Rational Design of Live Biotherapeutic Products for the Prevention of *Clostridioides difficile* Infection

Shanlin Ke[1], Javier A Villafuerte Gálvez[2], Zheng Sun[1], Yangchun Cao[2,3], Nira R Pollock[4,5], Xinhua Chen[2,#], Ciarán P Kelly[2,#], Yang-Yu Liu[1,6,#]

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.
[2]Division of Gastroenterology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.
[3]College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China.
[4]Division of Infectious Disease, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA.
[5]Department of Laboratory Medicine, Boston Children's Hospital, Boston, Massachusetts, USA.
[6]Center for Artificial Intelligence and Modeling, The Carl R. Woese Institute of Genomic Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA.

[#] Correspondence: Y.-Y.L. (yyl@channing.harvard.edu); C.P.K. (ckelly2@bidmc.harvard.edu); X.C. (xchen1@bidmc.harvard.edu).

Abstract

Clostridioides difficile infection (CDI) is one of the leading causes of healthcare- and antibiotic-associated diarrhea. While fecal microbiota transplantation (FMT) has emerged as a promising therapy for recurrent CDI, its exact mechanisms of action and long-term safety are not fully understood. Defined consortia of clonal bacterial isolates, known as live biotherapeutic products (LBPs), have been proposed as an alternative therapeutic option. However, the rational design of LBPs remains challenging. Here, we employ a computational pipeline and three independent metagenomic datasets to systematically identify microbial strains that have the potential to inhibit CDI. We first constructed the CDI-related microbial genome catalog, comprising 3,741 non-redundant metagenome-assembled genomes (nrMAGs) at the strain level. We then identified multiple potential protective nrMAGs that can be candidates for the design of microbial consortia targeting CDI, including strains from *Dorea formicigenerans*, *Oscillibacter welbionis*, and *Faecalibacterium prausnitzii*. Importantly, some of these potential protective nrMAGs were found to play an important role in the success of FMT, and the majority of the top protective nrMAGs can be validated by various previously reported findings. Our results demonstrate a computational framework for the rational selection of microbial strains targeting CDI, paving the way for the computational design of microbial consortia against other enteric infections.

Introduction

Clostridioides difficile infection (CDI) is one of the leading causes of healthcare- and antibiotic-associated diarrhea, affecting roughly 500,000 patients and leading to almost 30,000 deaths annually in the United States[1,2]. Exposure to toxinogenic *C. difficile* can lead to a spectrum of clinical outcomes, including asymptomatic colonization, mild diarrhea, and more severe disease syndromes such as pseudomembranous colitis, toxic megacolon, bowel perforation, sepsis, and death[3]. Antibiotics serve as the standard treatment for primary CDI[4,5]. However, CDI recurrence occurs in approximately a quarter of cases after antibiotic treatment[6,7]. Once CDI recurs, patients may get into a vicious cycle of antibiotic therapy and relapse[8]. Moreover, the use of antibiotics has been identified as the primary risk factor for developing CDI, and reports of strains with decreased sensitivity to vancomycin are becoming more frequent.

The human gut microbiome is critical in providing colonization resistance against exogenous pathogens through complex mechanisms such as nutrient competition, competitive metabolic interactions, niche exclusion, and induction of the host immune response[9]. Intestinal microbiota restoration, such as fecal microbiota transplantation (FMT), has been shown to be effective for CDI treatment as well as the restoration of colonization resistance against *C. difficile*[10,11]. While FMT has emerged as a promising therapy for recurrent CDI (rCDI), its exact mechanisms of action are not fully understood[12]. In addition, FMT has the potential to transmit undetected or emerging pathogens, which may result in hospitalization or even death[13,14]. Recently, the FDA has approved fecal microbiota products (e.g., Rebyota[15] and Vowst[16]) for the prevention of rCDI in individuals 18 years of age and older, following antibiotic treatment for rCDI. Rebyota is a room temperature shelf stable suspension of healthy donor stool[17], although its clinical effect size for the prevention of rCDI is modest (RR, 1.17; 95% CI, 0.99–1.39)[18] and its microbial composition is not predefined[19]. Although Vowst is a formulation of live fecal microbiota consisting of a highly purified collection of about 50 species of *Firmicutes* spores with a more robust clinical effect size (1.46; 95% CI, 1.21–1.75)[18], the ecological principle underlying the selection of these microbial strains is unclear.

The variability of biological properties among bacterial strains within the same species underscores the significance of conducting strain-level composition analysis to understand the role of the human microbiome in human health and disease[20]. For example, some strains from *Escherichia coli* (e.g., *E. coli O157:H7*) cause severe abdominal pain, bloody diarrhea, and vomiting[21]. In contrast, *E. coli Nissle 1917* is a non-pathogenic strain that has been utilized as a probiotic agent to treat gastrointestinal infections in humans[22,23]. Whole metagenome shotgun (WMS) sequencing is a rapid, cost-effective, and high-throughput technology for profiling microbial communities in human microbiome studies[24]. However, precise identification of microorganisms at the strain level remains challenging. Additionally, traditional strain-level profilers can only identify strains within the reference genome databases[25]. These databases are subject to limitations and biases and are unable to characterize microbes that do not have high-quality reference genomes. To resolve these limitations, an alternative strategy for WMS data analysis involves reconstructing metagenome-assembled genomes (MAGs) through *de novo* assembly and binning, offering the advantage of recovering genomes for uncultured microorganisms absent from current reference databases[26].

In this study, we leveraged a novel computational framework[27] we previously developed to rationally design a bacterial consortium against CDI (**Fig. 1**). The metagenome assembly and binning strategies were applied to reconstruct microbial population genomes directly from the microbiome samples of two independent CDI-related cohorts as well as the healthy controls from the Human Microbiome Project (HMP). Specifically, we sought to identify known and unknown taxa at the strain level, quantify the degree of donor strain engraftment, and design a candidate bacterial consortium against CDI.

Results

Study cohorts and metagenomic datasets

To rationally design microbial consortia against *C. difficile*, we aimed to infer species that may inhibit *C. difficile* from CDI-related microbiome samples. We first collected WMS sequencing data from our in-house clinical cohort (denoted as BIDMC-cohort hereafter) [28,29]. Our BIDMC cohort consists of 104 well-characterized recruited participants divided into four groups (**Table S1** and **Fig. 2a**): (1) Control (CON, n = 26); (2) Non-CDI Diarrhea (NCD, n = 14); (3) Asymptomatic Carriage of *C. difficile* (ASC, n = 17); (4) CDI (n = 47). Given the fact that the participants from the CON group were not healthy people, we retrieved an FMT study[30] (denoted as Verma-cohort hereafter) with publicly available data that assessed the microbiome composition of donors (n=21) and recipients (n=22, pre-and post-FMT) through WMS sequencing (**Methods** and **Fig. 2a**). In addition, we included two sets of randomly selected metagenome samples of healthy adults (n = 94) from the Human Microbiome Project (HMP)[31].

A high-quality microbial genome catalog

Following quality control, we performed metagenomic assembly and binning on those microbiome samples from three cohorts, yielding 7,769 MAGs. To evaluate the highest quality representative genomes, we dereplicated the 7,769 MAGs at an average nucleotide identity (ANI) threshold of 99%, resulting in a final set of 3,741 non-redundant MAGs (nrMAGs) with strain-level resolution. The nrMAGs were contributed by HMP, Verma-cohort, and the BIDMC-cohort in proportions of approximately 37%, 23%, and 40%, respectively **(Fig. 2b**). In particular, our findings indicate that recipients prior to FMT made a smaller contribution to the nrMAG collection compared to donors and recipients after FMT, suggesting a reduced microbial diversity **(Fig. S1**). These nrMAGs exhibited a mean completeness of 88%, mean contamination of 0.93%, mean genome size of 2.5 megabases (Mb), and mean N50 of 65.7 kilobases (kb) (**Fig. 1b-c** and **Fig. 2c-f**). Out of the 3,741 strain-level nrMAGs, 1,390 (37.16%) nrMAGs met medium-quality criteria (50% ≤ completeness < 90%, and ≤5% contamination), while 2,351 (62.84%) nrMAGs exhibited high-quality (≥ 90% completeness, and ≤ 5% contamination)[32,33] (**Fig. 1b**).

Using the Genome Taxonomy Database[34], these nrMAGs were taxonomically assigned to 17 phyla, 22 classes, 47 orders, 104 families, and 408 genera, spanning across 883 species. Most of them belonged to Firmicutes_A (60.97%), followed by Bacteroidetes

(10.18%), and Actinobacteria (10.00%). The phyla information of nrMAGs was summarized in **Fig. 1b-c**. Among those 883 species, *Agathobacter rectalis*, *Blautia_A wexlerae*, *Gemmiger formicilis*, *Fusicatenibacter saccharivorans*, and *Bifidobacterium longum* were the top five species with the highest strain-level diversity (i.e., number of nrMAGs identified within a specific species, **Fig. 2g**).

Microbial diversity

We initially investigated alpha diversity in the human microbiome at the nrMAG level. Alpha diversity measures (i.e., Richness and Shannon index) were compared among different groups from the same cohorts (**Fig. 2h-j**). No significant differences were found between the two randomly selected sets from the HMP (**Fig. 2h** and **Fig. S2a**). In accordance with the original study[30], we found that the Richness and Shannon indices of the gut microbiome in the recipients of pre-FMT were significantly lower than those in donors. After FMT, those recipients showed similar alpha diversity to donors (**Fig. 2i** and **Fig. S2b**). In the BIDMC-cohort, we found that only the CDI group showed significantly lower alpha diversity than the CON group (**Fig. 2j** and **Fig. S2c**). Participants from the ASC group only showed a significantly lower number of identified nrMAGs than CON group participants.

Principal coordinate analysis (PCoA) based on robust Aitchison distance, combined with PERMANOVA (permutational multivariate analysis of variance, a statistical method commonly used for testing the association between the microbiome and a covariate of interest), revealed no significant difference in the gut microbial community structure at the nrMAG-level between the two datasets from HMP (**Fig. 2k**). We found that the microbiomes of the donor and the recipients from pre-and post-FMT were compositionally distinct in the Verma-cohort (*P* = 0.0001, PERMANOVA **Fig. 2l**). Consistent with our previous study using 16S rRNA gene sequencing data[29], the overall microbial composition differed significantly among different groups in the BIDMC-cohort (*P* = 0.0001, PERMANOVA **Fig. 2m**).

Identify potential permissive and protective nrMAGs for the rational design of live biotherapeutics.

