):56-66.
2. Cassone M, Mody L. Colonization with Multi-Drug Resistant Organisms in Nursing Homes: Scope, Importance, and Management. Curr Geriatr Rep. 2015;4(1):87-95.
3. Gorrie CL, Mirceta M, Wick RR, Judd LM, Wyres KL, Thomson NR, 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):161-70.
4. Giannella M, Tedeschi S, Bartoletti M, Viale P. Prevention of infections in nursing homes: antibiotic prophylaxis versus infection control and antimicrobial stewardship measures. Expert Rev Anti Infect Ther. 2016;14(2):219-30.
5. Rooney PJ, O'Leary MC, Loughrey AC, McCalmont M, Smyth B, Donaghy P, et al. Nursing homes as a reservoir of extended-spectrum beta-lactamase (ESBL)-producing ciprofloxacin-resistant Escherichia coli. J Antimicrob Chemother. 2009;64(3):635-41.
6. Verhoef L, Roukens M, de Greeff S, Meessen N, Natsch S, Stobberingh E. Carriage of antimicrobial-resistant commensal bacteria in Dutch long-term-care facilities. J Antimicrob Chemother. 2016;71(9):2586-92.
7. van der Donk CF, Schols JM, Driessen CJ, Hagenouw RG, Meulendijks A, Stobberingh EE. Prevalence and spread of multidrug resistant Escherichia coli isolates among nursing home residents in the southern part of The Netherlands. J Am Med Dir Assoc. 2013;14(3):199-203.
8. van Dulm E, Tholen ATR, Pettersson A, van Rooijen MS, Willemsen I, Molenaar P, et al. High prevalence of multidrug resistant Enterobacteriaceae among residents of long term care facilities in Amsterdam, the Netherlands. PLoS One. 2019;14(9):e0222200.
9. Terveer EM, Fallon M, Kraakman MEM, Ormond A, Fitzpatrick M, Caljouw MAA, et al. Spread of ESBL-producing Escherichia coli in nursing home residents in Ireland and the Netherlands may reflect infrastructural differences. J Hosp Infect. 2019;103(2):160-4.
10. Verhoef L, Stobberingh E, Smid E, Kuijper EJ, De Greeff S, Heck M. Intestinal carriage of resistant bacteria and Clostridium difficile in nursing homes in the Netherlands—a point prevalence study. European Congress of Clinical Microbiology and Infectious Diseases; Vienna. 2017.
11. Arvand M, Moser V, Schwehn C, Bettge-Weller G, Hensgens MP, Kuijper EJ. High prevalence of Clostridium difficile colonization among nursing home residents in Hesse, Germany. PLoS One. 2012;7(1):e30183.
12. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev. 2019;83(3):e00007-19.
13. Araos R, Battaglia T, Ugalde JA, Rojas-Herrera M, Blaser MJ, D'Agata EMC. Fecal Microbiome Characteristics and the Resistome Associated With Acquisition of Multidrug-Resistant Organisms Among Elderly Subjects. Front Microbiol. 2019;10:2260.
14. Langelier C, Graves M, Kalantar K, Caldera S, Durrant R, Fisher M, et al. Microbiome and Antimicrobial Resistance Gene Dynamics in International Travelers. Emerg Infect Dis. 2019;25(7):1380-3.
15. Kim SG, Becattini S, Moody TU, Shliaha PV, Littmann ER, Seok R, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. Nature. 2019;572(7771):665-9.
16. Point prevalence survey of healthcare-associated infections and antimicrobial use in European long-term care facilities. April-May 2013. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/healthcare-associated-infections-point-prevalence-survey-longterm-care-facilities-2013.pdf. ECDC. 2014. Accessed 30 Nov 2020.
17. RIVM. WIP richtlijn Bijzonder resistente micro-organismen. Verpleeghuizen, woonzorgcentra en voorzieningen voor kleinschalig wonen voor ouderen. https://www.rivm.nl/publicaties?objectid=513c8b7b-189c4bcd-a124-cdeb80af520a&type=org&disposition=inline2014. Accessed 30 Nov 2020.
18. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters. version 73 http://www.eucast.org/clinical_breakpoints/. 2017. Accessed 30 Nov 2020.
19. NVMM. Laboratory detection of highly resistant microorganisms (HRMO). Revision 2017 https://www.nvmm.nl/media/1051/2012_hrmo_mrsa_esbl.pdf. 2012. Accessed 30 Nov 2020.
20. Terveer EM, Crobach MJ, Sanders IM, Vos MC, Verduin CM, Kuijper EJ. Detection of Clostridium difficile in Feces of Asymptomatic Patients Admitted to the Hospital. J Clin Microbiol. 2017;55(2):403-11.

21. Højsgaard S, Halekoh U, Yan J. The R Package geepack for Generalized Estimating Equations. J Stat Softw. 2005;15(2):1-11.
22. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
23. Johnston R, Jones K, Manley D. Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Qual Quant. 2018;52(4):1957-76.
24. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://wwwbioinformaticsbabrahamacuk/projects/fastqc. 2010. Accessed 12 Feb 2020.
25. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
26. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 2018;19(1):153.
27. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, et al. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads, eds Deng M, Jiang R, Sun F, Zhang X. Research in Computational Molecular Biology: 17th Annual International Conference, RECOMB 2013, Beijing, China. Springer, Berlin, Germany.
28. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-5.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.
30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.
31. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011;12(1):385.
32. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640-4.
33. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4):294-6.
34. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691-3.
35. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
36. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9.
37. 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):e00547-19.
38. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ, et al. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes[version 1; referees: 2 approved with reservations, 1 not approved]. F1000Research. 2016;5.
39. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.
40. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat Methods. 2013;10(1):57-9.
41. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7(1):84.
42. Kultima JR, Coelho LP, Forslund K, Huerta-Cepas J, Li SS, Driessen M, et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics. 2016;32(16):2520-3.
43. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, et al. MOCAT: a metagenomics assembly and gene prediction toolkit. PloS One. 2012;7(10):e47656-e.
44. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014.
45. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;preprint arXiv:1303.3997.
46. Schmitz D, Nooij S, Verhagen R, Janssens T, Cremer J, Zwagemaker F, et al. Jovian, user-friendly metagenomics. GitHub. Online at https://github.com/DennisSchmitz/Jovian/. 2020. Accessed 17 Mar 2020.
47. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food

security: soft-rotting enterobacterial plant pathogens. Anal Methods. 2016;8(1):12-24.
48. Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0.12. https://CRAN.R-project.org/package=pheatmap2019. Accessed 2 July 2020.
49. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.
50. Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome. 2018;6(1):32.
51. Lahti L, Shetty S. Tools for microbiome analysis in R. Microbiome package version 1.6.0. 2017.
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
53. Wickham H. ggplot2: Elegant Graphics for Data Analysis: New York: Springer Verlag; 2009. https://ggplot2.tidyverse.org.
54. Wickham H. tidyverse: Easily Install and Load the 'Tidyverse'. R package version 121 https://www.tidyverse.org/packages/. 2017. Accessed Mar 2020.
55. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67:1-48.
56. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw. 2017;82(13).
57. Fijan S. Microorganisms with claimed probiotic properties: an overview of recent literature. Int J Environ Res Public Health. 2014;11(5):4745-67.
58. Rogers BA, Sidjabat HE, Paterson DL. Escherichia coli O25b-ST131: a pandemic, multiresistant, community-associated strain. J Antimicrob Chemother. 2011;66(1):1-14.
59. Coque TM, Novais A, Carattoli A, Poirel L, Pitout J, Peixe L, et al. Dissemination of clonally related Escherichia coli strains expressing extended-spectrum beta-lactamase CTX-M-15. Emerg Infect Dis. 2008;14(2):195-200.
60. Broussier M, Gbaguidi-Haoré H, Rachidi-Berjamy F, Bertrand X, Slekovec C. Prevalence, genetic diversity of and factors associated with ESBL-producing Enterobacterales carriage in residents of French nursing homes. J Hosp Infect. 2019;104(4):469-75.
61. Arvand M, Moser V, Pfeifer Y. Prevalence of extended-spectrum-β-lactamase-producing Escherichia coli and spread of the epidemic clonal lineage ST131 in nursing homes in Hesse, Germany. J Antimicrob Chemother. 2013;68(11):2686-8.
62. Banerjee R, Johnston B, Lohse C, Porter SB, Clabots C, Johnson JR. Escherichia coli sequence type 131 is a dominant, antimicrobial-resistant clonal group associated with healthcare and elderly hosts. Infect Control Hosp Epidemiol. 2013;34(4):361-9.
63. Hilty M, Betsch BY, Bögli-Stuber K, Heiniger N, Stadler M, Küffer M, et al. Transmission dynamics of extended-spectrum β-lactamase-producing Enterobacteriaceae in the tertiary care hospital and the household setting. Clin Infect Dis. 2012;55(7):967-75.
64. Ma J, Zeng Z, Chen Z, Xu X, Wang X, Deng Y, et al. High prevalence of plasmid-mediated quinolone resistance determinants qnr, aac(6')-Ib-cr, and qepA among ceftiofur-resistant Enterobacteriaceae isolates from companion and food-producing animals. Antimicrob Agents Chemother. 2009;53(2):519-24.
65. Papagiannitsis CC, Kutilova I, Medvecky M, Hrabak J, Dolejska M. Characterization of the Complete Nucleotide Sequences of IncA/C(2) Plasmids Carrying In809-Like Integrons from Enterobacteriaceae Isolates of Wildlife Origin. Antimicrob Agents Chemother. 2017;61(9).
66. Kampmann C, Dicksved J, Engstrand L, Rautelin H. Composition of human faecal microbiota in resistance to Campylobacter infection. Clin Microbiol Infect. 2016;22(1):61.e1-.e8.
67. Annavajhala MK, Gomez-Simmonds A, Macesic N, Sullivan SB, Kress A, Khan SD, et al. Colonizing multidrug-resistant bacteria and the longitudinal evolution of the intestinal microbiome after liver transplantation. Nat Commun. 2019;10(1):4715.
68. Araos R, Montgomery V, Ugalde JA, Snyder GM, D'Agata EMC. Microbial Disruption Indices to Detect Colonization With Multidrug-Resistant Organisms. Infect Control Hosp Epidemiol. 2017;38(11):1312-8.
69. Rowan-Nash AD, Araos R, D'Agata EMC, Belenky P. Antimicrobial Resistance Gene Prevalence in a Population of Patients with Advanced Dementia Is Related to Specific Pathobionts. iScience. 2020;23(3):100905.
70. Arboleya S, Watkins C, Stanton C, Ross RP. Gut Bifidobacteria Populations in Human Health and Aging. Front Microbiol. 2016;7:1204.
71. Zmora N, Zilberman-Schapira G, Suez J, Mor U, Dori-Bachash M, Bashiardes S, et al. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome

Features. Cell. 2018;174(6):1388-405.e21.
72. Maldonado-Gómez MX, Martínez I, Bottacini F, O'Callaghan A, Ventura M, van Sinderen D, et al. Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome. Cell Host Microbe. 2016;20(4):515-26.
73. Ducarmon QR, Terveer EM, Nooij S, Bloem MN, Vendrik KEW, Caljouw MAA, et al. Short-read sequencing data of MDRO isolates from residents of a Dutch nursing home and short-read sequencing data of their gut microbiota with 16S rRNA gene amplicon and metagenomics sequencing. European Nucleotide Archive at EMBL-EBI under accession number PRJEB37898: Available from: https://www.ebi.ac.uk/ena/browser/view/PRJEB37898; 2020. Accessed Aug 2020.
74. Ducarmon QR, Nooij S. GitHub repository with all necessary data and RMarkdown files to reproduce analyses and figures from this manuscript. GitHub. 2020. Available from: https://github.com/qducarmon/nursing_home_MDRO. Accessed 27 Jan 2021.

Chapter 9

Asymptomatic gut colonization by extended-spectrum beta-lactamase-producing *Escherichia coli* is not associated with an altered gut microbiome or metabolome in Dutch adults

bioRxiv, 2021

Q. R. Ducarmon[a,b], R. D. Zwittink[a,b], R. P. J. Willems[c], A. Verhoeven[d], S. Nooij[a,b], F. R.M. van der Klis[e], E. Franz[e], J. Kool[e], M. Giera[d], C. M. J. E. Vandenbroucke-Grauls[c], S. Fuentes[e], E. J. Kuijper[a,b,e]

[a]Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands.
[b]Center for Microbiome Analyses and Therapeutics, Leiden University Medical Center, Leiden, The Netherlands.
[c]Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical Centers, Amsterdam Medical Center, Amsterdam Infection and Immunity Institute, Amsterdam, the Netherlands.
[d]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands.
[e]Center for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands.

bioRxiv. 2021 May 18; https://doi.org/10.1101/2021.05.18.444613.
Supplementary information available online.

Abstract

Background

Gut colonization by antibiotic resistant *E. coli* strains, including extended-spectrum beta-lactamase (ESBL)-producing *E. coli* is a risk factor for developing overt infection. The gut microbiome can provide colonization resistance against enteropathogens, but it remains unclear whether it confers resistance against potentially pathogenic ESBL-producing *E. coli*.

Materials

From a Dutch cross-sectional population study (PIENTER-3), feces from 2751 individuals were used to culture ESBL-producing bacteria. Of these, we selected 49 samples which were positive for an ESBL-producing *Escherichia coli* ($ESBL^+$), and negative for a variety of variables known to affect microbiome composition. These were matched in a 1:1 ratio to $ESBL^-$ samples based on age, sex, having been abroad in the past six months and ethnicity. Shotgun metagenomic sequencing was performed and taxonomic species composition and functional annotations (microbial metabolism and carbohydrate-active enzymes) were determined. Targeted quantitative metabolic profiling (^{1}H NMR-spectroscopy) was performed to investigate metabolomic profiles.

Results

No differences in alpha or beta diversity were observed, nor in relative abundance, between $ESBL^+$ and $ESBL^-$ individuals based on bacterial species level composition. Machine learning approaches based on microbiota composition did not accurately predict ESBL status (area under the receiver operating characteristic curve (AUROC)=0.53), neither when based on functional profiles. The metabolome did also not convincingly differ between ESBL groups as assessed by a variety of approaches, including machine learning through random forest (AUROC=0.61).

Conclusion

Using a combination of multi-omics and machine learning approaches, we conclude that asymptomatic gut carriage of ESBL-producing *E. coli* is not associated with an altered microbiome composition or function. This may suggest that microbiome-mediated colonization resistance against ESBL-producing *E. coli* is not as relevant as it is against other enteropathogens.

Introduction

Escherichia coli is a common gut commensal, but several strains possess virulence factors that enable them to cause gastrointestinal, urinary and extraintestinal infections[1, 2]. Colonization of the gut by multidrug-resistant organisms (MDRO), including extended-spectrum beta-lactamase (ESBL)-producing *E. coli* and carbapenem-resistant *E. coli*, often precede infections[3]. The gut microbiome can mediate colonization resistance against several enteric pathogens, but it remains unclear whether this is also the case for MDROs such as ESBL-producing *E. coli*, especially since many individuals harbor commensal *E. coli*. Colonization resistance can be conferred by the gut microbiome through nutrient competition, production of antimicrobial compounds, support of gut barrier integrity, bacteriophage deployment and through interaction with the immune system[4]. However, studies in humans have reported conflicting evidence regarding which bacterial genera or species within the gut microbiome could be of relevance in providing colonization resistance against ESBL-producing *E. coli* or ESBL-producing *Enterobacterales*. These conflicting results can, at least partially, be traced back to several confounding factors (e.g. medication) in those studies[5-8]. It was recently shown that unevenly matched case-controls studies with regard to lifestyle and physiological characteristics can produce spurious microbial associations with human phenotypes like disease, or in this case, colonization by ESBL-producing *E. coli*[9].

Here, we aimed to compare the gut microbiome and metabolome between individuals asymptomatically colonized with an ESBL-producing *E. coli* ($ESBL^+$) and individuals who are not ($ESBL^-$), determined by culture-based and molecular approaches. To avoid confounding factors from affecting study results, we selected samples from a large Dutch cross-sectional population study (PIENTER-3) for which 2751 fecal samples were used to culture ESBL-producing bacteria[10]. With this high number of samples available, we could apply stringent sample selection with regard to known confounders in microbiome studies such as antibiotic use, proton-pump inhibitor use, a variety of diets etc. Subsequently, we performed case control matching based on a variety of epidemiological and health related variables. We performed extensive functional and taxonomic profiling of the gut microbiome through metagenomics and metabolomics to investigate whether there are differences in the gut microbiome between matched $ESBL^+$ and $ESBL^-$ individuals.

Materials and methods

Sample collection

Samples were selected from a large Dutch population-wide study (PIENTER-3)[10]. This cross-sectional population study was carried out in 2016/2017, primarily designed

to obtain insight into age-specific seroprevalence of vaccine-preventable infectious diseases. Out of the 98 included samples for the current study, 95 were stored in the freezer within 15 minutes after defecation, one person did not provide information on this and two individuals took longer than one hour to store their sample in the freezer. Samples were kept on average for 2.97 days (±2.82) (six individuals did not indicate this information) in people's freezer before being delivered (on cold packs) to the mobile study team[10]. Fecal samples were kept on dry ice during transport to the National Institute for Public Health and the Environment and stored at -80°C the next day.

Detection of ESBL-producing *Enterobacterales*

Details of the microbiological methods have been described elsewhere (Willems RPJ, van Dijk K, Dierikx CM, Twisk JWR, van der Klis FRM, de Greeff SC, Vandenbroucke-Grauls CMJE. Gastric acid suppression, lifestyle factors and intestinal carriage of ESBL and carbapenemase-producing Enterobacterales: a nationwide population-based study [Submitted]). Briefly, stool specimens were enriched by tryptic soy broth with ampicillin (50 mg/L) and then cultured on selective agar plates (EbSA, Cepheid Benelux, Apeldoorn). Next, up to five oxidase-negative morphotypes were subcultured, identified to species level, and tested for antimicrobial susceptibility using standard procedures (VITEK 2 system, bioMérieux, Marcy-L'Étoile, France). Antimicrobial susceptibility was classified according to European Committee on Antimicrobial Susceptibility Testing clinical breakpoints[11]. ESBL production was screened for with combination disk diffusion and confirmed by polymerase chain reaction (PCR); PCR was performed for the $bla_{\text{CTX-M}}$, bla_{SHV} and bla_{TEM} groups[12]. ESBL testing was done according to the European Committee on Antimicrobial Susceptibility Testing guidelines[13]**.**

Sample selection

2751 fecal samples were cultured for ESBL- or CPE-producing bacteria, of which 198 samples were positive. For the purpose of our study, we selected samples positive for ESBL-producing *E. coli*, resulting in 176 potential samples. Next, we applied stringent exclusion criteria for all samples based on variables known to affect the gut microbiome. Individuals were excluded based on the following criteria: current proton-pump inhibitor use, antibiotic use in the last three months, diarrheal symptoms in the last month (defined as at least three thin stools within 24 hours), vomiting in the last month, blood in stool during the last month, abdominal pain or nausea during the last month, use of any pre- or probiotics, consumption of a special diet (vegetarian, cow's milk free diet, hen's egg protein-free diet, gluten free, nut and/or peanut-free, lactose limited diet, diabetes-related diet, limited protein diet, limited fat and/or cholesterol diet, enrichment of dietary fiber, caloric restriction, low in sodium, easily digestible, coloring agent-free, enriched in energy/protein, 'other diet') and whether stool was stored in the freezer after defecation (samples were excluded if not stored in the freezer). This selection resulted in 51 $ESBL^+$ samples for inclusion, which were subsequently matched to 51 $ESBL^-$

samples using the R MatchIt package (v3.0.2) with the "nearest" method in the *matchit* function. Subjects were matched based on age, sex, having been abroad during the last 6 months (yes/no) and ethnicity. $ESBL^-$ negative samples were selected using the same exclusion criteria. Three samples (1 $ESBL^-$ sample and 2 $ESBL^+$ samples) were further excluded as insufficient DNA was available for sequencing. One additional sample ($ESBL^-$) was excluded as we discovered afterwards that this individual had provided ambiguous answers regarding dietary habits. The final dataset for analysis contained 49 individuals in each group.

