THE UNIVERSITY OF CHICAGO

UNDERSTANDING THE DYNAMICS OF *CLOSTRIDIOIDES DIFFICILE* INFECTION IN HUMAN GUT MICROBIOMES USING MULTI-OMIC APPROACHES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:
IMMUNOLOGY

BY

SOPHIE SEULE SON

CHICAGO, ILLINOIS
JUNE 2025

Copyright © 2025 by Sophie SeulE Son

All Rights Reserved

To my friends, who shared

To my family, who taught me

To my dog, who has my heart

ACKNOWLEDGMENTS

I have had the immense fortune to have the support of so many people, without whom this journey would not have been possible.

First and foremost, I would like to thank my mentor, Eric Pamer, for his consistent encouragement and investment in my growth as a physician-scientist. I am so grateful for the immeasurable kindness and patience it took to see us all to the finish line.

To my committee members, Sam Light, Laurie Comstock, and Ran Blekhman, who have all been incredible mentors for me. Despite the twists and turns that my project took, their guidance lit the path forward even when I could not see it myself.

I would like to thank each member of the Pamer Lab: Qiwen Dong, Jerry Zhang, Rita Almeida de Oliveira, Alexander de Porto, Emma McSpadden, Cody Cole, and Rosemary Pope. Thank you and congratulations to Rosemary and Cody, my fellow graduate students who joined me on this journey from start to finish. From a global pandemic to the climatic finale of the lab, I could not have asked for better friends to share this experience. And a special thank you to Qiwen, whose guidance and support were indispensable for my development as a scientist. I found your consistent diligence and creative scientific thinking so inspiring and I look forward to your future endeavors.

Through the DFI, I was supported by so many collaborative and supportive individuals. Nearly everyone at the DFI has lent me their expertise and comradery throughout the years and I will keep these moments in the forefront as I look back to my times in the lab. To the Microbial Metagenomics Facility (MMF), the Host-Microbial Metabolomic Facility (HMMF), and bioinformatics cores, who were anchors for us in the lab. In particular, I would like to thank Anitha Sundararajan, who went above and beyond to plan through experiments with me, and Victoria Burgo, whose diligence and expertise I relied on for many essential parts of my project. And Ashley Sidebottom, whose door was always open to answer my spontaneous questions. I would also like to thank Ingrid Lerner, who wore so many hats at

the DFI and was always so enthusiastic to tackle any hurdles we encountered. And Noreen Bentley, who took on the incredible challenge of coordinating so many of our milestones. I am thankful to the clinical mentors I have been blessed with at the DFI: Matthew Odenwald, Angelica Moran, and Christopher Lehmann. All of whom were amazing physician-scientist role models who enriched my graduate studies with their invaluable clinical pearls and their natural curiosity. And a special thank you to Ram Ramaswamy, my partner-in-crime for many portions of my project, whose generosity and patience are only second to his curiosity and humor.

I would also like to extend my gratitude to the MSTP administration for their dedicated support for students. It is a herculean task to manage so many students yet they manage to accomplish so much.

Of course, I would like to thank my friends who have supported me, and more importantly, suffered through my jokes:

To my friends, colleagues, and peer mentors in the MSTP and Pritzker, thank you all for keeping me with you as you pursued your own training paths. I look forward to the day I will join you once again as colleagues.

To my friends from the BSD graduate programs (COI, COM, etc.) who filled my graduate years with cherished memories filled with bagels and drinks to celebrate our enumerable intramural losses. I expected the path towards the PhD to be a lonely experience but I was delighted to learn that my hypothesis was incorrect.

And to my beloved roommates, who made Chicago feel like home: Yujin Shin, Devika Jaishankar, Teresa Xiao, Maggie Liu, Willa Li, and Diana Li. All of the times you all pretended to understand my project were noticed and deeply appreciated.

Finally, I would like to thank my family, particularly my parents and sister. By carving out new lives across the world, they led by example how to harness resilience and courage to create their own opportunities. We have never been particularly expressive but we have

always been honest, earnest, and delightfully awkward. I am grateful to have your unwavering support as I head towards my next adventure.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

LIST OF FIGURES ix

LIST OF TABLES xi

ABBREVIATIONS xvii

ABSTRACT xviii

1 INTRODUCTION 1
- 1.1 *C. difficile* is a leading cause of nosocomial diarrheal illness 1
  - 1.1.1 Discoveries and developments in CDI clinical management 1
  - 1.1.2 *C. difficile* virulence factors 3
- 1.2 Clinical Microbiology Tools for *C. difficile* 4
- 1.3 Traditional *C. difficile* epidemiology and clinicogenomics 5
  - 1.3.1 Approaches for *C. difficile* epidemiology 5
  - 1.3.2 Next generation sequencing era of molecular epidemiology 7
  - 1.3.3 Taxonomic profiling using MetaPhlAn4 7
- 1.4 Metabolic perspective of *C. difficile* 8
  - 1.4.1 Host-microbe transformations of intestinal bile acids 10
  - 1.4.2 Short chain fatty acids 14
  - 1.4.3 *C. difficile* metabolism 15

2 MATERIALS AND METHODS 17

3 CULTURE-INDEPENDENT STRAIN TYPING OF *CLOSTRIDIOIDES DIFFICILE* IN CLINICAL FECAL SAMPLES 31
- 3.1 Introduction 31
- 3.2 Results 33

4 MULTI-OMIC INVESTIGATION OF *CLOSTRIDIOIDES DIFFICILE* COLONIZATION ACROSS DIVERSE PATIENT BACKGROUNDS 49
- 4.1 Introduction 49
- 4.2 Results 49

5 RESOLVING THE ROLE OF *CLOSTRIDIUM SCINDENS*, IN PROTECTION AGAINST *CLOSTRIDIOIDES DIFFICILE* COLITIS 69
- 5.1 Introduction 69
- 5.2 Results 70
  - 5.2.1 Physiological and inhibitory concentrations of bile acids 70
  - 5.2.2 *C. scindens* secondary bile acid production *in vitro* 71

5.2.3 4-member bacterial consortia is sufficient for reducing CDAD *in vivo* 71
5.2.4 *C. scindens* strain-dependent impact of nutrient competition *in vitro* 73
5.2.5 *C. scindens* monocolonization is not sufficient to protect against *C. difficile in vivo* . . . 74
5.2.6 Diversity in metabolic fitness within ST1 strains . . . 75

6 DISCUSSION . . . 85
6.1 Applications for strain-level metagenomic tracking in the clinic . . . 85
6.2 Multi-omic perspective on *C. difficile* . . . 88
6.3 Sub-inhibitory SBA production protects against severe CDI . . . 89
6.4 Strain-specific metabolomic protection against CDI . . . 90

7 SUPPLEMENTAL FIGURES . . . 91

REFERENCES . . . 111

LIST OF FIGURES

1.1 Brief timeline of major discoveries and developments in *C. difficile* clinical management (1935-2025). . . . 2
1.2 Laboratory tests for the diagnosis of CDI. . . . 4
1.3 Simplified model for MetaPhlAn4 and StrainPhlAn4. . . . 9
1.4 Biosynthetic pathway map of bile acids by human and/or murine (green) enzymes. 11
1.5 Simplified model of microbial transformations of cholic acid (CA) in the intestinal lumen. . . . 12
1.6 Biogeography of *C. difficile*-bile acid interactions throughout the *C. difficile* lifecycle. . . . 13

3.1 Benchmarking microbial profiling and *C. difficile* detection performance of MetaPhlAn4 *in silico* . . . 35
3.1 Benchmarking microbial profiling and *C. difficile* detection performance of MetaPhlAn4 *in silico* (cont.) . . . 36
3.2 Optimization of *C. difficile* detection by MetaPhlAn4 . . . 39
3.3 Optimized MetaPhlAn4 performance on patient samples . . . 42
3.3 Optimized MetaPhlAn4 performance on patient samples (cont.) . . . 43
3.4 *C. difficile* detection in simulated clinical samples using optimized StrainPhlAn4 44
3.4 *C. difficile* detection in simulated clinical samples using optimized StrainPhlAn4 (cont.) . . . 45
3.5 Improved within-species accuracy of StrainPhlAn4 *C. difficile* tracking . . . 48

4.1 Proportion of samples across *C. difficile* subcategories and fecal consistency groups 51
4.2 Survey of *C. difficile 16S rRNA* PCR positive patient fecal samples . . . 52
4.3 Survey of HD and *C. difficile 16S rRNA*$^-$ patient fecal samples . . . 53
4.3 Survey of HD and *C. difficile 16S rRNA*$^-$ patient fecal samples (cont.) . . . 54
4.4 Administration of antibiotics to study patients . . . 56
4.5 Microbial diversity and associated species across fecal samples . . . 57
4.6 Microbe-group associations with *C. difficile*$^+$ fecal samples and HD samples with MaAslin2 . . . 59
4.7 Microbe-group associations with prior antibiotics exposure with MaAslin2 . . . 61
4.8 Microbe-group associations with stool sample consistency with MaAslin2 . . . 62
4.9 Qualitative metabolite features differentially associated with the absence of *C. difficile* . . . 64
4.10 Qualitative metabolite features differentially associated with the absence of *C. difficile* . . . 65
4.11 Most predictive features from multi-omic integration . . . 68

5.1 Mechanisms driving *C. scindens*-mediated host protection against symptomatic CDI. . . . 70
5.2 Comparisons of physiological bile acid compositions . . . 77
5.3 Cholate above 300 uM in *C. scindens* CFS result in *C. difficile* vegetative growth inhibition . . . 78

5.4 Preliminary differences between protective and nonprotective gut metabolite profiles 79
5.5 *C. scindens* strain-dependent differences in *C. difficile* growth inhibition *in vitro* 80
5.6 *C. scindens* MSK.5.24 is insufficient to protect against *C. difficile in vivo* 81
5.7 *C. scindens* ATCC35704 is insufficient to protect against *C. difficile in vivo* . . 82
5.8 ST1-75 protection is attributable to faster amino acids depletion in nutrient-limited environment 83
5.9 ST1-75 depletes specific amino acids faster than R20291 84

S1 Model for construction of 150-member synthetic metagenome and *C. difficile* spike-in experiments 92
S2 *C. difficile* classification performance on the 150-member syncom metagenome. . 93
S3 Marker gene mapping and curation for CDC 30 *C. difficile* genomes. 94
S4 Patient sample inclusion and *C. difficile* isolation protocol 96
S4 Patient sample inclusion and *C. difficile* isolation protocol (cont.) 97
S5 Taxonomic profiles of healthy donor (HD) and culture positive (Cx pos) fecal metagenomic samples. 98
S6 Optimization of StrainPhlAn4 for the detection of *C. difficile* 99
S6 Optimization of StrainPhlAn4 for the detection of *C. difficile* (cont.) 100
S7 Model and patient selection for StrainPlAn4 *in silico* experiments (cont.) 101
S7 Model and patient selection for StrainPlAn4 *in silico* experiments 102
S8 Maximum-likelihood optimized StrainPhlAn4 phylogenetic tree output of CDC *C. difficile* isolate spike-ins into patient stool metagenomic samples (n=5). Genomes were separated into three sets as previously described (a): set 1 (green), set 2 (red), and set 3 (blue). For each spike-in, the *C. difficile* genome was randomly sampled at 30,000 (1x) or 60,000 (2x) reads while the reference genomes (gray) were sampled at 300,000 reads (10x) using optimized settings. 103
S9 Maximum-likelihood phylogenetic tree StrainPhlAn4 output of patient-derived *C. difficile* isolate spike-ins (red) into patient stool metagenomic samples (n=5). For each spike-in, the *C. difficile* genome were randomly sampled at 30,000 (1x) or 60,000 (2x) reads while the isolated culture genomes (blue) and the reference tree (gray) were sampled at 300,000 reads (10x) using optimized settings. 104
S10 Quantification of *C. difficile* from patient fecal samples using qPCR 105
S11 Quantification of *C. difficile* across *C. difficile* culture-positive patient fecal samples 106
S12 Quantified fecal metabolites from HDs and patient samples 107
S13 Quantified fecal metabolites from HDs and Cx positive patient samples 108
S14 Spread of training sample data set by *C. difficile* group inclusion on latent omic block component axes 109
S15 Quantification of *C. difficile* from patient fecal samples using qPCR 109
S15 Competition for proline and carbon sources impact *C. difficile* vegetative growth *in vitro* 110

LIST OF TABLES

S1 Human fecal sample shotgun sequencing read count summary statistics 91
S2 Curated MetaPhlAn4 *C. difficile* marker genes 95

ABBREVIATIONS

3-oxoLCA 3-oxolithocholic acid

5-AV 5-aminovalerate

5AR 5α-reductase

AIC Akaike Information Criterion

alloCA allocholic acid

alloisoLCA alloisolithocholic acid

ANI average nucleotide identity

ASV amplicon sequence variant

AUROC Area Under the Receiver Operating Curve

B6 C57BL/6

BA bile acid

BAAT bile acid-coenzyme A:amino acid N-acyltransferase

BACS bile acid-CoA synthetase

BCAA branch chain amino acid

BHI brain-heart infusion media

BSH bile salt hydrolase

CA cholic acid

CDAD *C. difficile* associated disease

CDC Centers for Disease Control

CDCA chenodeoxycholic acid

CFS cell-free supernatant

cfu colony-forming units

CI competitive index

CIA Coinertia analysis

CLR Centered Log Ratio

CRI Center for Research Informatics

CTA cytotoxicity assays

DCA deoxycholic acid

dpi days post-infection

EIA enzyme immunoassay

EIP Emerging Infections Program

FDA Food and Drug Administration

FMT fecal microbiota transplant

GC-MS gas chromatography-mass spectrometry

GCA glycocholic acid

GCDCA glycochenodeoxycholic acid

GF germ free

GGB Genus-level Genome Bin

GPCR G protein-coupled receptor

HD Healthy donor

HDAC histone deacetylase

HSDH hydroxysteroid dehydrogenases

IACUC Institutional Animal Care and Use Committee

IBAT ileal bile acid transporter

IBD Inflammatory Bowel Disease

isoDCA isodeoxycholic acid

isoLCA isolithocholic acid

IV intravenous

LC-MS liquid chromatography-mass spectrometry

LCA lithocholic acid

LP lamina propria

MBA murine bile acid

MCBA microbially conjugated bile acid

MIC minimum inhibitory concentration

MLST multilocus sequence typing

MLVA multilocus variable number tandem repeat analysis

mNGS metagenomic next generation sequencing

MP MetaPhlAn4

NAAT nucleic acid amplification test

NGS next generation sequencing

OD optical density

PA Pareto analysis

PaLoc pathogenicity locus

PBA primary bile acid

PBS phosphate buffered saline

PCR polymerase chain reaction

PFGE pulse-field gel electrophoresis

PMC pseudomembranous colitis

qPCR quantitative PCR

RA relative abundance

rCDI recurrent *C. difficile* infection

REA restriction enzyme endonuclease analysis

RI relative inhibition

SAM Sequence Alignment Map

SBA secondary bile acid

SCFA short chain fatty acid

SGB Species Genome Bin

SPF specific pathogen free

sPLSDA sparse Partial Least Squares Discriminant Analysis

ST sequence type

TBA tertiary bile acid

TCA taurocholic acid

TCDCA taurochenodeoxycholic acid

Treg regulatory T cell

TSS Total Sum Scaling

UCA ursocholic acid

UDCA ursodeoxycholic acid

VIF Variance Inflation Factors

VRE vancomycin-resistant *Enterococcus*

WGS whole genome sequencing

WLP Wood-Ljungdahl Pathway

WT wild type

ABSTRACT

C. difficile, a spore-forming anaerobe, is a leading infectious cause of nosocomial diarrhea. The clinical outlook for *C. difficile* associated disease (CDAD) is highly variable and includes asymptomatic carriage, mild diarrheal illness, and pseudomembranous colitis [1]. Symptomatic CDAD outcomes are associated with antibiotic-induced gut dysbiosis, so understanding the microbial ecology mediating colonization resistance against *C. difficile* may add valuable clinical context for infection control and treatment [2,3,4,5]. The clinical use of multi-omic approaches, integrating shotgun metagenomic sequencing and metabolomic profiling directly from patient fecal samples, may have the potential to track the spread of and investigate the dynamics of *C. difficile* in real time.

Despite this, the practical application of these tools in the clinic remains limited. As a step towards that goal, we optimized the powerful marker gene-based taxonomic profiling tool, MetaPhlAn4, and developed novel protocols for species level sensitivity tuning and strain-level *C. difficile* tracking. Using our locally optimized parameters, we investigated significant microbial and metabolic features discriminating CDI across diverse patient backgrounds. We then investigate the metabolic factors driving *C. difficile* inhibition by competing *Clostridia spp.* such as *C. scindens*, *C. symbiosum*, and *C. sporogenes*. Through these investigations, we observe strain-level variation in the inhibition of *C. difficile* by nutritional competition.

CHAPTER 1

INTRODUCTION

1.1 *C. difficile* is a leading cause of nosocomial diarrheal illness

1.1.1 Discoveries and developments in CDI clinical management

In 1935, *C. difficile* (then named *Bacillus difficile* but quickly renamed as *Clostridium difficile*) was identified as a "difficult to culture bacillus" in the normal neonate microbiota[6]. *C. difficile* remained understudied until the start of the '*C. difficile*'-era in the 1970s, identified as the causative agent of clindamycin-induced colitis and pseudomembranous colitis (PMC)[7,2]. *C. difficile* finally donned its current moniker of *Clostridioides* (*Clostridium*-like) *difficile* in 2016, due to the restriction of the *Clostridium* genus[8].

Traditional treatments for *C. difficile* have included the use of antibiotics that act locally in the gut such as oral route vancomycin and fidaxomicin. Although both treatments target gram-positives locally, fidaxomicin, a macrolide antibiotic, is currently preferred over oral vancomycin, a tricyclic glycopeptide, for recurrent *C. difficile* infection (rCDI) because of its narrow spectrum activity against *Clostridia*. Metronidazole was indicated as a first-line treatment in 2010 but has since fallen out of favor due to its broad activity against gram positives and rapid absorption in the jejunum. Due to growing concerns for gut microbial health, increasing rates of antibiotic resistance, and increased risk of CDI, alternative treatment options for enteric pathogens were needed. An innovative treatment, fecal microbiota transplant (FMT), demonstrated remarkable efficacy against rCDI[9,10] with better efficacy than fidaxomicin[11] and is associated with the recovery of microbiota diversity and of secondary metabolites in the gut lumen[12,13].

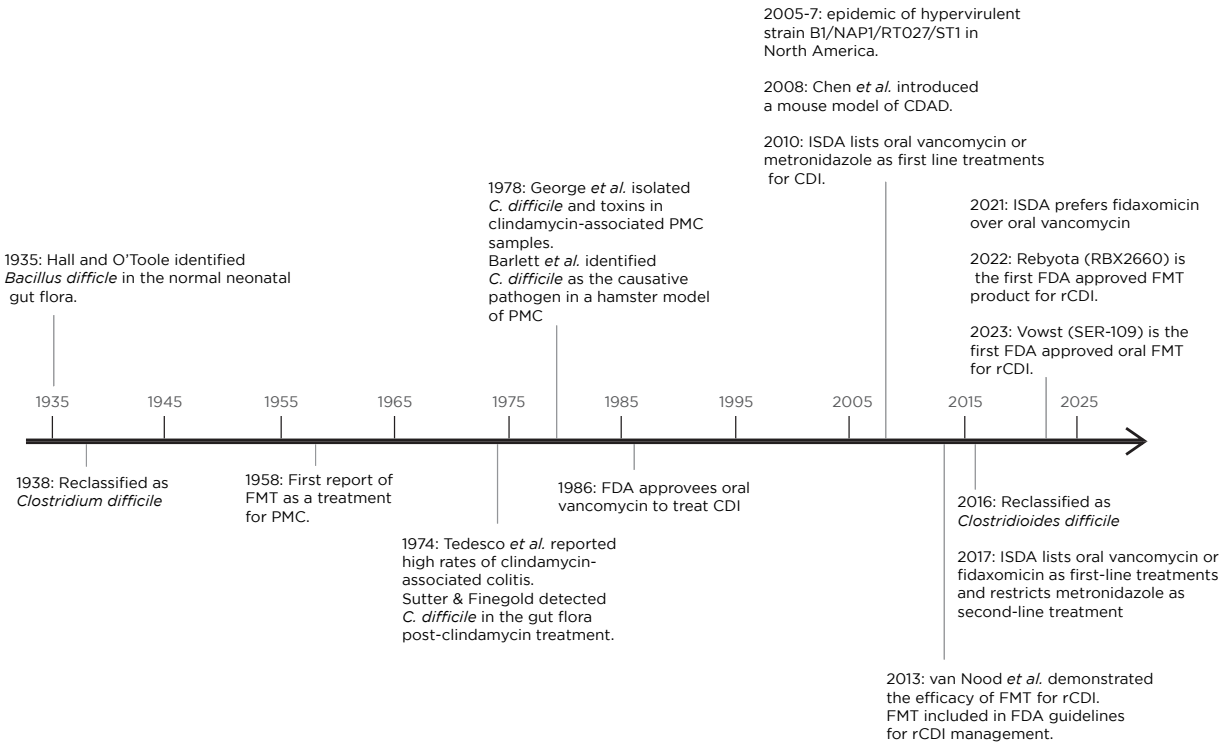

Figure 1.1: Brief timeline of major discoveries and developments in *C. difficile* clinical management (1935-2025).

* PMC = pseudomembranous colitis

1.1.2 C. difficile virulence factors

Symptomatic CDAD is attributed to the effects of two major virulence factors, Toxin A and Toxin B, large monoglycosyltransferases that disrupt the integrity of the gut epithelium[14]. These exotoxins, like other large *Clostridial* toxins (LCTs), glycosylate small GTPases on Ras and Rho[15], ultimately leading to cytoskeletal dysfunction and apopotosis[16]. The genes encoding these toxins, along with genes *tcdC*,*tcdE*, and *tcdR* are located on the pathogenicity locus (PaLoc), a 19.6 kb region of the chromosome. The expression of the PaLoc is positively regulated by *tcdR*, an RNA polymerase sigma factor[17], and *tcdC*, a negative regulator of TcdR[18]. While *tcdE*, a bacteriophage holin, is required for the secretion of the toxins[19]. The PaLoc is found at the same integration site across all toxigenic strains[20,21] but this locus is replaced by a highly conserved 115 or 75 bp region in nontoxigenic strains[14,22,21]. Initially, the spectrum of toxic effects of CDI were attributed to Toxin A[23] but more recent investigations have redefined Toxin B as the primary driver of CDI pathogenesis[24]. As a result, many clinical diagnostic assays for *C. difficile* target TcdB from patient fecal samples.

Some hypervirulent strains of *C. difficile*, such as the epidemic NAP1/BI/027/ST1 strains, encode an unrelated toxin, the binary toxin, on a 6.2 kb Cdt Locus (CdtLoc)[25]. The binary toxin, encoded by approximately 17-23% of *C. difficile* strains[26], is composed of the binding component (CdtB) that facilitates the internalization of the enzymatic ADP-ribosyltransferase (CdtA) into host colonocytes, leading to the dysregulation of the actin cytoskeleton[27,25] and is positively regulated by CdtR[25]. Comparatively, the binary toxin is less understood than the PaLoc but it is typically associated with increased virulence. However, recent studies have challenged the prevailing model that the production of Toxin A and B alone are sufficient for *C. difficile* virulence[28].

1.2 Clinical Microbiology Tools for *C. difficile*

Clinical diagnosis of CDI is challenging and requires clinical assessment combined with a choice of laboratory tests (fig. 1.2). Diagnostic approaches for CDI are broadly categorized as "culture-based" (e.g. anaerobic culture) or "molecular" tools (e.g. toxin EIA, toxin nucleic acid amplification test (NAAT), and cytotoxicity assays (CTA)) (fig. 1.2). The "gold standard" for *C. difficile* diagnosis combines anaerobic culture from fecal samples followed by toxin detection to identify toxigenic isolates. However, this approach is uncommon in hospital microbiology laboratories because it is time, labor, and resource-intensive[29]. As a result, many clinical laboratories forgo traditional culture-based methods for a multi-step targeted molecular approach[30], that while fast and cost efficient, loses epidemiological utility.

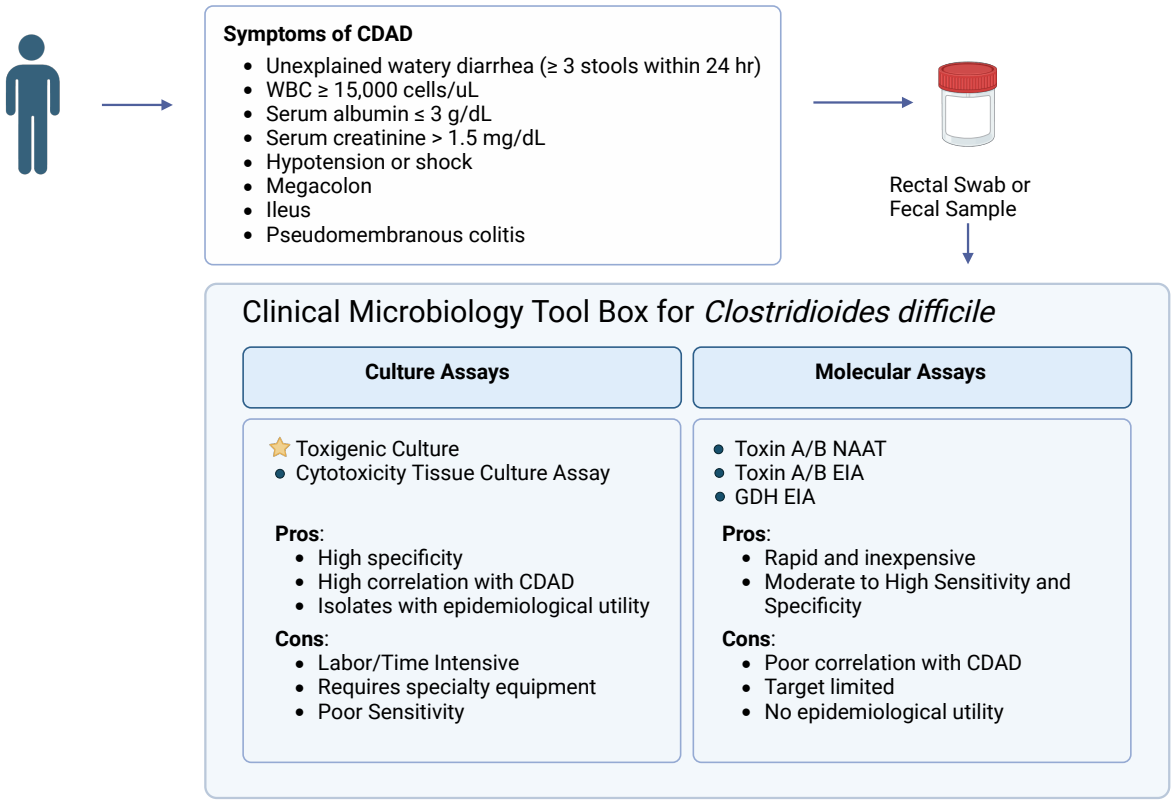

Figure 1.2: Laboratory tests for the diagnosis of CDI.

1.3 Traditional *C. difficile* epidemiology and clinicogenomics

Molecular microbial epidemiology has transformed our understanding of *C. difficile* with the rise of hypervirulent strains in North America. *C. difficile* is broadly classified across 5 major clades (Clade 1-5) and 3 more distantly related cryptic clades (C-I to C-III). The common ancestor of the major clades has been traced back 1.11 to 6.71 million years ago[31]. The within-host evolutionary rate was estimated to be approximately 1.4 mutations per genome annually (3.2×10^{-7} mutations per nucleotide per year)[32,33], a moderate to slow pace for bacterial evolution.