To identify candidate strains for the development of microbiota-derived biotherapeutics , we applied the generalized microbe-phenotype triangulation (GMPT) method, moving beyond the standard association analysis[27]. The GMPT relies on the following core hypothesis: species that are differentially abundant in most pairwise phenotype-based comparisons and whose abundances display a strong negative (or positive) correlation with the abundance of the pathogen tend to be causal preventive (or permissive) species that directly inhibit (or promote) the growth of the pathogen[27]. Our GMPT analysis incorporated microbiome data from the BIDMC-cohort, donor data from the Verma-cohort[30], and one set from HMP. Since we have two sets of randomly selected metagenome samples of healthy adults from the HMP, we systematically included one set of HMP microbiome data at a time to cross-validate the results between the two datasets. Then, we conducted pairwise comparisons for six individual phenotype groups, which encompassed CDI, ASC, NCD, and CON from the BIDMC-cohort, donors from the Verma-cohort, and a dataset from HMP.

Applying this approach to the data with the first set of HMP data, 15 pairwise differential abundance analyses generated a total of 1,349 nrMAGs present in at least one pairwise comparison (**Table S6**). To explore the potential relationship between those candidate nrMAGs and CDI, we calculated Spearman correlation coefficients between the average relative abundances of nrMAGs and pragmatic severity scores in a continuum of non-CDI controls and *C. difficile* colonized and infected subjects (i.e., HMP healthy controls: 0; Donor from Verma et al.[30]: 1; CON: 2; NCD: 3; ASC: 4 and CDI: 5) in different phenotypes. Similarly, we identified a total of 1,390 nrMAGs present in at least one pairwise comparison with the second set of HMP data (**Table S7**). Among the protective nrMAGs between the two runs with HMP data, 80.77% (525/650) and 81.14% (525/647) of them were overlapped, respectively. We then computed the average rank between two runs based on the frequency (**Table S8)**. Among the top 40 potential protective nrMAGs, the dominant species were *Dorea formicigenerans*, *Oscillibacter welbionis*, *Faecalibacterium prausnitzii*, *GCA-900066135 sp900066135*, *Bariatricus comes*, *Phocaeicola dorei*, *Anaerobutyricum hallii*, *Bacteroides ovatus*, *Blautia_A obeum*, *Mediterraneibacter faecis*, *Alistipes putredinis*, *Odoribacter splanchnicus*, *Streptococcus salivarius*, and *Dorea longicatena* (**Table 1**). Through a systematic review of literature, we found that most of our candidate strains have been reported to be protective from CDI or non-CDI antibiotic associated diarrhea at higher taxonomical levels (e.g., species and genus levels) across existing studies (**Table 1**). These findings support the validity of our methods.

The protective strains play an important role in FMT.

To further validate the potential role of the protective strains we identified from the GMPT pipeline, we systemically tracked the microbiome changes of the recipients who underwent FMT in the Verma-cohort. We aimed to investigate if those protective strains also play an important role in the success of FMT. Notably, the microbiome samples from the recipients in the Verma cohort were not included in our previous GMPT analysis.

First, we examined the gain and loss of microbial strains before and after FMT to assess the transfer and engraftment of the donor microbiome in the recipient. For donor, pre-FMT recipients, and post-FMT recipients, we identified 3,129, 2,093, and 3,054 nrMAGs, respectively. Notably, post-FMT recipients showed a loss of 33 nrMAGs (**Fig. 3a**), with the majority of the lost strains attributed to species such as *Anaeroglobus micronuciformis*, *Phascolarctobacterium faecium*, *Fusobacterium polymorphum*, and *Duodenibacillus sp003472385* from Firmicutes_C, Proteobacteria, and Fusobacteriota (**Fig. 3b**). On the contrary, all recipients exhibited a gain of 923 nrMAGs from their donors (**Fig. 3a**). The majority of these engrafted strains were taxonomically annotated to Actinobacteriota and Firmicutes_A, such as species like *Ruminococcus_D bicirculans*, *Faecalibacterium prausnitzii*, *Faecalibacterium prausnitzii_G*, *Agathobacter rectalis*, *Agathobacter faecis*, *Acetatifactor sp900066565*, *Bifidobacterium adolescentis*, and *Collinsella aerofaciens_G* (**Fig. 3c-d**).

For each donor-recipient pair, we then calculated the difference in their gut microbial community structure before and after FMT using the robust Aitchison distance. Our findings indicate that the distance between donors and recipients was significantly reduced after FMT compared to the pre-FMT state (**Fig. 3e**). Additionally, calculating the strain share rate for each donor-recipient pair before and after FMT revealed agreement with our previous finding that recipients gained more strains, shared a greater number of strains with the donor after FMT (**Fig. 3f**).

Changes in the microbiome induced by FMT not only indicate the transfer and engraftment of the donor microbiome but also involve alterations in the abundance of coexisting strains. To address this question, we conducted the differential abundance analysis among three groups. Consistent with the robust Aitchison distance and strain share rate analyses, we only identified less differential abundant strains between donor and post-FMT recipients (**Fig. 3g and Table S2**). We have identified 223 and 238 differential abundant nrMAGs from the comparison of donor vs. pre-FMT recipients (**Fig. 3g and Table S3**) and pre-FMT recipients vs. post-FMT recipients (**Fig. 3g and Table S4**), respectively. Among these differential abundant nrMAGs, we found 179 overlapped strains (**Fig. 3h, Table S5**), including strains from *Blautia_A wexlerae*, *Veillonella parvula_A*, *Veillonella parvula*, *Blautia_A sp900066165*, *Escherichia coli_D*, *Escherichia flexneri*, *Anaeroglobus micronuciformis*, *Blautia_A obeum*, *Lacticaseibacillus rhamnosus*, and *Veillonella dispar_A.* Specifically, we observed significant increases in some candidate protective strains following FMT. These include multiple strains from *Dorea formicigenerans*, *Mediterraneibacter faecis*, *Phocaeicola dorei*, *Blautia_A wexlerae*, and *Blautia_A obeum*. This finding further validated the potential role of protective strains in treating CDI.

Discussion

The growing interest in FMT as a therapeutic approach stems from its high success rate in treating recurrent CDI, leading to an exploration of its potential for addressing various human diseases[35]. However, FMT remains an unstandardized procedure with unclear mechanisms and long-term safety concerns[35,36]. Therefore, an advantage of microbial consortia over “whole stool” FMT is the introduction of a group of specific microbiota that can precisely target and effectively treat a disease while minimizing clinical risks. In this study, we used a computational pipeline to directly identify candidate bacterial strains from a diverse CDI-related metagenomic dataset, thereby facilitating the targeted development of microbial therapies and advancing our understanding of CDI pathogenesis and treatment.

By tracking the dynamic changes in gut microbiome data undergoing FMT, we identified significant shifts in the microbial structure of the recipients. Although we did not utilize the microbiome data from recipients before and after FMT in our GMPT pipeline, we found that some of the top ranked candidate protective strains showed significant increases after FMT, including multiple strains from *Dorea formicigenerans*, *Mediterraneibacter faecis*, *Phocaeicola dorei*, and *Blautia_A obeum*. This finding provides an additional layer of validity to our method. In addition, we performed a systematic literature review on the highest ranked candidate protective strains and found that the majority of them have been reported to have various protective roles at species or genus levels in the CDI continuum: negative association with *C. difficile* colonization, infection and severity. We found clustering of the main protective species within the families Lachnospiraceae, Bacteroidaceae, and Oscillospiraceae. For example, *F. prausnitzii*, a beneficial human gut microbe touted as a candidate for next-generation probiotics[37], was found to have reduced abundance in CDI patients, which was restored after FMT[38]. Interestingly, we have also identified a protective strain of the species *Dorea longicatena,* which is a component of a defined bacterial consortium (VE303) with encouraging Phase 2 clinical data, consisting of eight, nonpathogenic, nontoxigenic, commensal strains of Clostridia[39].

In addition to the potential protective strains, we also identified multiple permissive strains of *C. difficile*, including strains from *Enterococcus_B faecium* and *Eggerthella lenta*. This aligns with a previous study reporting that enterococci (including *E. faecium*) can enhance the fitness and pathogenicity of *C. difficile* via shaping the metabolic environment in the gut and reprograming *C. difficile* metabolism[40]. Additionally, *E. lenta* is an anaerobic gram-positive bacillus associated with polymicrobial intraabdominal infections[41]. Therefore, the potentially permissive strains that we identified from this study offer the opportunity to further understand how *C. difficile* interacts with the rich community of microorganisms in the colon. Moreover, in alignment with the variation in biological properties among bacterial strains within the same species[42], we have observed distinct roles played by different strains of *F. prausnitzii_D* in the context of CDI. This underscores the critical importance of conducting studies at the strain level.

The current study has some limitations. First, we leveraged metagenomic data from three independent datasets with technological variations, including differences in sequencing depth. Second, we did not pre-define a strict threshold to select potential protective strains from the candidate list for further experimental validations. Lastly, the inference of the efficacy of candidate protective strains against CDI is limited by the current computational algorithm. To test the efficacy of our proposed microbial consortia and gain a deeper understanding of exact mechanisms, the utilization of techniques of metabolomics and immunological approaches, along with direct *in vitro* and *in vivo* experiments, are necessary.

Taken together, our results provide compelling evidence for the rational design of microbial consortia against *C. difficile*. Many of the candidates detected here replicate previously reported findings, supporting the validity of our results. Importantly, our work paves the way for the design of LBPs against general microbiome-related diseases.

Methods

Study cohorts

Dataset I: BIDMC-cohort. The background and design of this cohort have been detailed in our previous studies[28,29]. This clinical cohort consists of 104 well-characterized recruited participants, who were divided into four groups associated with different *C. difficile* infection/colonization statuses: (1) *C. difficile* infection (CDI, n=47): Eligible patients were inpatients ≥ 18 years old with new-onset diarrhea, positive clinical stool NAAT (Xpert *C. difficile*/Epi) result, and a decision to treat for CDI; (2)

Asymptomatic Carriage (ASC, n = 17): Eligible patients were inpatients ≥ 18 years old, admitted for at least 72 hours, who had received at least one dose of an antibiotic within the past seven days, and did not have diarrhea in the 48 hours prior to stool specimen submission, and positive clinical stool NAAT result; (3) Non-CDI Diarrhea (NCD, n = 14): patients with diarrhea (confirmed using the same definition used for the CDI cohort) but had NAAT-negative stool on clinical C. difficile testing; and (4) Control (CON, n = 26): patients without diarrhea who had screened as eligible for the ASC cohort but were NAAT-negative on research stool testing. DNA of fecal samples (200 mg) were extracted using Mag-Bind® Universal Metagenomics Kit (Product# M5633-01, Omega Biotek) and DNeasy PowerSoil Kit (Catalog# 12888-100, Qiagen) according to manufacturer's instructions. The quality of the extracted DNA was measured by 1% agarose gel electrophoresis and Qubit® 3.0 Fluorometer (ThermoFisher). Subsequently, the extracted DNAs were used for shotgun metagenomic library construction, and sequencing was performed on the Illumina HiSeq X Ten platform, generating a 150 bp paired-end library for each sample.