DNA extraction for metagenomic shotgun sequencing

DNA was extracted by mechanical disruption (repeated bead-beating) and purified in a Maxwell RSC instrument (Promega Benelux BV, Leiden, The Netherlands). The Maxwell RSC Blood DNA extraction kit was according to manufacturer's instructions with several modifications, as follows. Fecal samples were thawed on ice and approximately 250 mg of well-homogenized fecal material was resuspended in S.T.A.R (stool transport and recovery buffer) buffer (Roche Diagnostics, Almere, The Netherlands), with 0.1 mm zirconia/silica beads and 2.5 mm glass beads. The fecal suspension was mechanically disrupted three times for one minute in a FastPrep-24 Instrument at room temperature and 5.5 oscillations, and maintained on ice after every cycle. Samples were further heated at 95°C for 15 minutes shaking at 300 rpm, and centrifuged for 5 minutes at full speed. Resulting supernatants (fecal lysates) were collected and the pellet was further resuspended in an additional 350 µl of S.T.A.R. buffer following the same procedure. Pooled fecal lysates were then transferred to the Maxwell RSC Instrument for further purification steps. Eluted sample was cleaned-up using the OneStep PCR Inhibitor Removal Kit (Zymo Research, Irvine, California), and DNA was quantified using a Quantus Fluorometer (Promega Corporation, Madison, WI, USA). Every extraction round included two negative DNA extraction controls (blank samples with S.T.A.R. buffer without any added fecal material) and two microbial mock communities as positive controls (ZymoBiomics Microbial Community Standards; Zymo Research, Irvine, California, USA).

Metagenomic shotgun sequencing

Shotgun metagenomic sequencing was performed by GenomeScan B.V. (Leiden, The Netherlands) using the NEBNext® Ultra™ II FS DNA Library Prep Kit (New England Biolabs, Ipswich, Massachusetts, USA) and the NextSeq 500 platform (paired-end, 150bp). Two positive sequencing controls (ZymoBiomics Microbial Community DNA Standards; Zymo Research, Irvine, California, USA) and two negative sequencing controls (sterile water) were included. Average number of raw reads (of 98 samples and four positive controls) is 4,747,908 (range 2,565,232 – 62,035,096) and a median of 4,142,237 paired-end reads. Raw shotgun sequencing reads were quality checked using

the FastQC (v0.11.9) and MultiQC (v1.8) tools, both before and after cleaning files for low-quality reads and human reads using the kneaddata (v0.7.10) tool with default parameters.

Taxonomic and functional annotation were performed on cleaned reads using the NGLess language (v1.2.0), associated tools and the Integrated Gene Catalog (IGC) database[14-18]. For taxonomic analysis, mOTUs (v2.5.1) was used with default parameters and unclassified reads (-1 category in mOTUs) were not included for downstream analyses[19]. Functional annotation was performed by aligning cleaned reads to the annotated IGC database (we annotated the IGC through eggNOG mapper v2.1.0 using default parameters and the "-m diamond" argument) using Burrows-Wheeler-Aligner MEM (BWA, v0.7.17)[17, 18, 20]. Unclassified reads were not taken into account for downstream analyses. Default parameters were used, apart from the 'normalization' argument, which was specified as normalization="scaled", which corrects for size of the feature (gene). Aligned reads were then aggregated using the Kyoto Encyclopedia of Genes and Genomes (KEGG), KEGG Orthology (KO) groups and Carbohydrate-active enzymes (CAZymes) annotations present in the IGC (features="KEGG_ko" or features= "CAZy" argument in NGLess)[21, 22].

Multi-locus sequence typing on *E. coli* was performed using the MetaMLST tool (default parameters). MetaMLST aligns sequencing reads against a database (which can be customized) of housekeeping genes to identify sequence types present in metagenomes. A custom *E. coli* database (Achtman MLST scheme) was created with MLST data from October 16th 2020 (https://pubmlst.org/bigsdb?db=pubmlst_ecoli_achtman_seqdef)[23]. No sequence types could be reliably detected in the samples, likely due to the very low relative abundance of *E. coli* and the corresponding low number of reads and coverage of *E. coli*.

Resistome profiling

To profile the antimicrobial resistance genes in the metagenomes, cleaned reads were aligned to the MEGARes database (v2.00) using BWA MEM with default settings[17]. The resulting SAM file was parsed using the ResistomeAnalyzer tool (https://github.com/cdeanj/resistomeanalyzer) and the default threshold of 80% was used, meaning an antibiotic-resistance determinant was only included if at least 80% of the gene is detected in a sample[24]. Read counts originating from alignments to housekeeping genes associated with antimicrobial resistance (AMR) (e.g. *rpoB* and *gyrA*) that require single nucleotide polymorphisms to confer resistance were filtered out of the count table before downstream analyses, as previously reported[25]. Gene level data (e.g. *tetO*, *tetQ* and *tetW*) were used for calculating alpha and beta diversity metrics and for differential abundance analysis. For visualization purposes, gene level outputs were aggregated at the mechanism level (e.g. beta-lactams, mupirocin).

Positive and negative controls for metagenomic sequencing

Eight mOTUs were detected in all four positive controls, exactly matching theoretical expectations. With regard to expected relative abundances, sequencing controls were, as expected, more accurate (average fold error of 1.14) than the DNA extraction controls (average fold error of 1.42 with underrepresentation of Gram-positive bacteria). The four included negative controls (two extraction controls and two sequencing controls) did not generate any reads. These results indicate good performance of sequencing, DNA extraction procedures and bioinformatic processing of the data.

Metabolomics

The method for NMR analysis of fecal samples was adapted from the protocol developed by Kim et al. with a few minor adaptations[26].

Sample preparation

Each feces-containing sample tube was weighed before sample preparation. To each sample tube 50 µl of 0.5 mm zirconium oxide beads (Next Advance, Inc.) and 750 µl of milli-Q water were added. Then, the tubes were subjected to bead beating for four sessions of one minute. The tubes were subsequently centrifuged at 18,000 g at 4°C for 15 minutes. For most samples, 600 µl of supernatant was transferred to new 1.5 ml Eppendorf tubes. In some cases the volume of available supernatant was slightly less. These tubes were centrifuged at 18,000 g at 4 °C for 1 hour. 270 µl of supernatant was added to 30 µl of pH 7.4 phosphate buffer (1.5 M) in 100% D_2O containing 4 mM TSP-d_4 and 2 mM NaN_3. A customized Gilson 215 liquid handler was used to transfer the samples to a 3.0 mm Bruker NMR tube rack. The original sample tubes were cleaned, dried and weighed again.

NMR measurements

^{1}H NMR data were collected using a Bruker 600 MHz Avance Neo/IVDr spectrometer equipped with a 5 mm TCI cryogenic probe head and a z-gradient system. A Bruker SampleJet sample changer was used for sample insertion and removal. All experiments were recorded at 300 K. A standard sample 99.8% methanol-d4 was used for temperature calibration before each batch of measurements[27]. One-dimensional (1D) ^{1}H NMR spectra were recorded using the first increment of a NOESY pulse sequence[28] with presaturation (γB_1 = 50 Hz) during a relaxation delay of four seconds and a mixing time of 10 ms for efficient water suppression[29]. Initial shimming was performed using the TopShim tool on a random mix of urine samples from the study, and subsequently the axial shims were optimized automatically before every measurement. Duration of 90° pulses were automatically calibrated for each individual sample using a homonuclear-gated mutation experiment[30] on the locked and shimmed samples after automatic tuning and matching of the probe head. 16 scans of 65,536 points covering 12,335 Hz were

recorded. J-resolved spectra (JRES) were recorded with a relaxation delay of 2 s and 2 scans for each increment in the indirect dimension. A data matrix of 40 × 12,288 data points was collected covering a sweep width of 78 × 10,000 Hz. Further processing of the raw time-domain data was carried out in the KIMBLE environment[31]. The Free Induction Decay of the 1D experiment was zero-filled to 65,536 complex points prior to Fourier transformation. An exponential window function was applied with a line-broadening factor of 1.0 Hz. The spectra were automatically phase and baseline corrected and automatically referenced to the internal standard (TSP = 0.0 ppm). A sine-shaped window function was applied and the data was zero-filled to 256 × 16,384 complex data points prior to Fourier transformation. In order to remove the skew, the resulting data matrix was tilted along the rows by shifting each row (k) by 0.4992× (128-k) points and symmetrized about the central horizontal lines.

Metabolite quantification

Metabolites were quantified using KIMBLE and the results were checked by quantifying the same metabolites both in the JRES and in the NOESY1D experiments and in 10 randomly chosen spectra using the Chenomx NMR Suite version 8.6 (Chenomx Inc., Edmonton AB, Canada).

Statistical analysis

Statistical software used for downstream analysis

Analyses and visualizations were performed in R (v4.0.4), using the following packages: phyloseq (v1.34.0), microbiome (v1.12.0), vegan (v2.5-7), tidyverse packages (v1.3.0), SIAMCAT (v1.10.0), table1 (v1.2.1) and ropls (v1.22.0)[32-38]. All analytical R code will be made publicly available upon acceptance of the manuscript. For all used tools, default parameters were used unless stated otherwise.

Community composition analysis of metagenomic data

We tested for differences in overall microbiota composition with permutational multivariate analysis of variance (PERMANOVA) using Bray-Curtis dissimilarity. As violation of the assumption of homogenous dispersions can lead to wrong conclusions regarding PERMANOVA, we first tested this assumption using the *betadisper* function of the vegan package. No heteroscedasticity was observed between the $ESBL^+$ and $ESBL^-$ group. To investigate both linear and non-linear patterns in the data, we performed dimension reduction using both principal coordinates analysis (PCoA) and t-distributed stochastic neighbor embedding (t-SNE), both based on Bray-Curtis dissimilarity. Alpha diversity indices were compared using independent t-tests.

Differential abundance analysis in metagenomic data

Differential abundance analysis of mOTUs, KO groups, CAZymes and resistance

genes between ESBL$^+$ and ESBL$^-$ samples was performed using SIAMCAT on relative abundance matrices. Features (mOTUs, KO groups or CAZymes) had to be present in at least 25% of samples to be included in the analysis. Regarding resistome analyses, a gene had to be present in 10% of samples to be included, as the 25% prevalence cut-off was too stringent resulting in only fourteen genes included in the analysis. To correct for false discovery rate, p-values were corrected in all tests using the Benjamini-Hochberg procedure[39].

Machine learning classifier on metagenomic data

We used obtained taxonomic and functional profiles for feature selection and construction of prediction models. To this end, least absolute shrinkage and selection operator (LASSO) logistic regression using the SIAMCAT package was performed to select predictive features and remove uninformative features based on species composition or functional profiles. Preprocessing was done by filtering mOTUs, KO groups, or CAZyme families which were present in at least 25% of samples. The vignette from SIAMCAT (https://siamcat.embl.de/articles/SIAMCAT_vignette.html) was followed[37]. In short, we performed data normalization using the "log.unit" method, 5-fold cross validation to split the data in several combinations of training and test data, trained the model using LASSO logistic regression ("lasso" parameter) and, lastly, made the predictions.

Metabolomics data

Metabolomic concentrations were first log10 normalized to reduce heteroscedasticity. Metabolite concentrations were subsequently centered and scaled to a mean of 0 and standard deviation of 1, as previously described[40]. Differences in concentrations between ESBL groups were tested using t-tests where p-values were corrected for multiple testing using two methods (to establish robustness of potential findings), namely Benjamini-Hochberg and Holm correction (with Holm correction being more conservative)[39, 41]. Next, we performed multivariate analyses using PCA and Partial Least-Squares Discriminant Analysis (PLS-DA). Lastly, random forest was applied to investigate whether ESBL$^+$ and ESBL$^-$ individuals could be accurately classified based on their respective metabolite profiles. As input to the random forest, normalized metabolite concentrations were used and, similarly as with metagenomic data, 5-fold cross validation was implemented in SIAMCAT.

Results

Participant and ESBL-producing *E. coli* isolates characteristics

The original sample selection contained 51 individuals in each group, but three samples were not suitable for metagenomic sequencing due to too low DNA concentrations after

extraction. One more individual had to be excluded due to ambiguous answers regarding dietary habits. Ultimately, this resulted in metagenomics data from 49 individuals per group. Demographic and participant characteristics were highly similar between the ESBL$^+$ and ESBL$^-$group and antibiotic use between the preceding three to twelve months was also evenly matched (Table 1). With regard to the ESBL-producing *E. coli* isolates that colonized our 49 ESBL$^+$ participants, 44 carried a CTX-M-type. The majority of these were CTX-M-1 (25) and CTX-M-9 (18) and one could not definitively be typed (CTX-M-1 or CTX-M-8). Isolates of four individuals were negative for CTX-M genes and for one participant it could not be determined. Additional information on antimicrobial susceptibility of the strains can be found in Supplementary Table 1.

Table 1: Characteristics of participants included in the study. P-values were obtained using an independent t-test (for numerical variables) or Fisher's exact test (for categorical variables).

	ESBL negative	ESBL positive	P-value
	(N=49)	(N=49)	
Age (years)			
Mean (SD)	44.1 (15.2)	46.6 (15.3)	0.43
Median [Min, Max]	45.0 [20.0, 74.0]	46.0 [21.0, 74.0]	
Sex			
Male	26 (53.1%)	23 (46.9%)	0.69
Female	23 (46.9%)	26 (53.1%)	
Abroad in last 6 months			
Yes	39 (79.6%)	37 (75.5%)	0.81
No	10 (20.4%)	12 (24.5%)	
Ethnicity			
Dutch	38 (77.6%)	36 (73.5%)	0.79
First generation other-Western	1 (2.0%)	0 (0%)	
Second generation other-Western	2 (4.1%)	3 (6.1%)	
First generation Suriname+Aruba+Dutch Antilles	3 (6.1%)	3 (6.1%)	
Second generation Suriname+Aruba+Dutch Antilles	1 (2.0%)	0 (0%)	
First generation other non-Western	4 (8.2%)	7 (14.3%)	
Antibiotic use in the prior 3 to 12 months			
Yes	6 (12.2%)	7 (14.3%)	0.77
No	43 (87.8%)	41 (83.7%)	
Do not know	0 (0%)	1 (2.0%)	

No differences between the ESBL$^+$ and ESBL$^-$ individuals in bacterial species composition or diversity parameters

We investigated potential differences in microbiota composition and diversity between ESBL$^+$ and ESBL$^-$ samples. A total of 1178 species (mOTUs) were detected in our cohort. Overall bacterial composition at the family and genus level are shown in Figure S1. The most abundant species and their average relative abundance in this cohort were *Bifidobacterium adolescentis* (4.6% ± 6.9%), *Ruminococcus bromii* (3.4% ± 4.8%)

undefined *Ruminococcaceae spp.* (2.9% ± 3.2%), *Eubacterium rectale* (2.7% ± 2.8%) and *Prevotella copri* (2.5% ± 5.7%). We did not observe differences in alpha diversity (observed mOTUs and Shannon index, Figure 1A and B), nor in beta diversity (PCoA and t-SNE, Figure 1C and D).

Next, we investigated whether there were differences in relative abundance between the study groups at the species level (mOTUs). Prior to differential abundance testing, mOTUs were filtered based on a prevalence of at least 25%, resulting in 261 mOTUs (representing 22.2% of the total observed mOTUs). No significant differentially abundant mOTUs were detected (all corrected p-values > 0.7). In order to elucidate whether microbiota composition is predictive of ESBL carriage, a machine learning classifier (LASSO logistic regression) was applied to the filtered mOTUs relative abundance matrix, which provided an AUROC value of approximately random classification (AUROC of 0.53, Figure 1E), indicating that mOTUs relative abundance does not allow for reliable prediction of ESBL status.

No differences in the resistome of individuals colonized by an ESBL-producing *E. coli* and ESBL$^-$ individuals

Of all cleaned reads, an average of 0.035% (±0.024%) reads per sample mapped against the MegaRes 2.0 database. There was no difference between ESBL groups in the average number of reads aligned to MegaRes 2.0 (independent t-test, p=0.84). A total of 98 unique antimicrobial resistance genes (ARGs) were detected with 17 different AMR mechanisms (e.g. beta-lactam), and the number of detected ARGs was not different between ESBL groups (independent t-test, p = 0.46) (Figure 2A). Overall ARGs profiles in the study groups assessed by plotting beta diversity, did not show a clear separation between ESBL groups (Figure 2B), which was confirmed by PERMANOVA (p=0.21). The most abundant ARGs and AMR mechanisms are visualized in Figure 2C and D. No differences in relative abundance of ARGs were found between the groups using differential abundance analysis (all corrected p-values > 0.4). Tetracycline resistance was most abundant in the resistomes (47.7% ± 24.7%, Figure 2C), followed by mupirocin resistance (33.7% ± 28.6%). Tetracycline resistance was conferred by several *tet* genes, while mupirocin resistance was conferred through the *ileS* gene. As it is known from literature that *Bifidobacterium* spp. can be intrinsically resistant to mupirocin through the *ileS* gene[42], we analyzed the correlation between the relative abundance of *Bifidobacterium* (at genus level) and the *ileS* gene, which was indeed high ($R=0.78$, $p<2.2x10^{16}$) (Figure S2). We then moved on to investigate functional profiles of our participants.

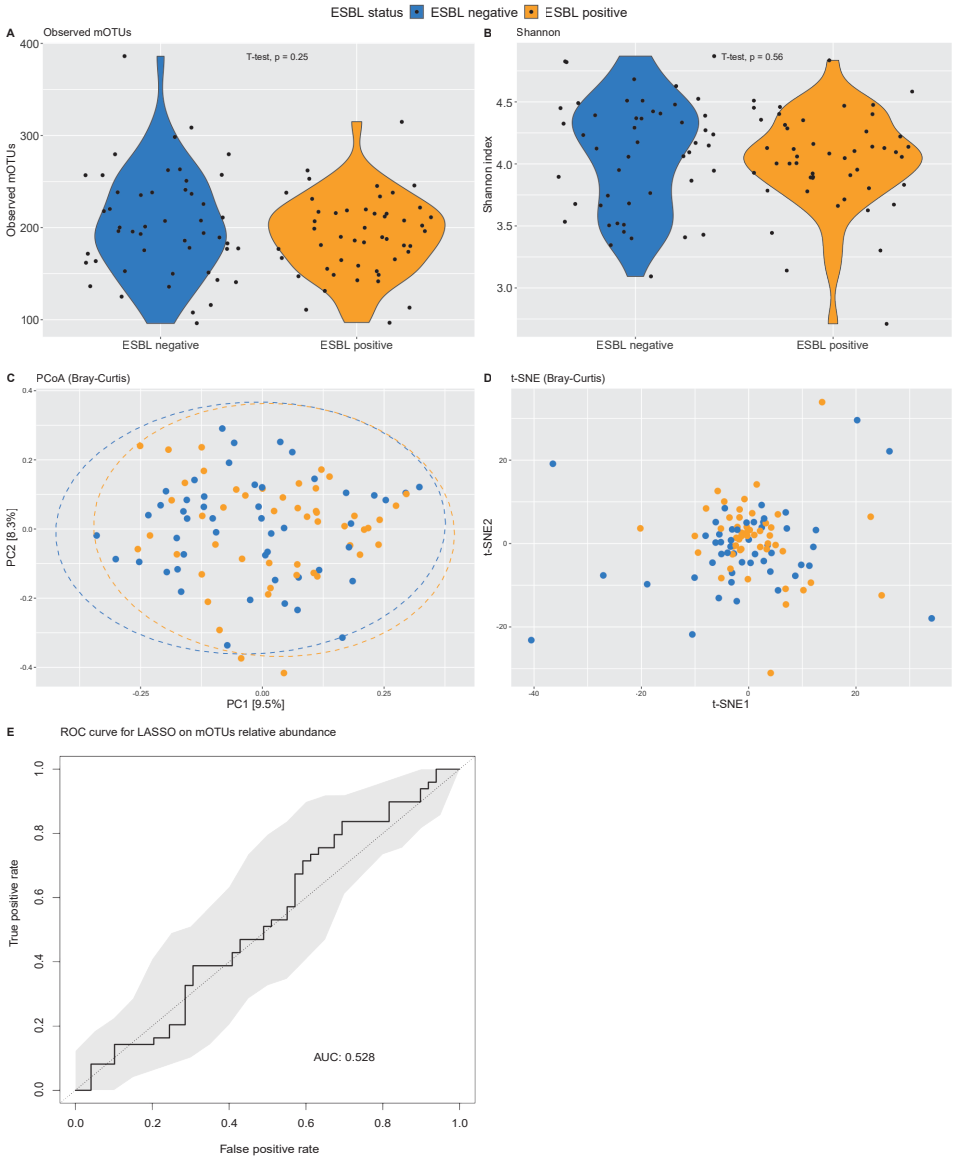


Figure 1: Taxonomic analyses between ESBL groups with comparisons of observed mOTUs (A) and Shannon index (B), unsupervised clustering using PCoA (C) and t-SNE (D) based on Bray-Curtis dissimilarity and the ROC curve for LASSO (E). The ROC curve shows the mean AUC value and its respective 95% CI.