Strain-level differences in *C. difficile* has been associated with variability in virulence[28,34] and metabolism[35,36]. Disease-associated toxigenic strains are present across most of the major clades: Clades 1, 2, 3, and 5[32,31]. *C. difficile* is a species with a highly mobile and variable genome[36,37]. In contrast to other *Firmicutes*, *C. difficile* has a large genome that typically ranges between 4.1-4.3 Mb with between 3,400-3,900 coding sequences[38,39]. Comparative phylogenomics estimates a core genome, a set of homologous genes that are present in nearly all ($\geq 99\%$) compared genomes from the species, of about $61.1 - 72.8\%$ (2317-2756 core genes)[40,41]. And *C. difficile* encodes a large, highly-plastic pangenome, a complete set of unique and shared genes including accessory genes, of over 9640 genes[42]. Due to the high genomic diversity and dispersal of toxigenic strains of *C. difficile*, strain-level epidemiology is critically important for public health.

1.3.1 Approaches for C. difficile epidemiology

Prior to the 1980s, only phenotypic methods, such as serotyping[43] and immunoblotting, were available for *C. difficile*. However, these phenotypic methods had low reproducibility, low typeability, and insufficient discriminatory power. The development of molecular typing overcame many of these shortcomings and can be summarized into banding pattern-based or sequence-based approaches.

In 1984, alternating electrical fields were shown to separate large DNA segments in an agarose gel[44]. Pulse-field gel electrophoresis (PFGE) is a molecular typing approach that discriminates subtypes using distinct banding patterns after restriction enzyme digestion of bacterial DNA. PFGE has long been regarded as the "gold standard" of fingerprinting bacteria and is still commonly used by authorities (e.g. US Food and Drug Administration (FDA) and Centers for Disease Control (CDC)) due to its high discriminatory power, reproducibility, and typeability, the ability to type isolates unambiguously. PFGE has been used to type *C. difficile* since 1993[45,46,47] and was instrumental in the discovery and management of the ST1 outbreak[48]. Restriction enzyme endonuclease analysis (REA) is another agarose-based technique that digests the *C. difficile* genome with the HindIII restriction enzyme. This technique offers high reproducibility but is more challenging to perform and to interpret than competing methods[49].

PCR ribotyping is an agarose-based technique that amplifies the *16S-23S* intergenic spacer region (ISR) and leverages the heterogeneity of these hypervariable regions for strain typing. *C. difficile* demonstrates particular intraspecies diversity in bacterial *rRNA (rrn)* operons[50], so is suitable for this technique[51]. As a result, PCR ribotyping has been the predominant typing strategy routinely used in Australia and Europe[52].

Multilocus sequence typing (MLST) is the most widely used DNA-based method for bacterial strain typing and is based on the amplication of 400-500bp regions spanning 7-8 highly conserved house-keeping genes and bins similar strains into sequence types (STs). The schema developed for *C. difficile* MLST[53] has been widely adopted for *C. difficile* epidemiological studies[54]. The discriminatory power of MLST is comparable to PCR ribotyping[55] and, as of June 2025, there are 1167 STs for *C. difficile*. It offers portability between different laboratories and is routinely used for hundreds of species through a centralized database[56].

1.3.2 Next generation sequencing era of molecular epidemiology

The lack of standardization across different traditional molecular typing strategies has made it difficult to compare the spread of *C. difficile* between different reference laboratories, hindering the efficacy of global epidemiology. Genomics-based approaches based on second (i.e. Illumina) and third-generation (i.e. PacBio) sequencing technology was the next leap forward in molecular epidemiology –opening the door for clinicogenomics, the use of genomic data for diagnostic, therapeutic, and epidemiological investigations[57]. Whole genome sequencing (WGS) has quickly been adopted for various investigative and epidemiological purposes since it provides a comprehensive overview of the entire genome. WGS offers unparalleled discriminatory power for strain discrimination that has been used to track the transmission of *C. difficile* in different clinical settings[58,59,60,61]. Metagenomics, the study of nucleotides from all organisms within a system, has become accessible with the use of short-read NGS technologies[57]. Initially, microbes were differentiated using variants found along hypervariable regions of microbial 16S ribosomal RNA. Alternatively, shotgun sequencing provides a broad, unbiased overview of all genomic material within a sample –providing information from eukaryotic, bacterial, fungal, and viral sources. Although the cost of shotgun metagenomic sequencing is significantly higher than *16S rRNA* metagenomic sequencing, steadily decreasing costs of sequencing, have made sequencing-based approaches more feasible. Compared to WGS, shotgun metagenomics is pathogen agnostic, provides information about the diversity of the gut flora, and does not rely on the successful culture of microbes. However, the practical use of shotgun metagenomics in clinical workflows, particularly with short-read sequencing, remains largely underdeveloped.

1.3.3 Taxonomic profiling using MetaPhlAn4

Accurate taxonomic profiling of metagenomic datasets is a major challenge and several strategies have been employed with varying costs and benefits. MetaPhlAn4 is a marker

gene-based taxonomic profiling tool that offers rapid and accurate results and is easily incorporated in bioBakery workflows, a suite of downstream microbiome analytic tools[62]. MetaPhlAn4 develops an internal database of species-specific marker genes by identifying at least 800 genes (450-4,500 bp) with a maximum "coreness" threshold, prevalence of the gene within a SBG (%), of 50-60% and then comparing these markers across all Species Genome Bins (SGBs) to identify up to 200 species-specific markers that align to less than 1% of other SGBs[63] (fig. 1.3). Taxonomic profiling using MetaPhlAn4 involves mapping sequencing short-reads to the marker gene database using Bowtie2[64]. One advantage of MetaPhlAn4 is the integration into bioBakery, a popular suite of tools for omics analysis[62] and incorporation of StrainPhlAn, for strain-level inferences[63]. StrainPhlAn4 is packaged with MetaPhlAn4, and uses the Sequence Alignment Map (SAM) files generated from mapping reads to the SGB marker genes database to determine the consensus marker gene[63]. Quality marker genes are then processed with metagenomics samples with PhyloPhlAn[65], for multiple sequence alignment (default = MAFFT[66]), mutation rate estimation, and maximum-likelihood phylogenetic tree building using RAxML[67].

1.4 Metabolic perspective of *C. difficile*

Metabolomics, the study of all small molecules (50-1,500 daltons (Da)) in a given organism or biological specimen, provides a static "snap-shot" of the dynamic metabolic processes. Compared to other omic datasets, metabolomics more closely reflects the biochemical phenotype of a particular system: aggregating transcriptional, translational, and post-translational regulation. Metabolomic studies have expanded our understanding of complex microbial interaction networks driving host-microbe interactions in the gut: incorporating (but not limited to) competition, co-feeding, and small molecule warfare.

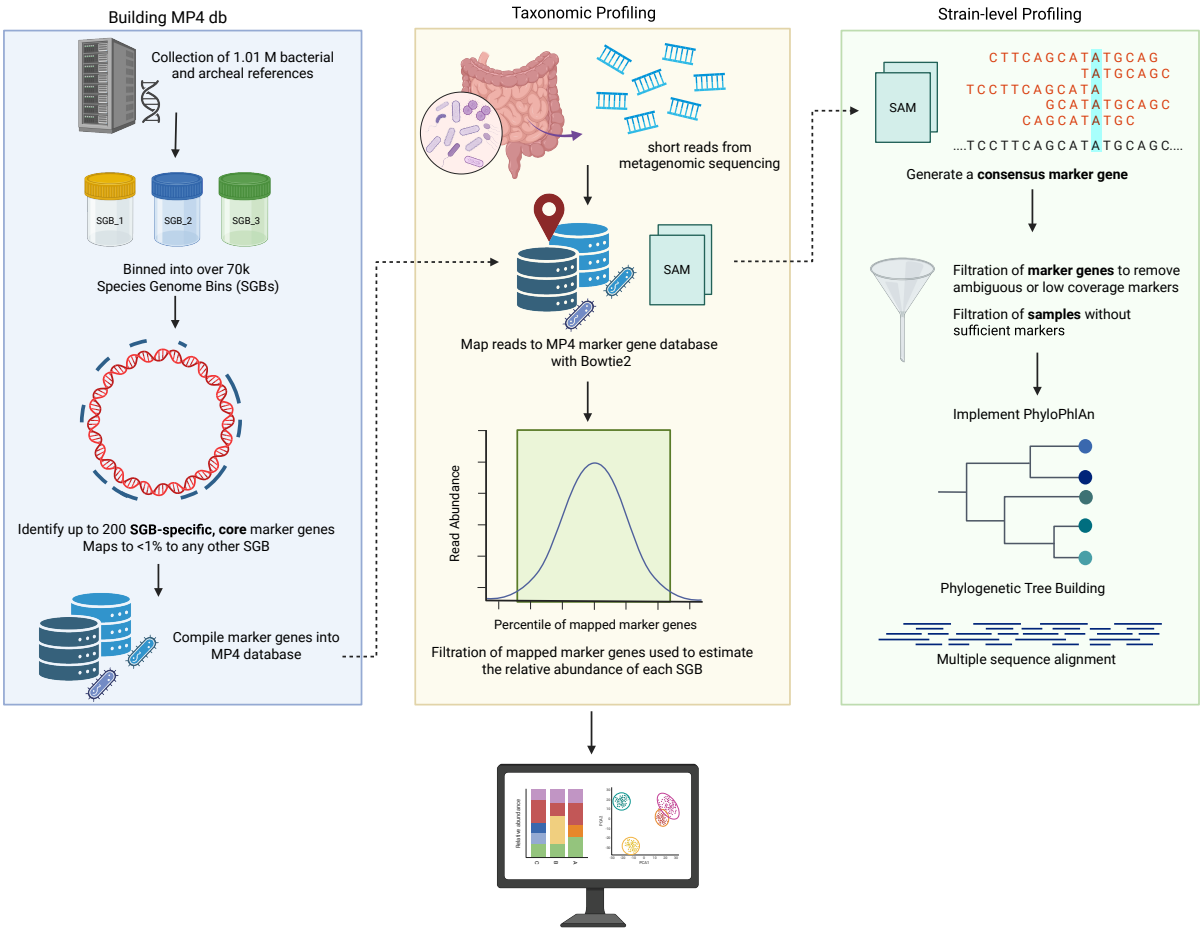

Figure 1.3: Simplified model for MetaPhlAn4 and StrainPhlAn4.

1.4.1 Host-microbe transformations of intestinal bile acids

Bile acids (BAs) are 24-carbon steroid detergents in the gut that enhance host absorption of dietary fats by forming mixed micelles[68]. Bile acid synthesis occurs in the liver in two pathways: the "classic pathway" and the "alternative pathway" (fig. 1.4). The classic (neutral) pathway initiated by liver cytochrome CYP7A1 produces 90% of the bile acid pool in humans. This pathway can produce both major primary bile acids (PBAs): cholic acid (CA) and chenodeoxycholic acid (CDCA)[69]. The alternative (acidic) pathway is initiated CYP27A21 and only produces CDCA. The acidic pathway predominantly occurs in the liver but can be present in other tissues including the lungs and brain[70,71]. PBAs are amidated with glycine or taurine by bile acid-coenzyme A:amino acid N-acyltransferase (BAAT) and bile acid-CoA synthetase (BACS) in the liver into conjugated PBAs[69]. Conjugated PBAs are stored in the gallbladder and is excreted into the duodenum as a response to meal consumption. The majority (95%) of bile acids are reabsorbed throughout the small intestine and through sodium-dependent ileal bile acid transporter (IBAT) on ileocytes, reentering hepatic circulation[69]. Recirculated SBAs can be additionally modified in the liver to produce tertiary bile acids (TBAs)[72].

Microbes in the gut contribute several reversible bile acid-modifying reactions, generating incredible diversity in the bile acid pool (fig. 1.5). bile salt hydrolase (BSH) deamidates the taurine or glycine from conjugated bile acids into PBAs (CA and CDCA). BSH activity is broadly present across the phyla in the gut and recent studies have started to appreciate the reconjugation of amino acids into microbially conjugated bile acids (MCBAs)[73,74]. Hydroxysteroid dehydrogenasess (HSDHs) facilitate regiospecific and stereospecific epimerization of a hydroxysteroid group into a ketone group[71]. The conversion of typical (kinked) BAs into their "allo" (flat) structure was attributed to 5α-reductase (5AR) or BaiA1 activity[75,76]. In addition, sulfation of LCA and other hydrophobic BAs by bacterial sulfatases can modify their solubility and overall cytotoxicity[77,78].

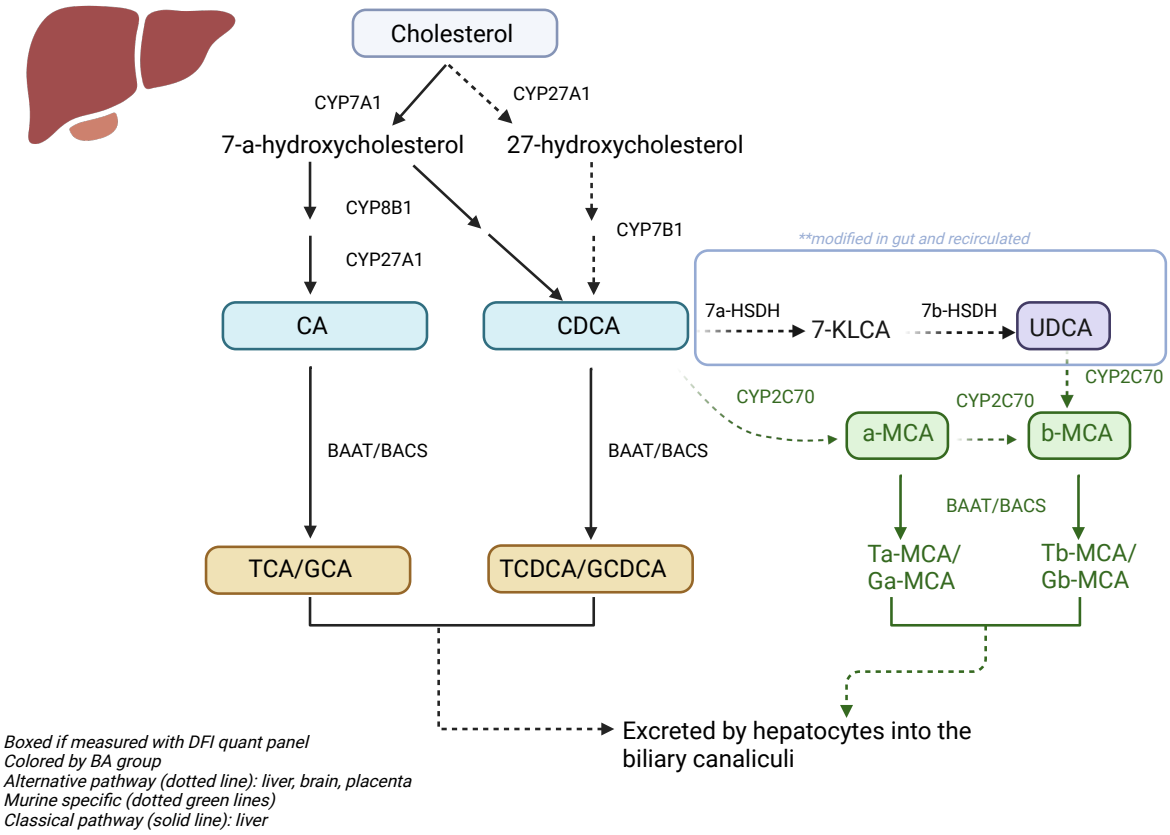

Figure 1.4: Biosynthetic pathway map of bile acids by human and/or murine (green) enzymes.

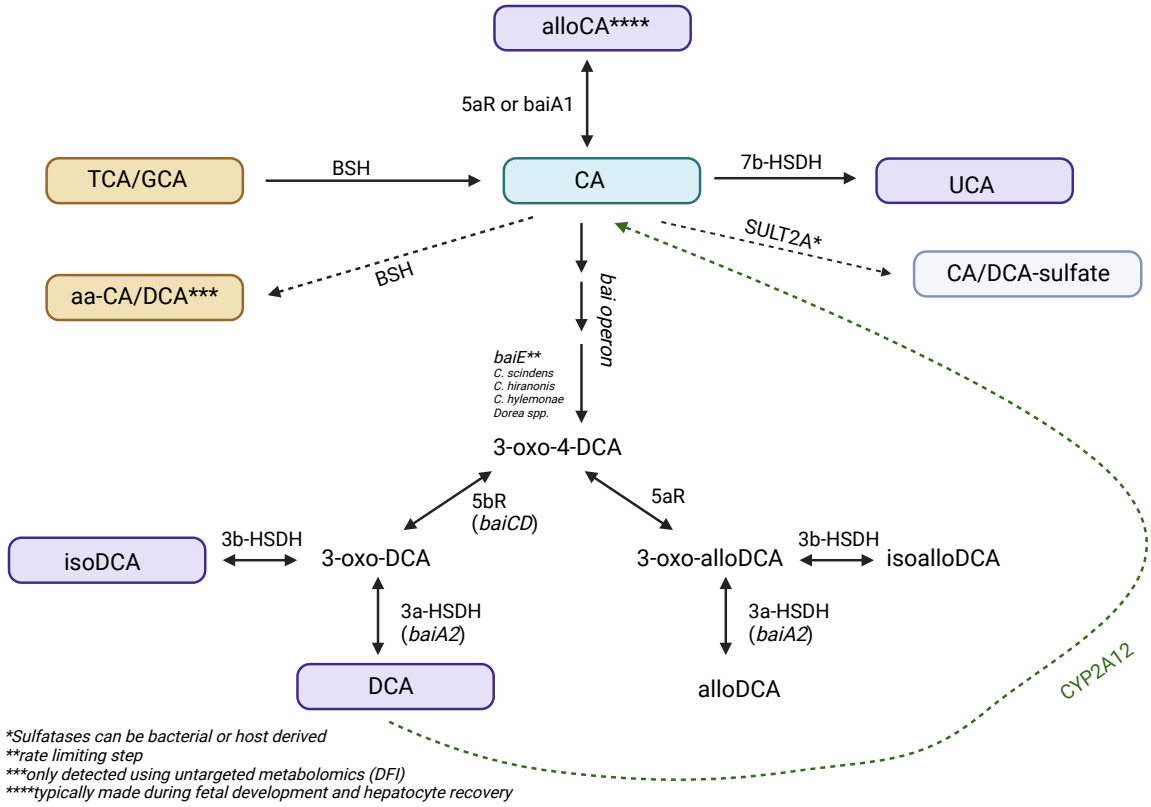

Figure 1.5: Simplified model of microbial transformations of cholic acid (CA) in the intestinal lumen.

Bile acids play different roles throughout the *C. difficile* life cycle (fig. 1.6). As *C. difficile* transits the gastrointestinal tract, ingested spores are stimulated to germinate into vegetative bacilli from co-germinants such as taurocholic acid (TCA), glsca, and glycine[79], while CDCA reduces overall germination. However, in the colon, the concentrations of SBAs have been shown to be inhibitory against vegetative *C. difficile* growth, particularly those with increased hydrophobicity and toxicity[80]. SBAs (particularly DCA and LCA) have been suggested to regulate *C. difficile* TcdA production[81], reduce motility[82,83], and modify the *C. difficile* metabolism[80].

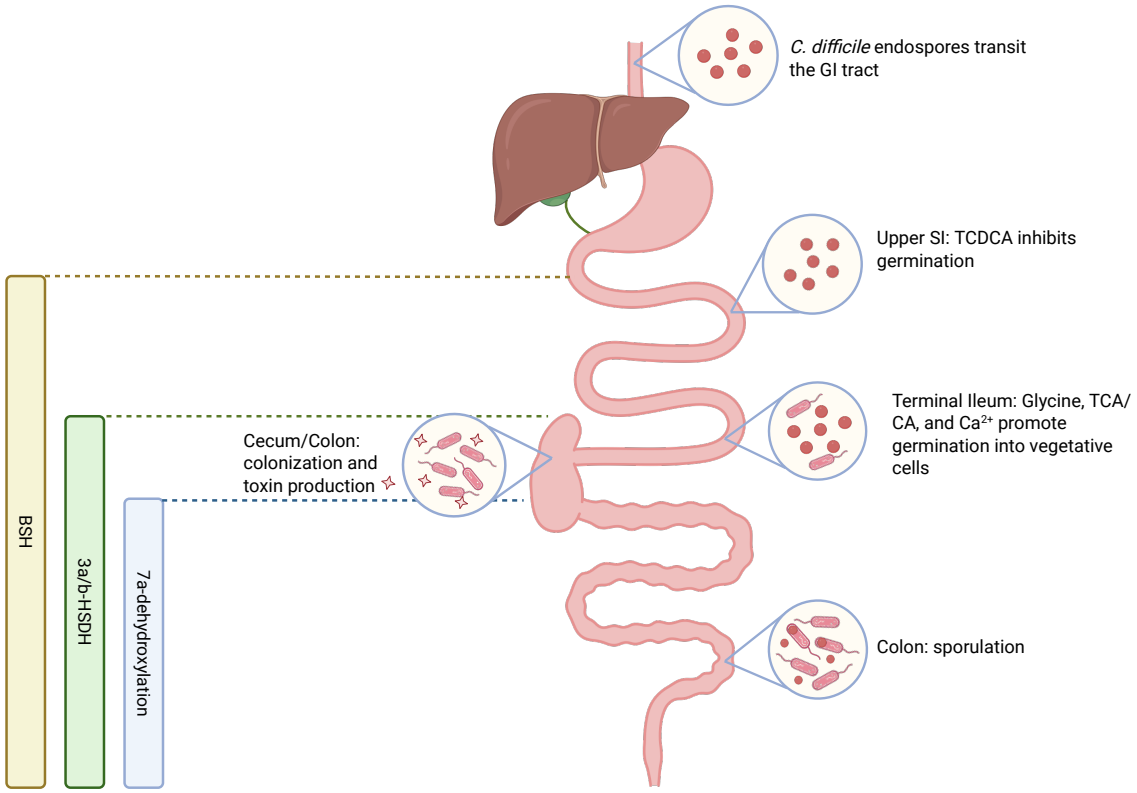

Figure 1.6: Biogeography of *C. difficile*-bile acid interactions throughout the *C. difficile* life-cycle.

1.4.2 Short chain fatty acids

Short chain fatty acids (SCFAs) are small molecules (1-6 carbons) produced by microbial fermentation of non-digestible fiber or branch chain amino acids (BCAAs) and are important drivers of host-microbe crosstalk in the gut. Primary fermenters of polysaccharides (e.g. members of phylum *Bacteroidetes*) produce intermediate compounds (e.g. lactate, acetate, or ethanol) that are consumed and modified into SCFAs (e.g. members of phylum *Firmicutes*). SCFAs are readily absorbed by the gut and circulated systemically promoting pleiotropic effects throughout the host. Notably, SCFAs can modulate the immune system through histone deacetylase (HDAC) inhibition and as G protein-coupled receptor (GPCR) ligands[84]. Acetate, butyrate, and propionate are the most common SCFAs in the human gut ($> 95\%$), while other SCFAs (e.g. valerate, formate, and caproate) constitute the remaining 5%.

Acetate is the most abundant SCFA in the colonic lumen and is a by-product of pyruvate and of the Wood-Ljungdahl Pathway (WLP). *C. difficile* is known to encode the complete set of genes for the WLP, but expresses these genes at lower rates than other acetogenic bacteria due to a preference for Stickland fermentation[85]. Acetate binds to the free fatty acid receptor 2 (FFAR2) receptor (found on neutrophils and ILC3s), inducing IL-1β production in an inflammasome-dependent manner, preventing bacterial translocation to the lymph nodes, and boosting IL-22 expression to promote anti-microbial peptide production[86]. Potassium acetate is also known to compete for histone deacetylase-6 (HDAC-6) binding, a receptor for Toxin A, leading to reduced Toxin A-mediated tubulin degradation[87]. Several members of *Clostridium cluster XIVa and IV* are significant producers of butyrate, most commonly through the formation of butyryl-CoA and subsequent removal of the CoA moiety[88]. Butyrate, a primary energy source for colonocytes (60-70%)[89], can inhibit *C. difficile* pathogenesis by increasing tight junction expression and stabilizing the intestinal integrity in a hypoxia-inducible Factor 1 (HIF-1)-dependent manner[90]. Butyrate-producing *Clostridium butyricum (C. butyricum)* and tributyrin, a butyrate precursor, also increased the number

of neutrophils, INF-γ, and IL-17A in the colonic lamina propria (LP)[91]. On the other hand, many strains of *C. difficile* are butyrogenic and have been known to use acetyl-CoA, 4-aminobutyrate, and succinate[92] as energy sources, resulting in butyrate[93].

1.4.3 C. difficile metabolism

Nutritional niche competition and Stickland fermentation

Several studies have indicated that the nutrient competition for amino acids can significantly impact *C. difficile* outgrowth *in vitro*[94]. Stickland fermentation was identified in 1934 in *C. sporogenes* and was found to be common pathway shared across *Clostridial sp.*; it is a hallmark of Cluster XI (which includes *C. difficile*)[95,96,97] and can be present in Cluster I (*C. sporogenes*), Cluster II, and Cluster XIVa (*C. scindens*)[97]. This reaction pairs amino acids as electron donors (deaminated or decarboxylated to produce NADH) or acceptors (reduced to produce NAD^+) as an alternative carbon and nitrogen source in carbohydrate-depleted environments. Although most amino acids can act as reductants (electron donors), proline and glycine are preferred oxidants (electron acceptors). Fermentable amino acids are relatively weak oxidants (reduction potential -190 mV to -10 mV) compared to molecular oxygen (+810 mV), but serve as important electron sinks in anaerobic environments[97]. An environment with high levels of available amino acids, including proline, glycine, cysteine, isoleucine, valine, leucine, and tryptophan, reflects favorable conditions for CDI[98]. *C. difficile* 630δerm was shown to preferentially deplete proline, cysteine, and leucine first, followed by other amino acids such as isoleucine, glycine, and valine, while aromatic amino acids, phenylalanine and tyrosine, are minimally metabolized[99].

Proline reduction is mediated by proline reductases on the *prd* operon (PrdB and PrdA): selenoenzymes generating 5-aminovalerate (5-AV) and NAD^+. Similarly, glycine is reduced by glycine reductases (GrdA and GrdB) encoded on the *grd* operon and generating acetate and NAD^+[100]. The presence of proline and cysteine was shown to reduce *C. difficile* toxin

production, specifically reducing the expression of *tcdA*[100]. Additionally, glycine is known to be a precursor for purine nucleotides and is a germination signal for *C. difficile* spores[101,79].

Transition metal competition

The competition of micronutrients, in particular transition metal ions (e.g. Fe, Mn, Zn), is known to impact the disease outcomes for CDI and other gut infections. The bioavailability of these metals is essential for all organisms as necessary co-factors involved in several metabolic pathways[102]. In low iron environments, the master iron regulator, ferric uptake regulator (Fur), regulates glycine reduction by repressing *grdB*[103,104]. However, the overabundance of intracellular iron is toxic due to Fenton chemistry, the oxidation of ferrous iron (Fe^{2+}) and hydrogen peroxide (H_2O_2) which produces highly oxidative free hydroxyl (OH^-) radicals[105]. As a result, the bioavailability of these metal ions is tightly regulated by host factors that weaponize both the sequestration and the influx of bioavailable iron[106]. It has been previously reported that bile acids with a 7α-hydroxyl or a 12α-hydroxyl group chelates iron and calcium ions and improve host absorption of these metals[107]. In addition, proteomic analysis showed that iron-sulfur clusters in *C. difficile* were disrupted when grown with CDCA, LCA, and DCA, but not CA[80]. Suggesting a potential relationship between iron uptake and bile acid stress in *C. difficile*.

CHAPTER 2

MATERIALS AND METHODS

Human fecal sample collection and selection

Patients were enrolled into multiple study cohorts at the University of Chicago Medical Center as previously described[108,109,110,111]. Stool samples from enrolled patients were collected by clinical nursing staff in inpatient wards and Intensive Care units. Samples were delivered to the clinical microbiology lab through the hospital pneumatic tube network and stored at 4°C until collected by the clinical research staff. Samples were collected twice daily and were aliquoted and stored at -80°C within 24 hours of sample production. Samples included for analysis were collected between June 2020 to June 2022 and were obtained under a protocol approved by the institutional review board at the University of Chicago (IRB20-1384, IRB21-0327, IRB20-0163, IRB20-11002). Inclusion criteria of patient fecal samples included: age > 18 with a sequencing depth of ≥ 4 million reads. To avoid sampling bias in our analyses, we selected, for each patient, the first sample with *C. difficile* detected by *C. difficile* 16S PCR or the first sample collected overall. All clinical data and patient demographic information were gathered from enrollment surveys administered by the clinical research staff and the University of Chicago Center for Research Informatics (CRI), a repository of all medical records.