Dataset II: Verma-cohort. In the study conducted by Verma et al[30], fecal samples were collected from 22 patients with recurrent CDI before and after FMT and their corresponding healthy donors (n=21, with one donor providing fecal samples for two different recipients). Eight-seven WMS human gut metagenomes were downloaded from this study via NCBI Sequence Read Archive (BioProject ID PRJNA705895). The clinical outcome in recurrent CDI patients after FMT was determined by the symptomatic resolution of CDI[30]. Clinical symptoms such as diarrhea, bloating, abdominal pain, and cramping were alleviated in all patients within 3–7 days following FMT[30].

Dataset III: Human Microbiome Project. Human gut metagenomes (Ninety-eight individuals) were randomly selected from HMP data (https://portal.hmpdacc.org/)[43]. All samples are from the HMP study[31] and are healthy adult subjects. In total, ninety-four human gut metagenomes were randomly selected based on the largest group size in our clinical cohort. To cross-validate the main findings, we randomly divided the HMP data into two sets in the downstream analyses.

Metagenome assembly and binning

Genome reconstruction of the human microbiome using metagenomic sequencing data was executed through the functional modules of metaWRAP (v1.3.2)[44]. All metagenomic sequencing data underwent quality control and removal of human contamination using metaWRAP-Read_qc. Clean reads were then assembled with the metaWRAP-Assembly module using metaSPAdes (v3.13.0)[45]. The assembled contigs were binned into bins using three metagenomic binning tools: MetaBAT (v2.12.1)[46], MaxBin (v2.2.6)[47], and CONCOCT (v1.0.0)[48]. The default minimum length of contigs used for constructing bins with MaxBin2 and CONCOCT was 1000 bp, and metaBAT2 was defaulted to 1500 bp[44]. The bins from each binning tool were integrated and refined with Bin_refinement module of metaWRAP with options "-c 50 -x 10", corresponding to the criterion of medium-quality draft MAGs[32]. CheckM (v1.0.12)[49] was used to estimate the completeness and contamination of the bins, and the minimum completion and maximum contamination were 50% and 10%, respectively.

De-replication of MAGs and genome annotation. All 7,776 MAGs underwent de-replication into non-redundant MAGs (nrMAGs) using dRep (v3.0.0) (≥50% genome completeness and ≤5% contamination)[50]. Initially, MAGs from three cohorts were divided into primary clusters using Mash at a 90% Mash ANI. Then, each primary cluster was used to form secondary clusters at the threshold of 99% ANI with at least 30% overlap between genomes[51]. Taxonomic annotation of all nrMAGs was conducted using GTDB-Tk (v.1.4.1)[52] based on the Genome Taxonomy Database (http://gtdb.ecogenomic.org/)[34], providing standardized taxonomic labels for subsequent analysis in this study.

Abundance estimation and phylogenetic analysis of nrMAGs

The metaWRAP-Quant_bins module coupled with Salmon (v0.13.1)[53] was employed to access the abundance of each nrMAGs in each metagenomic sample. The phylogenetic tree of the nrMAGs was constructed using PhyloPhlAn (v3.0.58)[54] and visualized through iTOL (https://itol.embl.de/)[55].

Statistical analysis

Microbial alpha diversity measures were calculated at the nrMAGs level using R vegan package (v2.5.7), and principal coordinates analysis (PCoA) plots were generated using robust Aitchison distance[56]. Differences in microbiome compositions across different groups were tested by the permutational multivariate analysis of variance (PERMANOVA) using the “adonis” function in R vegan package. All PERMANOVA tests were performed with 9999 permutations based on the robust Aitchison distance. We defined strain-sharing rates as the total number of shared strains between two samples divided by the number of common species identified from the two samples. Differences between the groups were analyzed using a Wilcoxon–Mann–Whitney test. For differential abundance analysis and GMPT (Generalized Microbe Phenotype Triangulation) pipeline[27], we used ANCOM (analysis of composition of microbiomes)[57], with a Benjamini–Hochberg correction at a 5% level of significance. All statistical analysis was performed with R (version 3.6.3).

Data availability

Metagenomic data from HMP are available via https://portal.hmpdacc.org. The metagenomic data from the study of Verma et al.[30] can be downloaded via NCBI Sequence Read Archive (BioProject ID PRJNA705895). The metagenomic data from the BIDMC-cohort is available in the NCBI Bioproject under accession code PRJNA1067975. Metagenome-assembled genomes for all samples are available on Figshare (https://doi.org/10.6084/m9.figshare.25355857).

Code availability

The code for the construction of the MAGs catalog and statistical analysis and visualization is available in the GitHub repository (https://github.com/ShanlinKe/CDI).

Acknowledgements

The authors thank all patients who participated in this study, as well as the technologists in the Beth Israel Deaconess Medical Center Clinical Microbiology Laboratory, for their help with sample collection. This study was supported by the China Scholarship Council

(201806305024 to Y.C.), National Key Research and Development Program (grants 2021YFD1300301 to Y.C.), Shaanxi Science Fund for Distinguished Young Scholars (grants 2024-JC-JCQN-25 to Y.C.), the Ningxia Key Project of Research and Development Plan (grants 2023BFC01036 Y.C.), the National Institute of Allergy and Infectious Diseases (grants R01AI141529 to Y.-Y.L., R01AI116596 to N.R.P. and C.P.K., and K23AI177749-01 to J.A.V.G.), Institute Merieux (grant to N. R. P. and C. P. K.), the Irving W. and Charlotte F. Rabb Award (to X.C.) and Milky Way Life Sciences (contract to C.P.K. and Y.-Y.L).

Author contributions

Y.-Y.L., X.C. and C.P.K. conceived and designed the project. N.R.P. and C.P.K. planned and performed the human studies and sample collections. S.K. performed all the data analysis. S.K., J.A.V.G., and Y.-Y.L. interpreted the results and prepared the manuscript. Z.S., X.C., and C.P.K. interpreted the results, reviewed and edited the manuscript. Y.C. acquired the raw sequencing data, reviewed and edited the manuscript. All authors have read and approved the manuscript.

Competing interests

C.P.K. has acted as a paid consultant to: Artugen Therapeutics, Facile Therapeutics, Finch, Fzata, Glaxo Smith Kline, Immunics Therapeutics, Recursion Pharmaceuticals, RVAC Medicines, Sanofi Pasteur, Seres Therapeutics, and Summit Therapeutics; C.P.K. has acted as a paid consultant and member of the Scientific Advisory Board to: Acurx Pharmaceuticals, Anokion, Ferring Pharma, Inova Diagnostics, Janssen Pharmaceuticals, Merck & Company, Milky Way Life Sciences, Pfizer, Takeda, and Vedanta Biosciences; C.P.K. has acted as an unpaid consultant and had private equity in Glutenostics; has acted as a paid consultant and has private equity in Cour Pharmaceuticals, and First Light Biosciences, Inc; and has acted as a study investigator for: Milky Way Life Sciences and Merck. X.C. is a paid consultant and board member of Milky Way Life Sciences and has acted as a consultant to Artugen Therapeutics and RVAC Medicines. X.C. is a co-founder with private equity in Milky Way Life Sciences and TaoTe Technology, which also co-founded Milky Way Life Sciences. Milky Way Life Sciences is a donor to BIDMC's Center for Nutritional Health and funds clinical trials at Beth Israel Deaconess Medical Center.

References

1 Magill, S. S. *et al.* Changes in Prevalence of Health Care-Associated Infections in U.S. Hospitals. *N Engl J Med* **379**, 1732-1744, doi:10.1056/NEJMoa1801550 (2018).

2 Sehgal, K. & Khanna, S. Gut microbiome and Clostridioides difficile infection: a closer look at the microscopic interface. *Therap Adv Gastroenterol* **14**, 17562848211994736, doi:10.1177/17562848211994736 (2021).

3 Rupnik, M., Wilcox, M. H. & Gerding, D. N. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* **7**, 526-536, doi:10.1038/nrmicro2164 (2009).

4 Nelson, R. L., Suda, K. J. & Evans, C. T. Antibiotic treatment for Clostridium difficile-associated diarrhoea in adults. *Cochrane Database Syst Rev* **3**, CD004610, doi:10.1002/14651858.CD004610.pub5 (2017).

5 Ooijevaar, R. E. *et al.* Update of treatment algorithms for Clostridium difficile infection. *Clin Microbiol Infect* **24**, 452-462, doi:10.1016/j.cmi.2017.12.022 (2018).

6 Rineh, A., Kelso, M. J., Vatansever, F., Tegos, G. P. & Hamblin, M. R. Clostridium difficile infection: molecular pathogenesis and novel therapeutics. *Expert Rev Anti Infect Ther* **12**, 131-150, doi:10.1586/14787210.2014.866515 (2014).

7 Bagdasarian, N., Rao, K. & Malani, P. N. Diagnosis and treatment of Clostridium difficile in adults: a systematic review. *JAMA* **313**, 398-408, doi:10.1001/jama.2014.17103 (2015).

8 Yokoyama, Y. *et al.* Risk factors of first recurrence of Clostridioides difficile infection. *Anaerobe* **75**, 102556, doi:10.1016/j.anaerobe.2022.102556 (2022).

9 Ducarmon, Q. R. *et al.* Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. *Microbiol Mol Biol Rev* **83**, doi:10.1128/MMBR.00007-19 (2019).

10 Laffin, M., Millan, B. & Madsen, K. L. Fecal microbial transplantation as a therapeutic option in patients colonized with antibiotic resistant organisms. *Gut Microbes* **8**, 221-224, doi:10.1080/19490976.2016.1278105 (2017).

11 Woodworth, M. H., Hayden, M. K., Young, V. B. & Kwon, J. H. The Role of Fecal Microbiota Transplantation in Reducing Intestinal Colonization With Antibiotic-Resistant

Organisms: The Current Landscape and Future Directions. *Open Forum Infect Dis* **6**, doi:10.1093/ofid/ofz288 (2019).