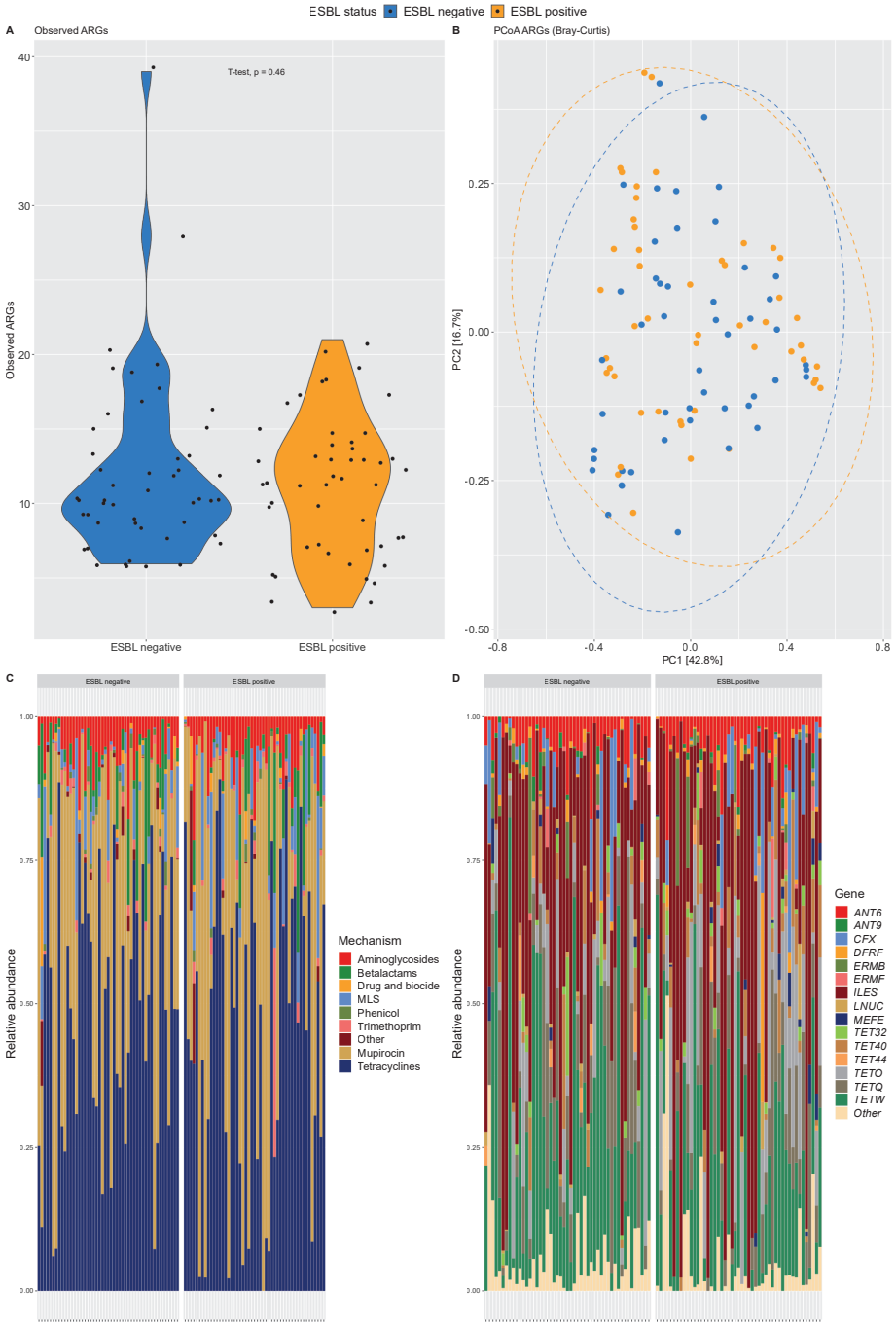


Figure 2: Resistome analyses with comparisons of the number of detected ARG (A), resistome diversity (B) and overviews of the most abundant resistance mechanisms (C) and resistance genes (D).

No differences between the ESBL$^+$ and ESBL$^-$ individuals in functional capacity of the microbiome

To compare the functionality of the gut microbiome between the study groups, cleaned reads were mapped against the annotated IGC database. On average, 95.8% (±1.7%) of reads aligned against the IGC, and the aligned number of reads was not different between ESBL groups (independent t-test, p=0.23). From the aligned reads, 49.2% (±2.2%) aligned against a gene annotated by a functional group (KO group) and this was not different between ESBL groups (independent t-test, p=0.13). There was no difference in overall functional profiles between the groups (PERMANOVA, p=0.19). 8450 KO groups were detected and after filtering on 25% prevalence, 5179 KO groups remained for differential abundance testing. No KO groups were significantly differentially abundant between ESBL groups (all corrected p-values > 0.2). To identify functional groups predictive of ESBL status, LASSO logistic regression was applied to the relative abundance matrix of KO groups. No accurate prediction model could be constructed (AUROC of 0.61), indicating that the functional groups do not contain information allowing for prediction of ESBL status.

No functional differences in Carbohydrate Active Enzymes (CAZymes) between the ESBL$^+$ and ESBL$^-$ group

From the aligned reads, 2.1% (±0.2%) aligned against a gene annotated to a CAZyme family and this was not different between ESBL groups (independent t-test, p=0.48). A total of 109 CAZyme families were detected with a mean of 77.7 (±5.7) per individual, with no differences between ESBL groups (independent t-test, p=0.34) (Figure 3A). The three most abundant CAZymes in our study were glycoside hydrolase (GH)13 (19.4% ± 3.3%), GH3 (11.4% ± 1.6%) and GH31 (6.2% ± 0.9%) (Figure 3C), corresponding to breakdown of starch and glycogen (GH13) and breakdown of plant cell wall glycans (GH3 and GH31)[43]. Variation in CAZyme relative abundance profiles could not be explained by ESBL group (PERMANOVA, p=0.57, Fig 3B). Compositional plots based on the top 20 most abundant CAZymes were highly similar between the ESBL groups (Figure 3C), and no differences in relative abundance of individual CAZyme families was observed (all corrected p-values > 0.6). To identify potential drivers of ESBL-producing *E. coli* colonization we used LASSO logistic regression on relative abundances CAZymes, which did not result in an accurate prediction model (AUROC of 0.56). This indicates there is only very low to no predictive power in relative abundances of CAZymes with regard to ESBL status.

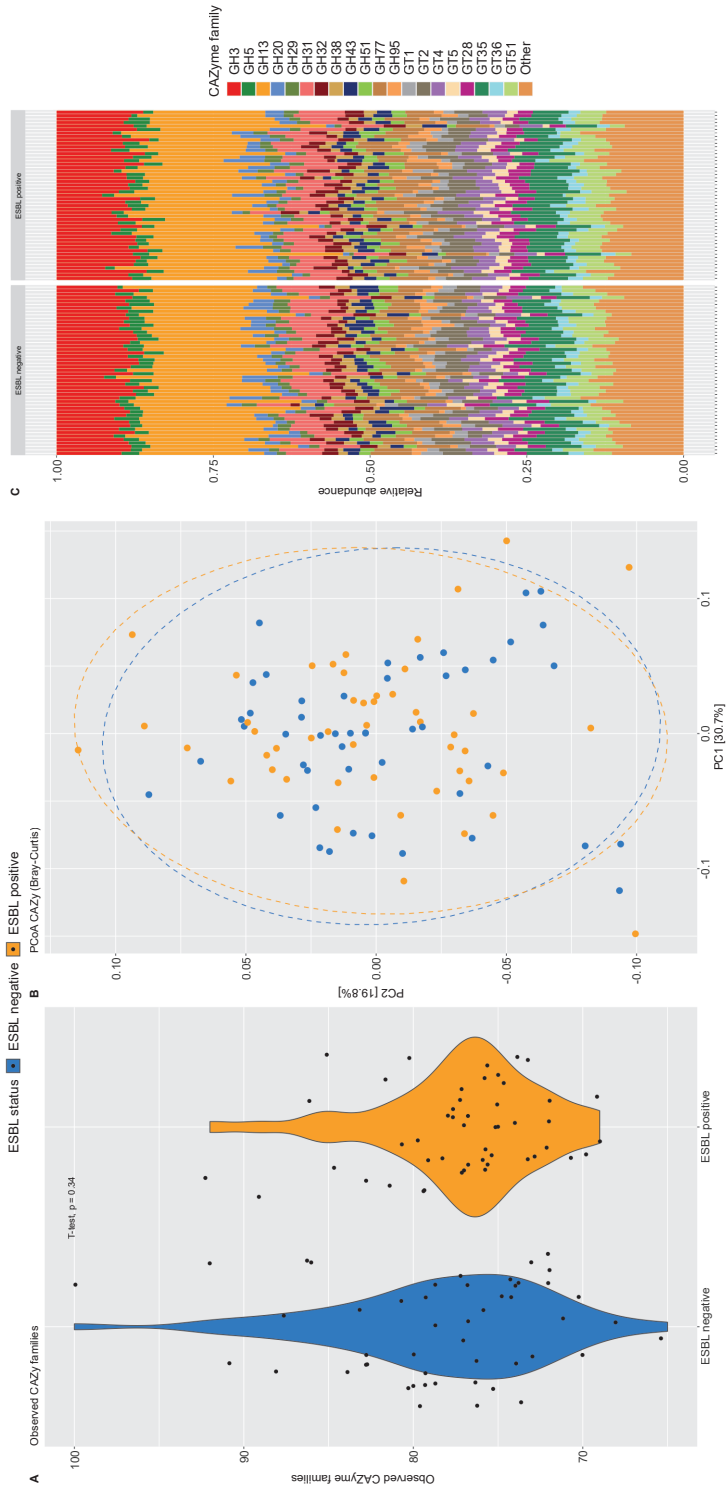


Figure 3: Overview of analyses based on CAZyme repertoire with a comparison of number of CAZyme families (A), PCoA based on Bray-Curtis dissimilarity (B) and a compositional plot to show the consistency of CAZyme families across participants. GH: Glycoside hydrolase, GT: glycosyl transferase.

Metabolomics profiling shows no clear differences between ESBL groups at the functional level

For metabolomic analysis we quantified metabolite concentrations in all individuals, except for one ESBL$^+$ sample that was excluded as a good quality NMR spectrum could not be recorded due to shimming problems. First, to investigate whether any differences in metabolite concentrations existed between ESBL groups, we performed univariate testing (independent t-tests). These results strongly depended on the method used for multi-error correction (11 metabolites were significantly different at p=0.048 with Benjamini-Hochberg, but none with Holm) (Figure S3 and Figure S4).

Unsupervised dimensionality reduction using PCA was performed to investigate whether any separation could be observed based on ESBL carriage (Fig 4A). Over 46% of the metabolome variation could be explain on the first principal component, with some separation of the study groups. However, supervised analysis using a PLS-DA indicates that no predictive value could be obtained for class separation based on two PLS components (Q2Y = -0.06). Lastly, we performed a random forest prediction model to investigate whether ESBL status could be predicted based on metabolite profiles, but this was not the case (AUROC = 0.61) (Figure 4B). Altogether, minor differences in metabolite concentrations could be detected using t-tests, but these were dependent on the method applied for correction for multiple testing. PCA between the ESBL groups showed a small overall signal, but no predictive value could be confirmed by both PLS-DA and random forest.

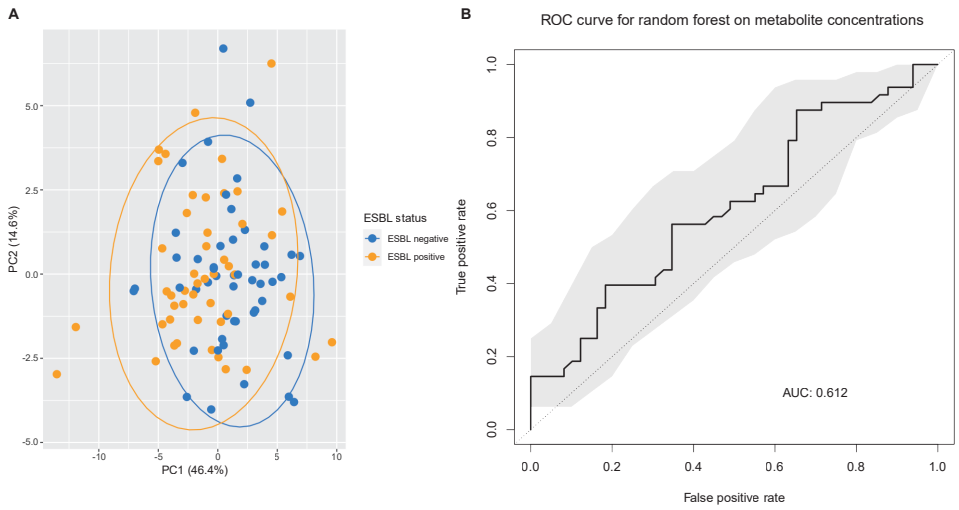


Figure 4: Metabolomic analyses with PCA (A) and the ROC curve of random forest based on metabolite profiles (B). The ROC curve shows the mean AUC value and its respective 95% CI.

Discussion

We present a unique study investigating differences in the gut microbiome and metabolome between individuals asymptomatically colonized by an ESBL-producing *E. coli* and matched non-colonized individuals. Importantly, in contrast to previous studies on this topic, we applied stringent inclusion criteria and matched $ESBL^+$ individuals with $ESBL^-$ individuals on important epidemiological variables, which minimized the chance for observing effects which could be attributed to confounding variables. The combination of metagenomics and metabolomics allowed for a deep molecular resolution of the gut microbiome, both at the taxonomic and functional level. We show that there is no difference in the gut microbiome of individuals asymptomatically colonized with an ESBL-producing *E. coli* as compared to individuals who are not colonized.

Confounding factors may, at least partially, be the reason for the previously reported differences in microbial signatures associated with protection from asymptomatic colonization by ESBL-producing bacteria and MDROs across different studies. It must be noted that these studies have mostly investigated vulnerable patient populations, such as nursing home residents and hospitalized patients. In such populations it is very complex to disentangle observed differences between colonized and non-colonized individuals from differences due to confounding variables (such as comorbidities and medication) between compared individuals[6, 8, 44-46]. In our study we excluded individuals based on many microbiome-influencing clinical factors, and performed matching on several clinical variables, as recently recommended for cross-sectional microbiome studies[9]. In this way, we could study the effect of colonization of ESBL-producing *E. coli* in isolation and convincingly show that no differences exist in the gut microbiome between colonized and non-colonized individuals.

In addition, previous research has generally not focused on species-specific colonization resistance, but rather on a broad category of MDROs (such as ESBL-producing *Enterobacterales*)[6, 8, 44-46]. Given the large genomic diversity within species[47], let alone within the order of *Enterobacterales*, it is highly unlikely that a common mechanisms exists which could prevent colonization of e.g. both ESBL-producing *Klebsiella pneumoniae* and ESBL-producing *E. coli*. Therefore in the current study we focused on a single species (*E. coli*), rather than a broad group of ESBL-producing *Enterobacterales*. Microbiome composition of individuals in our study population reflects that of other population cohorts in general. For example, *B. adolescentis* has been previously described in another Dutch cohort as the most abundant bacterial species, with an average relative abundance of 9.51% (±10.8%)[48]. In addition, *P. copri*, *R. bromii* and *E. rectale* were also highly abundant and prevalent, in line with the findings in the current study[48].

The resistome profiles identified in our study also corresponded well with what is generally described in literature, with tetracyline resistance being the most abundant resistance mechanism in the human gut[49-51]. The observed high relative abundance to mupirocin in our study could be explained by the intrinsic resistance of *Bifidobacterium spp*. to this, of which relatively high abundances were observed in this cohort.

We show that despite inter-individual variation in taxonomic profiles, the functionality of the microbiome as assessed by the relative abundance of CAZyme families, is highly consistent between individuals. These finding are in line with previous findings showing functional similarity at the metabolic level despite taxonomic diversity[52, 53].

This study is, to our knowledge, the first study to profile the gut metabolome in relation to colonization of ESBL-producing *E. coli*. We did not observe a relation between the metabolome, or any specific metabolite, and ESBL status. For other enteric pathogens, like *Salmonella enterica* serovar Typhimurium and *C. difficile*, specific metabolites have been shown to be strongly related to colonization resistance in rodent models[54, 55]. It should however be mentioned that these are infection models rather than asymptomatic colonization models, which would better represent our study.

A limitation of our study is that we do not have longitudinal data on the microbiome of these participants, and are therefore unable to make any statements about the duration of colonization of ESBL-producing *E. coli* and associations with the gut microbiome in time. This is particularly relevant considering the large variation in the duration of colonization between individuals[56, 57]. It could be speculated that individuals who are long-term colonized have a different gut microbiome than individuals who are only colonized for a short period of time, although there is no clear evidence for this in literature to our knowledge. Furthermore, longitudinal observations would allow us to identify changes occurring at the compositional and functional level when asymptomatic carriage turns into active infection or when people become decolonized. Lastly, one would have ideally have microbiome data of an individual shortly before an ESBL-producing *E. coli* would colonize and at time of colonization, so that microbiome changes within an individual can be investigated. Secondly, we do not have whole-genome sequencing data of the ESBL-producing *E. coli* isolates, which prevents us from placing these data into a broader epidemiological context. For example, if the majority of isolates would be sequence type (ST)131, an endemic ST, this would be valuable extra information and further extend the clinical relevance of our findings.

This study is however unique in the fact that $ESBL^+$ and $ESBL^-$ individuals were selected from a large Dutch cohort (n= 2751), and therefore we could apply stringent inclusion criteria and match the two groups on several demographic and clinical variables. To the

best of our knowledge, this is one of very few studies in the microbiome field that applied such a stringent study setup. This setup ensured that the potential effect of confounding factors was minimized. In addition, this study is the first to investigate differences in the gut microbiome and metabolome between individuals colonized by an ESBL-producing *E. coli* and non-colonized individuals using a combined approach of metagenomics and metabolomics. Therefore, it provides insight into both the composition and function of the gut microbiome.

Conclusions

Our study shows that there are no differences in the gut microbiome or metabolome of individuals who are, or are not, asymptomatically colonized by an ESBL-producing *E. coli*. We hypothesize that microbiome-mediated colonization resistance may therefore not be as relevant against ESBL-producing *E. coli* as it is for other enteric pathogens (like *C. difficile* and vancomycin-resistant *Enterococcus*), although longitudinal studies or controlled human colonization models are necessary to confirm this hypothesis.

Notes

Acknowledgements
We gratefully acknowledge all individuals who participated in the PIENTER-3 study.

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. EK is supported by an unrestricted grant from Vedanta Biosciences Inc. EK has performed research for Cubist, Novartis, and Qiagen and has participated in advisory forums of Astellas, Optimer, Actelion, Pfizer, Sanofi Pasteur, and Seres Therapeutics. The companies had no role in the study or writing of the manuscript.