Healthy donors were recruited from the University of Chicago community and screened for absence of antibiotic use and chronic illnesses. Participants were provided a sterile specimen collection container and collected samples were stored at 4°C until collected by research staff at the Duchossois Family Institute (DFI). Aliquoted samples were then stored at -80°C for further analysis. Reported demographic information (e.g. age, sex, race, BMI) was collected during study enrollment.

All human fecal samples underwent shotgun DNA sequencing, liquid chromatography-

mass spectrometry (LC-MS), and gas chromatography-mass spectrometry (GC-MS) for fecal metabolomics, cultured for *C. difficile*, and were tested for the presence and absolute abundance of *C. difficile* using PCR primers targeting *16S rRNA* genes and *tcdB*.

Clinical feature filtering

Cohort-specific metadata was stored on RedCap and clinical datasets were stored and maintained by the University of Chicago Center for Research Informatics (CRI). Statistical comparisons between clinical and sample-specific variables were initially compared for variance using the Kruskal-Wallis test. When significant, Fisher's exact tests with Benjamini-Hochberg multiple comparison correction identified significantly different groups.

Sample DNA extraction and library preparation

Isolated bacterial pellets or fecal samples were lysed and homogenized using a bead beater (BioSpec Product). DNA was extracted using the QIAmp PowerFecal Pro DNA Kit (Qiagen #51804). DNA was then purified using a spin column filter membrane and quantified using a Qubit 2.0 fluorometer. Libraries were created using 100-200 ng of genomic DNA using the QIAseq FX DNA Library Kit (Qiagen #180477). Briefly, DNA was fragmented enzymatically, and the desired insert size was achieved by adjusting fragmentation conditions. Fragmented DNA was end repaired using T4 DNA polymerase and "A's" were added to the 3' ends to stage inserts for ligation. During ligation, Illumina compatible Unique Dual Index (UDI) adapters were added to the inserts and the prepared libraries were amplified with PCR. Amplified libraries were purified, and quality control was performed with the Tapestation 4200 (Agilent Technologies).

Whole genome sequencing and genome assembly

Prepared libraries were sequenced using the Illumina MiSeq or NextSeq 500 platforms to generate 2x250bp or 2x150bp reads, respectively, targeted for 1 to 3 million pair-end reads per sample. Adapters were trimmed off raw reads and read quality was assessed using Trimmomatic (v.0.39). High-quality reads were assembled into contigs using SPAdes (v.3.14.0) and annotated using Prokka (v.1.14.6). *In silico* MLST[56] was determined for each draft genome using the PubMLST database (`https://pubmlst.org/organisms/clostridioides-difficile`).

Synthetic gut microbial community and in silico C. difficile challenge

150 isolates were selected from the DFI human gut commensal bacterial library. Clinically relevant *C. difficile* isolates were isolated from DFI patient stool samples or sourced from the CDC EIP *C. difficile* isolate bank (BioProject PRJNA577141)[112]. To simulate realistic genomic datasets, synthetic short-read sequencing reads were generated using readSimulator (v.0.0.1) to produce random 150 bp reads sampled randomly along the length of the genome[113]. To test the effect of read concentration on classification performance, *C. difficile* reads were randomly sampled for different read counts using seqtk (v.1.4)[114].

Metagenomic sequencing and taxonomic profiling assessment

Prepared libraries were sequenced using the Illumina NextSeq 500/NovaSeq 6000 platforms, producing an average of 7-8 million paired end reads of 150 bp per sample. Adapters were trimmed and human reads were removed using kneaddata (v.0.7.10)[115,116]. Taxonomy of synthetic communities or fecal metagenomic reads were classified MetaPhlAn4 (v.4.0.6) using the October 2022 release of the database. For synthetic metagenome samples, each sample was iterated over various stat_q parameters (0.0-0.2) at 0.05 step increments[63].

To enhance the sensitivity and specificity of *C. difficile* detection, we implemented a targeted tuning approach using the stat_q parameter in MetaPhlAn4. Specifically, we applied a local stat q adjustment of 0.05 exclusively to *C. difficile* while maintaining the default value (0.20) for all other taxa. This local tuning was benchmarked against a global stat_q adjustment where the value was uniformly set to 0.05 across all taxa. Performance comparisons included assessments of relative abundance accuracy, precision, recall, and overall microbial community preservation measured via Bray-Curtis dissimilarity. The local stat_q release of MetaPhlAn4 includes the '–config_stat_q' parameter that accepts a configuration file (e.g. metaphlan/stat_q.cfg) defining species from the MetaPhlAn database and its corresponding stat_q value per line. All SGBs corresponding to a listed species inherit the specified stat_q, while unlisted species receive the default (0.20).

To address relative abundance underestimation observed for clinically relevant but genetically heterogeneous ST-11 strains[117], we performed an iterative refinement of the MetaPhlAn4 marker gene database. Reads from the CDC EIP dataset, representing 30 C. difficile genomes covering 10 distinct sequence types (STs), were mapped against the 200 original MetaPhlAn4 *C. difficile* marker genes. Low-coverage markers were iteratively identified and removed to generate three progressively optimized databases (versions v1–v3). Relative abundance estimation accuracy was assessed across these databases at default (stat_q = 0.20) and optimized (stat_q = 0.05) thresholds.

Within sample taxonomic profiling accuracy was evaluated with purity (precision of the profile) or completeness (recall of the profile) as previously described[118,119]. The accuracy of species-level RA estimates per sample was calculated with L1 norm error and pairwise Bray Curtis dissimilarity distances. The accuracy of RA estimates per species in the synthetic metagenomes was measured with root mean squared error (RMSE), where N is the number

of mock samples, y_i is the true RA of species i, and y_i^* is the predicted RA of species i.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} ||y_i - y_i^*||^2}{N}}$$

Pathogenicity Locus diversity analysis

A set of 19,218 *C. difficile* genome genbank files was downloaded from the NCBI Pathogen database[120] on July 29, 2024. Protein annotations for toxin genes encoded in the PaLoc were extracted. Redundant sequences were identified and removed by clustering using CD-HIT (v4.8.1) at 100% clustering identity[121]. Non-redundant sequences for each gene were aligned using MAFFT (v7.526)[66,122] and used query each *C. difficile* isolate genome using HMMER (v3.4)[123]. Homologs were filtered to include those with E-value less than 1e-5 and a bias score lower than the overall bitscore. In addition, homologs were retained if their length is at least 75% of the length of the reference gene from *C. difficile* 630.

C. difficile 16S rRNA PCR and 16S rRNA qPCR

The presence of *C. difficile* in human fecal samples was detected by PCR using previously described *C. difficile*-specific primers for *16S rRNA* (Forward: 5'-TTGAGCGATTTACTT CGGTAAAGA-3', Reverse: 5'-CCATCCTGTACTGGCTCACCT-3') and *tcdB* (Forward: 5'-GAAAGTCCAAGTTTACGCTCAAT-3', Reverse: 5'-GCTGCACCTAAACTTACACC A-3')[61,124]. For each PCR reaction, 100 ng of DNA was amplified with Platinum Taq DNA polymerase (Invitrogen #10966083). The PCR reaction using the following parameters: Initial denaturation (94°C for 2 min), 30 cycles of denaturation (94°C for 30 sec), annealing (*16S rRNA* PCR: 58°C for 30 sec; *tcdB* PCR: 52°C for 30 sec), and extension (72°C for 1 min), and a final extension (72°C for 5 min).

PCR products were run on an Eppendorf Mastercycler and visualized on 2% SYBR safe agarose gels (Invitrogen #A42135) with a 100 bp ladder (Invitrogen #10488058). The

abundance of *C. difficile 16S rRNA* was measured by reverse transcription quantitative polymerase chain reaction (RT-qPCR) using these same primers with PowerTrack SYBR Green Master Mix (Applied Biosystems #A46012) on QuantStudio 6 Pro (ThermoFisher) under the following conditions: Initial Activation (95°C for 5 min), 40 cycles of denaturation (95°C for 15 sec) and Annealing/Extension (52°C for 60 sec), and a melt curve. To calculate the absolute abundance of *C. difficile 16S rRNA* genes, TOP10 chemically competent *E. coli* cells were transformed using the TOPO TA cloning kit for Sequencing (Invitrogen #K457502) with a vector (pCR 4-TOPO) containing a single copy of *C. difficile 16S rRNA* genes. A standard curve was created using known concentrations of linearized plasmid. *C. difficile*-specific curves were verified with melt curve analysis.

Clinical detection of C. difficile from human fecal samples

Frozen human fecal samples were thawed on ice before gently swabbed for 10 sec using the Cobas PCR Media Uni Swab Sample Kit (Roche #07958030190). Swabs in the storage buffer were vortexed for 10 sec before being stored at 4°C overnight. Samples were transported to the University of Chicago Clinical Microbiology Laboratory for processing using the Cobas *C. difficile* Test on the Cobas 4800 system.

Isolation of C. difficile from human fecal samples

Banked fecal samples that were positive for *C. difficile* 16S rRNA gene via PCR were thawed on ice and diluted 1:1 with ethanol under sterile conditions. The fecal slurries were vortexed to mix and incubated at room temperature for 2 hours to eliminate alcohol-sensitive species. Post-incubation, slurries were vortexed at 3600 rpm for 10 min and ethanol was removed from the pelleted fecal material. Samples were immediately transferred into the anaerobic chamber (Coylabs) and diluted with 1 mL sterile PBS. Diluted samples were diluted at 1:5, 1:10, or 1:20 into sterile PBS and plated on plates modified to allow *C. dif-*

ficile spore germination and *C. difficile* selection. Selective media (CCBHIS) was made with brain-heart infusion media (BHI) (BD Bacto #237500) supplemented with yeast extract (BD Bacto #212750), 1% w/v L-cysteine (Sigma-Aldrich #C7352), 16 mg/L cefoxitin sodium salt (Sigma-Aldrich #C4786), 0.25 g/L D-cycloserine (Sigma-Aldrich #C6880). Selective plates (CCBHISNRTA) were additionally supplemented 1% w/v taurocholate (Sigma-Aldrich #T4009), and 3 mg/L neutral red (ThermoFisher #J62643.18). Plated samples were incubated for 48 hours at 37°C under anaerobic conditions.

C. difficile colonies were selected in triplicate from plates based on typical colony morphology, yellow coloration, and fluorescence under UV light. Isolates were inoculated into fresh CCBHIS media in a sterile deep 96-well plate and incubated for 24 hours at 37°C before collection. The identity of *C. difficile* and the purity of the cultures were validated with *16S rRNA* amplicon sequencing. amplicon sequence variant (ASV) from the variable V4-V5 *16S rRNA* gene region were identified to the genus level using the Bayesian RDP classifier and identified to the species level using blastn against the RefSeq database. DNA libraries were prepared from a representative *C. difficile* colony per patient sample for whole genome sequencing.

Customized C. difficile database for strain-level profiling

Strain profiling of *C. difficile* was performed using StrainPhlAn4 with a custom database based on the genome catalog used for building MetaPhlAn4 databases[63]. We included genes that are 50% core among genomes of *C. difficile* (SGB6136) and are not found in genomes outside the Genus-level Genome Bin (GGB). The genus bin includes cryptic SGBs that are predominantly comprised of genomes derived from soil and water sources but share many genes with *C. difficile*, ultimately limiting the overall number of unique marker genes for *C. difficile* in the MetaPhlAn4 database. By allowing the genes to be shared within the genus, we were able to use 1,673 marker genes to enhance *C. difficile* strain resolu-

tion within the context of a gut microbiota, where the presence of cryptic, environmental SGBs is unlikely. The database has been deposited among the standard MetaPhlAn4 databases and can be implemented by specifying the index parameter in MetaPhlAn4 as '-x mpa_vFeb24_CDIFF_CHOCOPhlAnSGB_20240910'.

StrainPhlAn4 Parameter Tuning

The synthetic community samples challenged with *C. difficile* reads were prepared as described above. To ensure the quality of reads included in each iteration, the 'min_base_coverage' was set at 1 and the 'min_base_quality' at 30. Additionally, only marker genes present in at least 1 sample ('$marker_in_n_samples$' = 0), and samples with at least 20 marker genes were retained ('$sample_with_n_markers$' = 20). The samples were passed into the StrainPhlAn4 pipeline using the custom *C. difficile* database and were iterated over a range of 'breadth_threshold' values (50-80%) and 'min_reads_aligning' values (2-8).

Metabolomic data preparation

SCFAs, amino acids, and phenolic metabolites are derivatized with pentafluorobenzylbromide (PFBBr). 9 compounds (SCFA: acetate, butyrate, succinate, and propionate, Amino acids: proline, glycine, and cysteine) were quantified (mM) by negative ion collision induced-GC-MS. BAs (conjugated PBAs: glycocholic acid (GCA), glycochenodeoxycholic acid (GCDCA),TCA, and taurochenodeoxycholic acid (TCDCA), PBAs: CA, CDCA, SBAs: deoxycholic acid (DCA), lithocholic acid (LCA), ursocholic acid (UCA), ursodeoxycholic acid (UDCA), isodeoxycholic acid (isoDCA), alloisolithocholic acid (alloisoLCA), 3-oxolithocholic acid (3-oxoLCA), allocholic acid (alloCA)) were quantified (ug/mL) by negative mode liquid chromatography-electrospray ionization-quadrupole time-of-flight-MS ([-]LC-ESI-QTOF-MS, Agilent 6546). A panel of 85 compounds (50 PFBBr, 35 bile acids) was relatively quantified using normalized peak area relative to internal standards.

Single-omic analysis of human stool samples

Microbial community profiles were estimated using MetaPhlan4 with local stat_q of 0.05. Metabolomic datasets Metagenomic species-level and metabolomic features were filtered to exclude samples that were not defined at a species-level, prevalence below 0.1%, variance below 1×10^{-4}, and relative standard error (RSD) below 30%. Microbiome features were associated with clinical and sample categorical variables were detected using Maaslin2 (v.1.20.0)[125]. The mixed effects model associated TSS-normalized and log-transformed relative abundances or log-transformed metabolite values as a function of *C. difficile* presence, *C. difficile tcdB* subgroup, consistency of the sample, and exposure to any antibiotics within 8 weeks prior to sample collection and controlling for the group variability of the patient study cohort (i.e. LT, LD, HT, MICU, HD) as random effects with a minimum prevalence of 0 and a minimum abundance of 0.01%.

$$RA = (C.diff_{group}) + (toxin_{group}) + (stool_{consistency}) + (ABX_{exposure}) + (1|cohort)$$

Model fitting was assessed per feature by selecting for low Variance Inflation Factors (VIF) ($VIF < 5$) and minimized differences in Akaike Information Criterion (AIC) ($\Delta AIC < 10$). Relative abundances were log-transformed and normalized using Total Sum Scaling (TSS). Multiple comparisons were adjusted for using the Benjamini-Hochberg procedure and significant features ($p < 0.05$) were compared across patient groups.

Multi-omic analysis of human stool samples

A pseudocount of $\frac{min(RA>0)}{2}$ was applied across the metagenomic dataset as previously described[126]. Due to the compositional nature of MetaPhlAn4 profiling results, the relative abundance per species-level annotation was normalized with Centered Log Ratio (CLR), which divides each feature by the geometric mean of all compositions[127]. A pseudocount of

1×10^{-6} was applied across all metabolite features which were then log-transformed. The similarity of the taxonomic and metabolomic datasets was compared using Coinertia analysis (CIA) and Pareto analysis (PA). Highly-collinear ($Pearson > 0.9$) features were removed.

Correlations between the omic datasets and CDI outcomes were detected using multi-block sparse Partial Least Squares Discriminant Analysis (sPLSDA) from the MixOmics package (v.6.32.0)[128]. For sPLSDA integration, CLR-transformed relative abundance data and log-transformed metabolite data were scaled. The sPLSDA model was trained on 70% of a dataset that included the 374 *C. difficile 16S rRNA* negative samples and the 110 *C. difficile 16S rRNA* positive samples. Model performance was evaluated using evaluated using repeated cross-validation (e.g., 5-fold cross-validation repeated 10 times with random sampling), using the mixOmics::perf() function and with cumulative Area Under the Receiver Operating Curve (AUROC) using the mixOmics::auroc() function.

16S rRNA metagenomic sequencing

DNA was extracted from fecal samples using the QIAamp PowerFecal Pro DNA Kit and the variable $V_4 - V_5$ region of the 16S rRNA was amplified using polymerase chain reaction (PCR). Illumina libraries were generated using the Qiagen QIASeq 1-step amplicon kit and run on the Illumina MiSeq platform for 5,000-10,000 reads per sample. Raw reads are de-multiplexed and processed using dada2 into ASV. ASVs are classified to the genus level using the Bayesian Ribosomal Database Project (RDP) classifier and are mapped against RefSeq for species-level identification using blastn.

Bacterial Growth and Culture Conditions

All bacterial isolates (except *P. capsillosus*) were streaked on Columbia Blood Agar (CBA) for 48 hr under anaerobic conditions. All bacterial isolates (except *D. freteri*) were cultured in a rich media made with BHIs media (BD Bacto #237500) supplemented with

yeast extract (BD Bacto #212750), 1% w/v L-cysteine (Sigma-Aldrich #C7352). cell-free supernatant (CFS) were prepared from 24-48 hr cultures of *Lachnospiracaea sp.* sterilized through a 0.2 micron membrane filter. Rich media controls are sourced from the same batch and preparations as CFS starter culture media for each experiment.

Selective media (CCBHIS) was made with BHIS supplemented with 16 mg/L cefoxitin sodium salt (Sigma-Aldrich #C4786), 0.25 g/L D-cycloserine (Sigma-Aldrich #C6880). *C. difficile* strains are streaked onto CCBHIS plates supplemented with 0.1% taurocholate. *C. difficile* isolates were cultured in fresh BHIS in triplicate and incubated at 37°C overnight. *C. difficile* cultures were then diluted 1:10 and incubated at 37°C for 2-2.5 hours until an optical density (OD) of 0.4-0.5 at 600 nm (OD600). Then *C. difficile* was inoculated into experimental media at a starting OD of 0.01 and incubated for 24-48 hr at 37°C.

C. difficile spore preparation

C. difficile spores were prepared as described previously[129,130]. *C. difficile* isolates were plated on 70:30 sporulation plates and incubated under anaerobic conditions for 4-5 days. Spores were harvested from the plates into sterile H_2O and isolated using a 50% w/v sucrose gradient. Purified spores were then heated at 65°C for 20 minutes to remove vegetative cells. Spores were enumerated by calculating the difference in colony-forming units (cfu) on CCBHIS and CCHISTA plates. Spore stocks are stored at 4°C for long-term storage.

Nutrient Supplementation

The final concentration of proline (2 mM) and glycine (3 mM) was quantified from standard preparations of BHIS media using GC-MS. The final concentration of glucose (11 mM) was derived from previously reported defined media conditions for *C. difficile*[101]. Nutrients were dissolved in H_2O at 35-100x concentrations and sterile-filtered using a 0.2 micron pore membrane. Sterile nutrient stocks were reduced under anaerobic conditions overnight before

diluting to the final concentration in CFS media conditions.

Clostridial CFS relative inhibition (RI)

A 24 hr CFS preparation of *Clostridial spp.* was diluted 4:1 with fresh BHIs (80% CFS). BHIs from the same lot and fresh media preparation was used to mitigate batch effects. *C. difficile* strains were isolated in triplicate and cultured in the 80% CFS or the undiluted 100% CFS. RI was calculated as follows:

$$1 - (\frac{OD_{inhibited}}{OD_{unhibited}}) = 1 - (\frac{OD_{CFS}}{OD_{BHIS}})$$

Minimum Inhibitory Concentration Microbroth Assay

Bile acids were dissolved in DMSO at 100x the highest tested concentration and serially diluted by a factor of 2. Each bile acid solution was diluted 1:100 into 1 mL of fresh BHIS to the final concentration. *C. difficile* was inoculated from a mid-log culture to a starting OD600 of 0.01. The minimum inhibitory concentration was defined as the lowest concentration with vegetative growth inhibition after 24 hr incubation at 37°C under anaerobic conditions.

Animal Husbandry and Housing

Male and female C57BL/6 (B6) wild-type mice were bred and housed in gnotobiotic isolators within the University of Chicago Gnotobiotic Core Facility. Mice were age and sex-matched for each treatment group and transferred to BSL2+ housing conditions for each experiment. Mice are then single-housed. All mouse experiments were performed in compliance with institutional guidelines and were approved by the Institutional Animal Care and Use Committee (IACUC). SPF WT B6 mice were purchased from Jackson Laboratories.

Mouse Model of C. difficile Infection

Mice were infected with 200 spores of *C. difficile* VPI10463 and the progression of *C. difficile* disease was observed for 7 days post-infection. Disease severity was scored using 3 parameters: weight loss, diarrhea severity, and morbidity. Mouse fecal pellets were collected on days -7, 0, and 1 for *16S rRNA*-sequencing and metabolomic analysis. *C. difficile* was enumerated from fecal pellets 24 hours post-infection on CCBHISTA plates.

Colonization of gnotobiotic mice

Bacterial gavage stocks were prepared from mid-log bacterial cultures grown in BHIS or collected from lawns grown on CBA or BHIS plates in sterile PBS. These preparations were aliquoted 1:1 into a sterile preparation of PBS + 1% w/v L-cysteine for storage at -80°C. Stocks were combined in equal parts for a total of 10^8 cfu per mouse per dose. For monocolonization studies, mice were gavaged with 1 dose of commensal bacteria and allowed to colonize for 7 days. For multi-species community studies, mice were gavaged with a dose of commensal bacteria 7 days and 5 days prior to *C. difficile* infection.

Germ free cecal collection and filtration

Cecal contents from germ-free wild type B6 mice were harvested and resuspended in sterile water (200 mg cecal content mL^{-1} water). The cecal slurry was vortexed and centrifuged at 4,300 x g using an Allegra X-14R Centrifuge (Beckman Coulter) for 10 min to pellet solid waste material. The supernatant was collected and centrifuged 3 more times before filtration through a 0.2 micron filter. The filter-sterilized cecal supernatant media was aliquoted and frozen at -20 °C.

Amino acid supplementation

Fresh amino acid stocks (histidine, glycine, arginine, phenylalanine, methionine, threonine, alanine, lysine, serine, valine, isoleucine, leucine, proline, aspartate, glutamate, tyrosine, cysteine) were prepared by suspending 10 mg of each compound in 10 mL of 5N KOH and freezing 1 mL aliquots at -20 °eC. Amino acids were added to thawed germ-free cecal supernatant media at concentrations previously described to be sufficient for C. difficile growth[101]. Otherwise, an equal volume of sterile water was added to the cecal media. Then, each condition was pH adjusted to 8.4-8.5, filtered through a 0.2 micron filter, and reduced in an anaerobic chamber overnight.

Statistical Analysis

Statistical tests were performed using GraphPad Prism (v.10.2.2) or R (v4.4.2) in the RStudio (2024.12.1+563) integrated development environment. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

Data Availability

The datasets presented in this study can be found online in online repositories. Metagenomic information is publicly available from all study cohorts on NCBI under BioProjectID PRJNA838648 (healthy donor cohort), BioProjectID PRJNA912122 (liver disease patient cohort), BioProjectID PRJNA1039781 (liver transplant patient cohort), BioProjectID PRJNA845905 (heart transplant patient cohort), and BioProjectID PRJNA842425 (MICU patient cohort). The *C. difficile* whole genome sequences isolated in this study are available under BioProjectID PRJNA1251023. The MetaPhlAn4 (v4.1.1) implementation of local stat_q is available on GitHub under the 'local_stat_q' branch (`https://github.com/biobakery/MetaPhlAn/tree/local_stat_q`).

CHAPTER 3

CULTURE-INDEPENDENT STRAIN TYPING OF *CLOSTRIDIOIDES DIFFICILE* IN CLINICAL FECAL SAMPLES

3.1 Introduction

C. difficile is a spore-forming, anaerobic gram-positive bacillus and a leading cause of nosocomial diarrheal illness. *C. difficile* spores are acquired from the environment and, upon germination, produce exotoxins (TcdB and TcdA) in the gut lumen. A recent meta-analysis estimates the incidence of CDI at 8.3 cases per 10,000 patient days, resulting in as much as $5 billion in excess healthcare costs[131,132]. The epidemiology of *C. difficile* infection is complex, with some infections occurring in clusters that result from patient-to-patient transmission and other infections occurring in patients asymptomatically colonized with *C. difficile*. Strain-specific detection of *C. difficile* is necessary to distinguish focal outbreaks, that require enhancement of infection control practices, from sporadic occurrences of CDI.

The clinical course for CDI is highly variable, ranging from asymptomatic carriage to mild diarrheal illness to life-threatening colitis[1]. Antibiotic-induced gut dysbiosis predisposes patients to symptomatic *C. difficile* associated disease (*C. difficile*) and microbiome profiling may identify patients at risk for CDI[2,3,4,133]. Early next generation sequencing (NGS) studies of the fecal microbiota relied on the amplification of bacterial 16S rRNA genes to determine taxonomic compositions but were limited by potential amplification-based bias and the lack of genetic information beyond the targeted 16S sequence. Microbiome analyses have increasingly shifted to shotgun metagenomic sequencing of fecal DNA which provides a more comprehensive picture of metabolic pathways, toxins, antibiotic resistance genes, and strain level microbial compositions[134].

Taxonomic classification of metagenomic datasets is a critical step in microbiome analysis and multiple tools have been developed that use a range of strategies that include brute-

force sequence mapping approaches as well as strategies relying on k-mer matching or marker gene-based profiling[135]. These tools, however, have been optimized for broad applications across diverse microbial ecosystems and have been only occasionally exploited in clinical settings to identify intestinal pathogens.Clinical diagnosis of CDI is challenging and requires clinical insight combined with a choice of laboratory tests, that while useful, are recognized as imperfect. Diagnostic approaches for CDI are broadly categorized as "organism detection" (e.g. anaerobic culture or glutamate dehydrogenase enzyme immunoassay (EIA)) or "toxin detection" tools (e.g. toxin EIA, toxin nucleic acid amplification test (NAAT), and cytotoxicity assays (CTA)). The "gold standard" for *C. difficile* diagnosis combines anaerobic culture from fecal samples followed by toxin detection to identify toxigenic isolates. However, this approach is uncommon in hospital microbiology laboratories because it is time, labor, and resource intensive[29].metagenomic next generation sequencing (mNGS) has been used to detect pathogens in sterile body fluids, such as cerebrospinal fluid, synovial fluid, and peripheral blood[136,137,138,139]. The use of metagenomic sequencing for pathogen detection in fecal samples is more challenging because of the complexity of the fecal microbiota, the genomic overlap between pathogens and commensals, and the low relative abundance of many pathogens[140,141,142].

Early outbreak detection of *C. difficile* is critical for infection control due to the high prevalence in long-term care settings and the production of robust spores. Conventional surveillance recommendations combine laboratory diagnostics, such as TcdB NAATs, with retrospective medical chart review[143,144]. However, because these molecular diagnostic tests do not differentiate between bacterial strains, lack epidemiological utility[1]. Molecular strain typing has evolved from band-based approaches (i.e. REA,PFGE, polymerase chain reaction (PCR) ribotyping, and multilocus variable number tandem repeat analysis (MLVA)) to sequence-based approaches (i.e. MLST and WGS)[145]. But, the development of metagenomic strategies for epidemiology remains in its infancy.

Metagenomic sequencing of fecal samples has the potential to add a culture-free approach for real-time patient gut profiling and strain-level identification of enteric pathogens. To advance this goal, we tuned the MetaPhlAn4 and StrainPhlAn metagenomic profilers to enhance *C. difficile* species- and strain-level sensitivity and demonstrate that metagenomic sequencing of fecal samples from patients colonized with *C. difficile* enables accurate and reliable culture-independent strain identification.