12 Segal, J. P. *et al.* Mechanisms underpinning the efficacy of faecal microbiota transplantation in treating gastrointestinal disease. *Therap Adv Gastroenterol* **13**, 1756284820946904, doi:10.1177/1756284820946904 (2020).

13 DeFilipp, Z. *et al.* Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant. *N Engl J Med* **381**, 2043-2050, doi:10.1056/NEJMoa1910437 (2019).

14 Blaser, M. J. Fecal Microbiota Transplantation for Dysbiosis - Predictable Risks. *N Engl J Med* **381**, 2064-2066, doi:10.1056/NEJMe1913807 (2019).

15 Administration, T. U. S. F. a. D. FDA Approves First Fecal Microbiota Product. (2022).

16 Administration, T. U. S. F. a. D. FDA Approves First Orally Administered Fecal Microbiota Product for the Prevention of Recurrence of Clostridioides difficile Infection. (2023).

17 Dubberke, E. R. *et al.* Results From a Randomized, Placebo-Controlled Clinical Trial of a RBX2660-A Microbiota-Based Drug for the Prevention of Recurrent Clostridium difficile Infection. *Clin Infect Dis* **67**, 1198-1204, doi:10.1093/cid/ciy259 (2018).

18 Peery, A. F. *et al.* AGA Clinical Practice Guideline on Fecal Microbiota-Based Therapies for Select Gastrointestinal Diseases. *Gastroenterology* **166**, 409-434, doi:10.1053/j.gastro.2024.01.008 (2024).

19 Walter, J. & Shanahan, F. Fecal microbiota-based treatment for recurrent Clostridioides difficile infection. *Cell* **186**, 1087, doi:10.1016/j.cell.2023.02.034 (2023).

20 Liao, H., Ji, Y. & Sun, Y. High-resolution strain-level microbiome composition analysis from short reads. *Microbiome* **11**, 183, doi:10.1186/s40168-023-01615-w (2023).

21 Besser, M. D. R. E., Griffin, M. D. P. M. & Slutsker, M. D. M. P. H. L. Escherichia coliO157:H7 Gastroenteritis and the Hemolytic Uremic Syndrome: An Emerging Infectious Disease. *Annual Review of Medicine* **50**, 355-367, doi:10.1146/annurev.med.50.1.355 (1999).

22 Sonnenborn, U. Escherichia coli strain Nissle 1917-from bench to bedside and back: history of a special Escherichia coli strain with probiotic properties. *FEMS Microbiol Lett* **363**, doi:10.1093/femsle/fnw212 (2016).

23 Arribas, B. *et al.* A probiotic strain of Escherichia coli, Nissle 1917, given orally exerts local and systemic anti-inflammatory effects in lipopolysaccharide-induced sepsis in mice. *Br J Pharmacol* **157**, 1024-1033, doi:10.1111/j.1476-5381.2009.00270.x (2009).

24 Tu, Q., He, Z. & Zhou, J. Strain/species identification in metagenomes using genome-specific markers. *Nucleic Acids Res* **42**, e67, doi:10.1093/nar/gku138 (2014).

25 Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res* **27**, 626-638, doi:10.1101/gr.216242.116 (2017).

26 Ke, S., Weiss, S. T. & Liu, Y. Y. Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes. *Nat Commun* **13**, 5235, doi:10.1038/s41467-022-32991-w (2022).

27 Ke, S. *et al.* A computational method to dissect colonization resistance of the gut microbiota against pathogens. *Cell Reports Methods*, 100576, doi:https://doi.org/10.1016/j.crmeth.2023.100576 (2023).

28 Pollock, N. R. *et al.* Comparison of Clostridioides difficile Stool Toxin Concentrations in Adults With Symptomatic Infection and Asymptomatic Carriage Using an Ultrasensitive Quantitative Immunoassay. *Clin Infect Dis* **68**, 78-86, doi:10.1093/cid/ciy415 (2019).

29 Ke, S. *et al.* Integrating gut microbiome and host immune markers to understand the pathogenesis of Clostridioides difficile infection. *Gut Microbes* **13**, 1-18, doi:10.1080/19490976.2021.1935186 (2021).

30 Verma, S. *et al.* Identification and engraftment of new bacterial strains by shotgun metagenomic sequence analysis in patients with recurrent Clostridioides difficile infection before and after fecal microbiota transplantation and in healthy human subjects. *PLoS One* **16**, e0251590, doi:10.1371/journal.pone.0251590 (2021).

31 Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).

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

33 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542, doi:10.1038/s41564-017-0012-7 (2017).

34 Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996-1004, doi:10.1038/nbt.4229 (2018).

35 Wortelboer, K., Nieuwdorp, M. & Herrema, H. Fecal microbiota transplantation beyond Clostridioides difficile infections. *EBioMedicine* **44**, 716-729, doi:10.1016/j.ebiom.2019.05.066 (2019).

36 El-Matary, W. Fecal microbiota transplantation: long-term safety issues. *Am J Gastroenterol* **108**, 1537-1538, doi:10.1038/ajg.2013.208 (2013).

37 Bai, Z. *et al.* Comprehensive analysis of 84 Faecalibacterium prausnitzii strains uncovers their genetic diversity, functional characteristics, and potential risks. *Front Cell Infect Microbiol* **12**, 919701, doi:10.3389/fcimb.2022.919701 (2022).

38 Bjorkqvist, O. *et al.* Faecalibacterium prausnitzii increases following fecal microbiota transplantation in recurrent Clostridioides difficile infection. *PLoS One* **16**, e0249861, doi:10.1371/journal.pone.0249861 (2021).

39 Louie, T. *et al.* VE303, a Defined Bacterial Consortium, for Prevention of Recurrent Clostridioides difficile Infection: A Randomized Clinical Trial. *JAMA* **329**, 1356-1366, doi:10.1001/jama.2023.4314 (2023).

40 Smith, A. B. *et al.* Enterococci enhance Clostridioides difficile pathogenesis. *Nature* **611**, 780-786, doi:10.1038/s41586-022-05438-x (2022).

41 Ugarte-Torres, A., Gillrie, M. R., Griener, T. P. & Church, D. L. Eggerthella lenta Bloodstream Infections Are Associated With Increased Mortality Following Empiric Piperacillin-Tazobactam (TZP) Monotherapy: A Population-based Cohort Study. *Clin Infect Dis* **67**, 221-228, doi:10.1093/cid/ciy057 (2018).

42 Hilau, S., Katz, S., Wasserman, T., Hershberg, R. & Savir, Y. Density-dependent effects are the main determinants of variation in growth dynamics between closely related

bacterial strains. *PLoS Comput Biol* **18**, e1010565, doi:10.1371/journal.pcbi.1010565 (2022).

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

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

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

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

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

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

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

50 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**, 2864-2868, doi:10.1038/ismej.2017.126 (2017).

51 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).

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

53 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).

54 Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* **11**, 2500, doi:10.1038/s41467-020-16366-7 (2020).

55 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296, doi:10.1093/nar/gkab301 (2021).

56 Martino, C. *et al.* A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* **4**, doi:10.1128/mSystems.00016-19 (2019).

57 Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **26**, 27663, doi:10.3402/mehd.v26.27663 (2015).

58 Crobach, M. J. T. *et al.* The Bacterial Gut Microbiota of Adult Patients Infected, Colonized or Noncolonized by Clostridioides difficile. *Microorganisms* **8**, doi:10.3390/microorganisms8050677 (2020).

59 Mullish, B. H. *et al.* Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridioides difficile infection. *Gut*, doi:10.1136/gutjnl-2018-317842 (2019).

60 Berkell, M. *et al.* Microbiota-based markers predictive of development of Clostridioides difficile infection. *Nat Commun* **12**, 2241, doi:10.1038/s41467-021-22302-0 (2021).

61 Milani, C. *et al.* Gut microbiota composition and Clostridium difficile infection in hospitalized elderly individuals: a metagenomic study. *Sci Rep* **6**, 25945, doi:10.1038/srep25945 (2016).

62 Han, S. H., Yi, J., Kim, J. H., Lee, S. & Moon, H. W. Composition of gut microbiota in patients with toxigenic Clostridioides (Clostridium) difficile: Comparison between subgroups according to clinical criteria and toxin gene load. *PLoS One* **14**, e0212626, doi:10.1371/journal.pone.0212626 (2019).

63 Feuerstadt, P. *et al.* SER-109, an Oral Microbiome Therapy for Recurrent Clostridioides difficile Infection. *N Engl J Med* **386**, 220-229, doi:10.1056/NEJMoa2106516 (2022).

64 Tvede, M. & Rask-Madsen, J. Bacteriotherapy for chronic relapsing Clostridium difficile diarrhoea in six patients. *Lancet* **1**, 1156-1160, doi:10.1016/s0140-6736(89)92749-9 (1989).

65 Yoon, S. *et al.* Bile salt hydrolase-mediated inhibitory effect of Bacteroides ovatus on growth of Clostridium difficile. *J Microbiol* **55**, 892-899, doi:10.1007/s12275-017-7340-4 (2017).

66 Amrane, S. *et al.* Metagenomic and culturomic analysis of gut microbiota dysbiosis during Clostridium difficile infection. *Sci Rep* **9**, 12807, doi:10.1038/s41598-019-49189-8 (2019).

67 Hourigan, S. K. *et al.* Fecal Transplant in Children With Clostridioides difficile Gives Sustained Reduction in Antimicrobial Resistance and Potential Pathogen Burden. *Open Forum Infect Dis* **6**, ofz379, doi:10.1093/ofid/ofz379 (2019).

68 Douchant, K. *et al.* Defined microbial communities and their soluble products protect mice from Clostridioides difficile infection. *Commun Biol* **7**, 135, doi:10.1038/s42003-024-05778-6 (2024).

69 Francisco, D. M. A. *et al.* Risk Factors Associated with Severe Clostridioides difficile Infection in Patients with Cancer. *Infect Dis Ther* **12**, 209-225, doi:10.1007/s40121-022-00722-9 (2023).

70 Moelling, K. & Broecker, F. Fecal microbiota transplantation to fight Clostridium difficile infections and other intestinal diseases. *Bacteriophage* **6**, e1251380, doi:10.1080/21597081.2016.1251380 (2016).

71 Roychowdhury, S. *et al.* Faecalibacterium prausnitzii and a Prebiotic Protect Intestinal Health in a Mouse Model of Antibiotic and Clostridium difficile Exposure. *JPEN J Parenter Enteral Nutr* **42**, 1156-1167, doi:10.1002/jpen.1053 (2018).