Author contributions
QD, RZ and EK conceived and designed the study. RZ, SF and EK supervised the study. RW and CV-G performed the study on ESBL-producing bacteria in the open population in The Netherlands and analyzed all patient-related data. AV and MG performed metabolomics analysis and AV aided in statistical analysis of metabolomics. FK and EF coordinated sample collection. JK performed DNA extraction and related laboratory procedures. QD performed sample selection and analyzed epidemiological data, processed and analyzed metagenomic data, performed statistical analysis of

metabolomics data, created figures and wrote the manuscript. SN aided in metagenomic data processing and analysis. QD, RZ, RW, AV, SN, FK, EF, JK, AV, MG, CV, SF and EK discussed results and implications. All authors contributed to and approved the manuscript.

Code and data availability

All raw metagenomic data will be released under PRJEB44119 upon acceptance of the manuscript. All R code and necessary data files to reproduce the analyses and figures will be uploaded to GitHub upon acceptance of the manuscript.

References

1. Croxen MA, Law RJ, Scholz R, Keeney KM, Wlodarska M, Finlay BB. Recent advances in understanding enteric pathogenic Escherichia coli. Clin Microbiol Rev. 2013;26(4):822-80.
2. Pitout JD, Laupland KB. Extended-spectrum beta-lactamase-producing Enterobacteriaceae: an emerging public-health concern. Lancet Infect Dis. 2008;8(3):159-66.
3. Cassone M, Mody L. Colonization with Multi-Drug Resistant Organisms in Nursing Homes: Scope, Importance, and Management. Curr Geriatr Rep. 2015;4(1):87-95.
4. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev. 2019;83(3).
5. Huang YS, Lai LC, Chen YA, Lin KY, Chou YH, Chen HC, et al. Colonization With Multidrug-Resistant Organisms Among Healthy Adults in the Community Setting: Prevalence, Risk Factors, and Composition of Gut Microbiome. Front Microbiol. 2020;11:1402.
6. Le Bastard Q, Chapelet G, Birgand G, Hillmann BM, Javaudin F, Hayatgheib N, et al. Gut microbiome signatures of nursing home residents carrying Enterobacteria producing extended-spectrum β-lactamases. Antimicrob Resist Infect Control. 2020;9(1):107.
7. Leo S, Lazarevic V, Girard M, Gaïa N, Schrenzel J, de Lastours V, et al. Metagenomic Characterization of Gut Microbiota of Carriers of Extended-Spectrum Beta-Lactamase or Carbapenemase-Producing Enterobacteriaceae Following Treatment with Oral Antibiotics and Fecal Microbiota Transplantation: Results from a Multicenter Randomized Trial. Microorganisms. 2020;8(6).
8. Piewngam P, Quiñones M, Thirakittiwatthana W, Yungyuen T, Otto M, Kiratisin P. Composition of the intestinal microbiota in extended-spectrum β-lactamase-producing Enterobacteriaceae carriers and non-carriers in Thailand. Int J Antimicrob Agents. 2019;53(4):435-41.
9. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020.
10. Verberk JDM, Vos RA, Mollema L, van Vliet J, van Weert JWM, de Melker HE, et al. Third national biobank for population-based seroprevalence studies in the Netherlands, including the Caribbean Netherlands. BMC Infect Dis. 2019;19(1):470.
11. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters. Version 11.0, 2021. http://www.eucast.org/clinical_breakpoints/2021.
12. Reuland EA, Al Naiemi N, Kaiser AM, Heck M, Kluytmans JA, Savelkoul PH, et al. Prevalence and risk factors for carriage of ESBL-producing Enterobacteriaceae in Amsterdam. J Antimicrob Chemother. 2016;71(4):1076-82.
13. EUCAST. EUCAST guidelines for detection of resistance mechanisms and specific resistances of clinical and/or epidemiological importance. https://eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Resistance_mechanisms/EUCAST_detection_of_resistance_mechanisms_v1.0_20131211.pdf2013.
14. Coelho LP, Alves R, Monteiro P, Huerta-Cepas J, Freitas AT, Bork P. NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language. Microbiome. 2019;7(1):84.
15. Kultima JR, Coelho LP, Forslund K, Huerta-Cepas J, Li SS, Driessen M, et al. MOCAT2: a metagenomic assembly, annotation and profiling framework. Bioinformatics (Oxford, England). 2016;32(16):2520-3.
16. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, et al. MOCAT: a metagenomics assembly and gene prediction toolkit. PloS one. 2012;7(10):e47656-e.
17. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013;preprint arXiv:1303.3997.
18. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32(8):834-41.
19. Milanese A, Mende DR, Paoli L, Salazar G, Ruscheweyh H-J, Cuenca M, et al. Microbial abundance, activity and population genomic profiling with mOTUs2. Nat Commun. 2019;10(1):1014-.
20. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115-22.
21. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490-5.
22. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and

protein annotation. Nucleic Acids Res. 2016;44(D1):D457-62.
23. Zolfo M, Tett A, Jousson O, Donati C, Segata N. MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples. Nucleic Acids Res. 2017;45(2):e7.
24. Doan T, Worden L, Hinterwirth A, Arzika AM, Maliki R, Abdou A, et al. Macrolide and Nonmacrolide Resistance with Mass Azithromycin Distribution. N Engl J Med. 2020;383(20):1941-50.
25. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, et al. Impact of sequencing depth on the characterization of the microbiome and resistome. Sci Rep. 2018;8(1):5890.
26. Kim HK, Kostidis S, Choi YH. NMR Analysis of Fecal Samples. In: Giera M, editor. Clinical Metabolomics: Methods and Protocols. New York, NY: Springer New York; 2018. p. 317-28.
27. Findeisen M, Brand T, Berger S. A 1H-NMR thermometer suitable for cryoprobes. 2007;45(2):175-8.
28. Kumar A, Ernst RR, Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochemical and Biophysical Research Communications. 1980;95(1):1-6.
29. Price WS. Water Signal Suppression in NMR Spectroscopy. In: Webb GA, editor. Annual Reports on NMR Spectroscopy. 38: Academic Press; 1999. p. 289-354.
30. Wu PSC, Otting G. Rapid pulse length determination in high-resolution NMR. Journal of Magnetic Resonance. 2005;176(1):115-9.
31. Verhoeven A, Giera M, Mayboroda OA. KIMBLE: A versatile visual NMR metabolomics workbench in KNIME. Analytica Chimica Acta. 2018;1044:66-76.
32. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217.
33. Lahti L, Shetty S. Tools for microbiome analysis in R. Microbiome package version 1.12.0. 2021.
34. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4(43):1686.
35. Rich B. table1: Tables of Descriptive Statistics in HTML. R package version 1.2.1. https://CRAN.R-project.org/package=table1. 2020.
36. Oksanen J, Blanchet GF, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan. 2020.
37. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol. 2021;22(1):93.
38. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. J Proteome Res. 2015;14(8):3322-35.
39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
40. Vojinovic D, Radjabzadeh D, Kurilshikov A, Amin N, Wijmenga C, Franke L, et al. Relationship between gut microbiota and circulating metabolites in population-based cohorts. Nat Commun. 2019;10(1):5813.
41. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics. 1979;6(2):65-70.
42. Serafini F, Bottacini F, Viappiani A, Baruffini E, Turroni F, Foroni E, et al. Insights into physiological and genetic mupirocin susceptibility in bifidobacteria. Appl Environ Microbiol. 2011;77(9):3141-6.
43. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol. 2013;11(7):497-504.
44. Annavajhala MK, Gomez-Simmonds A, Macesic N, Sullivan SB, Kress A, Khan SD, et al. Colonizing multidrug-resistant bacteria and the longitudinal evolution of the intestinal microbiome after liver transplantation. Nat Commun. 2019;10(1):4715.
45. 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 Med. 2021;13(1):54.
46. Araos R, Tai AK, Snyder GM, Blaser MJ, D'Agata EMC. Predominance of Lactobacillus spp. Among Patients Who Do Not Acquire Multidrug-Resistant Organisms. Clin Infect Dis. 2016;63(7):937-43.
47. Van Rossum T, Ferretti P, Maistrenko OM, Bork P. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol. 2020;18(9):491-506.
48. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285):565-9.

49. Carr VR, Witherden EA, Lee S, Shoaie S, Mullany P, Proctor GB, et al. Abundance and diversity of resistomes differ between healthy human oral cavities and gut. Nat Commun. 2020;11(1):693.
50. Hu Y, Yang X, Qin J, Lu N, Cheng G, Wu N, et al. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. Nat Commun. 2013;4:2151.
51. van Schaik W. The human gut resistome. Philos Trans R Soc Lond B Biol Sci. 2015;370(1670):20140087.
52. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14.
53. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun. 2019;10(1):4505.
54. Jacobson A, Lam L, Rajendram M, Tamburini F, Honeycutt J, Pham T, et al. A Gut Commensal-Produced Metabolite Mediates Colonization Resistance to Salmonella Infection. Cell Host Microbe. 2018;24(2):296-307. e7.
55. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun. 2014;5:3114.
56. Overdevest I, Haverkate M, Veenemans J, Hendriks Y, Verhulst C, Mulders A, et al. Prolonged colonisation with Escherichia coli O25:ST131 versus other extended-spectrum beta-lactamase-producing E. coli in a long-term care facility with high endemic level of rectal colonisation, the Netherlands, 2013 to 2014. Euro Surveill. 2016;21(42).
57. Arcilla MS, van Hattem JM, Haverkate MR, Bootsma MCJ, van Genderen PJJ, Goorhuis A, et al. Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study. Lancet Infect Dis. 2017;17(1):78-85.

Chapter 10

General discussion and future perspectives

Enteric colonization with multi-drug resistant organisms (MDROs) or enteropathogenic bacteria such as *Clostridioides difficile* can precede development of an infection and is considered an important public health concern. It has long been suspected that the gut microbiome can provide protection against these potentially pathogenic bacteria and research into this field increased considerably in the last 10-15 years. The gut microbiome is able to confer resistance against colonization and infection by endogenous and exogenous microorganisms through a myriad of mechanisms, including nutrient competition, secretion of antimicrobial compounds such as short-chain fatty acids and bacteriocins, maintaining gut barrier integrity and its interaction with the host immune system. It is hypothesized that in the case of enteric colonization or infection, there is a lack of microbiome-mediated colonization resistance against the potentially pathogenic microorganism. Therefore, appropriate restoration of colonization resistance may prevent colonization or contribute to the eradication, before an infection can develop. The latter strategy may be especially valuable for vulnerable patient populations such as nursing home residents, stem cell transplant patients and patients admitted to the intensive care unit. The work in this thesis largely aimed at identifying gut bacteria involved in conferring microbiome-mediated colonization resistance against enteropathogens and MDROs. In addition, this thesis contributes to several technical challenges that the microbiome research field is currently facing, namely standardization of wet-lab and dry-lab procedures for clinical microbiome studies and development of novel computational tools for functional microbiome profiling. The work in this thesis is primarily computational and over the years we progressed from using 16S rRNA gene amplicon sequencing for microbiota profiling to studies using metagenomic sequencing and metabolomics, which present more challenges in data analysis. By building on the experience and knowledge gained in the first 2.5 years of this PhD, we designed and performed a very stringent cross-sectional study and combined multi-omics with machine learning approaches.

Technical challenges in the era of multi-omics

Research in the fields of (bio)medicine and biology is rapidly changing and is becoming increasingly quantitative of nature, with large and complex high-dimensional data being commonly used[1, 2]. This can especially be attributed to the arrival of the many -omics techniques which allow for deep resolution at the molecular level (DNA, RNA, proteins and metabolites) and thereby make a systems biology approach feasible[3]. These technical advancements are accompanied by decreasing costs for such measurements, which makes obtaining these large data sets easier, cheaper and more common[4]. However, the scientific community, including journals, funding bodies, education and software/ tool development do not always develop at the same pace, while this is imperative

for optimal usage of obtained data. In the sections below I will discuss the current challenges and opportunities of these technical issues.

The solution of the reproducibility crisis: findable, accessible, interoperable and reusable (FAIR) data

One of the main challenges in recent scientific research is the so-called "reproducibility crisis", which means that many studies cannot be reproduced, including studies in the gut microbiome research field[5-7]. One relatively simple approach that can help to resolve this crisis is that data reported in manuscripts should be FAIR[8]. One of the current main issues with regard to FAIR data is that not all study-related data are made available for the scientific community by researchers[9]. While this is understandable, as a lot of effort, time and money can be involved with collecting data from a large cohort, not sharing all data used for analysis can impede scientific advancement and hinders reproducibility of results[10]. Not all journals

are implementing stricter rules and guidelines for sharing microbiome data and it remains frequently stated that 'raw sequence data is available on request' or that the data cannot be shared due to potential privacy issues. The underlying explanation is that human reads are present in fecal metagenomes, which can in theory lead to identification of an individual[11]. The latter issue can be easily tackled by filtering out human reads prior to uploading the data to a central archive such as the European Nucleotide Archive (ENA). In this way, privacy would no longer be an argument for not sharing sequence data. However, the 'privacy issue' may actually be representative of an underlying aversion to data sharing. This could be one possible explanation for the fact that of manuscripts published in Nature and Science in which is stated that data is available upon request, only in less than half of the cases the data can actually be obtained[9]. This study by Tedersoo and colleagues evaluated data availability in 875 articles published between 2000 and 2019 and they contacted authors of 310 papers to investigate if data could be obtained[9]. An encouraging finding of this paper was that a yearly decay rate of 5.9% was found with respect to the 'data available upon request' statement, which implicates that data sharing is becoming more common. Unfortunately, sharing of metagenomic data in the microbiology field was found to be an exception to this trend, as its public availability has decreased over the past years[9]. The reason for keeping data within a research group may be to ensure a consistent stream of (high-impact) publications. The aforementioned issue of data being 'available upon request' is in general a way of complying with journal policies while not always having the intention of actually sharing raw data[9]. It should be stressed that this does not apply to all researchers with such a statement in their paper, as many of them are willing to share data without any specific requirement.

For the research conducted in this thesis, we have always made raw sequence data and associated metadata publicly available (**Chapter 6 – Chapter 9**). For several

manuscripts, we have also made all the applied statistical code and other necessary data files available (**Chapter 7** and **Chapter 8**), which further improves the reproducibility and transparency of our work. In view of this, it is a good development that journals are also changing their policies and require sharing and publishing of raw data and associated statistical code upon publication[12]. For example, a journal can instruct authors to make references to both the location of raw sequence data and to the location of all applied code, ensuring reproducibility and findability. Such a journal policy was also encountered in one of the studies in this thesis (**Chapter 8**). However, it is important that journals, similar to the funding bodies as discussed below, also control whether data is actually uploaded to public repositories in a FAIR manner. In any case, I hope that more journals will follow the encouraging trend of data and code sharing, which will allow the microbiome field to advance more quickly.

Improved sequence and metadata sharing will allow researchers to conduct higher-quality meta-analyses, something that has proven to be crucial for establishing robust disease-specific microbial signatures. For example, a large gut microbiome meta-analysis has led to identification of *Fusobacterium nucleatum* as more prevalent and abundant in colorectal cancer patients than in healthy controls, which was largely independent of geography or technical variation[13]. Also in this thesis (**Chapter 5**) we have been able to profit from publicly available data, as we re-analyzed metagenomic data from eight cross-sectional studies comparing gut microbiomes from colorectal cancer patients with controls. Also for **Chapter 6,** several studies have been published in the meantime that would allow for verification of our results and potentially a meta-analysis, but these were unfortunately not published at the time of publication[14, 15]. When consistent results are obtained from several cohort studies, this increases the degree of confidence of findings, which can serve as an incentive, foundation and guide for conducting follow-up experiments.

To tackle the aforementioned issues with regard to data sharing, there may be an important role for both funding agencies and journals/editors to enforce stricter rules, as outlined in editorials of the journals Nature in 2017 and Microbiome in 2018[11, 16]. For example, funding agencies may include a requirement to make all generated data publicly available immediately after publication. It is encouraging in this respect to see that one of largest Dutch funders in medical research, ZonMw, has adopted the FAIR guidelines for their funding calls and they strongly support open science[17]. They have even released a call in 2019 termed ‘Tackling antibiotic resistance by reusing data and increasing FAIRness’, where the goal was to get a better grip on tackling antibiotic resistance by reusing existing data resources[18]. Importantly, funding agencies will also have to dedicate resources to ensure compliance with data sharing and to provide technical support, as a mere data sharing requirement has shown to be insufficient due

to limited compliance[19]. A similar role is reserved for journal editors with regard to ensuring that authors comply with data sharing instructions, even though this may be a time-consuming effort. While time-consuming, it has been found that studies with publicly shared data get more citations on their work, an important metric for journals[20, 21].

The 'modern' biologist

The omnipresent availability of large data sets obtained from -omics technologies can pose researchers without a quantitative or computational background with a lot of challenges. In light of these developments, extensive collaboration between wet-lab and dry-lab researchers is becoming more important to obtain a detailed understanding of the generated data, which is also increasingly becoming clear in the microbiome field[22, 23]. It remains of crucial importance that researchers in biomedical research fields have some basic understanding of computational data processing for properly interpreting and judging -omics data described in scientific papers. This is not a one-way street though, as computational scientists also need to have at least a basic understanding of applied experimental methods for making optimal use of the data[24]. However, as biomedical research programs (BSc/MSc) at this point mostly offer traditional biomedical courses like physiology, genetics and biochemistry, during which experimental methods and wet-lab experiments are often part of the curriculum, it is necessary to incorporate more in-depth data science and statistics courses that focus on analyses often performed in -omics studies, such as principal component analysis[25-27].

Functional characterization of the microbiome

Microbiome researchers have, thus far, mostly investigated taxonomic profiles of microbial communities, e.g. through 16S rRNA gene amplicon sequencing and shotgun metagenomics. However, this has a large disadvantage of not obtaining information about the functional capacity of the microbial community, which is likely much more relevant than taxonomy when investigating the microbiome's relation with health and disease[28, 29]. For example, several studies have shown that the functional repertoire encoded in metagenomes displays much higher sensitivity to perturbations than the taxonomic profile[30-32]. One of several reasons for the lack of functional analyses in the gut microbiome field is the current scarcity of good and easy-to-use tools for this purpose[29]. With the decreasing costs of sequencing techniques[4], it becomes particularly important to have reliable and informative computational tools to determine the functional capacity of the microbiome. To the best of my knowledge, the first (and so far only) manual curation of functional information for the gut microbiome was performed on Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations by the Raes group, which were grouped into so-called gut metabolic modules[33]. These curations mainly focused

on microbial metabolism of carbohydrates, amino acids and lipids and are useful for obtaining better understanding of the utilization of these three macronutrients by the gut microbiome[33]. However, for these gut metabolic modules to be more readily applicable for (gut) microbiome research, modules with other metabolic pathways should be added, e.g. pathways involved in bile acid and short-chain fatty acid metabolism.

In this thesis we systematically curated known carbohydrate-active enzymes (CAZymes) involved in glycan breakdown and grouped these into several functional categories (**Chapter 5**). Such an annotation is not only much more intuitive and easier to interpret than a list of CAZyme families, but it can also reduce the number of features when analyzed at a different level. As the number of features (CAZymes in this case) is collapsed into approximately 100x fewer functional categories, this lowers the chance for finding false negatives due to multi-error correction. While multi-error correction may not be a major problem for CAZymes yet, since approximately ~700 CAZyme (sub) families are currently known, functional annotation using KEGG families can easily provide researchers with thousands of features and lists of genes from metagenomes can even result into millions of features per sample. Importantly, obtaining substrate information through CAZyme annotation allows for analyses that are not common in the microbiome field, but have proven their value in the transcriptomics field by detecting up or downregulated pathways in e.g. different types of leukemia, amongst many other discoveries[34-36]. An example of such an analysis technique is gene set enrichment analysis (GSEA), which can also be applied to the microbiome field (**Chapter 5**)[34]. GSEA mainly derives its strength from grouping genes into informative functional pathways (or in the case of CAZymes, into substrates). This has several major advantages over testing single features. For example, when testing individual genes there may be no significant differentially abundant genes, but all genes may show a trend towards a specific effect. When grouping these genes into functional pathways, a very significant effect at the pathway level may become apparent, which is in any case much more informative than gene-by-gene testing[34, 37]. The opposite may also happen, where a researcher observes many significantly differentially expressed genes, challenging the identification of commonalities across all the genes and only experts in the field may be capable of identifying and interpreting these correctly at the functional level. While GSEA(-like) tools and manually curated annotations are almost non-existent in the microbiome field, I am convinced that this is the future of functional metagenomics and will allow for a much more detailed understanding of the functional capacity of microbial communities and its relation to health and disease.