3.2 Results

Benchmarking microbial profiling and C. difficile detection performance in silico

MetaPhlAn4 is a MetaPhlAn is a well-documented, fast, and accurate taxonomic profiling tool[119,146,147,148] that has been used in investigations of *C. difficile* and the microbiome[149,150,151,152,153,154,126,155,156]. To assess the profiling accuracy and *C. difficile* detection of MetaPhlAn4, we constructed an *in silico* synthetic metagenome from 150 distinct human gut-derived microbial genomes derived from the DFI Symbiotic Bacterial Strain Bank[157,158] (fig. 3.1a, fig. S1a). NGS short reads were simulated, and 50,000 reads (0.67% RA) were randomly sampled from each genome resulting in a mock metagenomic sample of 7.5 million reads, comparable to the average read depth of fecal metagenomic samples in this study (table S1). Consistent with previous reports[147,119], MetaPhlAn4 demonstrated high overall classification purity and completeness from our mock metagenomic samples(figs. S2c and S2d). Out of 150 genomes included in our synthetic communities, 114 species were consistently detected (76%). The phylum *Bacillota* had the highest number of false negatives (FNs) (FN = 30 species (20%)) and the remainder were spread across the phyla *Actinomycetota* (FN = 2 species (1.3%)), *Bacteroidota* (FN = 2 species (1.3%)), and *Pseudomonadota* (FN = 2 species (1.3%)). Of our detected species, *M. torques* had the highest divergence

in predicted relative abundance (RA) (mean RA = 2.27%) (fig. 3.1a). Species more closely related to *C. difficile* in the genus *Lachnoclostridium* were consistently detected, including *C. scindens* (RA = 49.2%) and *C. sporogenes* (RA = 43.0%). We observed 25 false positives (FPs) including several ambiguous assignments, including *Clostridiaceae bacterium* with the highest estimated RA of 1.2% (fig. S2b).

While mNGS platforms have been used extensively to taxonomically classify complex microbial communities, their limited sensitivity and accuracy have limited their utility for targeted intestinal pathogen detection[159,160]. To assess *C. difficile* detection performance of MetaPhlAn4, we spiked the *in silico* synthetic metagenome of 150 commensal microbes with whole genome DNA sequences derived from 30 representative *C. difficile* isolates characterized by the CDC Emerging Infections Program (EIP) that represent the 10 most prevalent STs in the United States[112](fig. S1b). Given that *C. difficile* can be present in widely varying abundances in fecal samples, the number of short reads from each of the 30 *C. difficile* genomes introduced into the synthetic metagenome ranged from 1,000 (0.013% RA) to 50,000 reads (0.67% RA).

C. difficile was not detected by MetaPhlAn4 in the unspiked synthetic metagenome or in those spiked with 1,000 sequences and was over-reported in samples spiked with 5,000, 10,000, or 50,000 sequences. The predicted RAs of sequence spike-ins from ST-11 genomes (clade 5) were consistently reduced compared to spike-ins from strains belonging to clades 1 and 2, presumably resulting from genetic dissimilarities between these clades (ANI = 96%) (fig. S3)[117,161].

Optimization of C. difficile detection by MetaPhlAn4

Given MetaPhlAn4's overall strong classification performance with the synthetic metagenome, we focused on optimizing its sensitivity and accuracy for *C. difficile* detection. Our analysis identified three key limitations (1) failure to detect *C. difficile* at low read depths, (2) in-

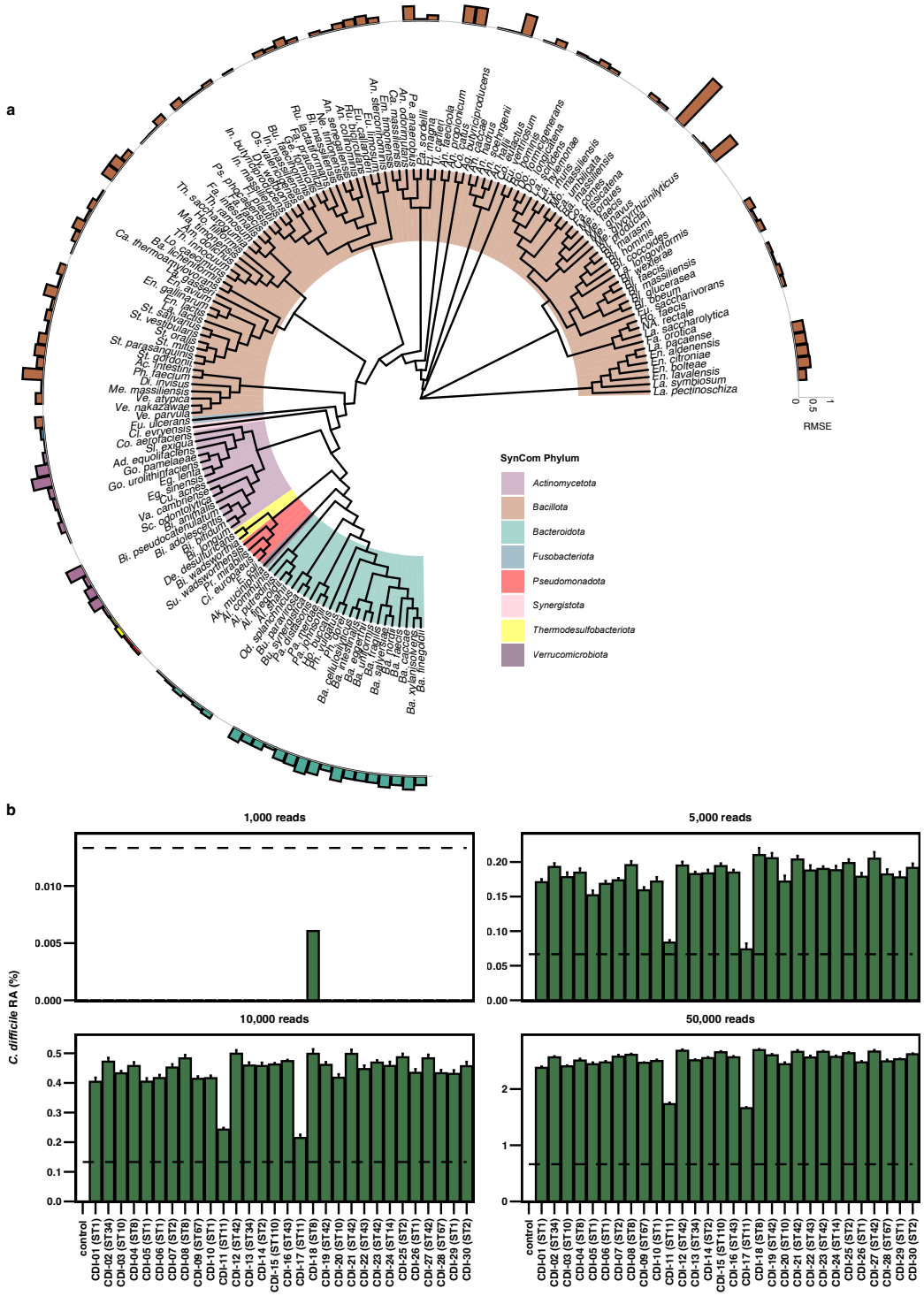

Figure 3.1: Benchmarking microbial profiling and *C. difficile* detection performance of MetaPhlAn4 *in silico*

Figure 3.1: Benchmarking microbial profiling and *C. difficile* detection performance of MetaPhlAn4 *in silico* (cont.)

(a) Neighbor joining phylogenetic tree featuring 150 human gut-derived commensal isolates constructed from genomic ANI distances. Root mean squared error of relative abundance (RA) estimates aligned for each species.

(b) Reported mean relative abundance of *C. difficile* by each classifier for representative *C. difficile* isolates at varying levels of spike in *C. difficile* reads (1,000 to 50,000). Error bars report the standard error from the mean. Dotted line represents the target relative abundance.

Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

consistencies in RA estimation for ST-11 strains, and (3) overestimation of *C. difficile* RA. Since RA accuracy depends on the total number of species detected in a sample, it does not directly map to the read depth. For this study, we consider an accuracy range within an order of magnitude acceptable, given zero false positives and consistent detection across strains.

MetaPhlAn4 estimates the clade coverage by calculating the robust average of marker gene coverage within each clade[162]. To enhance sensitivity, we adjusted the 'stat_q' parameter, which defines the quantiles included in the clade coverage calculation[62]. To optimize the stat_q threshold, we determined *C. difficile* detection in a range of parameter values between 0.20 (default) and 0.0. A threshold of 0.05 accurately detected all *C. difficile* strains in our synthetic metagenome at the spike in of 1,000 reads (fig. 3.2a, fig. S2a). Importantly, the increased detection of *C. difficile* did not compromise precision until stat_q was reduced to 0.0 and recall remained unchanged across all tested values (figs. S2c and S2d). These findings suggest that adjusting the global stat_q to 0.05 enhances the detection of *C. difficile* without substantially increasing the number of false positives. However, given that the changes to sensitivity impact all taxa, we cannot predict the broader impact on community profiling.

To address this concern, we implemented a local stat_q adjustment, where the optimized value (0.05) was applied only to *C. difficile*, while maintained at default (0.20) for all other

taxa. When comparing global and local stat_q adjustment at 0.05, we observed consistent improvement in *C. difficile* detection across all tested strains (fig. 3.2b). However, global tuning reported RA estimates closer to the expected values when compared to local tuning. This discrepancy is likely due to the increased false positives associated with global stat_q adjustment, which are not present with local adjustments. While the precision for global adjustment is significantly lower than those for either default or local adjustments, there are no significant differences in recall (figs. S2c and S2d). These results indicate that local stat_q adjustment enhances the *C. difficile* detection performance without introducing additional false positives as seen in the global adjustment. We further assessed if the local stat_q adjustment preserves the overall microbiome profile by comparing the pairwise Bray-Curtis distance between adjusted and default microbiome profiles. Global adjustment results in significantly higher Bray-Curtis dissimilarity scores than local adjustment, suggesting that local stat_q adjustment effectively maintains the overall microbiome profile (fig. 3.2c). Similarly, the species rank L1 norm error, the total error between true and predicted abundances of the species, of the default profiles were not significantly different from the local stat_q profiles (fig. S2e). Importantly, our targeted local stat_q tuning approach demonstrated substantial improvement in detection performance specifically for *C. difficile* without negatively affecting the accuracy of the overall microbiome profile, underscoring the potential to broadly adapt this strategy to enhance detection sensitivity for other clinically significant pathogens or microbial species of interest.

We observed consistent RA estimates for ST-11 strains (CDI-11 and CDI-17)(fig. 3.1b, fig. 3.2a, fig. S2a). The underestimated RAs are likely due to the extensive genomic diversity of *C. difficile*[31,54], particularly in the clinically important ST-11 strains[117], and the selection of marker genes in the MetaPhlAn4 database. To address the persistent challenges in accurate strain-level abundance estimation, we developed a novel iterative refinement strategy to remove unevenly distributed marker genes from the MetaPhlAn4 database, resulting in three

increasingly refined alternative databases (v1-v3) (table S2). Using the most refined database (v3), the mean RA of ST-11 strains converged towards the mean RA observed in other STs (fig. 3.2d). By establishing a framework to refine marker gene databases tailored explicitly to taxa with significant genomic heterogeneity, our method substantially enhances taxonomic classifier precision. Though initially validated with the 30 CDC epidemic *C. difficile* isolates (fig. S3), this strategy offers a scalable and simple method to enhance metagenomic classification accuracy across other clinically relevant species exhibiting similar genomic complexity.

Optimized MetaPhlAn4 performance on patient samples

To determine the ability of optimized MetaPhlAn4 to detect *C. difficile* in patient-derived fecal samples, we analyzed a subset of 484 patient-derived fecal samples collected between June 2020 to June 2022 from patients admitted to the University of Chicago Medical Center for heart transplant (HT), liver transplant (LT), liver disease (LD), or into the Medical Intensive Care Unit (MICU) at the University of Chicago Medical Center[108,110,111,109]. And 21 samples from Healthy donors (HDs), collected between August 2020 and August 2022, were included as a control. The dataset includes clinical information and shotgun metagenomic sequences from a fecal sample collected per patient (fig. S4a). Each sample was screened by PCR with primers specific for the *C. difficile* 16S rRNA gene[124], resulting in 374 negative samples and 110 positive samples. *C. difficile* abundance was determined by 16S rRNA gene quantitative PCR (qPCR) and fecal samples were cultured anaerobically on *C. difficile* selective media, yielding 73 isolates that were whole genome sequenced (fig. 3.3a, fig. S4b). The strain identity of these isolates was determined by pubMLST and average nucleotide identity (ANI) by FastANI (fig. 3.3b). In our isolate collection, the most frequent ST was ST1 (clade 2) (n=8, 11.0%), however, clade 1 was the most frequent clade (n=54, 74.0%) in our collection (fig. S4c).

The fecal samples from which the 73 isolates were cultured were all PCR positive for the

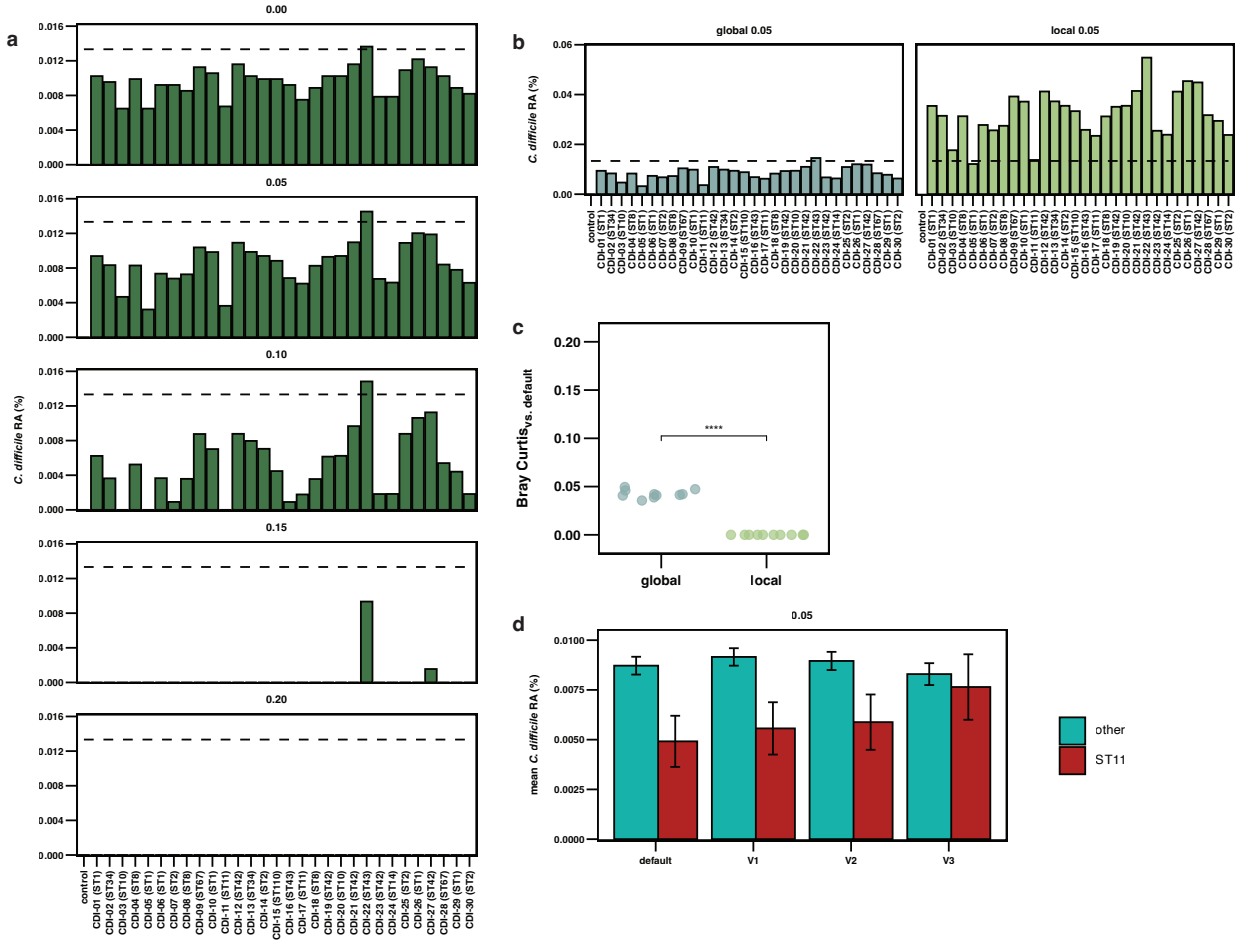

Figure 3.2: Optimization of *C. difficile* detection by MetaPhlAn4

(a) Reported RA of *C. difficile* for representative isolates of *C. difficile* (n = 30) across various stat_q parameters (0.0-0.20). Black dashed line represents the target RA for 1,000 *C. difficile* reads (0.0133%).

(b) Reported RA of *C. difficile* for representative isolates of *C. difficile* (n = 30) between global stat_q adjustment (0.05) and local stat_q adjustment (0.05). Black dashed line represents the target RA for 1,000 *C. difficile* reads (0.0133%).

(c) Pairwise Bray-Curtis Dissimilarity of the overall between stat_q adjustment (0.05) and default (0.20).

(d) Mean RA of ST-11 strains (pink) compared to the mean RA of the remaining 28 strains (white) for 1,000-50,000 reads.

Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

C. difficile 16S rRNA gene (fig. S5a). However, only 44 fecal samples tested positive for the *tcdB* gene by the Cobas *C. difficile* test (60.3%). Of the 29 fecal samples that tested negative for *tcdB*, 25 yielded *C. difficile* isolates (86.2%) that, upon genome sequencing, lacked the *tcdB* gene and all or the majority of PaLoc encoded genes.

MetaPhlAn4 detected *C. difficile* in 48 out of 73 culture-positive fecal metagenomes (65.8%) at a stat_q of 0.20 (median *C. difficile* RA). The local local stat_q of 0.05 detected *C. difficile* in 54 metagenomes (74.0%) (median *C. difficile* RA = 0.143%), confirming the greater sensitivity we demonstrated with the spiked synthetic communities (fig. 3.3c). Of the 6 additional *C. difficile* positive samples detected at a local stat_q of 0.05, 4 yielded strains that encoded a complete PaLoc while 2 lacked the PaLoc. 8 samples that were detected by Cobas were not detected by MetaPhlAn4, and yielded *C. difficile* isolates that encoded the complete PaLoc. 18 samples that tested positive by MetaPhlAn4 but negative by Cobas all yielded isolates that lacked a complete PaLoc. The 25 PaLoc deficient *C. difficile* isolates from the 73 patients were restricted to ST-3, ST-15, ST-26, ST-83, ST-100 and ST-205 of clade 1 and ST-37 and ST-39 of clade 4. Each of these isolates was cultured from patients who were hospitalized with different diagnoses and procedures (e.g. the 6 ST-3 isolates were isolated from patients undergoing heart transplantation or hospitalized for liver disease or for critical care) and thus are unlikely to represent in-hospital transmission. However, in our isolate collection, none of the PaLoc deficient isolates shared the same ST as PaLoc encoding isolates.

Additionally, we observe the same preservation of the microbiome profile while using local stat_q adjustment compared to global adjustment (fig. 3.3d). However, samples harboring *C. difficile* with lower copy numbers of *C. difficile* 16S rRNA genes ($< 10^5$ copies per mg sample) were undetected by MetaPhlAn4 (fig. 3.3e). In contrast, there was no significant difference in *C. difficile* 16S rRNA gene copy number between detection results by the Cobas test for *tcdB*. Our findings suggest that local optimization of MetaPhlAn4 provides modest

improvements in *C. difficile* detection without compromising overall profiling performance, though it remains limited by the pathogen density in the sample.

Optimization of C. difficile strain-level detection by StrainPhlAn4

The clinical importance of strain variation within the human gut microbiome is well appreciated –notably impacting the community composition, metabolic profile, and pathogenicity[163]. Recent studies have adopted WGS to track the transmission of *C. difficile* through different healthcare settings; however, these studies require isolation with anaerobic culture[164,165,59,166,58,167]. In contrast, metagenotyping potentially offers a culture-free alternative to identify strain transmission within communities and care settings[168].

Several methods have been developed for strain-level inferences from metagenomic data, but we focused on StrainPhlAn4[169] due to its compatibility with MetaPhlAn4. Two principal factors affect detection: the composition of the database and the sensitivity of the algorithm. To enhance the phylogenetic inference of *C. difficile* in clinical settings, we generated a custom database only consisting of *C. difficile* marker genes. This database consists of 1,673 genes that are present in at least 50% of *C. difficile* genomes and do not overlap with genomes outside of the genus.

To optimize sensitivity, we systematically evaluated the performance of *C. difficile* detection across varying parameter settings (fig. S6b). We fine-tuned the parameters for the minimum fraction of a marker gene coverage required for inclusion ('breadth_threshold') and the minimum number of aligned reads required for reliable marker calling ('min_reads_aligning'). Most strain profiling metagenomic algorithms, including StrainPhlAn4, struggle with low coverages typical of clinically relevant pathogens such as *C. difficile*[170]. To determine the impact of coverage limitations on the accuracy of strain determination, we randomly sampled various read concentrations from 30,000 (1x coverage; median RA 0.25%; IQR 0.18-0.27%) to 300,000 (10x coverage; median RA 2.48%; IQR 1.77-2.61%) sampled from the 30 CDC

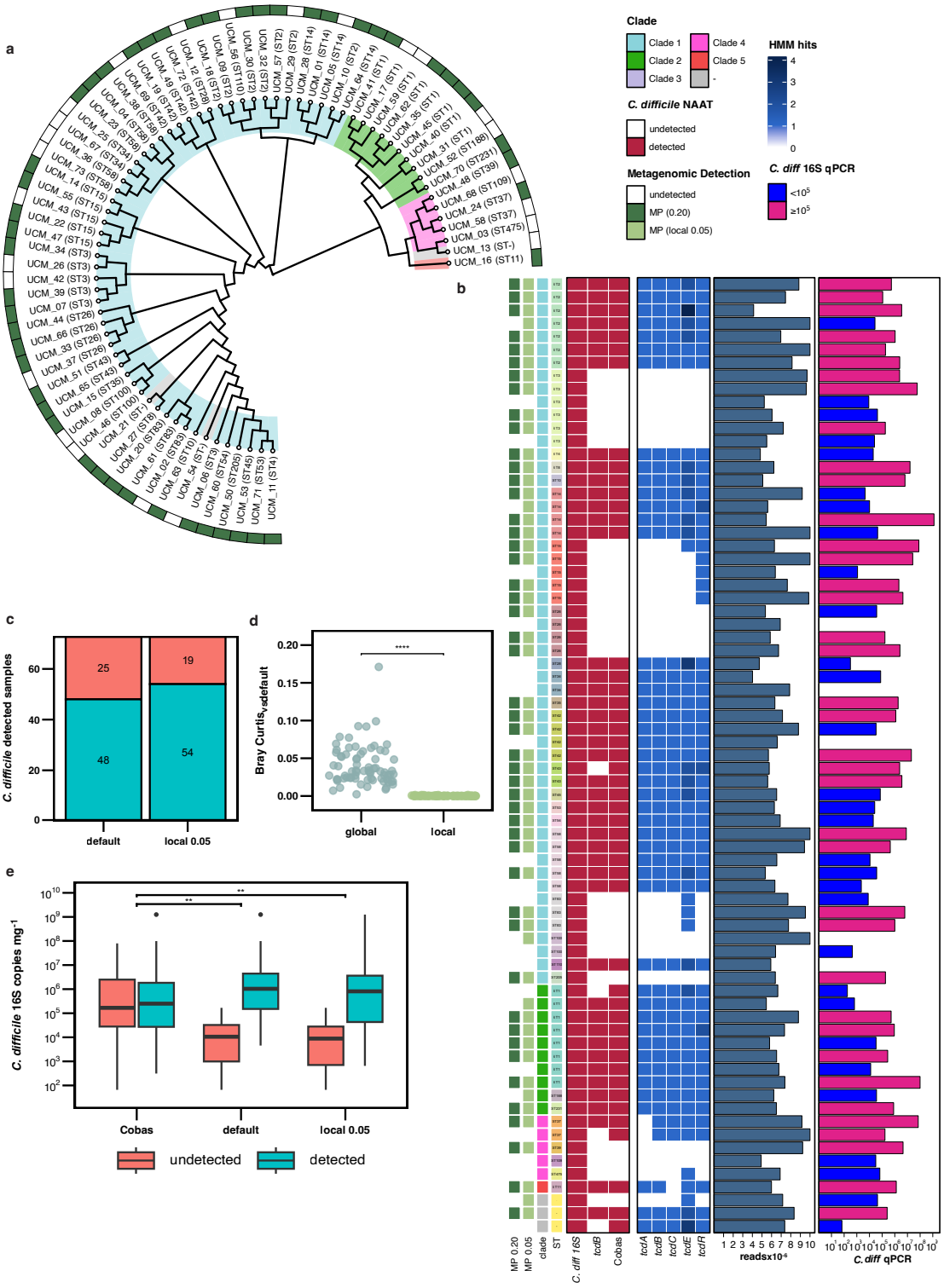

Figure 3.3: Optimized MetaPhlAn4 performance on patient samples

Figure 3.3: Optimized MetaPhlAn4 performance on patient samples (cont.)

(a) Neighbor joining phylogenetic tree of *C. difficile* isolates derived from patient fecal samples (n=73). The tree was constructed from pairwise genomic ANI distances. The detection of *C. difficile* from the fecal metagenomes by MetaPhlAn4 at stat_q of 0.20 decorates the tree and the sequence typing of each isolate was determined using MLST.

(b) Detection of *C. difficile* by PCR tests for *C. difficile* 16S rRNA gene, *tcdB*, and by the Cobas *C. difficile* Test used for patient diagnosis (n=73). Diversity of each isolate is characterized using MLST. The detection of *C. difficile* using MetaPhlAn4 is compared between default (0.20) and local stat_q adjustment (0.05). The read counts for each metagenomic sample are binned every 4 million reads and capped at 10^7 reads. The quantified number of *C. difficile* 16S copies per mg of fecal sample are binned into high ($\geq 10^5$) or low ($< 10^5$).

(c) Detection of *C. difficile* from the metagenomes of the patient-derived fecal samples between between local stat_q adjustment (0.05) and default (0.20).

(d) Pairwise Bray-Curtis Dissimilarity of the overall between local stat_q adjustment (0.05) and default (0.20).

(e) Copies of *C. difficile* 16S rRNA gene from qPCR compared to the detection by the Cobas *C. difficile* assay and MetaPhlAn4 for *C. difficile* culture positive metagenomes.

Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

isolates. These reads were added to sequences obtained from a healthy human microbiome *in silico* and processed through MetaPhlAn4 and StrainPhlAn4. Using default parameters, StrainPhlAn4 was unable to detect *C. difficile* at the lowest relative abundances (median 0.25-0.51%) for any isolate and did not detect ST11 isolates under any conditions (fig. S6a). Our results demonstrated that breadth threshold settings between 60-70 and minimum reads aligning settings of 4 improved the detection of all *C. difficile* strains (fig. S6b).