72 Vakili, B., Fateh, A., Asadzadeh Aghdaei, H., Sotoodehnejadnematalahi, F. & Siadat, S. D. Intestinal Microbiota in Elderly Inpatients with Clostridioides difficile Infection. *Infect Drug Resist* **13**, 2723-2731, doi:10.2147/IDR.S262019 (2020).

73 Vakili, B., Fateh, A., Asadzadeh Aghdaei, H., Sotoodehnejadnematalahi, F. & Siadat, S. D. Characterization of Gut Microbiota in Hospitalized Patients with Clostridioides difficile Infection. *Curr Microbiol* **77**, 1673-1680, doi:10.1007/s00284-020-01980-x (2020).

74 Shoaei, P. *et al.* Gut microbiota in burned patients with Clostridioides difficile infection. *Burns* **48**, 1120-1129, doi:10.1016/j.burns.2021.11.023 (2022).

75 Gu, X. *et al.* Gut Ruminococcaceae levels at baseline correlate with risk of antibiotic-associated diarrhea. *iScience* **25**, 103644, doi:10.1016/j.isci.2021.103644 (2022).

76 Dudik, B., Kinova Sepova, H., Greifova, G., Bilka, F. & Bilkova, A. Next generation probiotics: an overview of the most promising candidates. *Epidemiol Mikrobiol Imunol* **71**, 48-56 (2022).

77 Lee, Y. J., Yu, W. K. & Heo, T. R. Identification and screening for antimicrobial activity against Clostridium difficile of Bifidobacterium and Lactobacillus species isolated from healthy infant faeces. *Int J Antimicrob Agents* **21**, 340-346, doi:10.1016/s0924-8579(02)00389-8 (2003).

78 Solbach, P. *et al.* Microbiota-associated Risk Factors for Clostridioides difficile Acquisition in Hospitalized Patients: A Prospective, Multicentric Study. *Clin Infect Dis* **73**, e2625-e2634, doi:10.1093/cid/ciaa871 (2021).

Figures and Legends

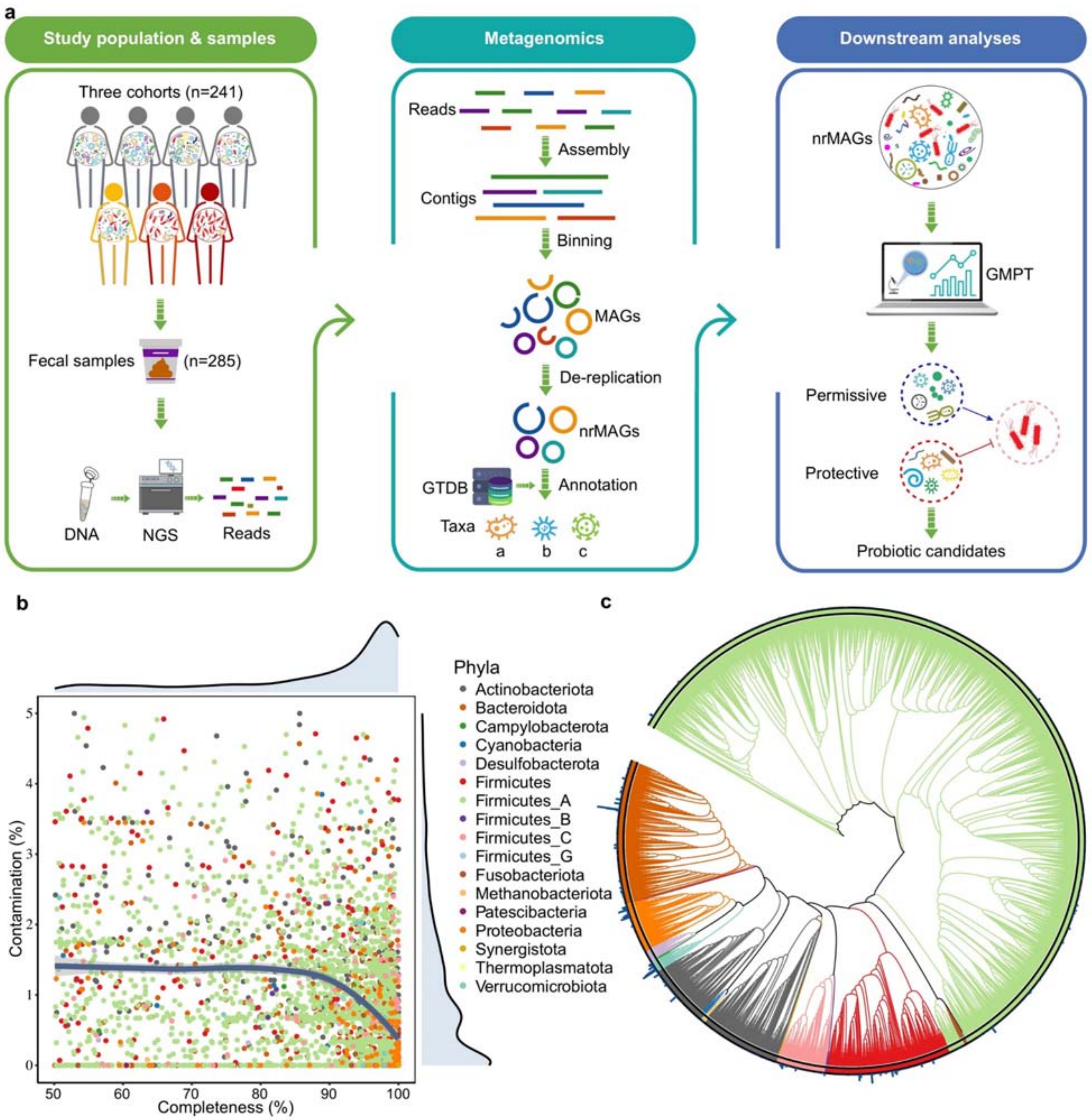


Fig. 1 Study workflow and the reconstruction of the microbial genome catalog. a. To rationally design microbial consortia against *C. difficile*, we sought to infer species that may potentially inhibit *C. difficile* from various metagenomic data. We collected a total of 285 shotgun metagenomic sequencing data from three independent cohorts. A total of 7,769 MAGs (≥50% completeness and ≤5% contamination) were constructed from all metagenomic sequencing data. The MAGs were then dereplicated to 3,741 non-redundant MAGs (nrMAGs, strain level) based on 99% of ANI. The taxonomy annotation and abundance estimation of nrMAGs were then conducted. We then

applied the generalized microbe-phenotype triangulation (GMPT) method to identify candidate strains for the development of microbiota probiotics. **b.** The distribution of completeness and contamination of nrMAGs is depicted, with the color of each point representing the respective phylum. Additionally, the size of each point corresponds to the genome size of the nrMAGs. **c.** A phylogenetic tree of nrMAGs was constructed using PhyloPhlAn. In this representation, the color of the outer cycle and clades signifies the phylum, while the bar plot within the cycle illustrates the average abundance across all microbiome samples.