In recent years, other -omics techniques such as metabolomics and, to a lesser extent, metatranscriptomics and metaproteomics have been used to get information on the actual activity of the microbiome (**Chapter 1**)[28, 38, 39]. We had the opportunity to perform

metabolomics in our study investigating the associations between enteric colonization of ESBL-producing *Escherichia coli* and the microbiome and metabolome (**Chapter 9**). While in this study the metabolomics data confirmed the results of the metagenomics (no differences between the groups), we previously experienced in a collaborative project on the effects of the herbicide glyphosate on the gut microbiome, that metabolomics can be very powerful to detect functional differences between groups at the level of metabolic pathways and may be more sensitive than metagenomics[40]. As we did not detect any differences between individuals colonized with ESBL-producing *E. coli* and non-colonized individuals with either metagenomics or metabolomics in **Chapter 9**, we did not take the next step of integrating both data sets, which could have provided more insight in case differences would have been found[41]. The importance of metabolomics for studying functionality of the microbiome is also reflected by the fact that multiple research groups have built tools that allow for prediction of metabolite concentrations based on metagenomic sequence data[42, 43]. With regard to metatranscriptomics and metaproteomics, we unfortunately did not apply these techniques during the research described in this thesis. These techniques have already been successfully used to increase understanding of microbial alterations in the gut in Type I Diabetes Mellitus (T1DM) patients on taxonomic and functional level[44] and to reveal differences in transcriptional activity in microbes over time in IBD patients[45]. Nevertheless, significant challenges remain for both techniques to be overcome before they can be more widely implemented in microbiome research in the coming years[38].

Microbiome-mediated colonization resistance

Microbiome-mediated colonization resistance is a relevant topic from both a therapeutic and evolutionary point of view, and its importance for health and disease is already recognized for many decades[46-49]. From a therapeutic perspective, obtaining a detailed understanding of microbiome-mediated colonization resistance against a microorganism can pave the way for targeted restoration of colonization resistance to eradicate or prevent colonization by this microorganism. With regard to the more evolutionary aspects, it is imperative for our understanding of microbial community dynamics in general to define what is necessary for a community to outcompete a specific microorganism. In this thesis we provided an extensive overview of the mechanisms through which the microbiome can confer colonization resistance (**Chapter 2**). A very recent, breakthrough finding that builds upon the concept of colonization resistance is that the resistance conferred by the gut microbiome to an infection can be strengthened by having contracted a prior infection[50]. It was shown that an initial infection of *Yersinia pseudotuberculosis* (a food-borne pathogen) could protect against subsequent colonization and infection by other potential pathogens, like *Klebsiella pneumoniae*. The initial infection rewired the

microbiome such that taurine, an intermediate in bile acid metabolism, was produced in higher amounts. Taurine was subsequently converted into antimicrobial sulfide, which in turn inhibited *K. pneumoniae* respiration by limiting access to oxygen through inhibition of enzymes involved in aerobic electron transport chains[50]. This effect could, importantly, be replicated in a mouse model infected with enteropathogenic *Citrobacter rodentium*, providing further support for the hypothesis that taurine can enhance colonization resistance by restricting pathogen respiration. It should however be noted that the gut microbiome-mediated effect could not be fully uncoupled from the host immune response[50].

In this thesis we aimed to identify bacteria that could play a role in providing colonization resistance against enteropathogens or MDROs (**Chapter 6 – Chapter 9**) (Figure 1 Step 1). We identified several bacterial taxa (e.g. *Fusicatenibacter* and *Eubacterium hallii*) associated with protection against *C. difficile* colonization (**Chapter 6**) and several bacterial taxa (*Dorea*, *Atopobiaceae* and *Lachnospiraceae ND3007 group*) associated with protection against MDRO colonization. However, these findings were not followed-up with more targeted experiments (Figure 1, Step 2 and Step 3), which leaves the question whether these bacteria actually play a causative role unanswered. In sharp contrast, we did not find associations pointing towards a role for the microbiome in mediating protection against infection caused by *N. americanus* (**Chapter 7**) or against asymptomatic colonization of ESBL-producing *E. coli* (**Chapter 9**).

With the successful implementation of fecal microbiota transplantation (FMT) for treating recurrent *Clostridioides difficile* infection (rCDI) almost a decade ago, it was expected that restoring colonization resistance through defined microbiome-based therapeutics would soon replace FMT as a treatment for rCDI[52]. Despite the extensive research that has been performed since, this has not yet led to widespread microbiome-based therapeutic intervention options (as is the case for the microbiome field in general)[53]. In my opinion this can mainly be attributed to study design and subsequent computational analyses issues and to the lack of follow-up wet-lab experiments and mechanistic research.

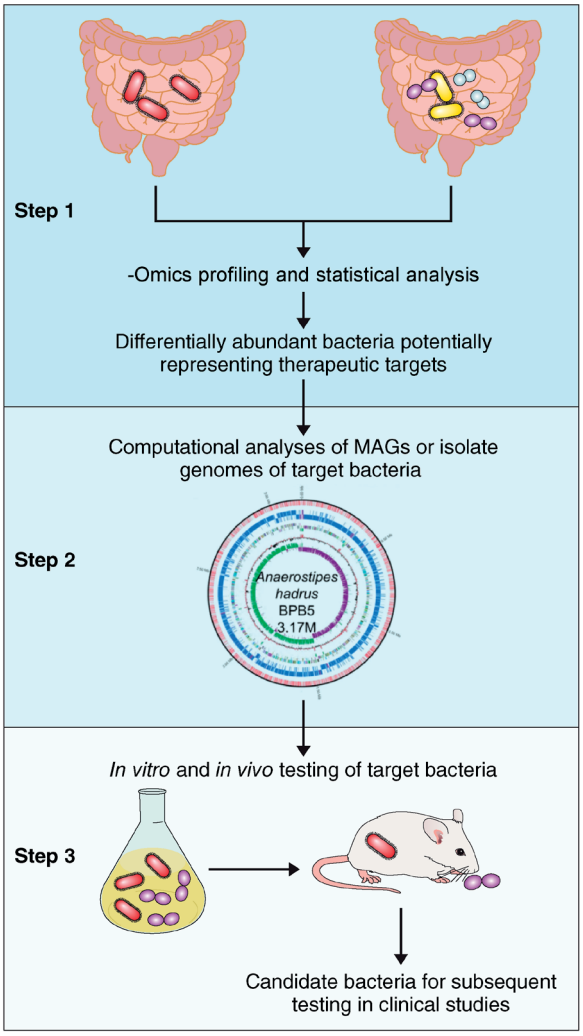


Figure 1: Workflow of studies in this thesis investigating microbiome-mediated colonization resistance and suggestions for follow-up research directions. In this thesis, studies were limited to the first step in the infographic, namely -omics profiling and identifying bacteria that may represent potential therapeutic targets, but no follow-up computational or wet-lab experiments have been performed. Follow-up computational experiments should involve analyses of the genomes of the bacterium of interest and understanding their encoded metabolic capacities (Step 2). After obtaining a detailed picture of such encoded capacities of the bacterium of interest, targeted in vitro and in vivo experiments can shed further light on the antagonistic actions against the potential pathogen. In vitro experiments can include co-culture of the bacterium of interest with the potential pathogen and in vivo experiments can include colonizing an animal with the potential pathogen and subsequently administering the bacterium of interest to investigate whether this leads to eradication of the potential pathogen (Step 3). Red bacteria indicate potential pathogenic bacteria, while yellow, purple and blue bacteria indicate potential bacteria of interest. The *A. hadrus* genome image was obtained from Zhang et al.[51].

The importance of study design and targeted follow-up computational analysis

Currently, there are too many studies which do not have an appropriate study design to make strong conclusions about which bacteria/microorganisms can potentially be protective against, or contribute to, a specific disease or pathogen colonization. It becomes increasingly clear that it is important to control for confounding factors in the study design, as many clinical variables have been shown to significantly affect gut microbiome composition[7, 54, 55]. A classic example illustrating the importance of uncoupling confounders is the original report that individuals with Type 2 Diabetes Mellitus (T2DM) had a very different gut microbiome as compared to healthy individuals[56]. However, a later publication showed that metformin, the first-line treatment for T2DM, usage was a very strong confounder and could explain the majority of the microbiome-modulating effect previously thought to be explained by having T2DM or not. The authors of this paper also strongly recommend future studies to disentangle effects of medication from effects of disease[57]. In this thesis, we could also not always fully dissociate effects of disease or colonization status from confounding factors (**Chapter 6** and **Chapter 8**). While in most epidemiological studies one can correct for confounding factors using (advanced) statistical models, there are no known statistical frameworks which can incorporate confounding variables for microbiome research. Ideally, one would have access to large enough cohorts so that sample selection can be applied to select for the phenotype of interest and to simultaneously exclude individuals positive for variables known to affect the microbiome. In cases of more severe disease, it is often not possible to exclude confounders such as medication use in the investigated population. In these cases it is possible to match cases against controls for more general confounders like age and sex, to at least minimize their effect on the obtained results[7]. Using samples from a large cross-sectional Dutch population cohort ("Pienter cohort"), we had the opportunity to select individuals who were negative for many microbiome confounders while positive for our phenotype of interest (ESBL-producing *E. coli* colonization) and could afterwards match them based on age, sex, travel history and ethnicity to individuals negative for our phenotype of interest (**Chapter 9**). While previous research has not taken confounders into account and showed several bacteria to be associated with MDRO protection (**Chapter 8** and [58, 59]), we demonstrated that there is no difference in the gut microbiome and metabolome between ESBL-producing *E. coli* colonized individuals versus those who are not colonized. This also suggests that microbiome-based therapeutics may not be as effective against ESBL-producing *E. coli* as they potentially are against other antibiotic-resistant bacterial species, such as vancomycin-resistant *Enterococcus* (VRE) (further discussed in the next paragraph).

After identifying bacteria that may be important in providing colonization resistance, follow-up research with more targeted computational and wet-lab analyses is warranted (Figure 1). To illustrate this point I will take the example of our recent finding that

the anaerobic and butyrate-producing bacterium *Anaerostipes hadrus*, amongst other bacterial species, was significantly associated with protection against *C. difficile* colonization (unpublished data, Figure 2).

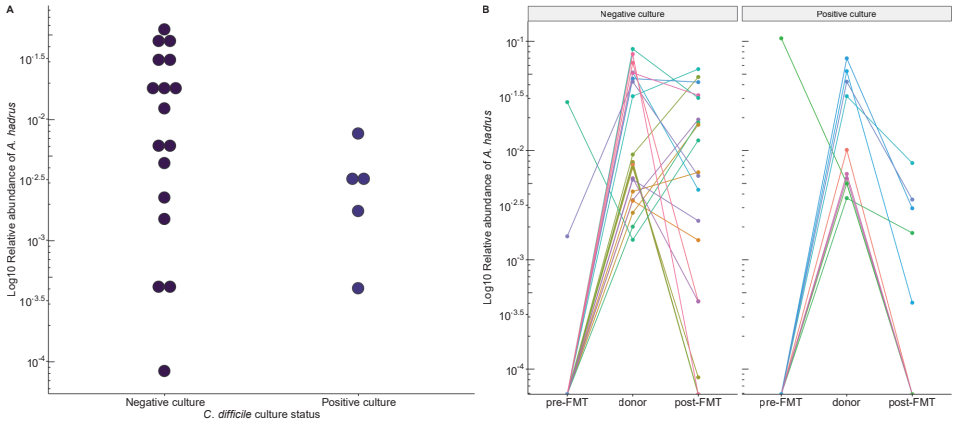


Figure 2: Comparison of *A. hadrus* relative abundance in n = 31 rCDI patients treated with FMT. Relative abundance on a log10 scale is indicated for post-FMT patients based on their status of being *C. difficile* culture positive or negative and only patients with detectable *A. hadrus* levels are shown. Seventeen of 21 samples culture-negative for *C. difficile* had detectable *A. hadrus* and 5/10 samples with a positive culture had detectable *A. hadrus*) (A). We further investigated whether this could be linked to a potential FMT donor effect, but this was not the case, as all patients were treated with donors whose microbiome contained *A. hadrus* (B). Each colored line represents the relative abundance of *A. hadrus* in an individual patient, pre- and post-FMT with the corresponding donor relative abundance of *A. hadrus*.

Ideally, the genomes and/or metagenome-assembled genomes (MAG)s of *A. hadrus* would be downloaded from public repositories and be extensively analyzed for potential antagonistic effects against *C. difficile* (Figure 1 Step 2). This could entail confirming or investigating mechanisms that are known to inhibit *C. difficile*, such as the presence of enzymatic machinery encoded for secondary bile acid conversion. A more generic approach could also be used. For example, one could annotate the *A. hadrus* genomes for biosynthetic gene clusters, which may encode for secondary metabolites that are involved in pathways or in mechanisms known to have an antagonistic function against *C. difficile*[60-62]. While such an approach sounds straightforward, this has long been impossible due to difficulties in culturing many common gut bacteria. In the past two years, however, several public repositories of MAGs have greatly expanded the availability of genomes of difficult-to-culture gut bacteria[63-65]. By leveraging this extensive genomic information, more targeted wet-lab experiments can be performed

to investigate and verify hypothesized mechanisms based on genomic data, as for many common gut bacteria very little is currently known with regard to their functional capabilities[66]. For example, if a putative bacteriocin-producing biosynthetic gene cluster in *A. hadrus* would be identified based on genomic data, one could attempt to purify the bacteriocin and directly test the effect of this bacteriocin on *C. difficile*.

Wet-lab research is necessary to elucidate a potential causative role for a bacterium in providing colonization resistance

After performing extensive computational analysis, mechanistic research in the wet-lab and in animal models is imperative to investigate if identified bacteria are really, and not only predicted to be, involved in providing protection against a given pathogen (Figure 1 Step 3). Such mechanistic research is non-trivial, as one will have to decide on many parameters such as pH, oxygen levels, incorporation of different cells occurring in the human gut (e.g. enterocoytes and goblet cells), growth of a functional mucus layer etc[67-69]. Ideally, conditions should be as representative of the human gut as possible. An important discovery in recent years is the capability of growing organoids from almost any human tissue, including the human gut[70, 71]. These organoids allow for more realistically mimicking the human gut and have been used to study the pathogenesis of various enteropathogenic bacteria, including *Salmonella enterica* and *C. difficile*[72, 73]. In addition, intestinal organoids have been used to identify a specific mutational signature caused by colibactin-producing *E. coli* and this mutational signature was subsequently detected in 10-20% of the investigated colorectal cancer genomes[74]. However, a complication in the organoid field is to incorporate the interactions of a complex human gut microbiome, which is currently one of the most difficult challenges to tackle[71, 75]. It should be mentioned that both the targeted computational analysis of genomes of potentially promising bacteria and the in vitro verification of hypothesized mechanisms involved in microbiome-mediated colonization resistance are unfortunately lacking in this thesis. However, we have started investigating antagonistic actions of *A. hadrus*, *Eubacterium rectale*, *Dorea longicatena* and *Butyricicoccus faecihominis* against *C. difficile* in vitro (unpublished data). Based on known functions of these bacterial species, we hypothesize that these antagonistic actions may be related to SCFA production and / or changes in gut pH and this is currently work in progress.

The group of Eric Pamer has performed various studies in the field of gut microbiome and colonization resistance at the mechanistic level. This research can serve as an example of moving beyond initial observations of differences in gut microbiota towards targeted development of bacterial consortia based on mechanistic understanding. Two of the early publications on VRE and the gut microbiome from this group identified that intestinal domination of VRE (as measured by enterococcal relative abundance by 16S rRNA gene amplicon sequencing) often preceded bloodstream infections

in patients undergoing allogenic hematopoietic stem cell transplantation[76, 77]. These findings were followed up by an elegant mouse study which showed that restoration of the gut microbiome of antibiotic-treated mice could restore colonization resistance against VRE, and that especially the *Barnesiella* genus was associated with clearance and reduction of VRE from the intestinal tract[78]. This was followed up by another mouse study which did actually not show an important role for a *Barnesiella* species, but for for *Blautia producta* and *Clostridium bolteae* instead. Mechanistically, it was further shown that a four-strain bacterial consortium (including *B. producta* and *C. bolteae*) was able to reverse antibiotic-induced susceptibility to VRE infection in mice[79]. An important aspect of the underlying mechanism of this cocktail was the production of a bacteriocin, a lantibiotic, by *B. producta*[80]. In patients at high risk of developing VRE infection, high abundance of this specific lantibiotic gene was associated with reduced *E. faecium* density[80]. Importantly, a patent has recently been filed for a consortium of bacteria including *B. producta* and *C. scindens* (see https://patents.justia.com/patent/20210000887) to reduce the burden of VRE infections in (vulnerable) patient groups. In addition, based on their findings, the Pamer group has performed translational research illustrating the importance of the gut microbiome against infection in stem cell transplant patients. In 2018, they published a study where auto-FMT was administered after successful stem cell engraftment[81]. This procedure was able to restore antibiotic-induced microbiome damage, accelerated restoration of neutrophil, monocyte and lymphocyte counts and was shown to be safe[81, 82]. Considering that the auto-FMT was generally shown to restore the gut microbiome (although this was not the case in all patients), it may therefore play a role in restoring colonization resistance against enteric pathogens and possibly MDROs. However, we are still awaiting results of whether auto-FMT resulted in lower infection rates and decreased graft versus host disease[81]. Noteworthy, none of the patients in the control group (not receiving an auto-FMT) recovered their gut microbiome in the same time period, even though they had a similar microbiota diversity at baseline and a similar drop during antibiotic treatment[81].

While this work is impressive and resulted in the recognition of bacterial consortia that are able to provide colonization resistance, it remains to be demonstrated whether such an intervention results in VRE eradication in immunocompromised patients, like stem cell transplant patients. Secondly, safety is a concern, as it has been previously reported that probiotic capsules containing *Lactobacillus*, which is generally considered a harmless commensal, can lead to a significantly higher chance of developing *Lactobacillus*-caused bacteremia in intensive care unit patients[83]. While safety trials for live biotherapeutic products (LBPs) generally do not report many adverse events, such trials have not yet been conducted in severely immunocompromised patients[84, 85].

Future therapeutic strategies for treating colonization and infection by bacterial enteropathogens and MDROs

Past and current therapeutic strategies for treating infection by bacterial enteric pathogens (e.g. *C. difficile*) and MDROs have largely been based on antibiotics. Asymptomatic colonization of MDROs is generally not treated, since the administration of antibiotics has various disadvantages and is not effective for MDRO eradication[77, 86]. There are some reports which it is described that individuals colonized with *C. difficile* have a higher risk to develop an infection and spread this bacterium to other patients[87-89]. However, it is not advised to treat asymptomatic *C. difficile* colonization with antibiotics, as this might also trigger an environment where *C. difficile* spores persist and are able to germinate and cause infection. However, patients with multiple recurrent CDI have a disturbed gut microbiome and an intervention with only antibiotic treatment fails. Since the landmark study of van Nood et al., FMT has become the best treatment option for rCDI patients, but so far rCDI is the only disease for which FMT is an accepted treatment[52]. All other indications for which FMT is investigated as a treatment are in an experimental or last-resort setting[90]. Also at the LUMC, FMT is applied in experimental and last-resort settings, with pilot studies being underway for IBD, for improving efficacy of checkpoint inhibitors in cancer patients, but also for metabolic disorders such as non-alcoholic fatty liver disease and non-alcoholic steatohepatitis and for neurological diseases such as hepatic encephalopathy and Parkinson's disease (dr. Liz Terveer and prof. dr. Ed Kuijper). One of the difficulties with using FMT as a therapeutic option, is that it is poorly defined as not all components are known and the quality and composition cannot be guaranteed to be consistent (**Chapter 3**)[91]. It is for this reason that in the coming years and decades it is expected that we will move away from administering undefined therapeutics (FMT) towards using live biotherapeutic products (LBPs), which are defined consortia of microorganisms (most often bacteria), or to other, well-defined microbiome therapeutics[91].