C. difficile detection in simulated clinical samples using optimized StrainPhlAn4

We then assessed the detection of *C. difficile* and the accuracy of post-optimized intraspecific inferences by spiking 30,000 *C. difficile* reads into five distinct patient fecal metagenomic backgrounds that were negative for *C. difficile* (*in silico* samples) (fig. S7a, fig. S7b). The phylogenetic output from StrainPhlAn4 also included the reference tree built from a curated collection of 260 *C. difficile* genomes and the 30 CDC isolates[31,112]. The pairwise phy-

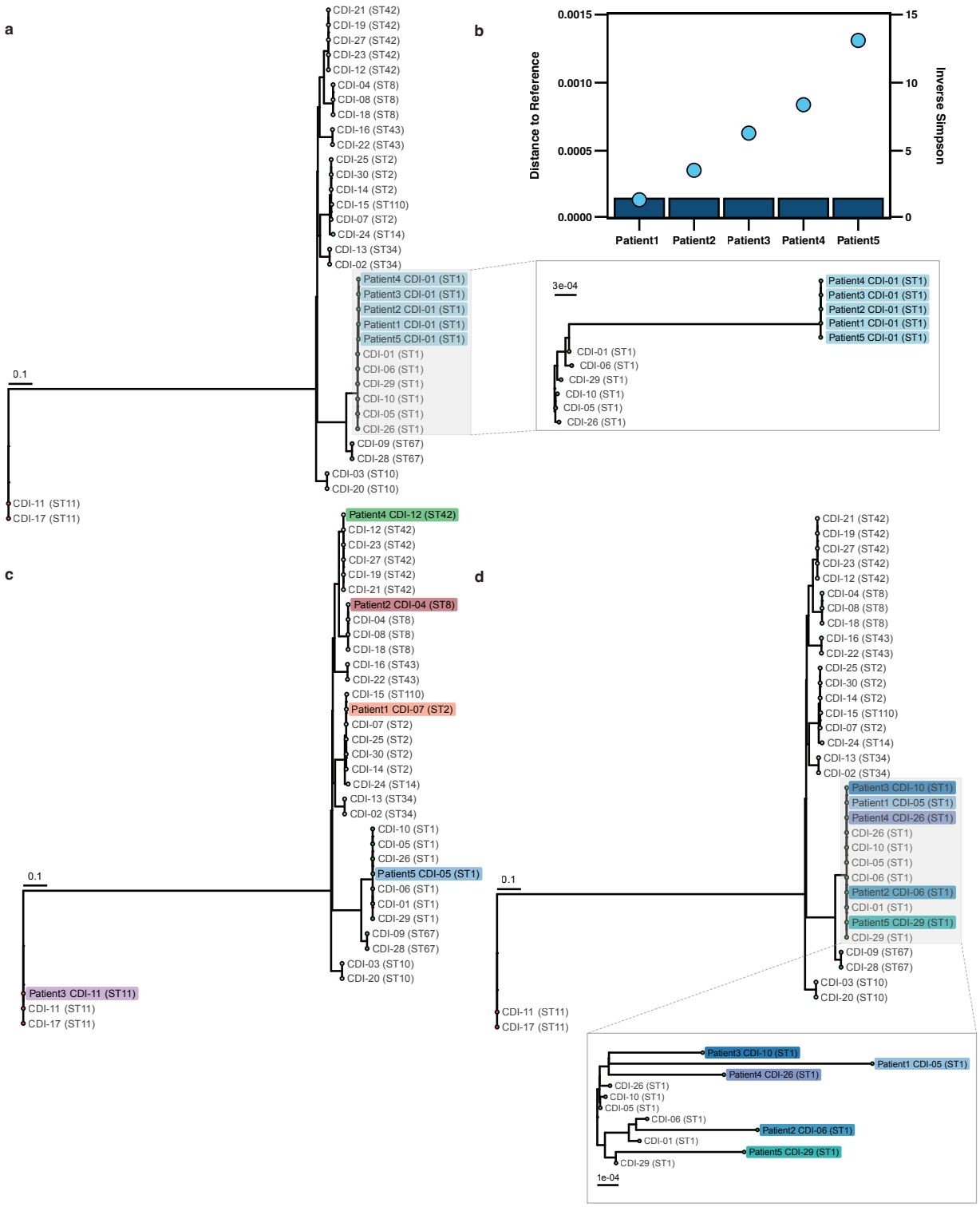

Figure 3.4: ***C. difficile*** **detection in simulated clinical samples using optimized StrainPhlAn4**

Figure 3.4: *C. difficile* detection in simulated clinical samples using optimized StrainPhlAn4 (cont.)

(a) Maximum likelihood optimized StrainPhlAn4 phylogenetic tree output of identical ST1 *C. difficile* strains spiked into patient stool metagenomic samples (n=5). For each spike-in, the *C. difficile* genome was randomly sampled at 30,000 reads while the reference genomes were sampled at 300,000 reads.

(b) Pairwise branch length distances between *in silico* samples and the reference genome (bars) compared with the species-level Inverse Simpson metric of the metagenomic background (points).

(c) Maximum likelihood optimized StrainPhlAn4 phylogenetic tree output of 5 different *C. difficile* STs (pink) spiked into patient stool metagenomic samples (n=5). For each spike-in, the *C. difficile* genome was randomly sampled at 30,000 reads while the reference genomes were sampled at 300,000 reads.

(d) Maximum likelihood optimized StrainPhlAn4 phylogenetic tree output of 5 different ST1 isolates spiked into patient stool metagenomic samples (n=5). For each spike-in, the *C. difficile* genome was randomly sampled at 30,000 reads while the reference genomes were sampled at 300,000 reads.

logenetic distances were extracted from the phylogenetic trees as previously described[168]. Default StrainPhlAn4 settings were unable to detect *C. difficile* in any of the patient backgrounds (fig. S6c). In contrast, optimization of StrainPhlAn4 enhanced the detection of *C. difficile* at 0.25% RA within every patient background (fig. S8). To investigate whether different metagenomic backgrounds influence detection efficiency, we tested *C. difficile* reads from the same genome (CDI-01, ST1) (fig. S7a). Each sample was accurately detected and all 5 clustered together despite varying levels of background diversity, suggesting that the background microbiota does not significantly confound strain tracking (fig. 3.4a and **??**).

MLST, the most widely used DNA-based method for bacterial strain typing, uses unique allele profiles of seven or more housekeeping genes to define distinct STs[171,172]. *In silico* typing using *C. difficile* specific MLST schemes is a highly reproducible and easily interpretable epidemiologic tool that has been widely adopted to characterize *C. difficile*[53,165,59,173,174,175]. To determine if optimized StrainPhlAn accurately typed *in silico* samples to the ST level, we introduced reads from 5 genomes representing 5 STs across 3 clades into patient fecal metagenomic backgrounds (fig. S7a). The *in silico* samples accurately clustered together with the reference genomes of the same clade and were clustered with the reference genomes of the same ST (fig. 3.4c). These findings demonstrate that the optimized StrainPhlAn

accurately predicts ST level phylogeny for *C. difficile*.

While the virulence of *C. difficile* varies between different STs, experimental studies of distinct ST1 isolates demonstrated variation in virulence, ranging from highly virulent to avirulent, that correlates with genomic differences[176,28]. MLST typing, in contrast to WGS, lacks the resolution to identify genomic differences that correlate with virulence or to distinguish patient-to-patient transmission from sporadic occurrences of CDI[171,33,55,175,177,178]. To determine if StrainPhlAn4 could distinguish *C. difficile* at the variant level, we introduced 30,000 reads from 5 different ST1 genomes into patient metagenomic backgrounds (fig. S7a). The resulting tree shows distinct clustering of *in silico* samples closely with their corresponding genomes. Notably, variants (CDI-05, CDI-10, and CDI-26) with high intraspecific genomic similarity (ANI $>$ 99.99%)[179], clustered closely together (fig. 3.4c). Despite the relatively low genomic coverage, our *in silico* samples clustered tightly with their reference genomes (fig. S6c). Therefore, optimized StrainPhlAn4 was sufficient to accurately resolve phylogenetic relationships beyond the ST level to the variant level, highlighting the potential application of StrainPhlAn4 in *C. difficile* outbreak surveillance.

Improved within-species accuracy of StrainPhlAn4 C. difficile tracking

We then sought to assess the performance of the optimized StrainPhlAn4 to track *C. difficile* directly from patient stool metagenomic samples. To do so, we compared the default and optimized StrainPhlAn4 outputs for 73 *C. difficile* positive patient metagenomes, each with a whole genome sequenced *C. difficile* isolate (fig. S9). The optimized StrainPhlAn improved the detection of *C. difficile* from 9 (12.32%) to 22 samples (30.14%) (fig. 3.5a). Notably, samples detected by StrainPhlAn4 represent STs across 4 conventional *C. difficile* clades, suggesting that the detection performance is not influenced by the phylogenetic inferences from StrainPhlAn. However, 35 samples were only detected by MetaPhlAn4, likely due to the higher number of reads required for reliable strain-level characterization. Indeed,

samples detected as *C. difficile* positive by MetaPhlAn4 only (n=35) show a significantly lower median *C. difficile* reads ($< 10^4$ reads) than samples that were detected as *C. difficile* positive by both MetaPhlAn4 and StrainPhlAn4 (fig. 3.5b). However, the read threshold estimate is impacted by the differences in the databases between the pipelines, as discussed earlier. These results underscore the importance of calibrating the sequencing depth in excess of the minimum read threshold for accurate detection even after parameter optimization.

Among metagenomic samples detected by StrainPhlAn4, cluster analysis revealed close associations with their corresponding ST and isolate genome controls (fig. 3.5c). The clustering of multiple samples is attributable to the high, within species level genomic similarity (ANI $> 99.99\%$)[179](fig. 3.5c and **??**). Collectively, these results demonstrate the enhanced performance of ST and variant level tracking of *C. difficile* directly from patient fecal metagenomic datasets.

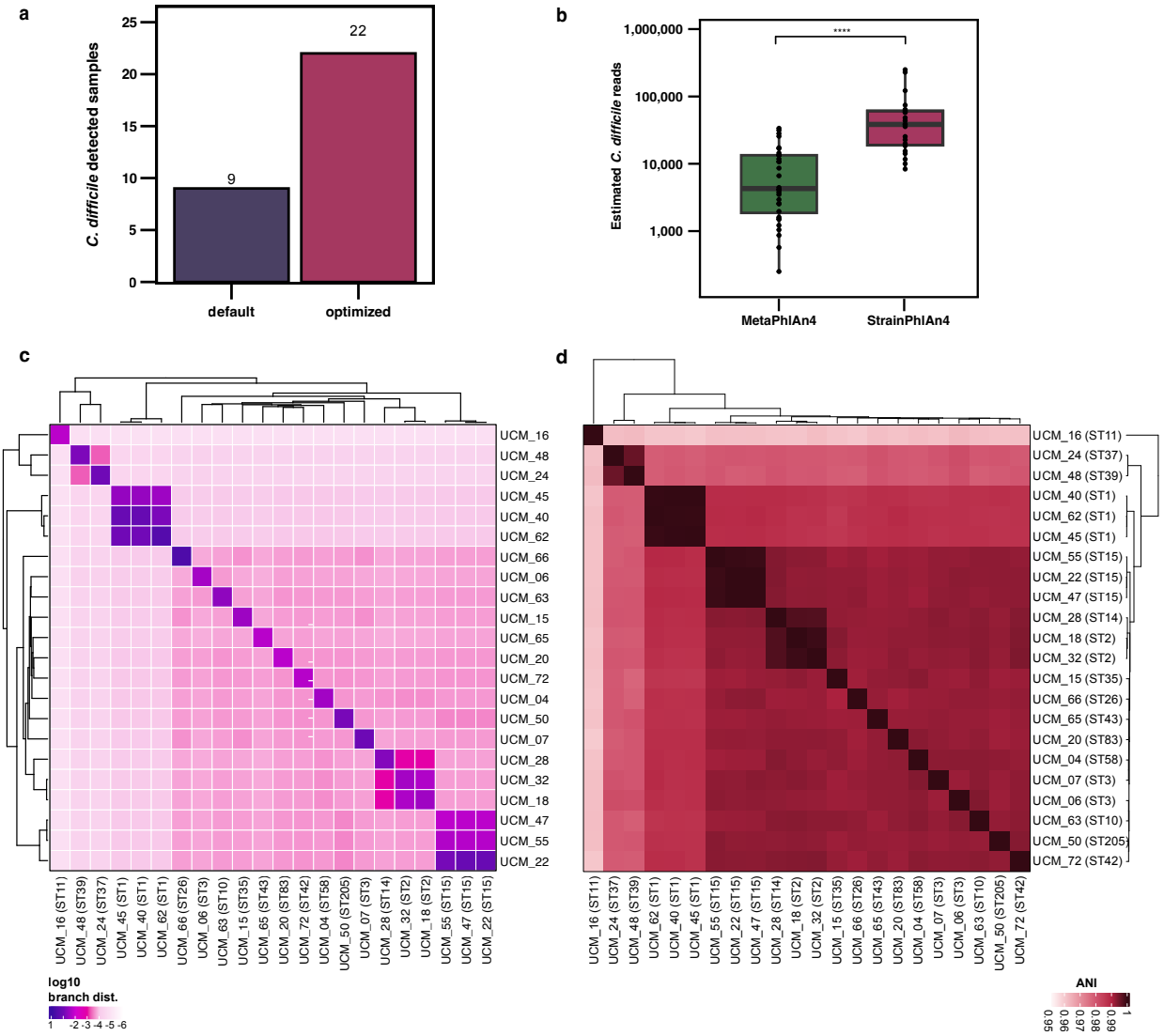

Figure 3.5: Improved within-species accuracy of StrainPhlAn4 *C. difficile* tracking

(a) Improved detection *C. difficile* from patient fecal metagenomic samples using optimized StrainPhlAn4 compared to default (n=73).

(b) Boxplots reporting the estimated *C. difficile* reads in samples detected as *C. difficile* positive by only MetaPhlAn4 or by MetaPhlAn4 and StrainPhlAn4. Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest.

(c) Heat map of pairwise branch length distances between patient metagenomic samples compared to cultured isolate genomes.

(d) Heatmap of genomic ANI distances between *C. difficile* cultured isolate genomes.

P values were adjusted for multiple comparisons using Benjamini-Hochberg procedure and represented as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

CHAPTER 4

MULTI-OMIC INVESTIGATION OF *CLOSTRIDIOIDES DIFFICILE* COLONIZATION ACROSS DIVERSE PATIENT BACKGROUNDS

4.1 Introduction

Multiple potential causes of microbial dysbiosis have been associated with increased risk for CDI including antibiotic exposure[180], Inflammatory Bowel Disease (IBD), or simply diarrhea[98,181]. In certain cases, these associations have been attributed to shifts in the composition of the gut microbial flora or of microbially produced metabolites. The physiological production of secondary metabolites such as SCFAs and SBAs is often associated with salutary impacts on immune modulation, barrier integrity, and pathogen defense. In addition, the landscape of bioavailable nutrients such as amino acids, carbohydrates, and transition metals have significant impacts on host-pathogen dynamics in the gut lumen[105]. This is particularly important for the control of enteric pathogens such as *C. difficile*. Utilizing tools tuned to enhance the sensitivity of *C. difficile* detection within short-read metagenomic datasets, we explore microbe-metabolite features associated with *C. difficile* colonization across various states of dysbiosis.

4.2 Results

Patient selection and clinical features

The severity of CDAD is associated with the complex integrations of several host-associated, microbe-associated, and environmental risk factors such as immune response and antibiotic treatment. Notably, the reduced diversity of the microbiome and the absence of microbially fermented metabolites have been associated with increased mortality and morbidity

for CDI patients[126,182,183]. We hypothesize that integrating metagenomic and metabolomic can identify features of CDI across a diverse collection of patient backgrounds. *C. difficile* often spreads throughout healthcare settings due to the production of aero-tolerant and robust spores. To capture a broad section of patients likely to have frequent and long-term exposure to a large hospital system, we retrospectively surveyed 484 patients across 4 different IRB-approved studies: 64 from HT, 130 from LD, 40 from LT, 196 from MICU, and 54 who were enrolled in multiple studies simultaneously. From each patient, the first *C. difficile* positive stool sample (*C. difficile*$^+$, n = 110), detected by *C. difficile*$^-$specific *16S rRNA* PCR, or the first overall sample collected (*C. difficile*$^-$, n = 374) was used for analysis to avoid potential sampling biases (fig. S4a). For comparisons to a "healthy" microbiome, we included samples from 21 HDs. Every subject provided a fecal sample which was then processed by short-read shotgun sequencing and mass-spectrometry for paired metagenomic and metabolomic datasets. From the culture positive *C. difficile*$^+$ samples, 73 *C. difficile* isolates were sequenced and subtyped by conventional MLST and the remaining 37 uncultured samples (uncultured, 34%) (fig. S4b). The distribution of *C. difficile* subtypes among patient-derived samples was generally evenly distributed except an increase of *tcdB*$^+$ LT samples compared to MICU (pairwise Fisher's Exact test p = 0.0145) (fig. 4.1a). Of these, 48 genomes (65%) encoded *tcdB* (*tcdB*$^+$) and include epidemic strains such as ST1 (clade 2), ST14 (clade 1), and ST11 (clade 5) while the 25 *tcdB*$^-$ included STs from clades 1 and 4 (fig. 4.2a).

The classic clinical manifestation of CDAD is unexplained watery diarrhea and some studies even suggest that diarrheal events are a risk factor for CDI, due to an increase in free amino acids[98,181]. Across the patient cohorts, the consistency of the fecal sample was generally consistent across LT, LD, and MICU cohorts but HT samples have significantly fewer "liquid" samples than LD (p = 0.00151), MICU (p= 5.48×10^{-6}), and Multi-study (p = 0.0195) groups when compared using pairwise Fisher's Exact test. However, we did not

observe a significant difference in sample consistency between the patient samples (fig. 4.1c). Overall, the cohorts are generally similar in regards to consistency and *C. difficile* subgroups but notable variation from the HT group will be accounted for in downstream analyses. Interestingly, the consistency of the sample was not necessarily correlated to the diversity found within the sample.

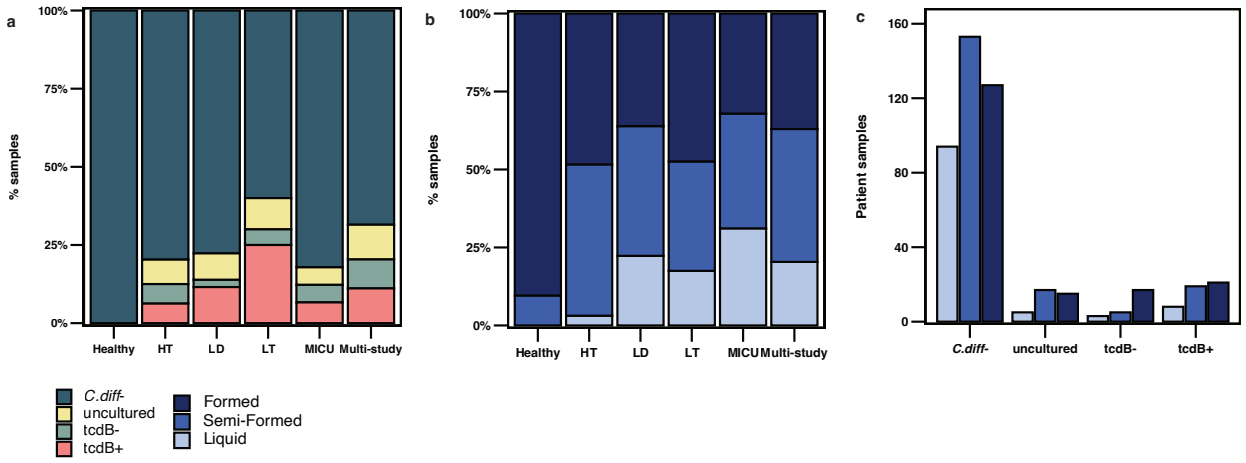

Figure 4.1: Proportion of samples across *C. difficile* subcategories and fecal consistency groups

(a) Percentage of study samples by cohort included into *C. difficile* categories: *C. difficile 16S rRNA*$^-$ (navy), uncultured *C. difficile 16S rRNA*$^+$ (yellow), cultured *C. difficile*$^+$ *tcdB*$^-$ (green), and cultured *C. difficile*$^+$ *tcdB*$^-$ (red).

(b) Percentage of study samples by cohort included into *C. difficile* categories: liquid, semi-formed, and formed.

(c) Total number of samples categorized by sample consistency between *C. difficile tcdB* subgroups (n= 484). Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests.

Quantification of C. difficile from fecal samples

Metagenomic microbial profiles were estimated using a custom *C. difficile* sensitivity-tuned application of MetaPhlAn4 introduced in Chapter 3. Within our *C. difficile*$^+$ patient samples, the range of *C. difficile* RA spans between imperceptible to a max of 7.57% (median = 0.034%, IQR = 0.319%). We found that the highest *C. difficile* RA estimated by

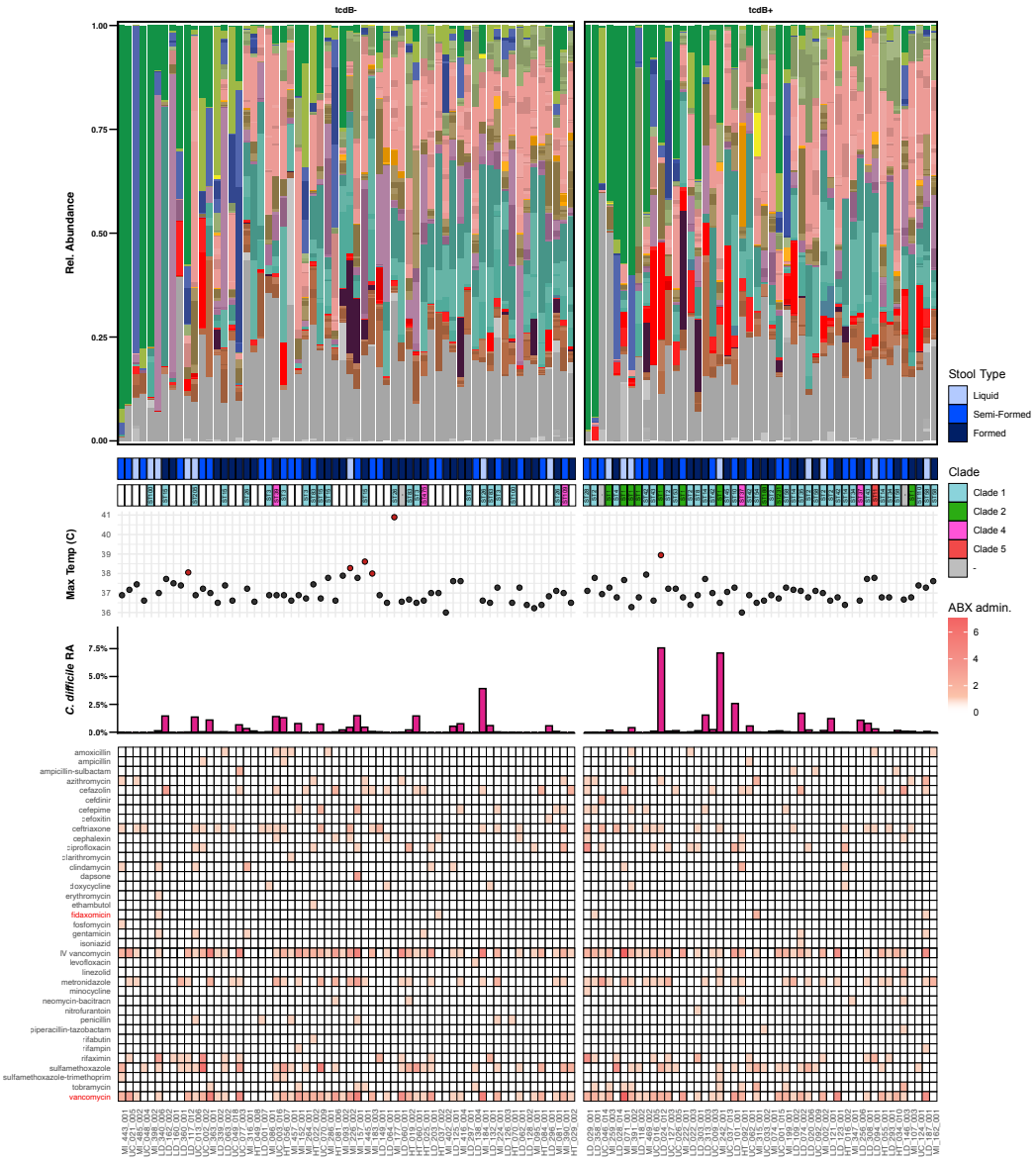

Figure 4.2: Survey of *C. difficile 16S rRNA* PCR positive patient fecal samples

(a) Genus level taxonomy of *tcdB*+ (n=48) and *tcdB*− (n=62) *C. difficile 16S rRNA*$^+$ samples (n=110). Annotated with stool consistency and MLST typing. Fever (red), body temperature (°C) $\geq$ 38, or the highest recorded temperature on the day of fecal sample collection. *C. difficile* relative abundance (RA%) estimates by MetaPhlan4 (local stat_q = 0.05). Samples ranked in order of increasing inverse Simpson.

(b) Number of antibiotics of *tcdB*$^+$ (n=48) and *tcdB*$^-$ (n=62) *C. difficile 16S rRNA*$^+$ samples (n=110) administered within 2 months before or after sample collection. First-line treatments for CDI are listed in red.

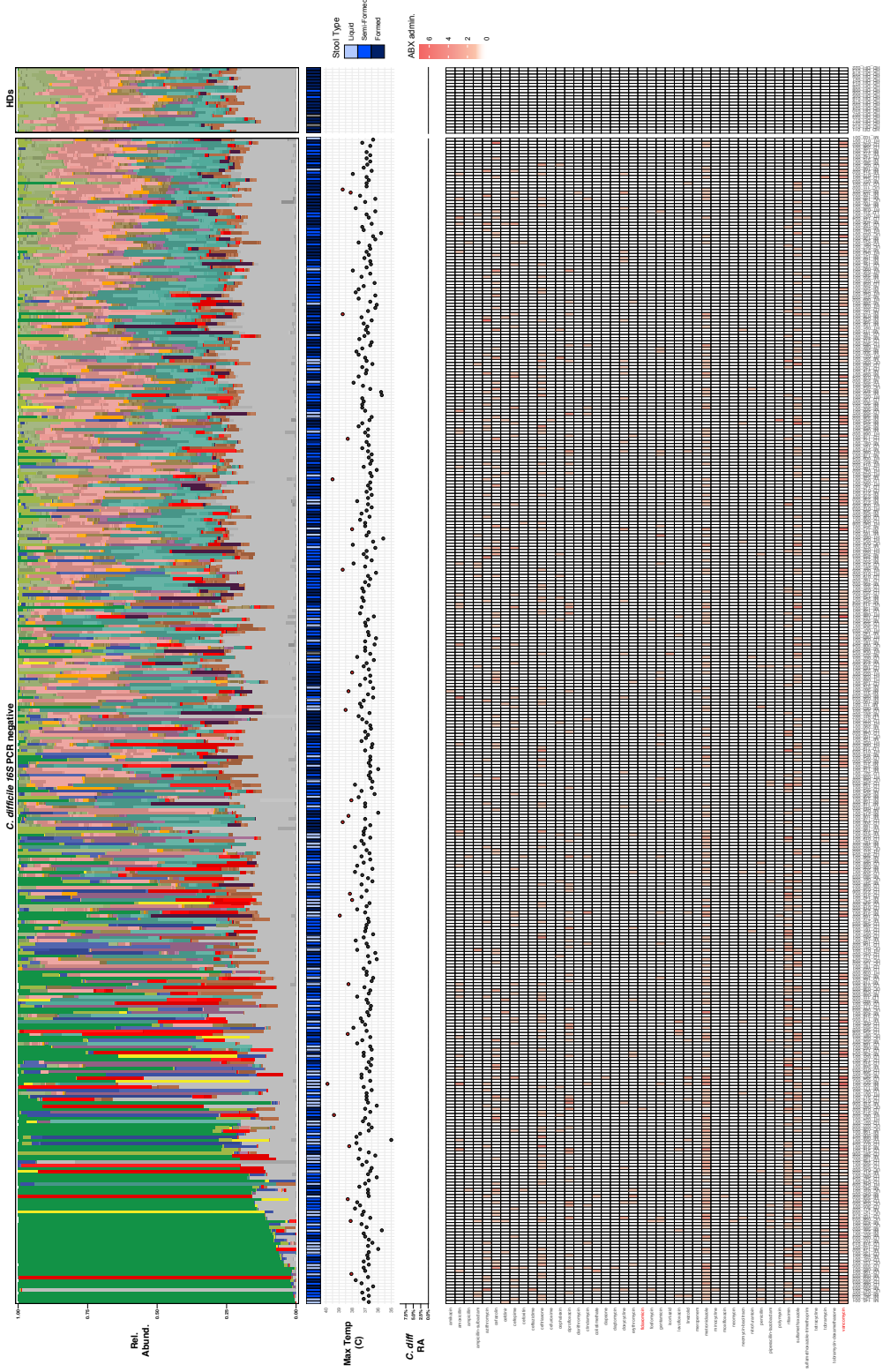

Figure 4.3: Survey of HD and *C. difficile 16S rRNA*$^-$ patient fecal samples

Figure 4.3: Survey of HD and *C. difficile 16S rRNA*$^-$ patient fecal samples (cont.)