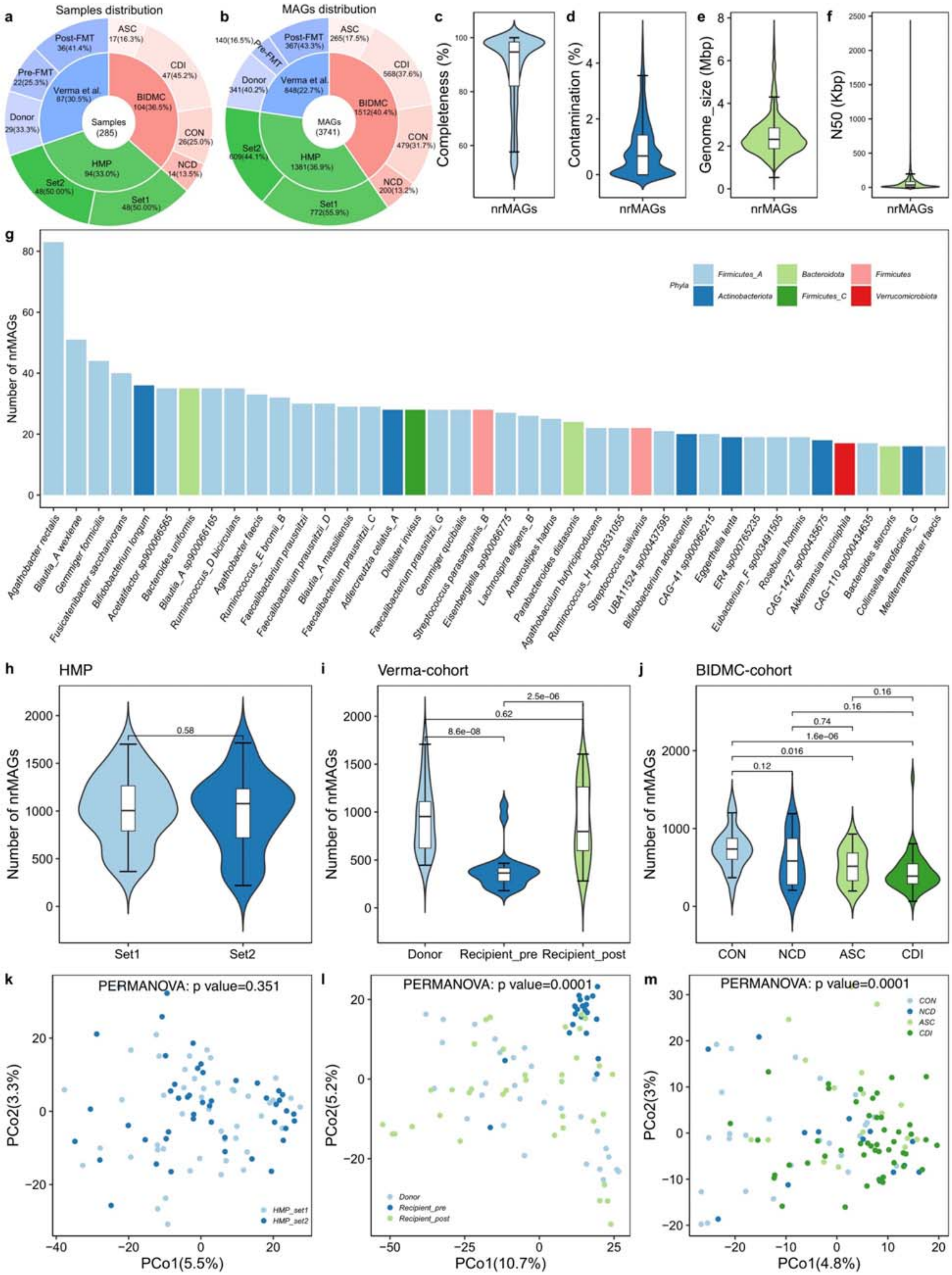
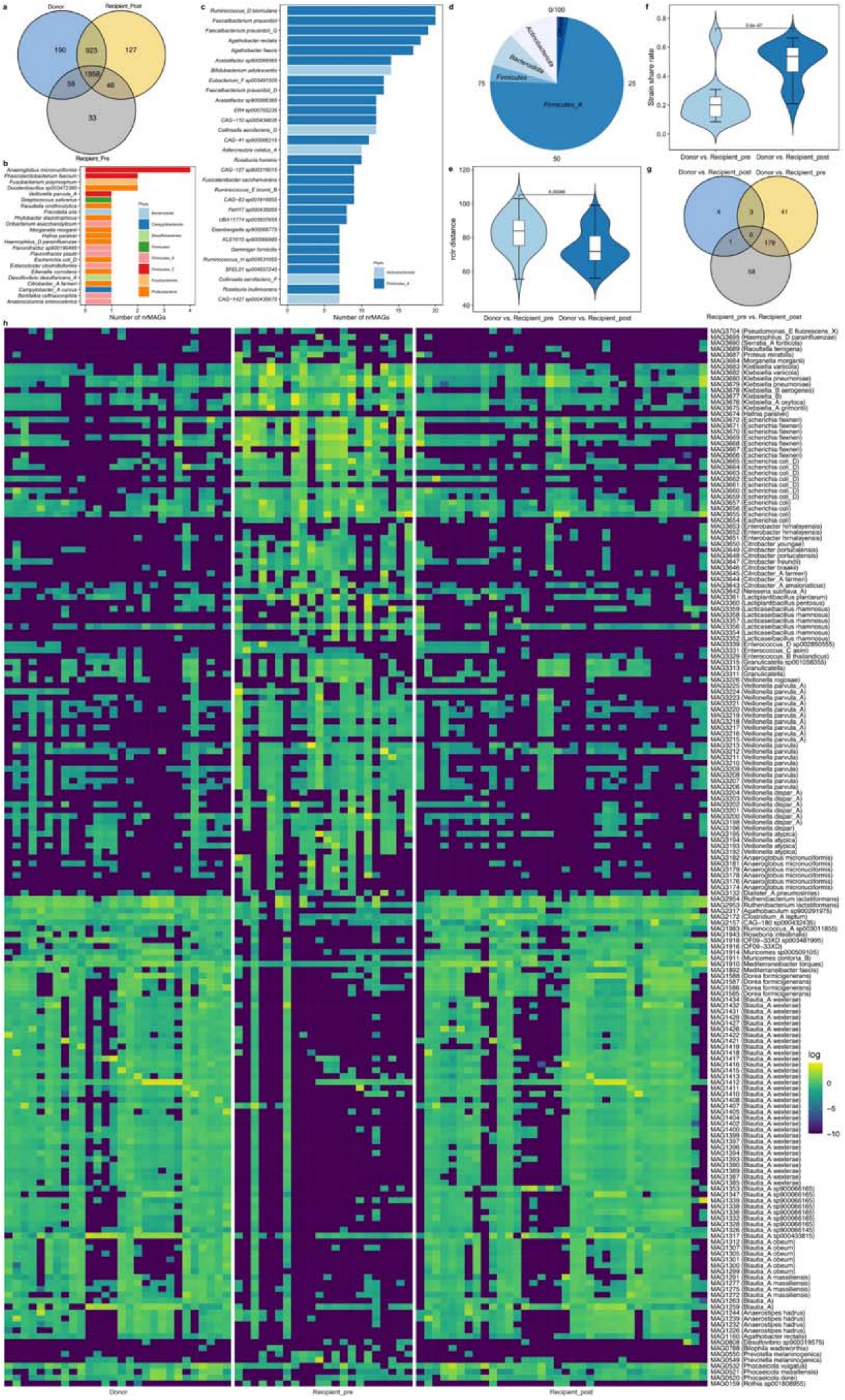

a
Samples distribution
Samples (285)
Verma et al. 87(30.5%)
BIDMC 104(36.5%)
HMP 94(33.0%)
Post-FMT 36(41.4%)
Pre-FMT 22(25.3%)
Donor 29(33.3%)
ASC 17(16.3%)
CDI 47(45.2%)
CON 26(25.0%)
NCD 14(13.5%)
Set2 48(50.00%)
Set1 48(50.00%)
b
MAGs distribution
MAGs (3741)
Verma et al. 848(22.7%)
BIDMC 1512(40.4%)
HMP 1381(36.9%)
Post-FMT 367(43.3%)
Pre-FMT 140(16.5%)
Donor 341(40.2%)
ASC 265(17.5%)
CDI 568(37.6%)
CON 479(31.7%)
NCD 200(13.2%)
Set2 609(44.1%)
Set1 772(55.9%)
c
Completeness (%)
100
80
60
nrMAGs
d
Contamination (%)
4
2
0
nrMAGs
e
Genome_size (Mbp)
8
6
4
2
0
nrMAGs
f
N50 (Kbp)
2500
2000
1500
1000
500
0
nrMAGs
g
Number of nrMAGs
80
60
40
20
0
Phyla
Firmicutes_A
Actinobacteriota
Bacteroidota
Firmicutes_C
Firmicutes
Verrucomicrobiota
Agathobacter rectalis
Blautia_A wexlerae
Gemmiger formicilis
Fusicatenibacter saccharivorans
Bifidobacterium longum
Acetatifactor sp900066565
Bacteroides uniformis
Blautia_A sp900066165
Ruminococcus_D bicirculans
Agathobacter faecis
Ruminococcus_E bromii_B
Faecalibacterium prausnitzii
Faecalibacterium prausnitzii_D
Blautia_A massiliensis
Faecalibacterium prausnitzii_C
Adlercreutzia celatus_A
Dialister invisus
Faecalibacterium prausnitzii_G
Gemmiger qucibialis
Streptococcus parasanguinis_B
Eisenbergiella sp900066775
Lachnospira eligens_B
Anaerostipes hadrus
Parabacteroides distasonis
Agathobaculum butyriciproducens
Ruminococcus_H sp003531055
Streptococcus salivarius
UBA11524 sp000437595
Bifidobacterium adolescentis
CAG-41 sp900066215
Eggerthella lenta
ER4 sp000765235
Eubacterium_F sp003491505
Roseburia hominis
CAG-1427 sp000435675
Akkermansia muciniphila
CAG-110 sp000434635
Bacteroides stercoris
Collinsella aerofaciens_G
Mediterraneibacter faecis
h
HMP
Number of nrMAGs
2000
1500
1000
500
0
0.58
Set1
Set2
i
Verma-cohort
Number of nrMAGs
2000
1500
1000
500
0
2.5e-06
0.62
8.6e-08
Donor
Recipient_pre
Recipient_post
j
BIDMC-cohort
Number of nrMAGs
2000
1000
0
0.16
0.16
0.74
1.6e-06
0.016
0.12
CON
NCD
ASC
CDI
k
PERMANOVA: p value=0.351
PCo2(3.3%)
20
0
−20
PCo1(5.5%)
−40
−20
0
20
HMP_set1
HMP_set2
l
PERMANOVA: p value=0.0001
PCo2(5.2%)
20
0
−20
PCo1(10.7%)
−50
−25
0
25
Donor
Recipient_pre
Recipient_post
m
PERMANOVA: p value=0.0001
PCo2(3%)
30
20
10
0
−10
−20
PCo1(4.8%)
−20
−10
0
10
20
CON
NCD
ASC
CDI

Fig. 2 The microbial genome catalog and microbial diversity. a. Sample distribution among different datasets and clinical groups. **b**. Number of MAGs recovered from different datasets and clinical groups. Violin plot of basic characteristics of nrMAGs on completeness (**c**), contamination (**d**), genome size (**e**), and N50 (**f**). **g**. The top-40 species with the highest strain-richness (i.e., number of nrMAGs) identified from the microbial genome catalog. The color of each bar signifies the phylum. Richness (number of identified nrMAGs) of the gut microbiome from HMP (**h**), Verma-cohort. (**i**), and BIDMC-cohort (**j**). Principal Coordinates Analysis (PCoA) plot based on robust Aitchison distance from HMP (**k**), Verma-cohort. (**l**), and BIDMC-cohort (**m**). All PERMANOVA tests were performed with 9999 permutations based on robust Aitchison distance, two-sided.