Live biotherapeutic products (LBPs)

At the start of my PhD trajectory in January 2018 (and already before that), it was expected that LBPs would quickly replace FMT as a therapeutic intervention against rCDI (and subsequently other diseases)[92]. LBPs have a theoretical advantage over the use of antibiotics for treating bacterial infections, as antibiotic treatment comes with the downside of selecting for antibiotic-resistant pathogens and negatively affecting the gut microbiome[91]. However, no LBPs have been implemented in patient care to date, but the expectations for their therapeutic application remain very high (**Chapter 3**)[91]. This is further illustrated by the fact that there are currently 17 clinical trials registered at clinicaltrials.gov, which investigate the potential of LBPs for highly divergent indications such as different types of cancer, rCDI, asthma, kidney disease and obesity.

I still expect that LBPs will ultimately replace FMT for treatment of rCDI with a similar cure rate of ~90%[93], but for other indications expectations should be tempered. While the cause of rCDI is strongly related to a disturbed gut microbiome, diseases like IBD are multifactorial with heavy involvement of the immune system and environmental factors[94]. This is the prime reason why I expect that LBPs for diseases with such a multifactorial etiology will be far less successful. This is currently also shown by the contradictory results for FMT trials against IBD and IBS[95-98], especially since current LBPs are developed based on a 'one-size fits all' approach. This one-size fits all approach is not likely to achieve a very high cure rate for such multifactorial diseases. It is also very likely that specific commensal bacteria will elicit differential immune responses across patients with different diseases[99]. In the oncology field, checkpoint inhibitors have revolutionized cancer treatment and the Nobel Prize of 2018 has been awarded to James Allison and Tasuku Honjo for their discovery[100]. However, 'only' 20% of patients with melanoma treated with checkpoint inhibitors achieves full remission[101]. It should be mentioned that this group of patients would likely not have survived long without checkpoint inhibitors, which is a crucial difference for defining success as compared to IBD or IBS. This is one of the reasons why the recent discovery that the gut microbiome may affect efficacy of immune checkpoint inhibitors was so important[102-104]. Although the exact mechanisms by which the microbiome potentially mediates this effect remain unknown, and different studies report different bacteria to be involved, there may be a role for the microbiome in favorably altering immune cell subsets[104]. Interestingly, in a study where FMT was administered to potentially improve checkpoint inhibitor efficacy in metastatic melanoma patients, donor feces were used from two patients who had previously been treated with checkpoint inhibitors and had achieved clinical remission for more than a year. Subsequent treatment success was only observed using one specific donor, but the underlying reason remained unclear[104]. To conclude, I believe that LBPs will eventually replace FMT for the treatment of rCDI, but we should not expect them to be miracle drugs capable of achieving cure rates of ~90% for other indications.

(Targeted) prebiotics

Prebiotics have been defined as "a substrate that is selectively utilized by host microorganisms conferring a health benefit" by a group of international experts[105]. Following from this definition, suppressing a pathogen may be possible through stimulating potentially antagonistic bacteria by providing the necessary substrates and thereby creating a gut environment unfavorable for enteropathogens, for example by increasing anaerobicity of the gut[106]. This could theoretically be achieved by providing 'good' bacteria with targeted substrates so that they will obtain a competitive advantage and thereby outcompete the pathogen or prevent the pathogen from colonizing in the first place. However, it must be noted that prebiotics (like probiotics) have not met the expectations yet and little to no evidence is available to support their use in

gastrointestinal disease[107, 108]. It is likely that for prebiotics to be more effective, the provided prebiotic should specifically be metabolized by bacteria in the gut that may contribute to reduction of disease burden. In almost all cases, prebiotics are dietary fibers which are not targeted to specific gut microbes. One interesting example where a dietary fiber (acetylated galactoglucomannan) was provided to match specific enzymatic machinery present in *Roseburia* and *Faecalibacterium* species was performed in pigs. This was not related to investigating suppression of a pathogen and rather serves as an example of how to match dietary fiber to bacterial enzymatic machinery[109]. *Faecalibacterium* relative abundance indeed increased post-intervention, but the effects on *Roseburia* relative abundance were less clear. To further investigate this, MAGs were assembled and it was noted that only a small subset of *Roseburia* MAGs contained the necessary enzymatic machinery for metabolizing acetylated galactoglucomannan. These MAGs indeed increased in abundance during the intervention, but this was not the case for *Roseburia* MAGs lacking the necessary enzymatic machinery. From a more general perspective, a 'butterfly' effect was observed whereby widespread community effects occurred through e.g. cross-feeding of products from dietary fiber fermentation[109]. This butterfly effect will likely occur in clinical settings in patients and therefore it remains unknown whether currently available prebiotics are potentially suitable for targeted intervention against pathogens, or can rather be used for inducing broad structural changes in the microbiome (or both). With the current scarcity of data with regard to the effect of targeted modulation of the gut microbiome through prebiotics, it is hard to foresee how this field will develop. Ideally, at least from a theoretical perspective, one would administer an LBP with specific prebiotics that are exclusively metabolized by members of that LBP and thereby create a competitive advantage to facilitate colonization.

Synbiotics

Administration of an LBP together with specific prebiotics, which are ideally metabolized by strains in the LBP, are called synbiotics. The theoretical advantage of synbiotics over LBPs or prebiotics is that the strains will gain a competitive advantage through substrate utilization and this would likely enhance the chance of successful engraftment, which has already been shown in rodent models[110, 111]. For example, Kearny et al. identified a resource, the edible seaweed nori, highly unlikely to be used by bacteria in the lab-mouse gut. They subsequently reasoned that if a microorganism would be introduced into the system during supplementation of seaweed, this microorganism would have a competitive advantage. Indeed, when administering a specific *Bacteroides plebeius* strain capable of porphyran (a polysaccharide present in seaweed) degradation in combination with seaweed supplementation, *B. plebeius* engrafted successfully long-term[111].

While synbiotics seem extremely promising, they are not trivial to produce from a technical point of view, as these strains will need to have very specific metabolic capacities. Preferably, one would like to isolate such strains from humans and not obtain them by genetically modifying bacteria, as this would pose a lot of extra regulatory hurdles[112]. On the other hand, by genetic modification of bacteria there is potential to cure a larger variety of diseases, including diseases that do not have a microbial origin, such as phenylketonuria[113-115]. Phenylketonuria patients are unable to metabolize the amino acid phenylalanine and prolonged consumption of this amino acid can result in severe neurological damage. The idea of using bacteria (next to a protein-restricted diet) to treat this condition is to administer bacteria that specifically metabolize this amino acid and thereby prevent its accumulation. In addition, the large inter-individual variation in gut microbiomes of humans will likely not allow for a 'one-size fits all' approach, as different microbiomes will compete differently with newly introduced strains and have different metabolic capacities. However, if major advances can be made and rare enough substrates can be identified to improve chances of colonization of introduced strains (see the seaweed example in the previous paragraph), there may be potential for development and implementation of synbiotics in the clinical setting.

Strengths and limitations

Before coming to the concluding remarks, I would like to take this opportunity to discuss some of the general strengths and limitations of the research described in this thesis.

Two major strengths of this thesis are 1) the versatility of topics and thereby the contributions to different branches of the microbiome field and 2) the progress to much more complex analysis techniques throughout the past ~3.5 years, which allowed me to more successfully extract information about the underlying biology. Most of the chapters in this thesis are devoted to the overarching theme of microbiome-mediated colonization resistance. The chapters range from studies aiming to identify bacteria and metabolites involved in providing colonization resistance, to studies describing the current understanding of the opportunities and challenges necessary for development of LBPs. These varied topics allowed me to obtain in-depth understanding of the many facets of the microbiome field including technical aspects, the implications for biology and medicine and the requirements for developing microbiome-based therapeutics. The fact that I was allowed to work on different potentially pathogenic (micro)organisms (*C. difficile*, the hookworm *N. americanus* and MDROs) enabled me to get a broader overview of infectious diseases than by concentrating on a single (micro)organism and this also facilitated extensive collaboration with different research groups. Furthermore, this thesis includes two technical-oriented chapters describing the effects of technical

variation on obtained microbiota profiles (**Chapter 4**) and the development of a new tool for profiling CAZymes from shotgun metagenomic data (**Chapter 5**). The research described in these chapters required a different approach than the clinical studies, as the underlying research questions were not directly related to biology and medicine, but were rather aimed at method optimization.

Over the course of conducting the research described in this thesis, the applied techniques moved from 16S rRNA gene amplicon analysis to shotgun metagenomics and metabolomics. The first two studies, chronologically speaking, performed during my PhD (**Chapter 4** and **Chapter 6**) merely involved 16S rRNA gene amplicon sequence data analysis and were limited to a single time point. We subsequently moved to longitudinal study designs (**Chapter 7** and **Chapter 8**), metagenomics (**Chapter 5**, **Chapter 8** and **Chapter 9**), metabolomics (**Chapter 9**) and machine learning approaches (**Chapter 5** and **Chapter 9**). The longitudinal studies allowed for investigating the consistency of microbiota patterns over time and to apply longitudinal analysis techniques. By integrating metagenomics and metabolomics with machine learning approaches, identification of potential biomarkers for a given phenotype becomes more likely and reliable than by only performing differential abundance analysis and taxonomic profiling through 16S rRNA gene amplicon analysis. Together, the variety of study designs required me to obtain knowledge of different statistical methods and their strengths and limitations, something which greatly contributed to my development and current knowledge.

There are also several weaknesses and things that I would have approached differently had I possessed the knowledge I currently have at the start of my PhD trajectory. The major weakness of the current thesis is that the mostly associative studies have not been followed up by more targeted genomic analyses of bacteria of interest nor by mechanistic wet-lab research. This holds particularly true for the findings described in **Chapter 6**, where bacterial taxa with the potential to inhibit *C. difficile* were identified. In light of this, we performed another in silico study where shotgun metagenomes of rCDI patients post-FMT were compared and where one group remained colonized by *C. difficile* and patients in the other group fully eradicated *C. difficile* (unpublished data). *A. hadrus*, *E. rectale*, *B. faecihominis* and *D. longicatena* were identified to be more abundant in the non-colonized group. Fortunately, these bacteria are currently being further investigated or their potential antagonistic effect against *C. difficile* at the Experimental Bacteriology group of LUMC. Second, I would have liked to include machine learning approaches in **Chapter 6** to investigate whether *C. difficile* colonization status could reliably be predicted. If the conducted differential abundance analysis and desired feature selection through machine learning shows the same bacterial taxa to be important in protection against *C. difficile* colonization, this would have made our findings more robust. A third

weakness of this thesis is that our studies were exclusively focused on the interactions of enteropathogens or MDROs with the gut microbiome without taking into account any parameters about host immunity function, as this was outside the scope of this thesis. Interactions between host immunity and the gut microbiome, an enteropathogen or an MDRO are complex but necessary for a detailed understanding of infectious diseases processes[50, 116-118]. Fourth, in this thesis we have always relied on fecal samples to investigate the gut microbiome. As was discussed in **Chapter 7**, fecal samples may not be representative for what happens at the mucosal surface of the intestinal tract and this is especially relevant when studying microorganisms which cause infections in other parts of the intestine than the colon. Indeed, the fecal microbiome can differ a lot from the microbiome in other locations of the intestine[119-121]. Lastly, I would like to touch upon my experience of analyzing microbiome data from studies that were not always specifically set up for conducting microbiome analyses (**Chapter 6 – Chapter 8**). For future studies with microbiome analyses, it will be important to involve a microbiome researcher as early as possible at the design of the study. By doing so, appropriate research questions can be formulated a priori, and most importantly, the microbiome researcher can aid in deciding on the appropriate study design to answer these research questions. In addition, advice can be provided on more practical issues like sample collection, storage and subsequent processing steps (**Chapter 4**).

Concluding remarks

The projects described in this thesis are diverse, ranging from methodology optimization to investigating the gut microbiome in clinical cohorts with the goal of finding bacteria associated with providing microbiome-mediated colonization resistance against enteropathogens and MDROs.

While many issues in microbiome research need to be addressed and numerous open biological questions remain, I excitedly look forward to the future of microbiome research and hopefully towards the implementation of the first rationally designed microbiome-based therapeutics into the clinic in the coming years.

References

1. Dunn MC, Bourne PE. Building the biomedical data science workforce. PLoS Biol. 2017;15(7):e2003082.
2. Wickware P. Next-generation biologists must straddle computation and biology. Nature. 2000;404(6778):683-4.
3. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18(1):83.
4. Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, et al. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 2016;17:53.
5. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol. 2015;16:276.
6. Wen H, Wang HY, He X, Wu CI. On the low reproducibility of cancer studies. Natl Sci Rev. 2018;5(5):619-24.
7. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020;587(7834):448-54.
8. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
9. Tedersoo L, Küngas R, Oras E, Köster K, Eenmaa H, Leijen Ä, et al. Data sharing practices and data availability upon request differ across scientific disciplines. Sci Data. 2021;8(1):192.
10. Eckert EM, Di Cesare A, Fontaneto D, Berendonk TU, Bürgmann H, Cytryn E, et al. Every fifth published metagenome is not available to science. PLoS Biol. 2020;18(4):e3000698.
11. Langille MGI, Ravel J, Fricke WF. "Available upon request": not good enough for microbiome data! Microbiome. 2018;6(1):8.
12. Christian TM, Gooch A, Vision T, Hull E. Journal data policies: Exploring how the understanding of editors and authors corresponds to the policies themselves. PLoS One. 2020;15(3):e0230281.
13. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25(4):679-89.
14. Berkell M, Mysara M, Xavier BB, van Werkhoven CH, Monsieurs P, Lammens C, et al. Microbiota-based markers predictive of development of Clostridioides difficile infection. Nat Commun. 2021;12(1):2241.
15. Ke S, Pollock NR, Wang XW, Chen X, Daugherty K, Lin Q, et al. Integrating gut microbiome and host immune markers to understand the pathogenesis of Clostridioides difficile infection. Gut Microbes. 2021;13(1):1-18.
16. Empty rhetoric over data sharing slows science. Nature. 2017;546(7658):327.
17. ZonMw. https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/. 2021.
18. ZonMw. https://www.zonmw.nl/nl/subsidies/openstaande-subsidieoproepen/detail/item/call-for-submission-of-project-ideas-tackling-antibiotic-resistance-by-reusing-data-and-increasing/. 2019.
19. Couture JL, Blake RE, McDonald G, Ward CL. A funder-imposed data publication requirement seldom inspired data sharing. PLoS One. 2018;13(7):e0199789.
20. Christensen G, Dafoe A, Miguel E, Moore DA, Rose AK. A study of the impact of data sharing on article citations using journal policies as a natural experiment. PLoS One. 2019;14(12):e0225883.
21. Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. PLoS One. 2020;15(4):e0230416.
22. Penders B, Horstman K, Vos R. Walking the Line between Lab and Computation: The "Moist" Zone. BioScience. 2008;58(8):747-55.
23. Muller EE, Glaab E, May P, Vlassis N, Wilmes P. Condensing the omics fog of microbial communities. Trends Microbiol. 2013;21(7):325-33.
24. Leonelli S. Data Interpretation in the Digital Age. Perspect Sci. 2014;22(3):397-417.
25. Kruchten AE. A Curricular Bioinformatics Approach to Teaching Undergraduates to Analyze Metagenomic Datasets Using R. Front Microbiol. 2020;11:578600.
26. Advising Biology Majors about Career Choices: Resources & Information for Biology Instructors. The American Biology Teacher. 2017;79(1):14-21.
27. Rosenwald AG, Pauley MA, Welch L, Elgin SC, Wright R, Blum J. The CourseSource Bioinformatics Learning Framework. CBE Life Sci Educ. 2016;15(1):le2.
28. Heintz-Buschart A, Wilmes P. Human Gut Microbiome: Function Matters. Trends Microbiol. 2018;26(7):563-74.
29. Langille MGI. Exploring Linkages between Taxonomic and Functional Profiles of the Human Microbiome. mSystems. 2018;3(2).

30. Pérez-Cobas AE, Gosalbes MJ, Friedrichs A, Knecht H, Artacho A, Eismann K, et al. Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. Gut. 2013;62(11):1591-601.
31. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med. 2011;3(106):106ra.
32. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. Cell. 2013;152(1-2):39-50.
33. Vieira-Silva S, Falony G, Darzi Y, Lima-Mendez G, Garcia Yunta R, Okuda S, et al. Species-function relationships shape ecological properties of the human gut microbiome. Nat Microbiol. 2016;1(8):16088.
34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
35. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. JAMA. 2010;304(24):2706-15.
36. Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E, et al. Acquired initiating mutations in early hematopoietic cells of CLL patients. Cancer Discov. 2014;4(9):1088-101.
37. Pan W, Kwak IY, Wei P. A Powerful Pathway-Based Adaptive Test for Genetic Association with Common or Rare Variants. Am J Hum Genet. 2015;97(1):86-98.
38. Zhang X, Li L, Butcher J, Stintzi A, Figeys D. Advancing functional and translational microbiome research using meta-omics approaches. Microbiome. 2019;7(1):154.
39. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018;50(6):790-5.
40. Mesnage R, Teixeira M, Mandrioli D, Falcioni L, Ducarmon QR, Zwittink RD, et al. Use of Shotgun Metagenomics and Metabolomics to Evaluate the Impact of Glyphosate or Roundup MON 52276 on the Gut Microbiota and Serum Metabolome of Sprague-Dawley Rats. Environ Health Perspect. 2021;129(1):17005.
41. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, et al. Learning representations of microbe–metabolite interactions. Nat Methods. 2019;16(12):1306-14.
42. Mallick H, Franzosa EA, McLver LJ, Banerjee S, Sirota-Madi A, Kostic AD, et al. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nat Commun. 2019;10(1):3136.
43. Xie J, Cho H, Lin BM, Pillai M, Heimisdottir LH, Bandyopadhyay D, et al. Improved Metabolite Prediction Using Microbiome Data-Based Elastic Net Models. 2021:2021.07.01.450697.
44. Heintz-Buschart A, May P, Laczny CC, Lebrun LA, Bellora C, Krishna A, et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. Nat Microbiol. 2016;2:16180.
45. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol. 2018;3(3):337-46.
46. Bohnhoff M, Miller CP, Martin WR. Resistance of the mouse's intestinal tract to experimental Salmonella infection. I. factors which interfere with the initiation of infection by oral inoculation. J Exp Med. 1964;120(5):805-16.
47. Ducarmon QR, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. Microbiol Mol Biol Rev. 2019;83(3).
48. Sorbara MT, Pamer EG. Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. Mucosal Immunol. 2019;12(1):1-9.
49. van der Waaij D, Berghuis-de Vries JM, Lekkerkerk L-v. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. J Hyg (Lond). 1971;69(3):405-11.
50. Stacy A, Andrade-Oliveira V, McCulloch JA, Hild B, Oh JH, Perez-Chaparro PJ, et al. Infection trains the host for microbiota-enhanced resistance to pathogens. Cell. 2021;184(3):615-27.e17.
51. Zhang Q, Wu Y, Wang J, Wu G, Long W, Xue Z, et al. Accelerated dysbiosis of gut microbiota during aggravation of DSS-induced colitis by a butyrate-producing bacterium. Sci Rep. 2016;6(1):27572.
52. 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):407-15.
53. Young VB. Unexpected Results From a Phase 2 Trial of a Microbiome Therapeutic for Clostridioides difficile Infection: Lessons for the Future. Clin Infect Dis. 2021;72(12):2141-3.
54. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560-4.
55. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science.