(a) Genus level taxonomy of healthy donor samples (n=21) and *C. difficile 16S rRNA*$^-$ samples (n=374). Annotated with stool consistency and MLST typing. Fever (red), body temperature (C) $\geq$ 38, or the highest recorded temperature on day of fecal sample collection. *C. difficile* relative abundance (%) estimates by MetaPhlan4 (local stat_q = 0.05). Samples ranked in order of increasing inverse Simpson.

(b) Number of antibiotics of healthy donor samples (n=21) and *C. difficile 16S rRNA*$^-$ samples (n=374) administered within 2 months (56 days) before or after sample collection. First-line treatments for CDI are listed in red.

MetaPhlAn4 are from liquid samples with *tcdB*$^-$ *C. difficile*. However, we did not observe a significant difference in the MetaPhlAn4 RA for *C. difficile* or in the absolute copies of *C. difficile 16S rRNA* between *C. difficile* subgroups (fig. S11a-S11b).

Antibiotic use and diversity

Antibiotics are essential and life-saving therapies for infections but the broad use of antibiotics has been well established to deplete the microbial diversity in the gut and is an important risk factor for CDI. We defined antibiotic exposure as antibiotics administered within 8 weeks (56 days) prior to sample collection and subsequent antibiotic treatment within 8 weeks after sample collection.Overall, antibiotic exposure prior to and treatment post-sample collection was common across all patients (*C. difficile*$^-$ = 324 (86.6%); *C. difficile*$^+$ = 97 (88.2%)), excluding the HDs (HD = 0) (fig. 4.2b, fig. 4.3b). Since 2017, vancomycin and fidaxomicin have been listed as current first-line treatments for CDI by the Infectious Disease Society of America (ISDA) (fig. 1.1). Both antibiotics are poorly absorbed from the gastrointestinal tract so oral/rectal administration can act locally to target *C. difficile*. As expected, a higher proportion of patients associated with *C. difficile*$^+$ samples received these antibiotics prior to and post-sample collection (fig. 4.4a). Inversely, intravenous (IV) vancomycin was the most commonly administered antibiotic with an average of 46.1% of patients (sd = 2.11%) receiving this treatment prior to sample collection but is not secreted into the gut lumen at appreciable concentrations so is unlikely to dramatically impact the

microbiota. Metronidazole was listed as a first-line treatment for *C. difficile* until 2017 and is indicated for a broad range of anaerobes. Although metronidazole was also frequently administered, the percentage of patients given metronidazole prior to (mean = 33.56%, sd = 2.46%) and post-sample collection (mean = 15.40%, sd = 1.36%) was even across all subgroups and is unlikely to significantly change our results. Notably, the subset of patients associated with uncultured *C. difficile*$^+$ samples were administered CDI-targeting antibiotics more days before other groups (median = 17 days), but this difference is not reflective of differences in *C. difficile* quantification, microbial diversity, or metabolite profiles (fig. 4.4b). This is likely due to an overall slow recovery time of the microbiota, particularly in patients receiving other antibiotics and treatments. In summary, antibiotic use across the patients was common but generally similar across study groups while the use of CDI antibiotics may be useful to distinguish adverse outcomes of *C. difficile* colonization.

Since the antibiotic exposure of patients was remarkably similar, we hypothesized that aggregate diversity metrics such as species richness and alpha diversity were unlikely to differentiate *C. difficile*$^+$ from *C. difficile*$^-$ samples. Indeed, the total number of species detected by MetaPhlAn4 and the α-diversity measured by inverse Simpson were not significantly different between *C. difficile*$^+$ and *C. difficile*$^-$ samples (fig. 4.5a- 4.5b). However, there was a significant difference in β-diversity measured by Bray-Curtis distance between *C. difficile* samples (PERMANOVA p = 0.001), *tcdB* subgroups (p = 0.001), antibiotic exposure prior to collection (p = 0.001)(fig. 4.5c). But the low $R^2 = 0.02 - 0.046$, suggests that these factors only minimally explain the variance in the overall inter-sample diversity.

Microbial associations

Due to the significant yet small influence of multiple grouping factors on the metagenomic β-diversity, we hypothesized that significant associations between microbes and these factors require analysis at a higher resolution. We used a multi-variate mixed-effects regres-

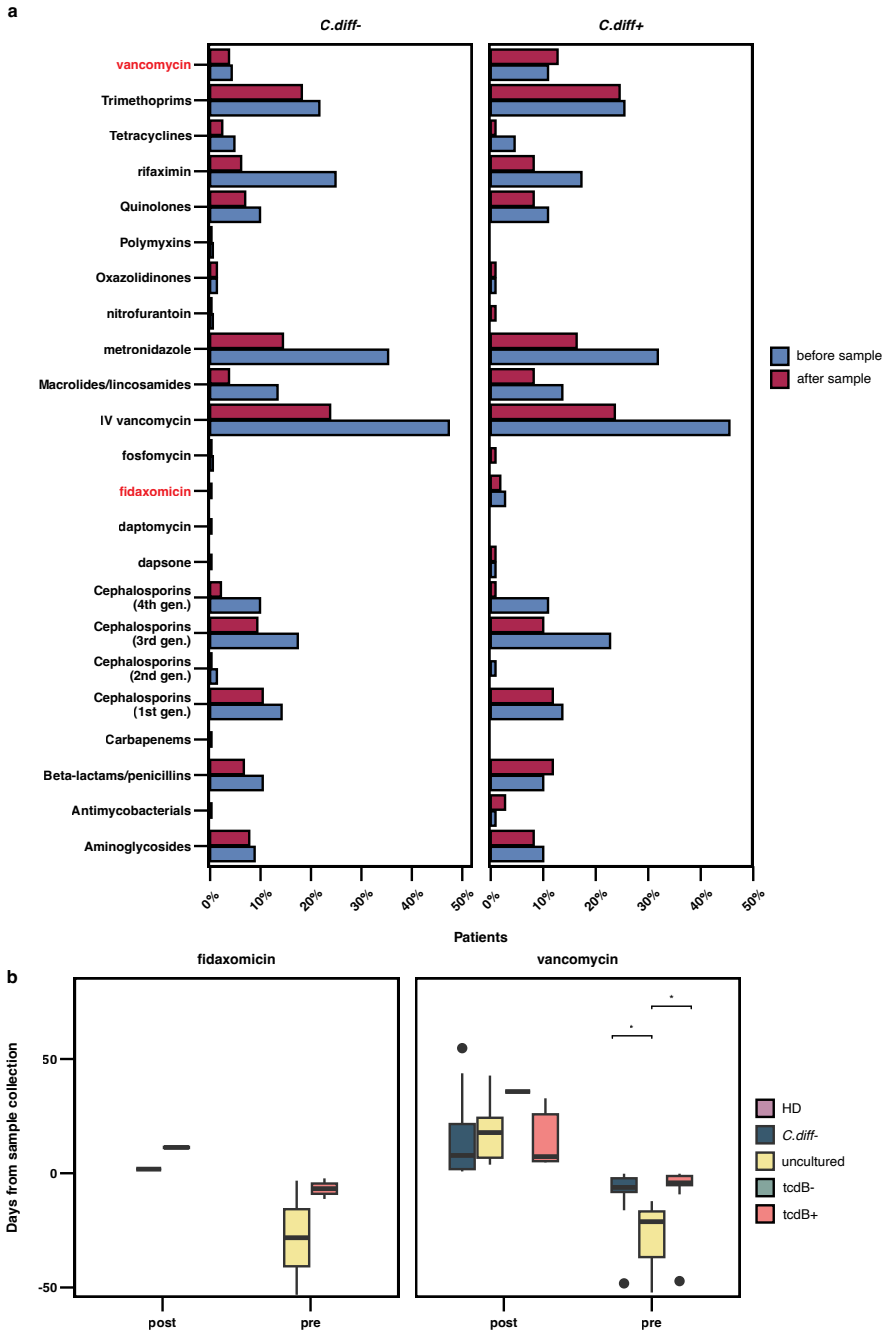

Figure 4.4: Administration of antibiotics to study patients

(a) Percent of patients that received antibiotics within 8 weeks (56 days) prior to (blue) or after (red) the date of patient sample collection (n=484). First-line treatments for CDI are highlighted in red.

(b) Comparison of the length of time (days) of first-line CDI therapies to the date of sample collection between patient groups.

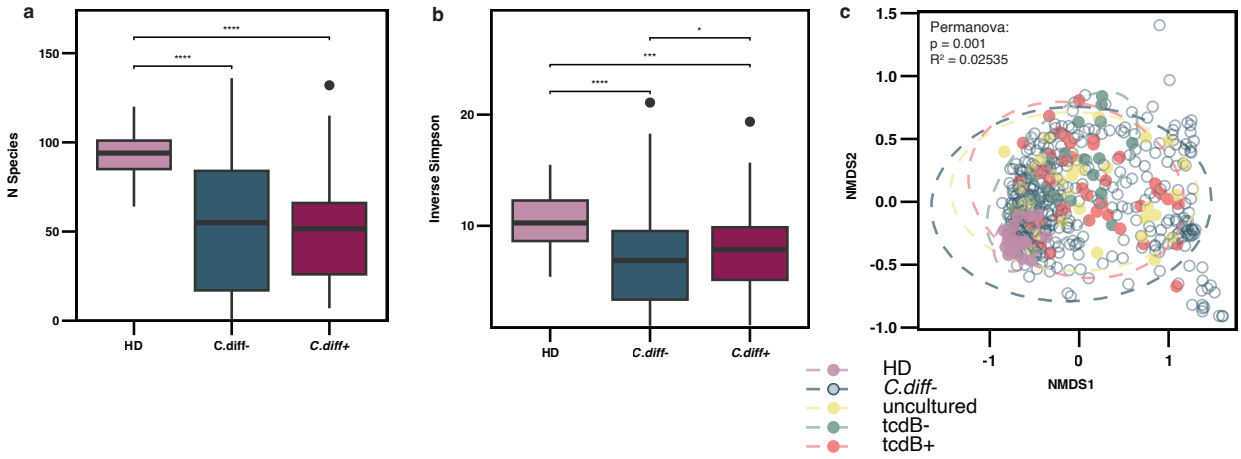

Figure 4.5: Microbial diversity and associated species across fecal samples

(a) Species richness compared between HDs (n=21), *C. difficile*$^{-}$ (n=374), and *C. difficile*$^{+}$ fecal samples (n=110).

(b) Inverse Simpson diversity compared between HDs, *C. difficile*$^{-}$, and *C. difficile*$^{+}$ fecal samples.

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

(c) Non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances of fecal metagenomes. Ellipses assume a multivariate t-distribution around each sample group centroid. Statistical comparison of variance between sample groups analyzed with PERMANOVA.

sion model with MaAslin2[184], to reveal species-level features associated with each grouping variable. Overall, the model identified 161 significantly associated species.

Microbe-*C. difficile* colonization associations

Predictably of 7 significantly associated species, the strongest predictor of *C. difficile*$^+$ samples was *C. difficile* (β coef = 4.93, p = 6.48×10^{-54}) (fig. 4.6). However, a group of *Bacillota* (formerly *Firmicutes*) were predictive of *C. difficile*$^-$ samples. This group includes *Hungatella hathewayi (H. hathewayi)*, *Erysipelatoclostridium ramosum (E. ramosum)*, *Blautia producta (B. producta)*, *Enterocloster clostridioformis (E. clostridioformis)*, and *C. symbiosum*. *H. hathewayi* and *E. ramosum*, are typically present in the normal human microbiota but have been associated with opportunistic bacteremia and colorectal cancer (CRC)[185]. *E. coli* was also predictive of *C. difficile*$^+$ samples, which corroborates the association of gram-negative co-enrichment as seen with vancomycin-resistant *Enterococcus* (VRE)[186,187]. The microbes associated with HD status (n = 39) include but are not limited to *Anaerostipes hadrus (A. hadrus)*, *Anaerobutyricum hallii (A. hallii)*, *Dorea longicatena (D. longicatena)*, and *Eubacterium rectale (E. rectale)*.

Microbe-antibiotic exposure associations

The model also identified 3 taxa independently associated with prior antibiotic exposure: *Enterococcus faecium (E. faecium)*, *Lacticaseibacillus rhamnosus (L. rhamnosus)*, and *Lacticaseibacillus paracase (L. paracase)*. Notably, *E. faecium* has been implicated with adverse health outcomes[110,109], co-colonization with *C. difficile*[188], but was not associated with *C. difficile*$^+$ samples in the patient samples. In addition, *L. rhamnosus* and *L. paracase*, common probiotics associated with dairy production, likely expanded post-antibiotic treatment due to reduced overall microbial diversity. Samples without prior antibiotic exposure (unexposed) were associated with 85 species. Symbiotic species such as *Phocaeicola vul-*

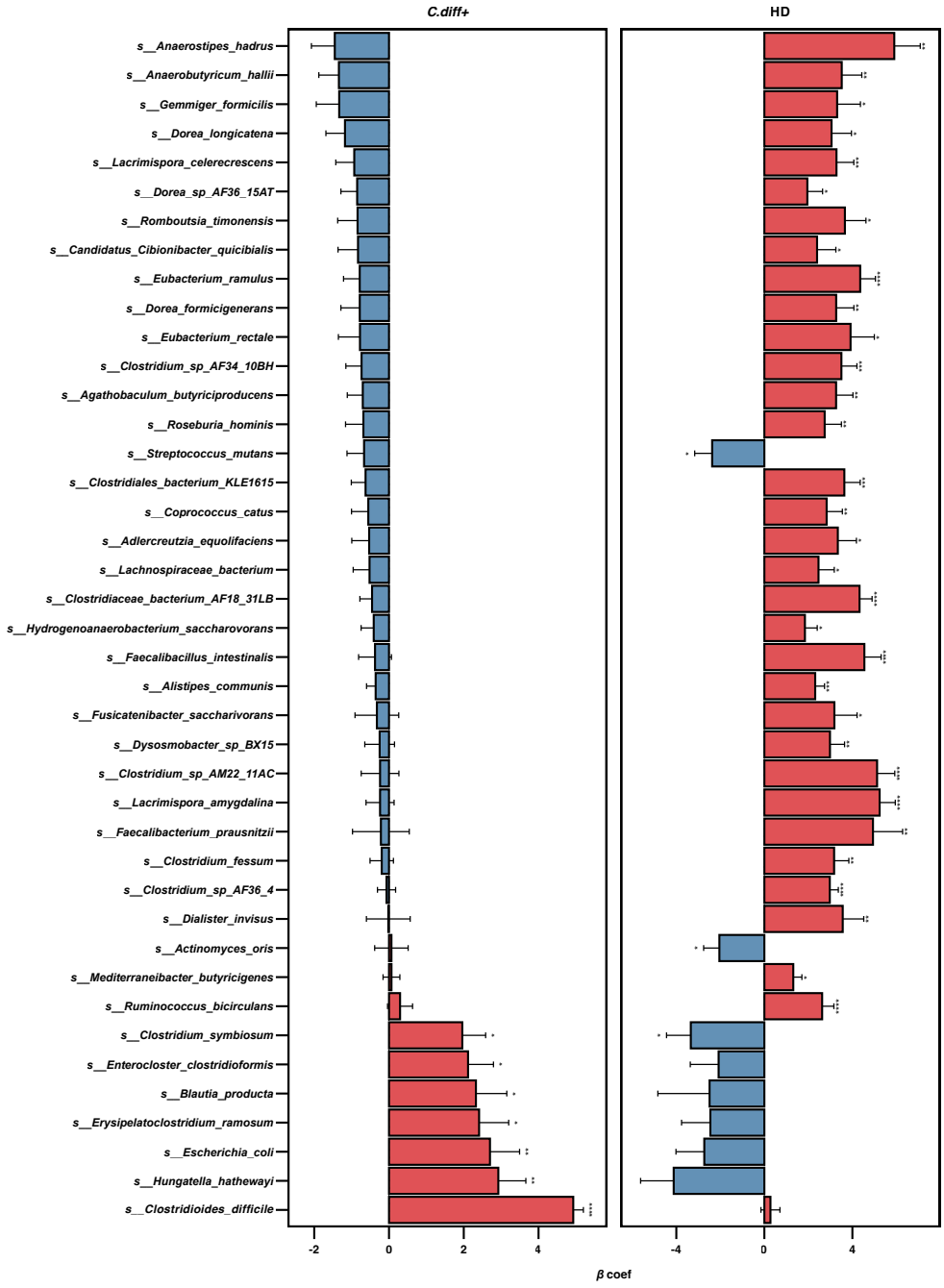

Figure 4.6: Microbe-group associations with *C. difficile*$^+$ fecal samples and HD samples with MaAslin2

Significant species-level features ($p < 0.05$; n = 46) associated with *C. difficile*$^+$ fecal samples and HD samples using mixed effects models with *C. difficile*$^-$ samples as the reference. Positive (red) and negative (blue) effect size of each association (β coefficient). Multiple comparisons were adjusted for using the Benjamini-Hochberg procedure and are represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$..

gatus (P. vulgatus), *Bacteroides ovatus (B. ovatus)*, *Bacteroides uniformis (B. uniformis)*, produce SCFAs (acetate and propionate), deconjugate BAs, and compete for fermentable amino acids[189,190,191], have been associated with colonization resistance against mucin degrading species and *C. difficile* pathogenesis[192]. Interestingly, *Akkermansia muciniphila (A. muciniphila)*, a mucin-degrading species, has been previously associated with enhancing CDAD[193] and conversely with carbohydrate competition[194], was also negatively associated with antibiotic treatment (β coef = -1.35, p = 2.49×10^{-2}).

Microbe-diarrhea associations

Stool consistency was typically associated with the increased loss of several species (n = 150) with increasingly loose stool. These include major sources of SCFAs: *Bifidobacterium longum (B. longum)*, *Clostridium leptum (C. leptum)*, *Enteroclosterbolteae (E. bolteae)*, and *Anaerostipes caccae (A. caccae)*. Liquid samples were positively associated with *Candida albicans (C. albicans)*, a fungal species, and *Staphylococcus epidermidis (S. epidermidis)*, a common skin commensal species. This result is likely reflective of the lower species diversity in liquid samples (inverse Simpson median = 4.43, IQR = 6.32), leading to spurious associations.

Metabolite associations

Metabolomics offers a practical functional profile of the microbial community, potentially revealing biochemical activities driving host-pathogen interactions in the gut environment. Other studies have recognized the importance of the microbially-derived secondary metabolites on the pathogenesis of *C. difficile*[195,192]. We hypothesize that metabolic biomarkers can distinguish samples harboring *C. difficile*. We initially directly compared the log-fold change of measured targeted metabolite values to identify differentially abundant compounds compared to *C. difficile*$^-$ samples (fig. 4.9). We observe that 20 compounds were enriched

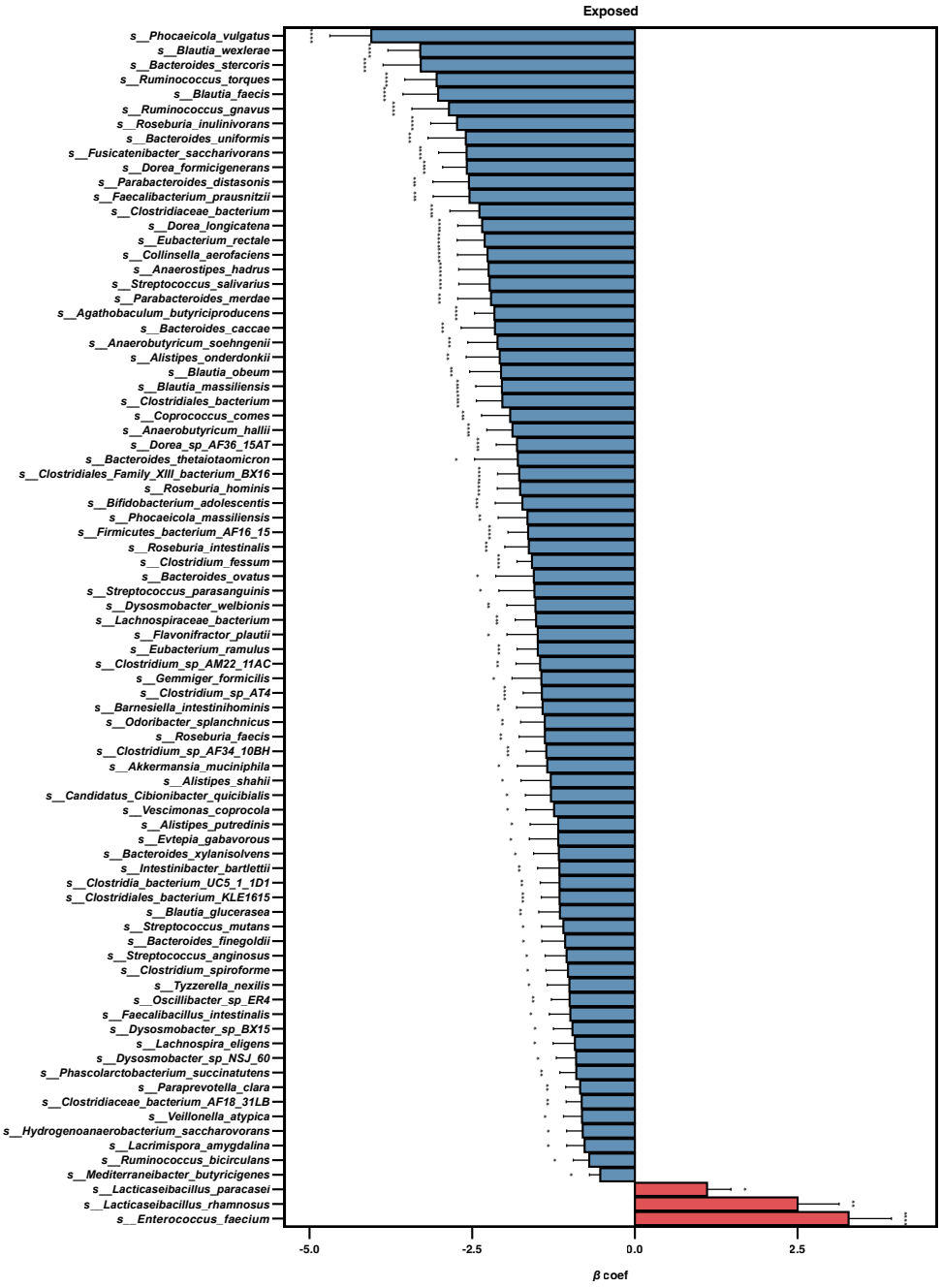

Figure 4.7: Microbe-group associations with prior antibiotics exposure with MaAslin2

Significant species-level features ($p < 0.05$; n = 89) associated with any exposure to antibiotics within 8 weeks prior to sample collection using mixed effects models with unexposed samples as the reference. Positive (red) and negative (blue) effect size of each association (β coefficient). Multiple comparisons were adjusted for using the Benjamini-Hochberg procedure and are represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

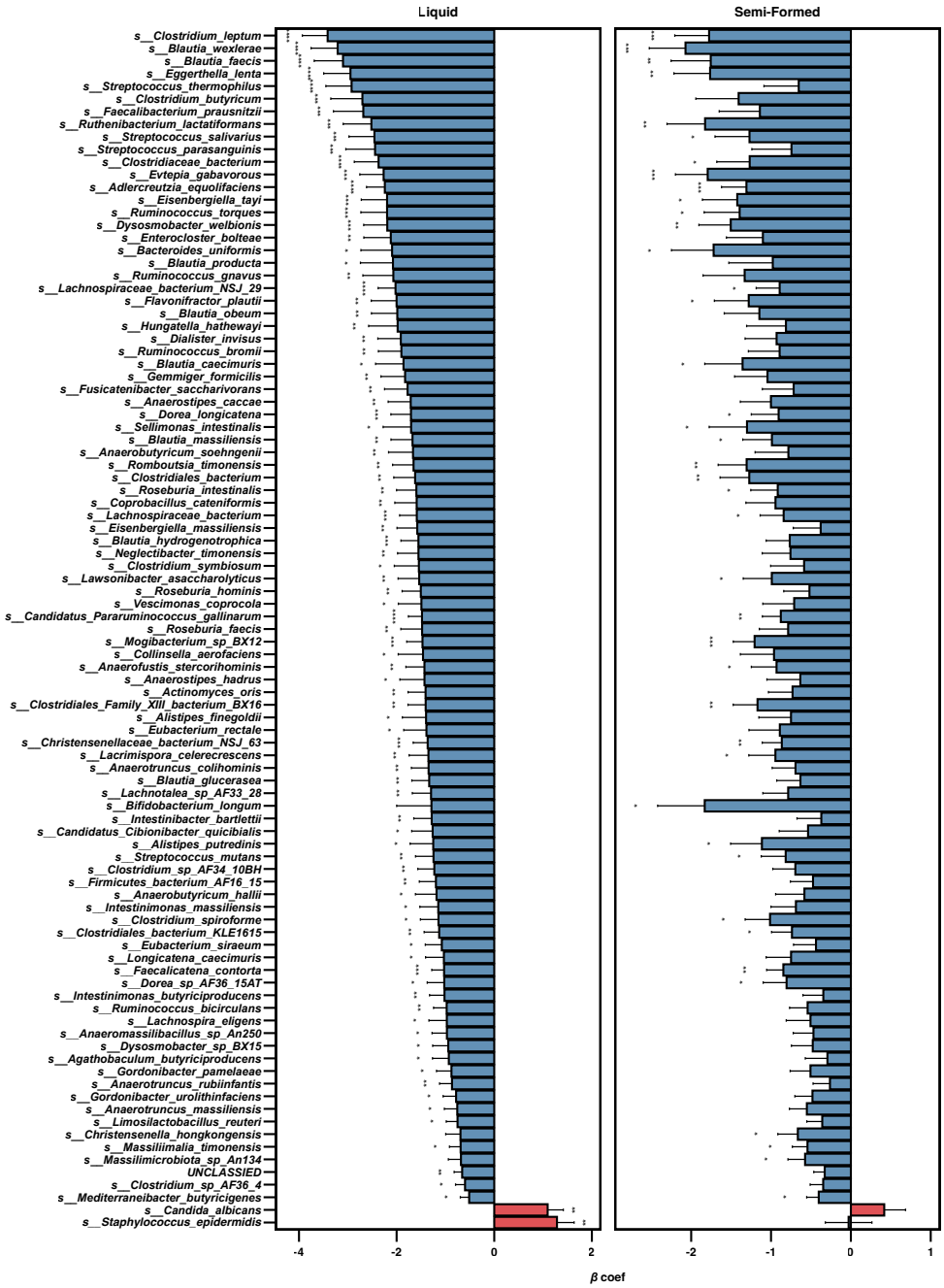

Figure 4.8: Microbe-group associations with stool sample consistency with MaAslin2

Significant species-level features ($p < 0.05$; n = 152) associated with liquid or semi-formed samples using mixed effects models with formed samples as the reference. Positive (red) and negative (blue) effect size of each association (β coefficient). Multiple comparisons were adjusted for using the Benjamini-Hochberg procedure and are represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

in *C. difficile*$^-$ samples. These include: 4 SBAs, 4 phenolic aromatic amines, 4 branch chain fatty acids, 3 MBAs, 3 aminated fatty acids, phenylacetate and acetate. Of the enriched SBAs, ω-muricholate (3α,6α,7β-trihydroxy cholanoic acid, oMCA), was previously described to be depleted in cirrhotic patients but enriched in those treated with rifampin[?], anti-tuberculosis antibiotics with noted hepatotoxicity, possibly due to induction of 6-α-hydroxylation[196]. This suggested that the increased representation of *C. difficile*$^+$ samples from LT patients (fig. 4.1) was a confounding variable in these results.