a
Donor
Recipient_Post
Recipient_Pre
190
923
127
1958
58
46
33
b
Anaerogiobus micronuciformis
Phascolarctobacterium faecium
Fusobacterium polymorphum
Duodenibacillus sp003472385
Veillonella parvula_A
Streptococcus salivarius
Raoultella ornithinolytica
Prevotella oris
Phytobacter diazotrophicus
Oribacterium asaccharolyticum
Morganella morganii
Hafnia paralvei
Haemophilus_D parainfluenzae
Flavonifractor sp900199495
Flavonifractor plautii
Escherichia coli_D
Enterocloster clostridioformis
Eikenella corrodens
Desulfovibrio desulfuricans_A
Citrobacter_A farmeri
Campylobacter_A curvus
Bordetella caffrasonophila
Anaerococcus ammovalentus
Phyla
Bacteroidota
Campylobacterota
Desulfobacterota
Firmicutes
Firmicutes_A
Firmicutes_C
Fusobacteriota
Proteobacteria
0
1
2
3
4
Number of nrMAGs
c
Ruminococcus_D bicirculans
Faecalibacterium prausnitzii
Faecalibacterium prausnitzii_G
Agathobacter rectalis
Agathobacter faecis
Acetatifactor sp900066565
Bifidobacterium adolescentis
Eubacterium_F sp003491505
Faecalibacterium prausnitzii_D
Acetatifactor sp900066365
ER4 sp000765235
CAG-110 sp000434635
Collinsella aerofaciens_G
CAG-41 sp900066215
Adlercreutzia celatus_A
Roseburia hominis
CAG-127 sp900319515
Fusicatenibacter saccharivorans
Ruminococcus_E bromii_B
CAG-83 sp001916855
Part17 sp000435555
UBA11774 sp003507655
Eisenbergiella sp900066775
KLE1615 sp900066985
Gemmiger formicilis
Ruminococcus_H sp003531055
SFEL01 sp004557245
Collinsella aerofaciens_F
Roseburia inulinivorans
CAG-1427 sp000435675
Phyla
Actinobacteriota
Firmicutes_A
0
5
10
15
20
Number of nrMAGs
d
0/100
25
50
75
Actinobacteriota
Bacteroidota
Firmicutes
Firmicutes_A
e
120
100
80
60
40
rclr distance
0.00066
Donor vs. Recipient_pre
Donor vs. Recipient_post
f
0.8
0.6
0.4
0.2
0.0
Strain share rate
2.6e-07
Donor vs. Recipient_pre
Donor vs. Recipient_post
g
Donor vs. Recipient_post
Donor vs. Recipient_pre
Recipient_pre vs. Recipient_post
4
3
41
0
1
179
58
h
MAG3704 (Pseudomonas_E fluorescens_X)
MAG3695 (Haemophilus_D parainfluenzae)
MAG3690 (Serratia_A fonticola)
MAG3689 (Raoultella terrigena)
MAG3687 (Proteus mirabilis)
MAG3684 (Morganella morganii)
MAG3683 (Klebsiella variicola)
MAG3682 (Klebsiella variicola)
MAG3680 (Klebsiella pneumoniae)
MAG3679 (Klebsiella pneumoniae)
MAG3678 (Klebsiella_B aerogenes)
MAG3677 (Klebsiella_B)
MAG3676 (Klebsiella_A oxytoca)
MAG3675 (Klebsiella_A grimontii)
MAG3674 (Hafnia paralvei)
MAG3672 (Escherichia flexneri)
MAG3671 (Escherichia flexneri)
MAG3670 (Escherichia flexneri)
MAG3669 (Escherichia flexneri)
MAG3668 (Escherichia flexneri)
MAG3667 (Escherichia flexneri)
MAG3666 (Escherichia flexneri)
MAG3665 (Escherichia coli_D)
MAG3664 (Escherichia coli_D)
MAG3663 (Escherichia coli_D)
MAG3662 (Escherichia coli_D)
MAG3661 (Escherichia coli_D)
MAG3660 (Escherichia coli_D)
MAG3659 (Escherichia coli_D)
MAG3657 (Escherichia coli)
MAG3656 (Escherichia coli)
MAG3655 (Escherichia coli)
MAG3654 (Escherichia coli)
MAG3653 (Enterobacter himalayensis)
MAG3652 (Enterobacter himalayensis)
MAG3651 (Enterobacter himalayensis)
MAG3650 (Citrobacter youngae)
MAG3649 (Citrobacter portucalensis)
MAG3648 (Citrobacter portucalensis)
MAG3647 (Citrobacter freundii)
MAG3646 (Citrobacter braakii)
MAG3645 (Citrobacter_A farmeri)
MAG3644 (Citrobacter_A farmeri)
MAG3643 (Citrobacter_A amalonaticus)
MAG3642 (Neisseria subflava_A)
MAG3361 (Lactiplantibacillus plantarum)
MAG3360 (Lactiplantibacillus pentosus)
MAG3359 (Lacticaseibacillus rhamnosus)
MAG3358 (Lacticaseibacillus rhamnosus)
MAG3357 (Lacticaseibacillus rhamnosus)
MAG3356 (Lacticaseibacillus rhamnosus)
MAG3354 (Lacticaseibacillus rhamnosus)
MAG3352 (Lacticaseibacillus rhamnosus)
MAG3339 (Enterococcus_D sp002850555)
MAG3331 (Enterococcus_C asini)
MAG3329 (Enterococcus_B thailandicus)
MAG3315 (Granulicatella sp001058355)
MAG3313 (Granulicatella)
MAG3311 (Granulicatella)
MAG3226 (Veillonella rogosae)
MAG3225 (Veillonella parvula_A)
MAG3224 (Veillonella parvula_A)
MAG3223 (Veillonella parvula_A)
MAG3221 (Veillonella parvula_A)
MAG3220 (Veillonella parvula_A)
MAG3219 (Veillonella parvula_A)
MAG3218 (Veillonella parvula_A)
MAG3217 (Veillonella parvula_A)
MAG3216 (Veillonella parvula_A)
MAG3215 (Veillonella parvula_A)
MAG3213 (Veillonella parvula)
MAG3212 (Veillonella parvula)
MAG3211 (Veillonella parvula)
MAG3210 (Veillonella parvula)
MAG3209 (Veillonella parvula)
MAG3208 (Veillonella parvula)
MAG3207 (Veillonella parvula)
MAG3206 (Veillonella parvula)
MAG3204 (Veillonella dispar_A)
MAG3203 (Veillonella dispar_A)
MAG3202 (Veillonella dispar_A)
MAG3201 (Veillonella dispar_A)
MAG3200 (Veillonella dispar_A)
MAG3198 (Veillonella dispar_A)
MAG3196 (Veillonella dispar)
MAG3195 (Veillonella atypica)
MAG3194 (Veillonella atypica)
MAG3193 (Veillonella atypica)
MAG3192 (Veillonella atypica)
MAG3182 (Anaeroglobus micronuciformis)
MAG3181 (Anaeroglobus micronuciformis)
MAG3179 (Anaeroglobus micronuciformis)
MAG3178 (Anaeroglobus micronuciformis)
MAG3176 (Anaeroglobus micronuciformis)
MAG3174 (Anaeroglobus micronuciformis)
MAG3132 (Dialister_A pneumosintes)
MAG2954 (Ruthenibacterium lactatiformans)
MAG2953 (Ruthenibacterium lactatiformans)
MAG2317 (Agathobaculum sp900291975)
MAG2172 (Clostridium_A leptum)
MAG2157 (CAG-180 sp000432435)
MAG1980 (Ruminococcus_A sp003011855)
MAG1943 (Roseburia intestinalis)
MAG1918 (OF09-33XD sp003481995)
MAG1916 (OF09-33XD)
MAG1914 (Muricomes sp000509105)
MAG1911 (Muricomes contorta_B)
MAG1910 (Mediterraneibacter torques)
MAG1892 (Mediterraneibacter faecis)
MAG1588 (Dorea formicigenerans)
MAG1587 (Dorea formicigenerans)
MAG1586 (Dorea formicigenerans)
MAG1585 (Dorea formicigenerans)
MAG1434 (Blautia_A wexlerae)
MAG1432 (Blautia_A wexlerae)
MAG1431 (Blautia_A wexlerae)
MAG1429 (Blautia_A wexlerae)
MAG1427 (Blautia_A wexlerae)
MAG1426 (Blautia_A wexlerae)
MAG1422 (Blautia_A wexlerae)
MAG1421 (Blautia_A wexlerae)
MAG1419 (Blautia_A wexlerae)
MAG1418 (Blautia_A wexlerae)
MAG1417 (Blautia_A wexlerae)
MAG1416 (Blautia_A wexlerae)
MAG1415 (Blautia_A wexlerae)
MAG1413 (Blautia_A wexlerae)
MAG1412 (Blautia_A wexlerae)
MAG1411 (Blautia_A wexlerae)
MAG1410 (Blautia_A wexlerae)
MAG1408 (Blautia_A wexlerae)
MAG1407 (Blautia_A wexlerae)
MAG1405 (Blautia_A wexlerae)
MAG1404 (Blautia_A wexlerae)
MAG1402 (Blautia_A wexlerae)
MAG1400 (Blautia_A wexlerae)
MAG1399 (Blautia_A wexlerae)
MAG1397 (Blautia_A wexlerae)
MAG1396 (Blautia_A wexlerae)
MAG1394 (Blautia_A wexlerae)
MAG1393 (Blautia_A wexlerae)
MAG1390 (Blautia_A wexlerae)
MAG1389 (Blautia_A wexlerae)
MAG1387 (Blautia_A wexlerae)
MAG1385 (Blautia_A wexlerae)
MAG1353 (Blautia_A sp900066165)
MAG1347 (Blautia_A sp900066165)
MAG1339 (Blautia_A sp900066165)
MAG1338 (Blautia_A sp900066165)
MAG1336 (Blautia_A sp900066165)
MAG1332 (Blautia_A sp900066165)
MAG1328 (Blautia_A sp900066165)
MAG1326 (Blautia_A sp900066145)
MAG1317 (Blautia_A sp000433815)
MAG1312 (Blautia_A obeum)
MAG1307 (Blautia_A obeum)
MAG1305 (Blautia_A obeum)
MAG1301 (Blautia_A obeum)
MAG1300 (Blautia_A obeum)
MAG1299 (Blautia_A obeum)
MAG1291 (Blautia_A massiliensis)
MAG1277 (Blautia_A massiliensis)
MAG1275 (Blautia_A massiliensis)
MAG1272 (Blautia_A massiliensis)
MAG1263 (Blautia_A)
MAG1259 (Blautia_A)
MAG1244 (Anaerostipes hadrus)
MAG1239 (Anaerostipes hadrus)
MAG1232 (Anaerostipes hadrus)
MAG1226 (Anaerostipes hadrus)
MAG1160 (Agathobacter rectalis)
MAG0808 (Desulfovibrio sp900319575)
MAG0788 (Bilophila wadsworthia)
MAG0550 (Prevotella melaninogenica)
MAG0549 (Prevotella melaninogenica)
MAG0532 (Phocaeicola vulgatus)
MAG0521 (Phocaeicola massiliensis)
MAG0520 (Phocaeicola dorei)
MAG0159 (Rothia sp001808955)
log
0
-5
-10
Donor
Recipient_pre
Recipient_post

785

Fig. 3 The changes in recipients' microbiome after FMT. a. The distribution of nrMAG among donors, pre-FMT recipients, and post-FMT recipients. **b**, The distribution of lost nrMAGs after FMT at the species level, and the color of each bar represents the phylum. **c**. The distribution of engrafted nrMAGs after FMT at the species level, and the color of each bar represents the phylum. **d**. The distribution of engrafted nrMAGs after FMT at the phylum level. **e**, The robust Aitchison distance between donor and recipient pairs before and after FMT. **f**, The nrMAG share rate between donor and recipient pairs before and after FMT. **g**. The differential abundant nrMAG distribution among three pair-wise comparisons between donors, pre-FMT recipients, and post-FMT recipients. **h**. The heat map showed the abundance distribution of overlapped nrMAGs identified from the comparisons of donor vs. pre-FMT recipients and pre-FMT recipients vs. post-FMT recipients. These nrMAGs were taxonomically annotated using GTDB-Tk based on the Genome Taxonomy Database.