2016;352(6285):565-9.
56. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60.
57. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature. 2015;528(7581):262-6.
58. Araos R, Tai AK, Snyder GM, Blaser MJ, D'Agata EMC. Predominance of Lactobacillus spp. Among Patients Who Do Not Acquire Multidrug-Resistant Organisms. Clin Infect Dis. 2016;63(7):937-43.
59. Annavajhala MK, Gomez-Simmonds A, Macesic N, Sullivan SB, Kress A, Khan SD, et al. Colonizing multidrug-resistant bacteria and the longitudinal evolution of the intestinal microbiome after liver transplantation. Nat Commun. 2019;10(1):4715.
60. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 2019;47(W1):W81-w7.
61. Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Res. 2019;47(18):e110.
62. Ren H, Shi C, Zhao H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. iScience. 2020;23(1):100795.
63. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol. 2021;39(1):105-14.
64. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell. 2019;176(3):649-62.e20.
65. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. Nature. 2019;568(7753):499-504.
66. Thomas AM, Segata N. Multiple levels of the unknown in microbiome research. BMC Biol. 2019;17(1):48.
67. Goers L, Freemont P, Polizzi KM. Co-culture systems and technologies: taking synthetic biology to the next level. J R Soc Interface. 2014;11(96).
68. Pearce SC, Coia HG, Karl JP, Pantoja-Feliciano IG, Zachos NC, Racicot K. Intestinal in vitro and ex vivo Models to Study Host-Microbiome Interactions and Acute Stressors. Front Physiol. 2018;9:1584.
69. Shah P, Fritz JV, Glaab E, Desai MS, Greenhalgh K, Frachet A, et al. A microfluidics-based in vitro model of the gastrointestinal human-microbe interface. Nat Commun. 2016;7:11535.
70. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009;459(7244):262-5.
71. Puschhof J, Pleguezuelos-Manzano C, Clevers H. Organoids and organs-on-chips: Insights into human gut-microbe interactions. Cell Host Microbe. 2021;29(6):867-78.
72. Leslie JL, Huang S, Opp JS, Nagy MS, Kobayashi M, Young VB, et al. Persistence and toxin production by Clostridium difficile within human intestinal organoids result in disruption of epithelial paracellular barrier function. Infect Immun. 2015;83(1):138-45.
73. Lawrence AE, Abuaita BH, Berger RP, Hill DR, Huang S, Yadagiri VK, et al. Salmonella enterica Serovar Typhimurium SPI-1 and SPI-2 Shape the Global Transcriptional Landscape in a Human Intestinal Organoid Model System. mBio. 2021;12(3).
74. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+) E. coli. Nature. 2020;580(7802):269-73.
75. Min S, Kim S, Cho SW. Gastrointestinal tract modeling using organoids engineered with cellular and microbiota niches. Exp Mol Med. 2020;52(2):227-37.
76. Ubeda C, Taur Y, Jenq RR, Equinda MJ, Son T, Samstein M, et al. Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. J Clin Invest. 2010;120(12):4332-41.
77. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis. 2012;55(7):905-14.
78. Ubeda C, Bucci V, Caballero S, Djukovic A, Toussaint NC, Equinda M, et al. Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. Infect Immun. 2013;81(3):965-73.
79. Caballero S, Kim S, Carter RA, Leiner IM, Sušac B, Miller L, et al. Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. Cell Host Microbe. 2017;21(5):592-602.e4.

80. Kim SG, Becattini S, Moody TU, Shliaha PV, Littmann ER, Seok R, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. Nature. 2019;572(7771):665-9.
81. Taur Y, Coyte K, Schluter J, Robilotti E, Figueroa C, Gjonbalaj M, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med. 2018;10(460).
82. Schluter J, Peled JU, Taylor BP, Markey KA, Smith M, Taur Y, et al. The gut microbiota is associated with immune cell dynamics in humans. Nature. 2020;588(7837):303-7.
83. Yelin I, Flett KB, Merakou C, Mehrotra P, Stam J, Snesrud E, et al. Genomic and epidemiological evidence of bacterial transmission from probiotic capsule to blood in ICU patients. Nat Med. 2019;25(11):1728-32.
84. Kao D, Wong K, Franz R, Cochrane K, Sherriff K, Chui L, et al. The effect of a microbial ecosystem therapeutic (MET-2) on recurrent Clostridioides difficile infection: a phase 1, open-label, single-group trial. Lancet Gastroenterol Hepatol. 2021;6(4):282-91.
85. Henn MR, O'Brien EJ, Diao L, Feagan BG, Sandborn WJ, Huttenhower C, et al. A Phase 1b Safety Study of SER-287, a Spore-Based Microbiome Therapeutic, for Active Mild to Moderate Ulcerative Colitis. Gastroenterology. 2021;160(1):115-27.e30.
86. Tacconelli E, Mazzaferri F, de Smet AM, Bragantini D, Eggimann P, Huttner BD, et al. ESCMID-EUCIC clinical guidelines on decolonization of multidrug-resistant Gram-negative bacteria carriers. Clin Microbiol Infect. 2019;25(7):807-17.
87. Hung YP, Tsai PJ, Hung KH, Liu HC, Lee CI, Lin HJ, et al. Impact of toxigenic Clostridium difficile colonization and infection among hospitalized adults at a district hospital in southern Taiwan. PLoS One. 2012;7(8):e42415.
88. Clabots CR, Johnson S, Olson MM, Peterson LR, Gerding DN. Acquisition of Clostridium difficile by hospitalized patients: evidence for colonized new admissions as a source of infection. J Infect Dis. 1992;166(3):561-7.
89. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med. 2013;369(13):1195-205.
90. Wortelboer K, Nieuwdorp M, Herrema H. Fecal microbiota transplantation beyond Clostridioides difficile infections. EBioMedicine. 2019;44:716-29.
91. Ducarmon QR, Kuijper EJ, Olle B. Opportunities and Challenges in Development of Live Biotherapeutic Products To Fight Infections. J Infect Dis. 2021.
92. Kelly CP. Fecal microbiota transplantation--an old therapy comes of age. N Engl J Med. 2013;368(5):474-5.
93. 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):1236-47.
94. de Souza HSP, Fiocchi C, Iliopoulos D. The IBD interactome: an integrated view of aetiology, pathogenesis and therapy. Nat Rev Gastroenterol Hepatol. 2017;14(12):739-49.
95. Paramsothy S, Paramsothy R, Rubin DT, Kamm MA, Kaakoush NO, Mitchell HM, et al. Faecal Microbiota Transplantation for Inflammatory Bowel Disease: A Systematic Review and Meta-analysis. J Crohns Colitis. 2017;11(10):1180-99.
96. Costello SP, Hughes PA, Waters O, Bryant RV, Vincent AD, Blatchford P, et al. Effect of Fecal Microbiota Transplantation on 8-Week Remission in Patients With Ulcerative Colitis: A Randomized Clinical Trial. Jama. 2019;321(2):156-64.
97. Xu D, Chen VL, Steiner CA, Berinstein JA, Eswaran S, Waljee AK, et al. Efficacy of Fecal Microbiota Transplantation in Irritable Bowel Syndrome: A Systematic Review and Meta-Analysis. Am J Gastroenterol. 2019;114(7):1043-50.
98. El-Salhy M, Hatlebakk JG, Gilja OH, Bråthen 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):859-67.
99. Pamer EG. Fecal microbiota transplantation: effectiveness, complexities, and lingering concerns. Mucosal Immunol. 2014;7(2):210-4.
100. Huang PW, Chang JW. Immune checkpoint inhibitors win the 2018 Nobel Prize. Biomed J. 2019;42(5):299-306.
101. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. Nat Commun. 2020;11(1):3801.
102. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science. 2018;359(6371):91-7.
103. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science. 2018;359(6371):97-103.
104. Baruch EN, Youngster I, Ben-Betzalel G, Ortenberg R, Lahat A, Katz L, et al. Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients. Science. 2021;371(6529):602-9.

105. Gibson GR, Hutkins R, Sanders ME, Prescott SL, Reimer RA, Salminen SJ, et al. Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. Nat Rev Gastroenterol Hepatol. 2017;14(8):491-502.
106. Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science. 2018;362(6418).
107. Ford AC, Harris LA, Lacy BE, Quigley EMM, Moayyedi P. Systematic review with meta-analysis: the efficacy of prebiotics, probiotics, synbiotics and antibiotics in irritable bowel syndrome. Aliment Pharmacol Ther. 2018;48(10):1044-60.
108. Limketkai BN, Iheozor-Ejiofor Z, Gjuladin-Hellon T, Parian A, Matarese LE, Bracewell K, et al. Dietary interventions for induction and maintenance of remission in inflammatory bowel disease. Cochrane Database Syst Rev. 2019;2(2):Cd012839.
109. Michalak L, Gaby JC, Lagos L, La Rosa SL, Hvidsten TR, Tétard-Jones C, et al. Microbiota-directed fibre activates both targeted and secondary metabolic shifts in the distal gut. Nat Commun. 2020;11(1):5773.
110. Shepherd ES, DeLoache WC, Pruss KM, Whitaker WR, Sonnenburg JL. An exclusive metabolic niche enables strain engraftment in the gut microbiota. Nature. 2018;557(7705):434-8.
111. Kearney SM, Gibbons SM, Erdman SE, Alm EJ. Orthogonal Dietary Niche Enables Reversible Engraftment of a Gut Bacterial Commensal. Cell Rep. 2018;24(7):1842-51.
112. Plavec TV, Berlec A. Safety Aspects of Genetically Modified Lactic Acid Bacteria. Microorganisms. 2020;8(2).
113. Charbonneau MR, Isabella VM, Li N, Kurtz CB. Developing a new class of engineered live bacterial therapeutics to treat human diseases. Nat Commun. 2020;11(1):1738.
114. Isabella VM, Ha BN, Castillo MJ, Lubkowicz DJ, Rowe SE, Millet YA, et al. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. Nat Biotechnol. 2018;36(9):857-64.
115. Puurunen MK, Vockley J, Searle SL, Sacharow SJ, Phillips JA, Denney WS, et al. Safety and pharmacodynamics of an engineered E. coli Nissle for the treatment of phenylketonuria: a first-in-human phase 1/2a study. Nat Metabol. 2021;3(8):1125-32.
116. Stevens EJ, Bates KA, King KC. Host microbiota can facilitate pathogen infection. PLoS Pathog. 2021;17(5):e1009514.
117. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Res. 2020;30(6):492-506.
118. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. Cell. 2014;157(1):121-41.
119. Montassier E, Valdés-Mas R, Batard E, Zmora N, Dori-Bachash M, Suez J, et al. Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner. Nat Microbiol. 2021;6(8):1043-54.
120. Vuik F, Dicksved J, Lam SY, Fuhler GM, van der Laan L, van de Winkel A, et al. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. United European Gastroenterol J. 2019;7(7):897-907.
121. Yuan C, Graham M, Staley C, Subramanian S. Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship. mSystems. 2020;5(3).

Nederlandse samenvatting
Acknowledgements
List of publications
PhD Portfolio
Curriculum vitae

Nederlandse samenvatting

Onderzoek naar het darmmicrobioom, de verzameling van micro-organismen in de darmen, is de afgelopen jaren enorm toegenomen door de vele ontdekkingen die de cruciale rol van het microbioom op de menselijk gezondheid benadrukken. Het darmmicrobioom is bijvoorbeeld belangrijk voor het produceren van bepaalde essentiële vitamines, het fermenteren van voedingsvezels die door de mens zelf niet kunnen worden afgebroken en het trainen van het immuunsysteem. Bovendien is het verlenen van kolonisatieresistentie een zeer belangrijke functie die het microbioom kan vervullen. Microbioom-gemedieerde kolonisatieresistentie houdt in dat het microbioom voorkomt dat potentiële pathogenen of exogene micro-organismen zich in de darm vestigen en een infectie veroorzaken. Er zijn vele mechanismen bekend waarmee het microbioom kolonisatieresistentie kan verlenen. Voorbeelden hiervan zijn competitie om plaats en nutriënten, het produceren van antimicrobiële stoffen en de interactie met het immuunsysteem. Echter, deze mechanismen, en daarmee de beschermende rol van het microbioom tegen pathogenen, kunnen negatief worden beïnvloed door verschillende factoren, waaronder het gebruik van antibiotica. Hoewel antibioticagebruik miljoenen levens heeft gered door het succesvol bestrijden van bacteriële infecties, heeft het een keerzijde vanwege mogelijk ernstige beschadiging van het darmmicrobioom.

Door het veelvuldige gebruik van antibiotica in de afgelopen decennia komt er ook steeds meer aandacht voor het gevaar van bijzonder resistente micro-organismen (BRMO). Dit uit zich ook in de vorm van speciale onderzoeksprogramma's met het thema antibioticaresistentie die nationale onderzoeksinstituten momenteel hebben. Behalve dat BRMO's ernstige en moeilijk te behandelen infecties kunnen veroorzaken, is er ook een grote groep mensen die asymptomatisch gekoloniseerd is met een BRMO en daarmee niet alleen bijdraagt aan ongemerkte verspreiding, maar ook een verhoogd risico loopt op het ontwikkelen van een infectie veroorzaakt door de BRMO. Een mogelijke verklaring voor het feit dat het grootste deel van de asymptomatisch gekoloniseerde mensen geen infectie oploopt, is dat het microbioom in deze situatie voorkomt dat de BRMO kan uitgroeien, hoewel andere factoren, zoals het immuunsysteem, tenminste even belangrijk zijn. Het is van belang meer inzicht te krijgen in dit proces. Idealiter zou men begrijpen welke micro-organismen of welke functies in het microbioom van asymptomatisch gekoloniseerde mensen ervoor zorgen dat zij geen infectie ontwikkelen, terwijl dat bij anderen wel gebeurt. Dit zou op termijn kunnen leiden tot de ontwikkeling van specifieke microbioom-gebaseerde medicijnen die zulke infecties, of zelfs asymptomatische kolonisatie, kunnen genezen of voorkomen.

Het onderzoek beschreven in dit proefschrift is uitgevoerd met als doel darmbacteriën te identificeren die een rol kunnen spelen in het verlenen van kolonisatieresistentie tegen

potentiële pathogenen en BRMO's in het maag-darmkanaal. Dit proefschrift bestaat uit drie delen. In deel 1 van dit proefschrift wordt microbioom-gemedieerde kolonisatieresistentie besproken vanuit een theoretisch perspectief en wordt samengevat wat momenteel de uitdagingen en mogelijkheden zijn voor het ontwikkelen van microbioom-gebaseerde therapieën. In deel 2 wordt dieper ingegaan op de methodiek voor zowel laboratorium-gebaseerde als computer-gebaseerde technieken om microbioomanalyses optimaal uit te kunnen voeren. Ten slotte worden in deel 3 verschillende klinische studies beschreven die zijn uitgevoerd om mogelijk beschermende darmbacteriën te identificeren.

Deel 1: Mechanismen van microbioom-gemedieerde kolonisatieresistentie en hoe microbioom-gebaseerde therapieën te ontwikkelen

Het concept dat het darmmicrobioom belangrijk is ter verdediging tegen darminfecties met enteropathogene micro-organismen is al lange tijd bekend. In **hoofdstuk 2** worden de verschillende mechanismen die bijdragen aan microbioom-gemedieerde kolonisatieresistentie in detail besproken, zoals de productie van antimicrobiële stoffen en competitie om nutriënten. Een belangrijk nieuw deel van het microbioomonderzoeksveld is het effect van medicijnen op de werking van het microbioom (en andersom ook het effect van het microbioom op de effectiviteit van medicijnen). Dit nog zeer nieuwe onderzoeksveld wordt *pharmacomicrobiomics* genoemd. Om deze redenen worden de effecten van verschillende, veelgebruikte medicijnen, zoals metformine en maagzuurremmers, op kolonisatieresistentie beschreven. Hoewel het belangrijk is om, vanuit het perspectief van het microbioom, onderliggende werkingsmechanismen die leiden tot kolonisatieresistentie te begrijpen, is het ook cruciaal om te weten hoe enteropathogenen deze mechanismen kunnen omzeilen. Om deze reden wordt voor acht veel voorkomende enteropathogene bacteriën uitgebreid beschreven hoe zij deze verdedigingsmechanismen ontwijken.

In **hoofdstuk 3** wordt ingegaan op wat er nodig is voor het ontwikkelen en produceren van microbioom-gebaseerde therapieën, met de nadruk op consortia van levende bacteriën (*live biotherapeutic products*; LBP's). Hiervoor is samengewerkt met één van de leiders in dit veld, het Amerikaanse biotechbedrijf Vedanta Biosciences. Met dit bedrijf heeft de afdeling Medische Microbiologie een wetenschappelijk samenwerking om een optimaal LBP voor patiënten met recidiverende *Clostridioides difficile* infectie te ontwikkelen die de feces microbioom transplantatie (FMT) kan vervangen. Er is een product ontwikkeld van acht bacteriesoorten die op dit moment (2021) in een fase 2/3 onderzoek wordt getest. De twee voornaamste voordelen van het potentieel gebruik van LBP's ten opzichte van FMT is geen risico van overdracht van ongewenste, onbekende en potentieel schadelijke bacteriën met metabole producten en de toediening van een goed omschreven en gekarakteriseerd product. Met name de overdracht van multiresistente en (entero)pathogene bacteriën zijn door de Amerikaanse FDA herkend

en vereisen speciaal ingerichte en gespecialiseerde centra (zoals de Nederlandse Donor Feces Bank; NDFB) die deze behandeling zonder risico's kunnen aanbieden.

Deel 2: Optimaliseren en standaardiseren van computationele en laboratorium-technieken voor microbioomonderzoek

In de afgelopen jaren is gebleken dat de technieken die worden gebruikt voor microbioomanalyses, zowel laboratorium- als computer-gebaseerde technieken, het uiteindelijke microbioomprofiel sterk kunnen beïnvloeden. In **hoofdstuk 4** wordt beschreven wat het effect is van verschillende DNA-extractiemethoden en verschillende analysesoftware op het microbioomprofiel. Hoewel de analyseprogramma's ongeveer gelijk presteerden, werden er wel verschillen in prestatie gevonden tussen DNA-extractiemethoden. Deze studie heeft geleid tot een standaard DNA-extractiemethode die nog steeds in onze onderzoeksgroep wordt gebruikt. Een nog enigszins onderbelicht probleem is het includeren van negatieve controles wanneer de microbioomsamenstelling van weefsels of vloeistoffen onderzocht wordt waarin weinig bacteriën aanwezig zijn, zoals bijvoorbeeld urine en tumorweefsel. Er bestaat hierdoor een risico dat het verkregen microbioomprofiel bestaat uit een mix van contaminatie (bijvoorbeeld door DNA-moleculen die op reagentia aanwezig zijn, maar niet in het monster zelf) en biologisch signaal. Zonder negatieve controles zijn deze microbioomprofielen lastig te interpreteren. In dit hoofdstuk tonen wij het cruciale belang aan van negatieve controles, omdat enkele van de onderzochte weefsels geen biologisch microbioom bevatte, maar zo goed als identiek waren aan het microbioomprofiel verkregen uit negatieve controles. In **hoofdstuk 5** wordt de ontwikkeling van nieuwe analysesoftware beschreven waarvan het doel is om koolhydraat-actieve enzymen (CAZymes) te karakteriseren uit complexe sequentiedata (waarbij al het DNA in een monster wordt afgelezen). Ten eerste zijn er *machine learning*-technieken toegepast om deze CAZymes zo betrouwbaar mogelijk te kunnen detecteren uit sequentiedata en vervolgens zijn analysemethoden toegepast om die CAZymes uit een metagenoom te kunnen kwantificeren. Ten tweede is er een nieuw annotatieschema ontwikkeld waarin de substraten van CAZymes zijn gegroepeerd in functioneel informatievere groepen, zoals bijvoorbeeld een groep CAZymes die voedingsvezels kunnen afbreken. Tot slot hebben wij deze nieuw ontwikkelde tool toegepast op beschikbare data van verschillende cohorten waarin het microbioom van darmkankerpatiënten vergeleken is met controlepersonen. Hierin konden wij aantonen dat het microbioom van darmkankerpatiënten meer CAZymes bevat die een substraat van dierlijke oorsprong verwerken (glycosaminoglycanen) en minder CAZymes bevat die een rol spelen bij het verwerken van voedingsvezels.