To control for the study group variation and also to delineate the associations by grouping variable, we applied our mixed effects model on the metabolomic dataset (fig. 4.10). The model identified 4 metabolic features that independently differentiated *C. difficile*$^+$ samples from *C. difficile*$^-$ samples: isobutyrate, 5-AV, ω-muricholate, and 4-methylvalerate (fig. 4.10a). Isobutyrate, 5-AV, and 4-methylvalerate are all fermentation byproducts of amino acid fermentation: reductive L-valine, L-proline, L-leucine catabolism[97,197]. These results contribute to the growing body of work highlighting the importance of fermentable amino acids for *C. difficile* metabolism[198,199].

Antibiotic exposure was generally associated with a loss of microbially-derived secondary metabolites such as SCFAs and SBAs (fig. 4.10b). CA, an unconjugated PBA, was enriched in samples exposed to antibiotics, indicating the loss of *bai*-operon encoding species such as *C. scindens*. Similarly, the metabolite associations across different stool consistencies reveal an increasing reduction of SCFAs, SBAs, and several amino acids. Overall, metabolic end products of amino acid reduction can differentiate *C. difficile*$^-$ and *C. difficile*$^+$ samples across clinically distinct patient backgrounds. However, mixed effects models predict each feature independently at a time, which may add bias to the results[184].

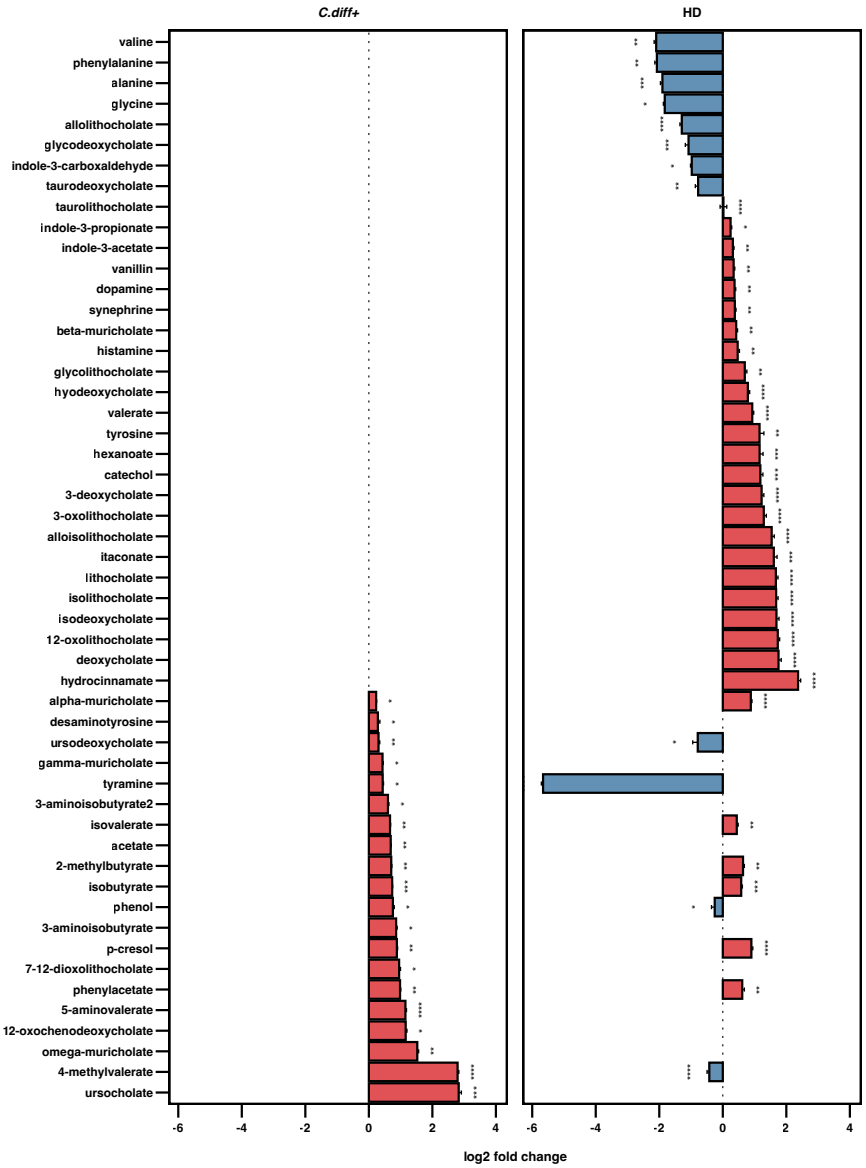

Figure 4.9: Qualitative metabolite features differentially associated with the absence of *C. difficile*

Log-fold change of significant metabolites compared between *C. difficile*$^-$ samples (n=374) and *C. difficile*$^+$ patient samples (n=110) (left). Log-fold change of significant metabolites compared between *C. difficile*$^-$ samples (n=374) and HD samples (n=21) (right).

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

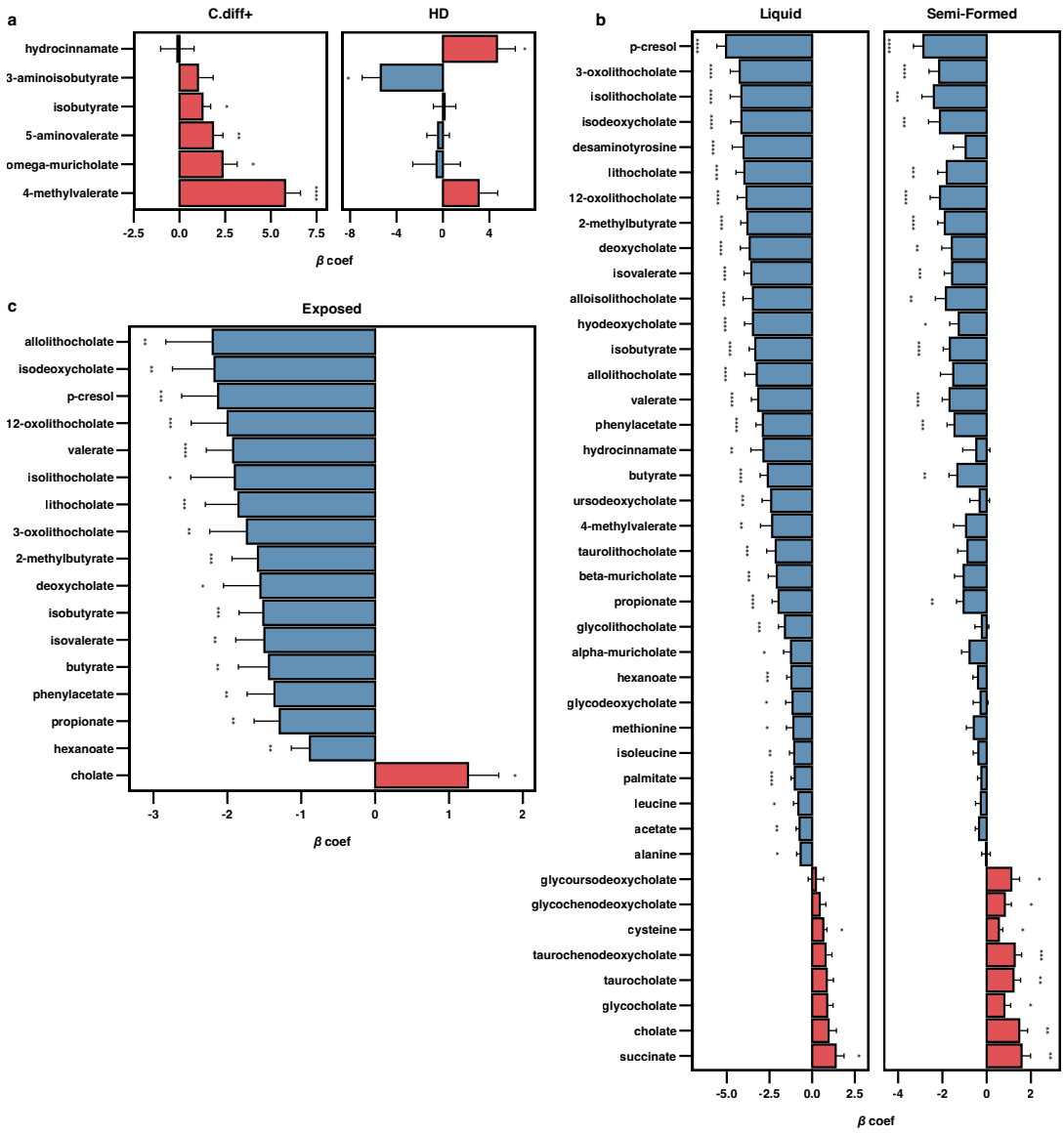

Figure 4.10: Qualitative metabolite features differentially associated with the absence of *C. difficile*

Significant metabolic features associated with: (a) *C. difficile*$^{+}$ fecal samples and HD samples with *C. difficile*$^{-}$ samples as the reference ($p < 0.05$; n = 6), (b) any exposure to antibiotics within 8 weeks prior to sample collection with unexposed samples as the reference ($p < 0.05$; n = 17), and liquid or semi-formed samples with formed samples as the reference ($p < 0.05$; n = 59).

Positive (red) and negative (blue) effect size of each association (β coefficient). Multiple comparisons were adjusted for using the Benjamini-Hochberg procedure and are represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Multi-omic predictions of CDI outcomes

To avoid overfitting on a large number of variables, we also performed a multi-block sPLSDA analysis[200] using maximum distances to reveal the most predictive features across both metagenomic and metabolomic datasets using the methods defined in the MixOmics package[128] (fig. 4.11). Partial Least Squares (PLS) regression maximizes covariance with outcome variables whilst minimizing colinearity and is useful for predicting outcomes from large arrays of predictor variables. Sparsity, reduction in the number of irrelevant variables, is introduced by iteratively adding "soft-threshold" penalties onto the latent components/loading vectors[201]. The greatest variation in the dataset occurs within the *C. difficile*$^-$ population, associated with low microbial diversity (enriched in *E. faecalis* and *S. epidermidis* and depleted in LCA, that is irrespective of *C. difficile* presence (fig. S14), which may be improved with additional tuning. However, the performance of the model demonstrates reasonable AUROC of each *C. difficile* group (fig. S15). The taxonomy and metabolite loading values which discriminate *C. difficile*$^+$ samples lie on the second and third latent components (fig. 4.11a, fig. S14, fig. S14). *C. difficile* was also positively associated with *R. gnavus* and *E. coli* on component 2 (fig. 4.11b). *R. gnavus* is a mucolytic commensal, that scavenges sialic acid from the colonic mucus[202], has been previously associated with biofilms observed in IBD[203,204,205] and CDI[206,207]. In combination with *E. coli* and other pathobionts from IBD patients, *R. gnavus* may have a role in recruiting pro-inflammatory Th1 and Th17 cells to the lamina propria[208].

Notably, evidence of Stickland fermentation (4-methylbutyrate, isovalerate, isobutyrate, 5-AV, etc.) which was identified by the mixed effects model, was significant on component 2 (fig. 4.10a). The presence of *C. difficile* was positively associated with both the oxidative (driven by *vorC1*, *iorA*, *ptb1*, *buk*) and reductive arms (driven by *hadABC*, *ldhA*, and *acdB*) of L-leucine metabolism. Serine was the only amino acid significantly associated with *C. difficile*, which may be a sign of serine biosynthesis to fuel cysteine metabolism[99]. In addition,

the abundance of UDCA, succinate, para-cresol, and acetate was positively associated with *C. difficile* while alloisoLCA, a potent anti-microbial against gram-negative pathogens[76] and inducer of colonic regulatory T cells (Tregs) in an NR4A1-dependent manner[209], was negatively associated.

The abundance of 3-aminoisobutyrate (β-aminoisobutyrate, BAIBA), the product of valine and thymine catabolism, is anti-inflammatory and was associated with protection against CDI post-treatment with antimicrobial steroids[210], had the highest variation on component 3, discriminating the HDs samples (fig. 4.11a). Using the sPLSDA method recovered several metabolites previously associated with HDs and *C. difficile*$^+$ samples in the log-fold change analysis (fig. 4.9). Integrating the inferences gleaned across our models, we were able to identify key metabolic features such as 5-AV and acetate, which also differentiated *C. difficile*$^-$ and *C. difficile*$^+$ patients from our quantitative metabolite panel (fig. S12, fig. S13). Acetate, a common by-product of many microbial metabolic pathways, was enriched in diverse HD samples compared to patient samples (fig. S12b). In contrast, *C. difficile*$^+$ samples had the highest concentration of 5-AV and HD samples had the lowest concentration (fig. S12c). This observation may be attributed to various factors including but not limited to: reduced growth of 5-AV producers and crossfeeding with 5-AV-fermenting species[211]

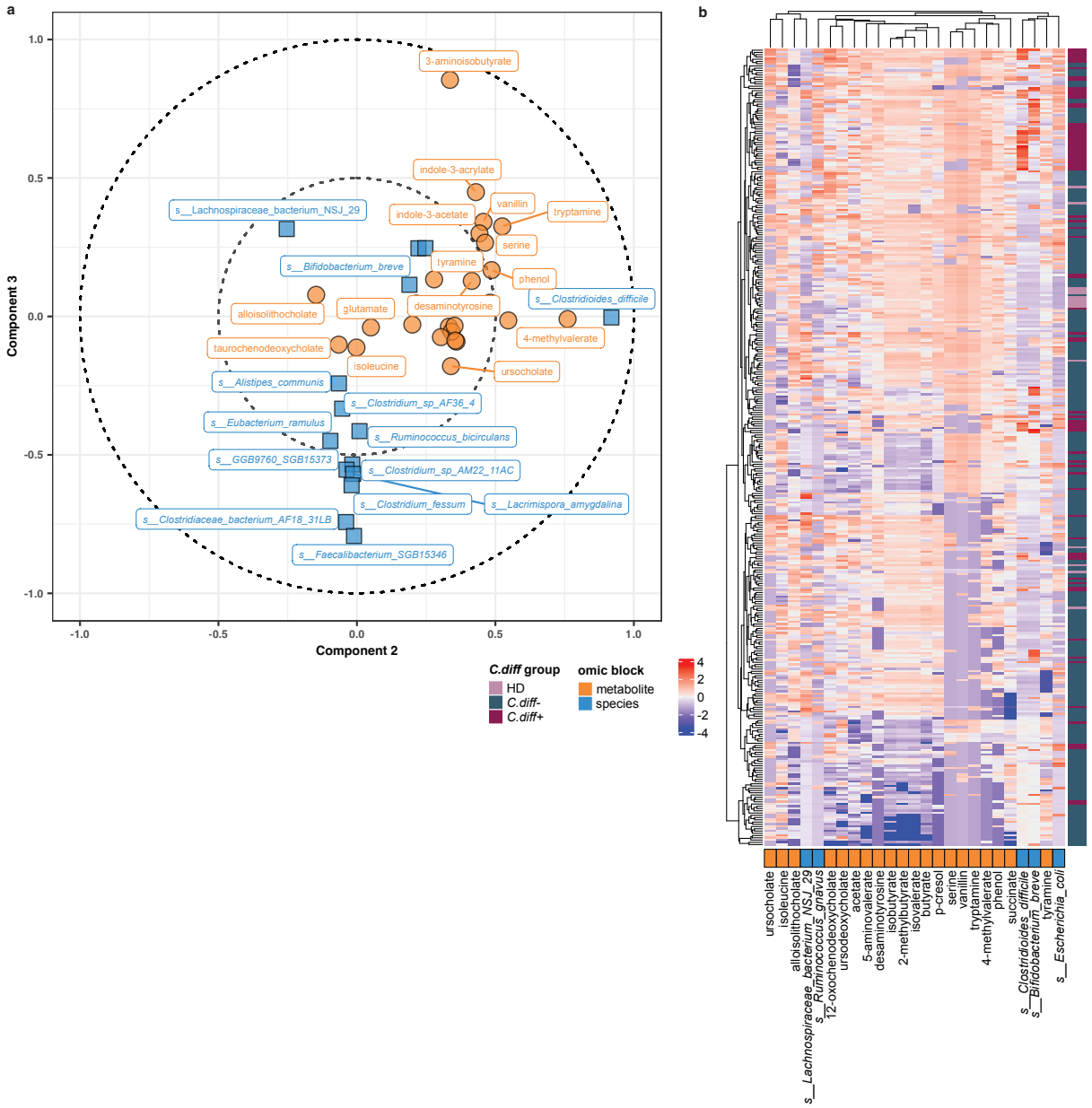

Figure 4.11: Most predictive features from multi-omic integration

(a) Correlation circle plot from multiblock sPLSDA performed on stool samples (n = 505). Coordinates plot the correlation of each feature with the second and third components of each omic block dataset.

(b) Clustered image map (cim) of features loaded on the second latent component of the integrated omics model. Maximal correlations of features to samples provided by the mixOmics::cimDiablo() function.

CHAPTER 5

RESOLVING THE ROLE OF *CLOSTRIDIUM SCINDENS*, IN PROTECTION AGAINST *CLOSTRIDIOIDES DIFFICILE* COLITIS

5.1 Introduction

The clinical presentation of CDI can vary wildly from asymptomatic carriage to diarrhea to life-threatening PMC. The severity of CDAD is strongly associated with gut dysbiosis typical of antibiotic treatment, chemotherapy, and autoimmune conditions. In contrast, a "healthy" microbiota can limit the outgrowth of vegetative *C. difficile* by various mechanisms which include competition, production of antibacterial molecules, and production of secondary metabolites, like SCFAs[212,86,90,213] and secondary bile acids (SBAs)[214].[214] identified the protective role of SBAs, DCA and LCA, produced by the 7α-dehydroxylating gut commensal, *C. scindens*. However, recent studies have challenged the role of SBAs production as a primary driver in the *C. scindens*-dependent protection against CDAD[199,215]. We sought to investigate the bile acid-independent mechanisms and bile acid-dependent mechanisms of protection against CDI by *C. scindens* and other commensals *in vitro* and *in vivo*.

Microbiota replacement therapies, particularly FMT, have been effective at preventing the recurrence of CDI[10] and are available through recent FDA-approved treatments (fig. 1.1). However, cautious and rational probiotic design is required to ensure the safety and efficacy for clinical use[194] which requires a comprehensive understanding of the complex mechanisms driving microbial interactions in the gut. The probiotic use of *Clostridia spp.* is promising due to salutary, health-promoting, effects mediated by the production of secondary metabolites, nutrient niche competition, and strengthening effects on the intestinal barrier[216]. We found that strain-specific differences in *C. difficile* and *C. scindens* mediate the efficiency of

bile acid-dependent and bile acid-independent mechanisms of protection against CDAD.

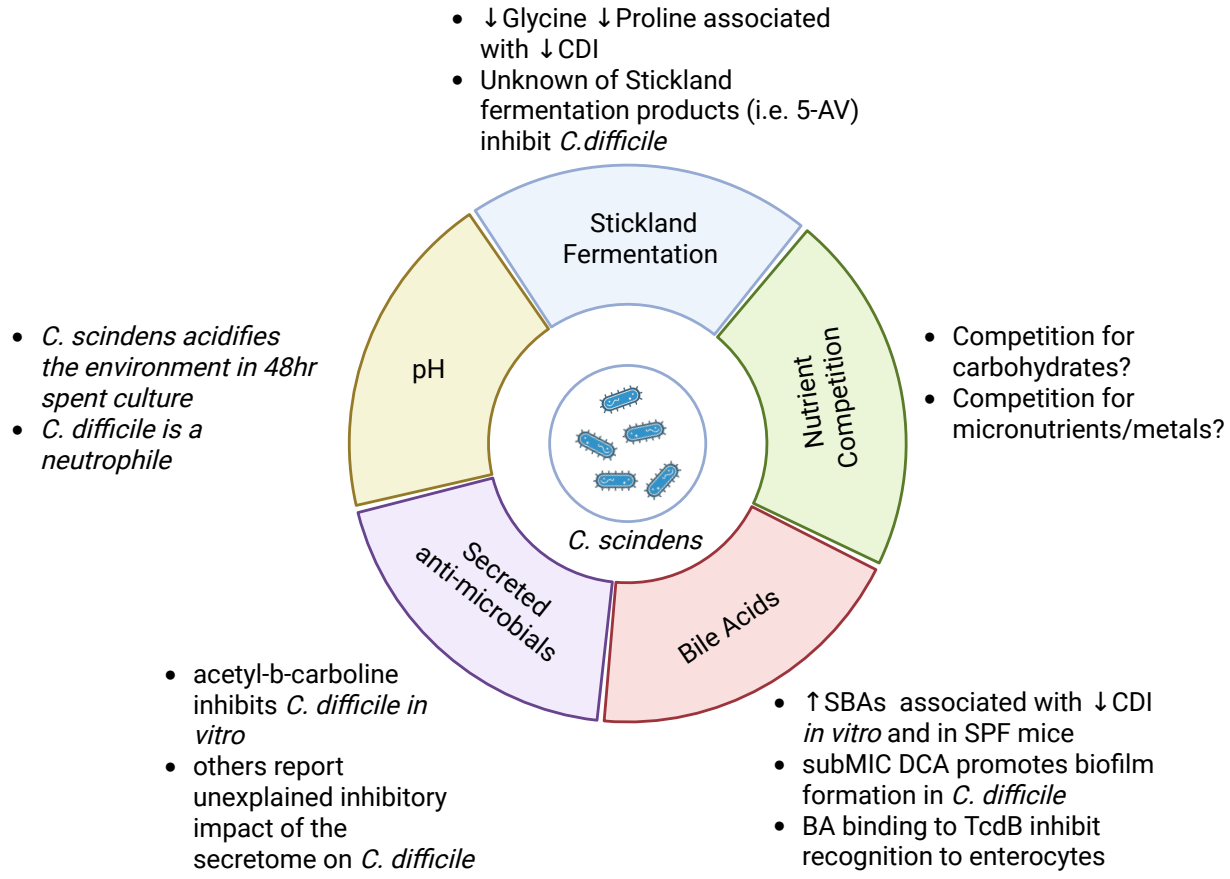

Figure 5.1: Mechanisms driving *C. scindens*-mediated host protection against symptomatic CDI.

5.2 Results

5.2.1 Physiological and inhibitory concentrations of bile acids

Naturally, mice and humans have notable differences in bile acid physiology therefore murine models are imperfect approximations to study human bile interactions (fig. 1.4). Notably, rodents produce (3,6,7-hydroxy) murine bile acid (MBA) using CYP2C70 such as α-muricholic acid, β-muricholic acid, and ω-muricholic acid[217]. In addition, rodent CYP2A12 activity can revert SBAs back into PBAs[218]. The fecal bile acids of specific pathogen free

(SPF) wild type (WT) B6 mice have high levels of MBAs and a higher TPBA/GPBA ratio compared to healthy HDs ($n = 21$). Notably, the fecal bile acid profile of germ free (GF) mice ($n = 11$) is dominated by TPBAs (fig. 5.2a). In this study, we are using a gnotobiotic mouse model using the WT B6 background due to the lack of GF humanized mouse lines[199,76].

5.2.2 C. scindens secondary bile acid production in vitro

C. scindens is the canonical model system for the production of SBAs (DCA and LCA) in part due to its high 7α-dehydroxylation activity[219]. Previous studies reported strain-based differences in bile-dependent inhibition of *C. difficile* by *C. scindens in vitro*[34]. To validate potential strain differences in SBA production, we measured the growth of *C. difficile* overtime in *C. scindens* CFS cultured across a concentration gradient of CA (fig. 5.3). Using quantitative LC-MS, we observed that *C. scindens* MSK.5.24 had a higher concentration of DCA at 600 uM CA (fig. 5.3a). In addition, the vegetative growth of *C. difficile* R20291 was inhibited when grown in the 24 hr *C. scindens* CFS with 300-610 uM CA and nearly complete inhibited at 1.2 mM CA (fig. 5.3b). In addition, exogenous CA was not shown to be inhibitory against *C. difficile* growth below 2.5 mM.

5.2.3 4-member bacterial consortia is sufficient for reducing CDAD in vivo

The production of SBAs is a coordinated effort of integrating the activity of several microbial enzymes (fig. 1.5). Earlier work in the lab identified 4 gut species associated with host protection in human and ABX murine gut including *C. scindens*, *P. capsillosus*, *B. hansenii*, and *D. freteri* (formerly *Barnesiella intestihominis*)[214]. This 4-member consortia (B4mix) demonstrated enhanced host protection against CDAD than *C. scindens* alone but the mechanism driving this protection was undefined. To test if the B4mix was sufficient to ameliorate CDAD, we compared the disease outcomes of gnotobiotic mice colonized with B4mix or PBS (fig. 5.4a). We did not observe a difference in pathogen burden but there was a significant

difference in overall survival and body weight loss (fig. 5.4b-5.4d).

SBA epimers such as 3-oxoLCA and alloisoLCA have been reported to modify colonic Th17s and Tregs[220,221,222,223]. AlloisoLCA was also a feature negatively associated with *C. difficile* detected by the sPLSDA model. In addition, these SBA epimers were 10-100x more potent inhibitory activity than DCA (fig. 5.2b). A preliminary blastp of bile acid modification genes identified a putative *bsh* in *D. freteri* and 3α-HSDH (baiA2) in *P. capsillosus* which may produce 3oxoLCA (not shown). We hypothesized that the B4mix can produce 3oxoLCA but we did not detect a significant amount of 3oxoLCA in the feces (fig. 5.4e). This could be attributable to the higher level of DCA, which has been suggested to negatively regulate murine HSDH activity[224].

As expected, B4mix colonized mice also higher average DCA concentrations (mean = 213.87 uM, SE = 29.62 uM, median = 266.28 uM) but lower LCA (mean = 6.21 uM, SE = 0.98 uM, median = 6.3 uM) concentrations (fig. 5.4e). These concentrations show that a minimal 4 member consortia can achieve the same level of DCA production as a rich SPF microbiome (mean = 221.54 uM, median = 222.227 uM, IQR = 40.757 uM) and approaching SPF levels of LCA (mean =36.8 uM, median = 32.83 uM, IQR = 11.2)(fig. 5.2a). The inhibitory effect of SBAs on *C. difficile* is concentration-dependent but some suggest that physiological concentrations of SBAs are too low to reasonably expect *C. difficile* inhibition[214,199]. Our measured minimum inhibitory concentration (MIC) for DCA in BHIS was relatively high at 1.2 mM, which is comparable to MIC measurements from other studies[199,225]. This suggests that the production of DCA alone is not sufficient to reach the MIC as seen with consistent pathogen burden between PBS and B4mix mice (fig. 5.4b). However, the mean and median fecal DCA concentration (mean = 1.54 mM, median = 903.89 uM, IQR = 451.98 uM) in human fecal samples exceeded the MIC value but displayed high variability. In addition, we observe that the median and mean concentration of 3-oxoLCA (mean = 102.71 uM, median = 64.50 uM, IQR = 28.51 uM) and alloisoLCA (mean = 29.10

uM, median = 12.22 uM, IQR = 9.10475 uM) to be above the measured MIC of *C. difficile* and 10-100x more inhibitory than DCA (fig. 5.2b). LCA is highly insoluble so an accurate MIC measurement could not be measured. In addition, we found that CA was not inhibitory within our tested concentrations but others report a high MIC at 10mM. This suggests that a healthy human gut may have inhibitory levels of SBAs likely depending on various factors including microbiota composition and diet. Future studies using humanized murine models for bile acids can ascertain if the B4mix can reproduce human levels of SBAs and if it is sufficient the reducing the burden of *C. difficile*[226,227].