Table. 1 Summary of the literature evidence regarding the potential role of protective species in CDI identified through our computational pipeline. The top 25 potential protective species were selected based on the overlapped protective strains identified from the GMPT results with two sets of HMP microbiome data.

Family	Rank	Species	Taxonomy ID alternate names	PMID Author reference *Sponsor*	Study size and Groups	Effect
Lachnospiraceae	1	*Dorea formicigenerans*	39486 *Eubacterium formicigenerans*	34252073[π] Verma[30]	22 rCDI (pre/post FMT) Healthy donors	*D. formicigerans* is top 5 engrafter in rCDI patients after FMT from donor
	2	*GCA-900066135 sp900066135*	2830660 *Lachnospiraceae bacterium Marseille-Q4251*			
	6	*Anaerobutyricum hallii*	39488 *Eubacterium hallii*	32384826 Crobach[58]	41 CDI, 41 colonized, 43 controls	↑ *A. hallii*: controls vs. colonized (↑6.6x AB+, ↑ 3.8x AB-)
	8	*Dorea longicatena*	88431 *Eubacterium sp. III-35*	37060545 Louie[39] *Vedanta*	Phase 2 RCT VE303 (8-organism consortium) 79 rCDI 1:1:1 placebo (PBO): low (LD): high-dose (HD) VE303	*D. longicatena* is part of consortium. 8-week CDI recurrence: PBO 45.5%, LD 37.0%, HD 13.8% (p=0.006 HD vs. PBO)
	9	*Blautia obeum*	40520 *Ruminococcus obeum*	30816855 Mullish[59] *Finch*	14 rCDI receiving FMT (pre, post: 1, 4, 12 weeks) 5 healthy donors	Pre-FMT: ↓ bile salt hydrolase (BSH) activity and genes (*bsh*/*baiCD*), ↑ primary bile acids (taurocholic) Post-FMT: ↑ *bsh*/*baiCD*, ↑*B. obeum* (BSH producer) Culture supernatant of *B. obeum* (& 3 other BSH+ species) attenuates CDI in mouse model
				33854066 Berkell[60]	AB+: 14 CDI, 64 AAD, 669 no diarrhea (ND)	*B. obeum* is 21.5% of OTU30 (oligotyping) ↑OTU30 3.8x: (AAD+ND) vs. CDI ↑OTU30 3.7x: AAD vs. CDI
	10	*Bariatricus comes*	GBIF 10828568			
	15	*Mediterraneibacter faecis*	592978 *Ruminococcus faecis*			
	17	*Anaerostipes hadrus*	649756 *Eubacterium hadrum*			
	18	*Lachnospira sp900316325*				
	20	*Agathobacter rectalis*	39491 *Pseudobacterium / Roseburia / Eubacterium rectale* *Bacteroides rectalis*			
	21	*CAG-81 sp900066535*				
	23	UBA7182 sp003480725	1952150			
	25	*Blautia wexlerae*	418240	33854066 Berkell[60]	See *B. obeum*	*B. wexlerae* is 57.9% of OTU30 (oligotyping)
	Family level effects (human studies only)			27166072 Milani[61]	29 Non-CDI, AB+ 30 Non-CDI, AB- 25 CDI	Lachnospira relative abundance: CDI: 0.31%, AB+: 1.22%, AB-: 3.28% (p<0.005 CDI vs. AB-)

				30785932 Han[62]	*Cd tcdB*: 79 NAAT+ 20 NAAT-	Lachnospira relative abundance: NAAT + 9.00%, NAAT – 16.51% p=0.003
				35045228 Feuerstadt[63] *Seres*	Phase 3 RCT 182 rCDI 1:1 PBO: SER-109	SER-109 contains, among others, the following genera of Lachnospiraceae: *Anaerobutyricum*, *Anaerostipes*, *Bariatricus*, *Blautia*, *Dorea*, *Lachnospira*, *Mediterraneibacter* 8-week CDI recurrence: PBO 40%, SER-109 12% (p<0.001)
Bacteroidaceae	5	*Phocaeicola dorei*	357276 *Bacteroides dorei*			
	13	*Phocaeicola vulgatus*	821 *Bacteroides vulgatus*	2566734 $^{\pi}$ Tvede[64]	RCT 6 rCDI, rectally instilled: 2 donor feces vs. 4 bacterial strain mix	All patients had clinical resolution at 24h. No recurrence within 1 year. *P. vulgatus*: one of 10-strain in mix, inhibited *Cd* growth in vitro
	14	*Bacteroides ovatus*	28116 *Pasteurella ovata, Pseudobacterium ovatum Bacteroides fragilis subsp. ovatus*	2566734 $^{\pi}$ Tvede[64]	See *P. vulgatus*	*B. ovatus*: one of 10-strain in mix, inhibited *Cd* growth in vitro
				29076071 $^{\pi}$ Yoon[65] *KoBio Labs*	In vitro study of susceptibility of *Cd* cultures to supernatants of different bacterial organisms	*B. ovatus* SNUG40239 supernatant inhibited *Cd* growth in a bile acid dependent manner.
				30816855 Mullish[59] *Finch*	See *B. obeum*	Post-FMT: ↑*B. ovatus* (BSH producer) Culture supernatant of *B. ovatus* (& 3 other BSH+ species) attenuates CDI in mouse model
				31488869 Amrane[66]	11 CDI 8 healthy donors	*B. ovatus* among top 3 cultivable organisms absent in CDI and present in >75% of controls.
				31660343 Hourigan[67]	9 children with CDI/rCDI receiving FMT (pre/post) Donor stool	↑ *B. ovatus*: recipients post vs. pre FMT (p=0.03) donors vs. recipients pre-FMT (p=0.04)
				38280981 Douchant[68]	Murine model of CDI 18 & 4-strain synthetic microbial communities	*B. ovatus* part of synthetic microbial community: Protects mice from CDI (RT 027, 078) Effect persists with bacteria-free supernatant. *In vitro Cd* toxin proteolysis
	24	*Bacteroides uniformis*	820	36443547 Francisco[69]	200 CDI with cancer 42 severe/fulminant 158 non-severe	*B. uniformis* top species associated with non-severe CDI by IDSA/SHEA criteria (effect size 2.5, p<0.05)
Oscillospiraceae	3	*Oscillibacter welbionis*	--			
	4	*Faecalibacterium prausnitzii*	853 *Fusobacterium prausnitzii*	28090385 Moelling[70]	N=1, rCDI cured with FMT Followed for 4.5 years	*F. prausnitzii* and *A. municiphila* only two species engrafting at end of follow-up
				29385239 $^{\pi}$ Roychowdhury[71]	CDI mouse model, 13 mice/group, orally received: *F. prausntizii* (Fp), Fp + Potato Starch (PS), PS, Supernatant of Fp + PS, and saline	*F. prausnitzii* + PS vs. saline: ↓ *Cd* DNA (day 3) ↓ IL-1β & IL-8 mRNA (day 5 vs. 1) ↑ IL-10 mRNA (day 5 vs. 1)
				31660343	See *B. ovatus*	↑ *F. prausnitzii:*

				Hourigan[67]		donors vs. recipients pre-FMT (p=0.008)
				32296918 Vakili[72]	28 CDI 56 non-CDI	↑ *F. prausnitzii:* Non-CDI vs. CDI (p=0.015)
				32801806 Vakili[73]	50 CDI 50 healthy controls	↑ *F. prausnitzii:* Controls vs. CDI (p<0.05)
				33836037 Björkqvist[38]	15 rCDI receiving FMT (pre, post: 2 weeks, 2-4 months) 9 healthy donors	↑ *F. prausnitzii:* Donors vs. rCDI pre-FMT (p<0.01) rCDI post-FMT vs. pre-FMT (p<0.001)
				34924229 Shoaei[74]	69 Burn unit patients (23 CDI, 46 non-CDI) 46 healthy controls	↑ *F. prausnitzii:* Non-CDI vs. CDI (p<0.001) Controls vs. CDI (p=0.003)
				35005566 Gu[75] *Finch*	30 healthy adults, amoxicillin-clavulanate: (13 AAD, 17 non-AAD) Days: 0, 1, 2, 3, 7, 14, 28	↑ *F. prausnitzii:* On day 0, ↓2.33x risk of AAD Non-AAD vs. AAD (Days 1-7, p<0.05)
				35477270 Dudik[76]	Narrative review	*F. prausnitzii* in top 2 of most promising 'next-generation probiotics'
	22	*Lawsonibacter sp900066645*				
	Family level effects (human studies only)			35045228 Feuerstadt[63] *Seres*	See Lachnospiraceae	SER-109 contains, among others, the following genera of Oscillospiraceae ~[Ruminococcaceae]: *Faecalibacterium*, *Lawsonibacter*
Streptococcaceae	7	*Streptococcus salivarius*	1304 *Lactobacillus salivarius*	12672580 Lee[77]	In vitro study: 102 lactic acid producing bacteria from 32 healthy infants.	*S. salivarius* among 12/32 strains with anti-*Cd* activity in co-culture
Rikenellaceae	11	*Alistipes putredinis*	28117			
unclassified Bacillota	12	UBA1191	1947933 Firmicutes bacterium UBA1191			
Odoribacteraceae	16	*Odoribacter splanchnicus*	28118 *Bacteroides splanchnicus*	32589701 Solbach[78]	1506 hospitalized patients 139 colonized on admission 16 new *Cd* through admission	↑ *O. splanchnicus* on admission associated with absence of Cd colonization through admission (LDA 3.4, FDR<0.05)
Peptostreptoco-ccaceae	19	*Romboutsia timonensis*	1776391 *Romboutsia sp. DR1*			

PMID: Pubmed ID, *Cd*: *Clostridioides difficile*, CDI: *Cd* Infection, rCDI: recurrent CDI, AB+: antibiotic-exposed, AB-: non-antibiotic-exposed, FMT: Fecal Microbiota Transplantation, NAAT: Nucleic Acid Amplification Test, *tcdB*: Cd Toxin B gene, RCT: Randomized Controlled Trial, OUT: Operational Taxonomic Unit, RT: Ribotype, AAD: Antibiotic Associated Diarrhea, LDA score: Linear Discriminant Analysis score, FDR: False Discovery Rate

Π denotes strain-level information available in the reference (*Dorea formicigenerans* ATCC 27755, *Phocaeicola vulgatus* A33-14, *Bacteroides ovatus* A40-4, *Bacteroides ovatus* SNUG40239, *Faecalibacterium prausnitzii* ATCC 27766)