Deel 3: Het identificeren van micro-organismen die geassocieerd zijn met kolonisatieresistentie in klinische studies

Het laatste, en belangrijkste deel van dit proefschrift, bestaat uit klinische studies

waarin het doel was om bacteriën te identificeren die geassocieerd kunnen worden met kolonisatieresistentie. Dit zou op termijn kunnen leiden tot het gebruik van deze beschermende bacteriën in een LBP.

In **hoofdstuk 6** hebben wij het darmmicrobioom van drie patiëntgroepen vergeleken, een groep patiënten met *C. difficile*-infectie, een groep patiënten die asymptomatisch gekoloniseerd is met *C. difficile* en een controlegroep van patiënten die niet gekoloniseerd is door *C. difficile*. Zodoende hebben wij verschillende bacteriën (voornamelijk *Eubacterium hallii* en *Fusicatenibacter*) gevonden die mogelijk bescherming bieden tegen *C. difficile*-kolonisatie. Een andere belangrijke bevinding was dat asymptomatisch gekoloniseerde patiënten een lagere microbioomdiversiteit hadden ten opzichte van niet-gekoloniseerde patiënten. De mate waarin de twee eerdergenoemde bacteriesoorten ook echt beschermend zouden kunnen zijn tegen *C. difficile* zal nader onderzocht moeten worden, aangezien we met deze data alleen een associatie hebben aan kunnen tonen en geen oorzakelijk verband.

Hoofdstuk 7 beschrijft een studie waarin 20 gezonde vrijwilligers in het LUMC geïnfecteerd werden met *Necator americanus*, een mijnworm die voornamelijk in tropische gebieden voorkomt en die buikpijn, bloedarmoede en een eiwittekort kan veroorzaken. Door het microbioom voor en tijdens de infectie in kaart te brengen, konden wij de veranderingen over de tijd bestuderen. Eerder onderzoek liet wisselende resultaten zien op dit gebied. Hoewel er in onze studie geen consistente mijnworm-geïnduceerde effecten konden worden gedetecteerd, was er wel een verschil in de stabiliteit van het microbioom tussen patiënten met verschillende klinische symptomen na de besmetting. Deze resultaten tonen aan dat er wel een verband bestaat tussen mijnworm-geïnduceerde symptomen en de stabiliteit van het microbioom, maar er zijn geen specifieke bacteriën gevonden die mogelijk beschermen tegen besmetting en het ontstaan van symptomen.

Verpleeghuisbewoners hebben een verhoogd risico om in het darmstelsel gekoloniseerd te worden met een BRMO. Zo gebruiken zij veelvuldig antibiotica, wonen en leven in een relatief kleine ruimte met patiënten die uit ziekenhuizen overgeplaatst zijn en hebben intensief contact met verzorgend personeel en medebewoners. In **hoofdstuk 8** worden verschillende aspecten beschreven van de aanwezigheid van BRMO's in een Nederlands verpleeghuis, zoals de risicofactoren voor besmetting met een BRMO, de mogelijke verspreiding van BRMO's en de mogelijke rol van het darmmicrobioom op het verlenen van kolonisatie resistentie tegen BRMO's. Wij vonden dat antibioticagebruik in de twee maanden voor de monsterafname en een ziekenhuisopname in het voorafgaande jaar risicofactoren waren voor BRMO-dragerschap. Vervolgens hebben wij door het *sequencen* van de genomen van de BRMO's kunnen aantonen dat de BRMO-genomen

van bewoners op een bepaalde afdeling zo goed als identiek waren, wat sterk suggereert dat er binnen deze afdeling van het verpleeghuis zeer waarschijnlijk verspreiding van een BRMO heeft plaatsgevonden. Voor de microbioomanalyse hebben wij de bewoners opgesplitst in twee groepen: een groep die op ten minste één tijdspunt was gekoloniseerd met een BRMO en een groep die nooit gekoloniseerd was. In de groep die nooit gekoloniseerd was, werden hogere relatieve hoeveelheden van *Dorea*, *Atopobiaceae* en *Lachnospiraceae* groep ND3007 aangetroffen. Deze bacteriële taxa zijn dus mogelijk geassocieerd met bescherming tegen BRMO-kolonisatie. Tot slot observeerden wij in het microbioom van de meerderheid van verpleeghuisbewoners een hoge hoeveelheid *Bifidobacterium*, een bacterieel genus dat normaal in hoge aantallen wordt aangetroffen in de ontlasting van baby's en zeer jonge kinderen. Om uit te sluiten dat deze bevinding terug te leiden was naar probioticagebruik, hebben wij d.m.v. *metagenomic sequencing* de genomen van *Bifidobacterium* species vergeleken tussen verschillende bewoners. De *Bifidobacterium* genomen verschilden duidelijk van elkaar. We concludeerden dat er geen duidelijke gemeenschappelijke bron of overdracht tussen de verpleeghuisbewoners bestond. Aangezien wij de eerste onderzoeksgroep zijn die zo'n hoge hoeveelheid *Bifidobacterium* in de darmen van verpleeghuisbewoners hebben gevonden en we hier nog geen goede verklaring voor hebben kunnen vinden, blijven de oorzaak en het gevolg hiervan voorlopig onbekend.

Als vervolg op **hoofdstuk 8** is in samenwerking met het RIVM een studie opgezet waarin is gekeken naar het darmmicrobioom en -metaboloom van mensen die asymptomatisch gekoloniseerd zijn met *extended-spectrum beta-lactamase* (ESBL)-producerende *Escherichia coli* (**hoofdstuk 9**), een antibioticumresistente bacterie die door de Amerikaanse Centers for Disease Control and Prevention als een bedreiging voor de volksgezondheid wordt bestempeld. Een groot voordeel van deze samenwerking is dat wij data konden gebruiken uit het Pienter cohort van het RIVM. In dit cohort zijn data verzameld van duizenden volwassenen, waardoor er strenge inclusiecriteria toegepast konden worden en veel factoren konden worden uitgesloten waarvan bekend is dat ze een groot effect te hebben op de samenstelling van het microbioom. Gecombineerd met de kweekdata van ESBL-producerende bacteriën uit ongeveer 2750 personen, konden wij gebalanceerde groepen maken van ESBL-producerende *E. coli*-gekoloniseerde en niet-gekoloniseerde mensen op basis van klinische en epidemiologische variabelen. Uiteindelijk konden wij geen verschillen aantonen tussen het microbioom en metaboloom van gekoloniseerde en niet-gekoloniseerde mensen, wat suggereert dat microbioom-gemedieerde kolonisatieresistentie geen rol speelt tegen ESBL-producerende *E. coli*. Dit impliceert vervolgens dat microbioom-modificerende therapieën misschien niet geschikt zijn om voor dekolonisatie van deze bacterie uit de darm te zorgen.

Toekomstige microbioom-modificerende therapieën

Het uiteindelijke doel van onderzoek naar microbioom-gemedieerde kolonisatieresistentie is het ontwikkelen van therapieën die middels modificatie van het microbioom infecties kunnen voorkomen of genezen. Na een revolutionaire studie die in 2013 werd gepubliceerd, waarin FMT zeer effectief bleek tegen recidiverende *C. difficile*-infectie, was de verwachting dat FMT relatief snel vervangen zou worden door een consortium van goed gedefinieerde bacteriën, een LBP. Het grote voordeel van het toedienen van een LBP is dat men exact weet wat de patiënt krijgt en risico's daarmee geminimaliseerd kunnen worden, terwijl er bij een FMT ook mogelijk schadelijke micro-organismen of andere stoffen kunnen worden getransplanteerd. Het ontwikkelen van LBP's is echter lastiger gebleken dan verwacht en momenteel zijn er nog geen goedgekeurde therapieën gebaseerd op zulke consortia, niet voor recidiverende *C. difficile*-infectie en niet voor andere ziektes. Verschillende redenen hiervoor zijn uitgebreid besproken in **hoofdstuk 3**. Andere potentiële microbioom-modificerende therapieën zijn prebiotica (stoffen en voeding die de groei van specifieke bacteriën kunnen bevorderen) en synbiotica (het combineren van gedefinieerde bacteriële consortia met prebiotica), hoewel de klinische implementatie van deze twee middelen voor het bestrijden van infecties nog ver weg lijkt. Echter, dankzij de steeds betere mechanistische inzichten in de werking van het microbioom en het verlenen van kolonisatieresistentie, verwacht ik dat in de komende jaren de eerste LBP's zullen worden gebruikt in de kliniek om infecties effectiever te kunnen voorkomen en bestrijden. Op langere termijn zou deze toenemende kennis zelfs kunnen leiden tot meer gepersonaliseerde behandelingen, waarbij op basis van het microbioom van de patiënt en de bijbehorende ziekte verschillende LBP's gebruikt zouden kunnen worden.

Acknowledgements

This thesis is the result of extensive collaborations with many different groups and researchers. I would therefore like to acknowledge and thank all people, groups and institutes that contributed to and co-authored the chapters in this thesis. I would furthermore like to thank everyone from the Medical Microbiology department who has helped me during the last four years, the PhD students, technicians, postdocs, scientific staff, professors and last, but definitely not least, the secretaries. In addition, I would like to especially thank my students, the CPM and Parasitology departments for several enjoyable collaborations and the Zeller group members for making my stay in Heidelberg as nice as possible during COVID-19 times.

A few people I would like to thank more extensively:
Ed, you have truly been a wonderful promotor and I really could not have wished for a better one. You have always supported and challenged me to improve my science, explore new possibilities and you always provided feedback to manuscripts, abstracts and whatever else, at day and night. What always amazed me most was your never-ending enthusiasm for everything related to microbiology and I am sure this was an inspiration for many scientists, including myself. Thank you for your trust and belief in me.

Romy, like Ed, you have been instrumental in many of the chapters in this PhD thesis and for my development as a scientist in the past four years. You have been the ideal co-promotor for me during this PhD trajectory, as you always gave me complete freedom to explore my own interests and research directions, while still providing critical and useful feedback. In addition, I could always freely discuss any concerns I had with you. Thank you for all the support, ideas and the great supervision.

Bastian, while you were only my supervisor for 1.5 years, I still learned a great deal from you, especially with regard to critically evaluating research results. The different types of controls in microbiome research come to mind…

Anoe and Sam, I am very happy that you will be my paranymphs! Anoe, you were my ‘partner in crime’ when I started my PhD and continued to be so for almost my entire PhD time. Thank you for all the wet-lab support and the many chats both in and outside the lab. Sam, you joined the lab while I was 1.5 years into my PhD, but we quickly started working together quite a lot, something which has already led to some very nice publications.

Alvaro and Elys, thank you for all the support, especially during the time Paola was ill. I will not forget all the times you drove me back home late at night after another day in the hospital, even though you both must have been exhausted as well. Fabianna, we had many discussions about topics that are very far away from this thesis and I hope we'll have some more in the future!

Pap en mam, er is eigenlijk geen beginnen aan om jullie te bedanken voor alles wat jullie voor mij doen en hebben gedaan, daarom houd ik het kort. Bedankt voor al jullie steun, interesse in mijn onderzoek en het altijd klaar staan voor wat dan ook. Renzo, hoewel het niet de bedoeling was, heb ik door jouw vele verslagen en thesis ook nog van een soort parttime opleiding in accountancy kunnen genieten tijdens mijn PhD ;)

Paola, we did not have the easiest ride during my PhD time. There were far too many hours and days that we spent in the hospital, but I am extremely grateful and happy that you have recovered so well. I am very proud of, and impressed with, how you got your life back on track after such a trajectory. To many more happy and healthy years together!

List of Publications

1. Nooij S, ***Ducarmon QR***, Laros JFJ, Zwittink RD, Norman JM, Smits WK, Verspaget HW, Keller JJ, Terveer EM, Kuijper; working group of the Netherlands Donor Feces Bank. Faecal microbiota transplantation influences procarcinogenic *Escherichia coli* in recipient recurrent *Clostridioides difficile* patients. *Gastroenterology*. 2021 Oct;161(4):1218-1228.e5. doi: 10.1053/j.gastro.2021.06.009. Epub 2021 Jun 11.
2. Mesnage R, Teixeira M, Mandrioli D, Falcioni L, Ibragim M, ***Ducarmon QR***, Zwittink RD, Amiel C, Panoff JM, Bourne E, Savage E, Mein CA, Belpoggi, F, Antoniou MN. Multi-omics phenotyping of the gut-liver axis reveals metabolic perturbations from a low-dose pesticide mixture in rats. *Commun Biol.* 2021 Apr 14;4(1):471. doi: 10.1038/s42003-021-01990-w.
3. ***Ducarmon QR****, Terveer EM*, Nooij S, Bloem MN, Vendrik KEW, Caljouw MAA, Sanders IMJG, van Dorp SM, Wong MC, Zwittink RD*, Kuijper EJ*. Microbiota-associated risk factors for asymptomatic gut colonisation of multi-drug-resistant organisms in a Dutch nursing home. *Genome Med*. 2021 Apr 7;13(1):54. doi: 10.1186/s13073-021-00869-z. *These authors contributed equally
4. ***Ducarmon QR***, Kuijper EJ, Olle B. Opportunities and Challenges in Development of Live Biotherapeutic Products To Fight Infections. *J Infect Dis*. 2021 Jun 16;223(Supplement_3):S283-S289. doi: 10.1093/infdis/jiaa779.
5. Mesnage R, Teixeira M, Mandrioli D, Falcioni L, ***Ducarmon QR***, Zwittink RD, Mazzacuva F, Caldewell A, Halket J, Amiel C, Panoff JM, Belpoggi, F, Antoniou MN. Use of Shotgun Metagenomics and Metabolomics to Evaluate the Impact of Glyphosate or Roundup MON 52276 on the Gut Microbiota and Serum Metabolome of Sprague-Dawley Rats. *Environ Health Perspect.* 2021 Jan;129(1):17005. doi: 10.1289/EHP6990. Epub 2021 Jan 27.
6. Golloso-Gulbat MJ, ***Ducarmon QR***, Tan RCA, Zwittink RD, Kuijper EJ, Nacis JS, Santos NLC. Gut microbiota and dietary intake of normal-weight and overweight Filipino children. *Microorganisms*. 2020;8(7):E1015. Published 2020 Jul 8. doi:10.3390/microorganisms8071015
7. ***Ducarmon QR***, Hornung BVH, Geelen AR, Kuijper EJ, Zwittink RD. Towards standards in clinical microbiota studies: comparison of three DNA extraction methods and two bioinformatic pipelines. *mSystems.* Feb 2020, 5 (1) e00547-19; doi: 10.1128/mSystems.00547-19.
8. Crobach MJT*, ***Ducarmon, QR****, Terveer, EM, Harmanus, C, Sanders, IMJG, Verduin, KM, Kuijper, EJ*, Zwittink, RD*. The Bacterial Gut Microbiota of Adult Patients Infected, Colonized or Noncolonized by *Clostridioides difficile*. *Microorganisms*. 2020 May, 6;8(5):677. doi: 10.3390/microorganisms8050677. *These authors contributed equally
9. ***Ducarmon QR***, Hoogerwerf MA, Janse JJ, Geelen AR, Koopman JPR, Zwittink RD,

Goeman JJ, Kuijper EJ, Roestenberg M. Dynamics of the bacterial gut microbiota during controlled human infection with *Necator americanus* larvae. *Gut Microbes*. 2020 Nov 9;12(1):1-15. doi: 10.1080/19490976.2020.1840764.

10. Vendrik KEW, Ooijevaar RE, de Jong PRC, Laman JD, van Oosten BW, van Hilten JJ, ***Ducarmon QR***, Keller JJ, Kuijper EJ, Contarino MF. Fecal microbiota transplantation in neurological disorders. *Front Cell Infect Microbiol*. 2020;10:98. Published 2020 Mar 24. doi:10.3389/fcimb.2020.00098
11. ***Ducarmon QR***, Zwittink RD, Hornung BVH, van Schaik W, Young VB, Kuijper EJ. Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. *Microbiol Mol Biol Rev*. 2019;83(3):e00007-19. Published 2019 Jun 5. doi:10.1128/MMBR.00007-19
12. Hornung BVH, Zwittink RD, ***Ducarmon QR***, Kuijper EJ. Response to: 'Circulating microbiome in blood of different circulatory compartments' by Schierwagen *et al.* Gut. 2020 Apr;69(4):789-790. doi: 10.1136/gutjnl-2019-318601. Epub 2019 Apr 6.

PhD Portfolio

PhD student:	Quinten R. Ducarmon
LUMC department:	Medical Microbiology
Promotor:	prof. dr. Ed J. Kuijper
Copromotor:	dr. Romy D. Zwittink
Research programme:	Molecular basis of bacterial pathogenesis, virulence factors and antibiotic resistance
Title of Thesis:	**Microbiome-mediated colonization resistance**: defense against enteropathogens and multi-drug resistant organisms

PhD training	Year	Hours
Mandatory courses		
PhD Introductory Meeting	2018	5
BROK Course	2018	42
Basic Methods and Reasoning in Biostatistics	2018	42
Generic/disciplinary courses		
Data and Projectmanagement	2018	28
Practical Linux	2018	11
MGC Technology Facilities course	2018	42
Code and data management with Git	2018	14
Using R for data analysis	2018	32
Open & Reproducible Microbiome Data Analysis	2018	28
RNASeq course data analysis	2018	56
How to make an academic poster	2018	8
ICME 10 (Hands-on training in Prokaryotic and Eukaryotic metagenomics)	2019	40
Analysis of Repeated Measurements	2019	42
Negotiation for PhDs	2020	8
Machine learning for bioinformatics & systems biology	2019	84
Multiomics Data Integration in R	2020	56
Attended lectures and participation in meetings		
Lecture Series Micropia	2018	10
Sentinal Surveillance voor *Clostridium difficile* infectie meeting	2018	4
Gut Day	2018	8
Gut Day	2019	8
Congress attendance and poster or oral presentations		
Scientific Spring Meeting KNVMM & NVMM	2018	16
Scientific Spring Meeting KNVMM & NVMM	2019	16
ECCMID	2019	32

PhD training	Year	Hours
Congress attendance and poster or oral presentations		
Exploring Human Host-Microbiome Interactions in Health and Disease	2019	24
Scientific Spring Meeting KNVMM & NVMM	2021	16
ECCMID	2021	32
Grants and Awards		
FEMS Travel Grant		
FEMS Research & Training Grant		
Poster Prize (Runner-up) at Exploring Human Host-Microbiome Interactions in Health and Disease, 2019.		

Curriculum vitae

Quinten Ducarmon was born on March 8th, 1994 in Terneuzen, the Netherlands. After graduating from 'S.S.G. De Rede' in 2011, he started the Pre-Med program at University College Roosevelt (UCR) and graduated *cum laude* in January 2015. During these years, he also became an International Master in chess and participated, amongst others, in the World Junior Championship in 2014 in Pune, India. He obtained an MSc degree in Nutrition and Health at Wageningen University & Research (WUR) with a specialization in Nutritional Physiology. During his MSc, he performed an internship at the Systems Biology group at Netherlands Organisation for Applied Scientific Research (TNO) which focused on the effect of a nutritional intervention on the gut microbiome. This is where his interest in the role of the microbiome in health and disease was really triggered. In January 2018 he started his PhD research at LUMC's Center for Microbiome Analyses and Therapeutics (CMAT) under supervision of prof. dr. Ed Kuijper and dr. Romy Zwittink. Here, Quinten focused on studying the role of the microbiome in providing colonization resistance against enteropathogens and multi-drug resistant organisms and this topic allowed for various collaborations, both in and outside of the LUMC. In early 2021, Quinten spent four months at the European Molecular Biology Laboratory (EMBL) in Heidelberg in the group of dr. Zeller, where he developed the first computational tool capable of directly profiling carbohydrate-active enzymes from human gut metagenomes. He will continue his academic career as a postdoctoral researcher in the group of dr. Zeller. Together with Paola Flores, Quinten lives in Oegstgeest.