In addition, sub-inhibitory concentrations of SBAs have been reported to have notable impacts on *C. difficile* virulence, which may be possible bile acid-oriented explanations for the increased survival we observed in the B4mix colonized mice (fig. 5.4c). Sub-inhibitory levels of DCA were reported to promote biofilm formation in *C. difficile*, which may contribute to enhanced pathogen persistence in the gut[82,228]. Alternatively, sub-inhibitory levels of LCA, isolithocholic acid (isoLCA), and alloisoLCA reduces *C. difficile tcdA* expression and exerts cytotoxic stress on *C. difficile* while sparing gram-negative commensals[229]. Combined, this adds another layer of complexity regarding the physiological importance of sub-inhibitory concentrations of SBAs for CDI host-pathogen dynamics. Future studies are required to untangle these sub-inhibitory effects with other potential causes of *C. difficile* inhibition by the bacterial constituents of the B4mix.

5.2.4 C. scindens strain-dependent impact of nutrient competition in vitro

C. difficile is auxotrophic for several amino acids including proline, cysteine, leucine, isoleucine, tryptophan, and valine[101]. Since *C. scindens* strains have strain-level differences in their amino acid utilization, we sought to investigate the impact of nutrient competition on *C. difficile* vegetative growth *in vitro*. We compared the 24 hr growth of *C. difficile* in 24 hr CFS with the growth in 4 parts of 24 hr *C. scindens* CFS supplemented with 1 part

fresh media (80% CFS)(fig. 5.5a, Section 2). We tested 3 *C. scindens* strains in addition to other *Clostridial spp.*: *C. symbiosum* MSK.7.21 (Stickland Fermentation$^-$ *baiE*$^-$) and *C. sporogenes* ATCC15579 (Stickland Fermentation$^+$ *baiE*$^-$).

We observed that the RIs of R20291 and *C. difficile* VPI10463 were significantly rescued for the 80% CFS, indicating a greater impact of nutrient competition in *C. difficile* inhibition by *C. scindens* VPI12708 than *C. scindens* ATCC35704. In contrast, growth in *C. scindens* ATCC35704 and *C. scindens* MSK.5.24 CFS was not significantly rescued with media replenishment (fig. 5.5b). In addition, we observe differences in RI between R20291 and *C. difficile* VPI10463 when grown in *C. symbiosum* MSK.7.21 and *C. sporogenes* ATCC15579. However, these results reflect the metabolism and growth rates in BHIs and may not be broadly applicable to other media or in the gut lumen as nutrient-dependent inhibition impact of *C. scindens* was shown to be ablated in different carbohydrate environments[230].

Although the importance of amino acid availability is generally well known, the direct impact of Stickland fermentation byproducts on *C. difficile* dynamics is less clear. We found that 5-AV does not inhibit *C. difficile* vegetative growth except at incredibly supraphysiological concentrations (1 M)(fig. S15).

5.2.5 C. scindens monocolonization is not sufficient to protect against C. difficile in vivo

A study from 2021 suggested that monocolonization of Stickland fermenting species, such as *C. scindens*, *Clostridium hiranonis (C. hiranonis)*, *Clostridium leptum (C. leptum)*, and *Paraclostridium bifermentans (P. bifermentans)*, was sufficient to improve disease outcomes of CDI with *C. difficile* VPI10463 due to competition for amino acids such as proline and glycine[199]. In addition, we observed strain-level differences in *C. scindens* and *C. difficile* contributed to the variance in the *C. difficile* inhibition *in vitro* (fig. 5.5b). The combination of host factors and scarcity of free amino acids in the gut lumen may additionally change

these dynamics.

To validate test if different strains of *C. scindens* are sufficient for protection against CDI, we monocolonized GF B6 mice with *C. scindens* MSK.5.24 or *C. scindens* ATCC35704 7 days prior to infection with 200 spores of *C. difficile* VPI10463 (fig. 5.6a). We did not find a significant improvement in pathogen burden, survival, or weight loss for *C. scindens* MSK.5.24 monocolonized mice (fig. 5.6b-5.6d, fig. 5.7a-5.7c). The depletion of proline and production of 5-AV in *C. scindens* MSK.5.24 mice confirmed that Stickland fermentation was active in these mice (fig. 5.6e, fig. 5.7d). As expected, these mice were unable to produce SBAs due to the lack of bile salt hydrolase (BSH) encoding bacteria (fig. 5.6f) These results suggest that competition for proline and glycine was not sufficient for inhibition of *C. difficile in vivo*.

5.2.6 Diversity in metabolic fitness within ST1 strains

Metabolic phenotypes can vary across different *C. difficile* ribotypes[35] and commensal backgrounds[230], as we observed in fig. 5.5b. In Dong et al. 2025, we found this to be true beyond the ST level. Colonization of avirulent strains was shown to be protective against toxigenic CDI, as seen with avirulent ST1-75 (*tcdB*$^+$ clade 2 ST1) against R20291 (*tcdB*$^+$ clade 2, ST1)[28,25,151]. ST1-75 encodes an in-frame 69 bp deletion of *cdtR*, which regulates the production of the binary toxin (CDT) and the expression of the PaLoc[28,25]. Despite this, ST1-75 and R20291 are very closely related strains (ANI $> 99.8\%$). Notably, ST1-75 was able to confer protection at a 1:1 ratio, a stark contrast to prior studies only observed similar protection at higher administered doses of nonvirulent strains[231,232].

Surprisingly, the competition advantage of ST1-75 seen *in vivo* was not recapitulated *in vitro* using rich BHIs media (fig. 5.8a) but was observed in the context of filtered GF cecal contents, a low nutrient environment(fig. 5.8b). In contrast, the CRISPR-edited CdtR mutant R20291 strain (R_cdtR$*$), did not outcompete R20291 in either the rich media

or cecal extract conditions, indicating that the loss of CdtR alone was not sufficient to explain this phenotype (fig. 5.8a, fig. 5.8b). Comparing the rates of nutrient consumption from filtered GF cecal contents, ST1-75 strains were shown to consume glycine, alanine, valine, methionine, isoleucine, and phenylalanine at a faster rate than WT R20291 or in the R_cdtR* mutant (fig. 5.8c, fig. 5.9a). In contrast, ST1-75 and R20291 were equally efficient at consuming preferred Stickland substrates (proline, leucine, and cysteine) (fig. 5.9b) and equally did not consume aspartate, glutamate, and lysine (fig. 5.9c).

Glycine is a common Stickland oxidant[97] and consumption is mediated by glycine reductase genes (*grdABC*). The expression of glycine reductases is regulated by the availability of other amino acids, notably proline, by the PrdR regulator[100]. While proline is the preferred electron sink for *C. difficile*, time tracking of *C. difficile* metabolism revealed that glycine consumption is upregulated during stationary phase[99]. Alanine is preferentially oxidized[99] but phenylalanine, can be oxidized and reduced throughout the entire growth curve for *C. difficile*, by co-opting reductases used for fermenting leucine (*hadABC*, *ldhA*, and *acdB*)[99]. However, phenylalanine, like other aromatic amino acids, is minimally metabolized by stationary phase[99]. Methionine is reduced into 4-(methylthio)butyrate. Isoleucine and valine are BCAAs that can act as Stickland reductants which generate 2-methylbutyrate and isobutyrate respectively[233]. Supplementing the cecal media with a cocktail of 18 amino acids known to be essential for optimal *C. difficile* growth[101] ablated the competitive advantage of ST1-75 over R20291 (fig. 5.8e).

In accordance with other studies[98,234], our patient samples typically had higher (but not significant) levels of essential amino acids (glycine and proline) compared to HDs (fig. S12a), correlating to the loss of amino acid competing species in antibiotic exposed patients (fig. 4.10b). The variability of nutrient consumption between closely related $tcdB^+$ *C. difficile* strains may explain differences in CDI outcomes across patients.

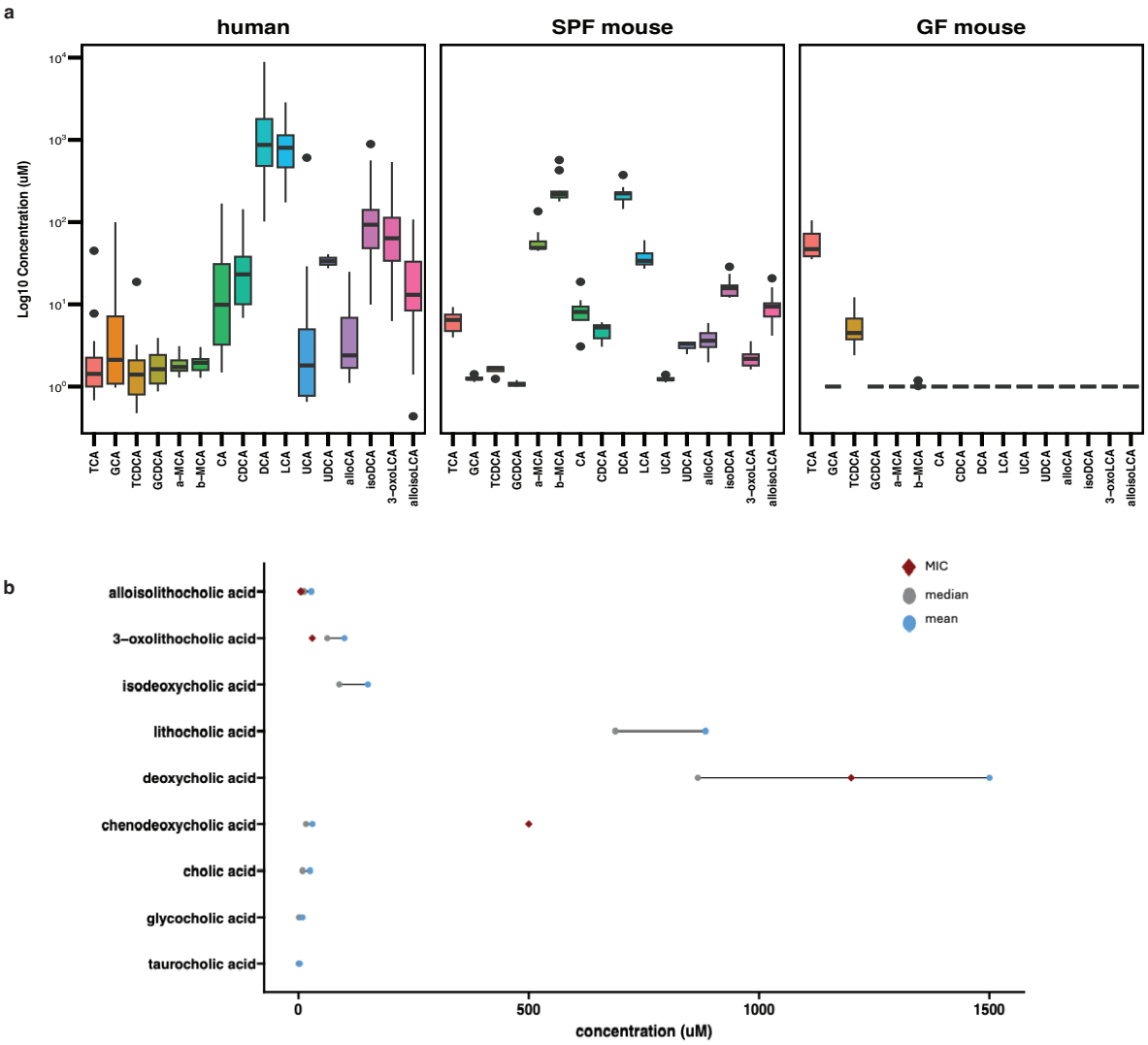

Figure 5.2: Comparisons of physiological bile acid compositions

(a) Physiological bile acid compositions across HDs fecal samples (n = 21), SPF mice (n = 9), and GF mice (n = 25).

(b) Comparison of the minimum inhibitory concentration (MIC) of bile acids on *C. difficile* R20291 with the mean and median concentrations (uM) in HD samples.

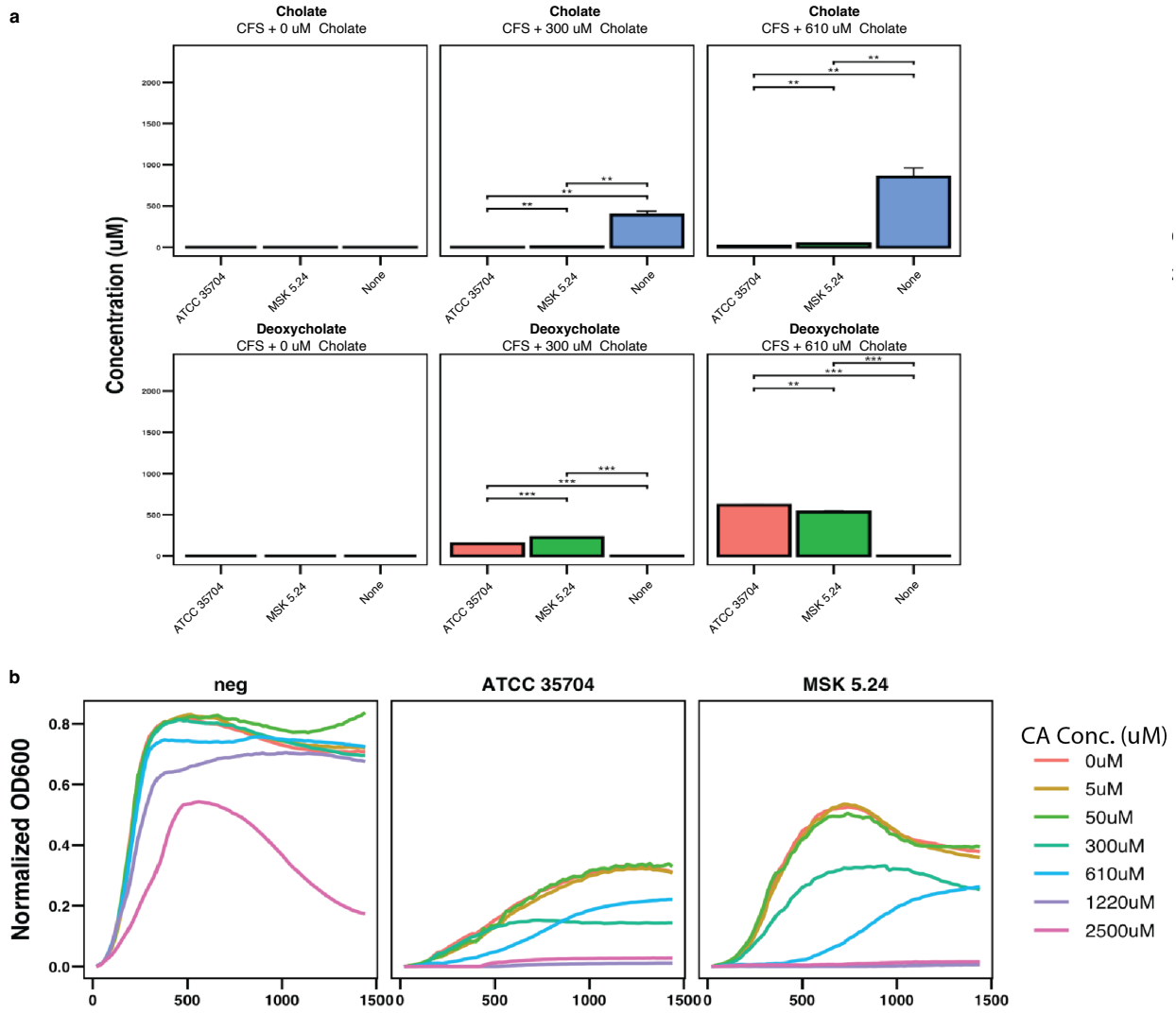

Figure 5.3: Cholate above 300 uM in *C. scindens* CFS result in *C. difficile* vegetative growth inhibition

(a) Concentration of bile acids produced by *C. scindens* strains in BHIS media supplemented with exogenous CA (0-610uM).

(b) Vegetative growth of *C. difficile* R20291 in CFA from *C. scindens* grown in the presence of exogenonous cholate for 24 hours.

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests. P values for all statistical tests were adjusted for multiple comparisons using the Benjamini-Hochberg procedure and represented as follows: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

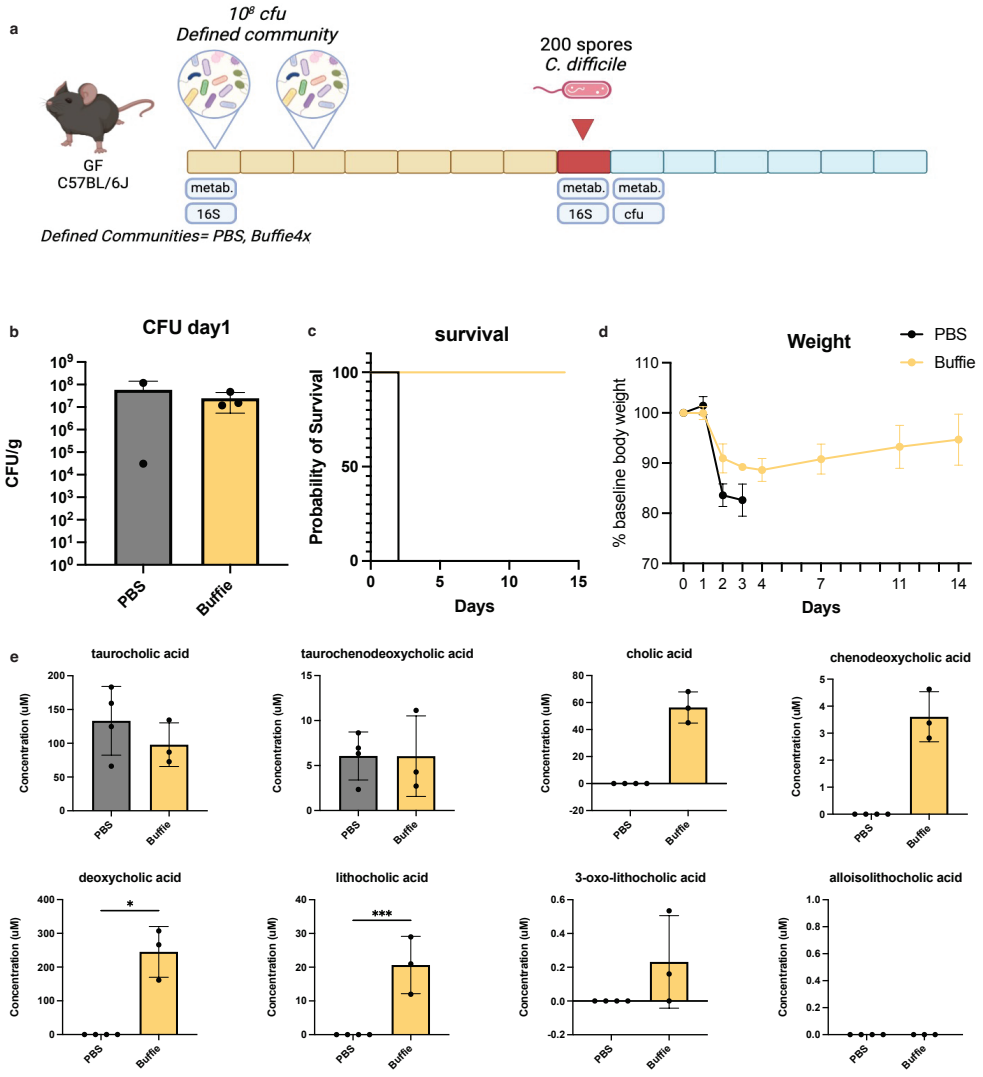

Figure 5.4: Preliminary differences between protective and nonprotective gut metabolite profiles

(a) GF B6 mice were inoculated with total inoculum of 10^8 cfu of a minimal consortium (B4mix) or PBS 7 days prior to infection (d-7) via oral gavage. Mice were infected with 200 spores of *C. difficile* VPI10463 (d0) and monitored for disease severity for 7 days. Fecal samples were taken for metabolomics and *16S rRNA sequencing* on day -7, 0, and 1.

(b) Quantified *C. difficile* cfu per gram of feces compared between PBS or B4mix colonized mice.

(c) Survival of mice post-*C. difficile* infection.

(d) Body weight loss (%) between PBS or B4mix colonized mice post-*C. difficile* infection.

(e) Fecal concentration (ug/mL) of bile acids between PBS or B4mix colonized mice on 1 days post-infection (dpi).

Experiment designed and analyzed by S.S.S and Q.D. Data collected by Q.D.

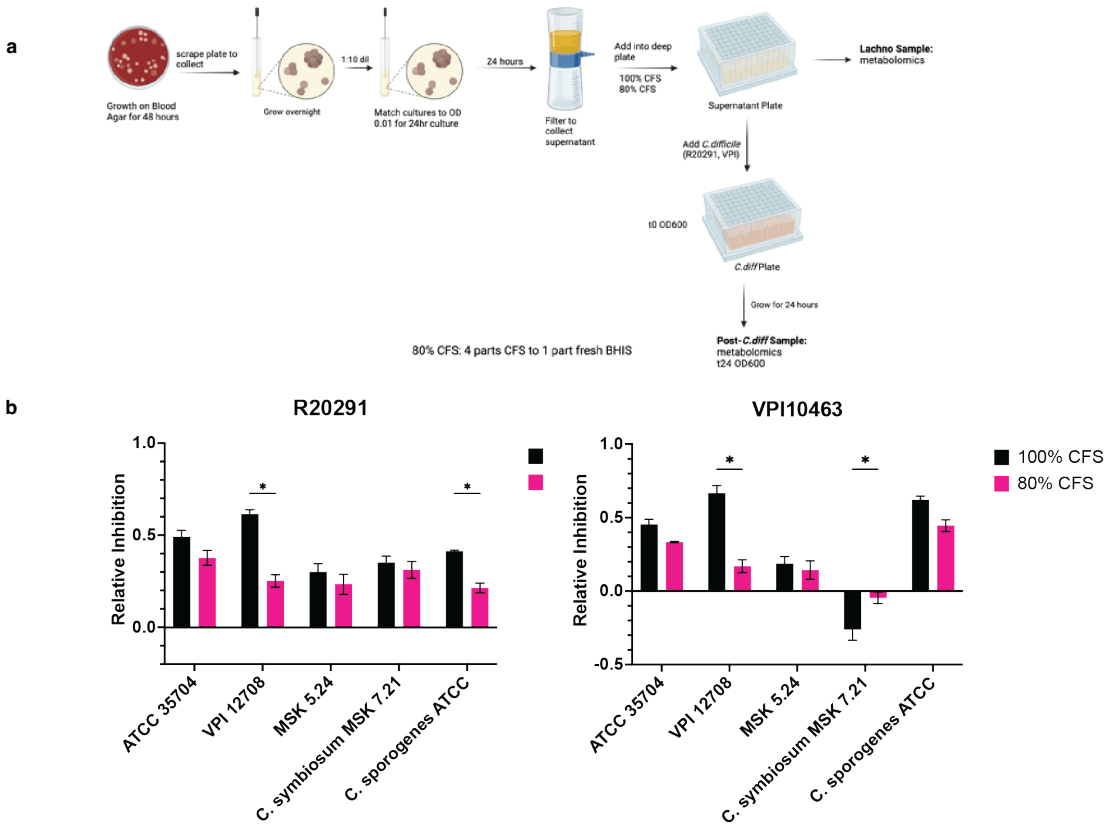

Figure 5.5: *C. scindens* strain-dependent differences in *C. difficile* growth inhibition *in vitro*

(a) Triplicate colonies from two *C. difficile* strains (R20291 and VPI10463) were isolated and cultured in 100% *C. scindens*-derived CFS or 80% CFS from BHIs media.

(b) RI of *C. difficile* when cultured in different *C. scindens*-derived CFS.

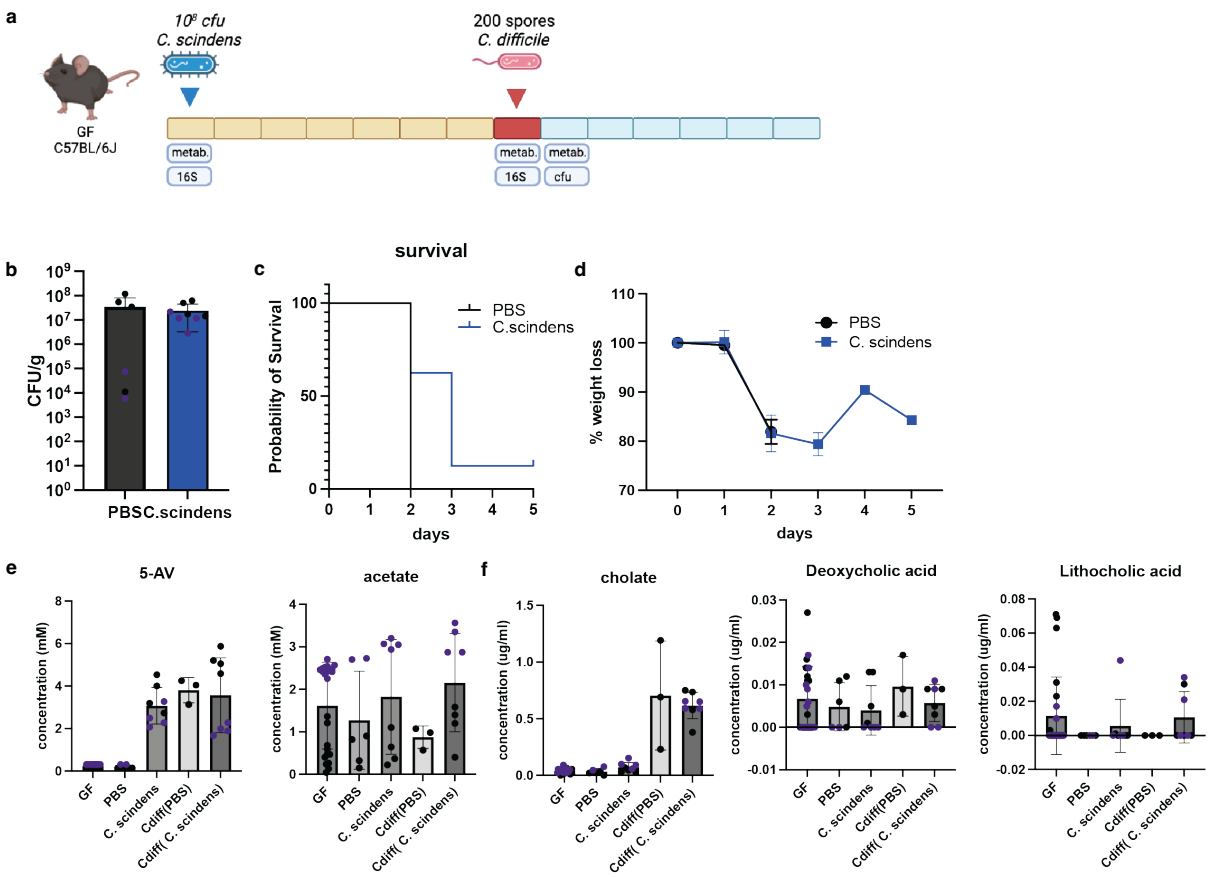

Figure 5.6: *C. scindens* MSK.5.24 is insufficient to protect against *C. difficile in vivo*

(a) GF B6 mice were inoculated with 10^8 cfu of *C. scindens* MSK.5.24 or PBS 7 days prior to infection (d-7). Mice were infected with 200 spores of *C. difficile* VPI10463 (d0) and monitored for disease severity for 7 days. Fecal samples were taken for metabolomics and *16S rRNA sequencing* on day -7, 0, and 1 via oral gavage.

(b) Quantified *C. difficile* cfu per gram of feces compared between PBS or *C. scindens* MSK.5.24 colonized mice.

(c) Survival of mice post-*C. difficile* infection.

(d) Body weight loss (%) between PBS or *C. scindens* MSK.5.24 colonized mice post-*C. difficile* infection.

(e) Fecal concentration (mM) of 5-AV and acetate between PBS or *C. scindens* MSK.5.24 colonized mice on 1 dpi.

(f) Fecal concentration (ug/mL) of CA, DCA, and LCA between PBS or *C. scindens* MSK.5.24 colonized mice on 1 dpi.

Experiment designed and analyzed by S.S.S and Q.D. Data collected by Q.D.

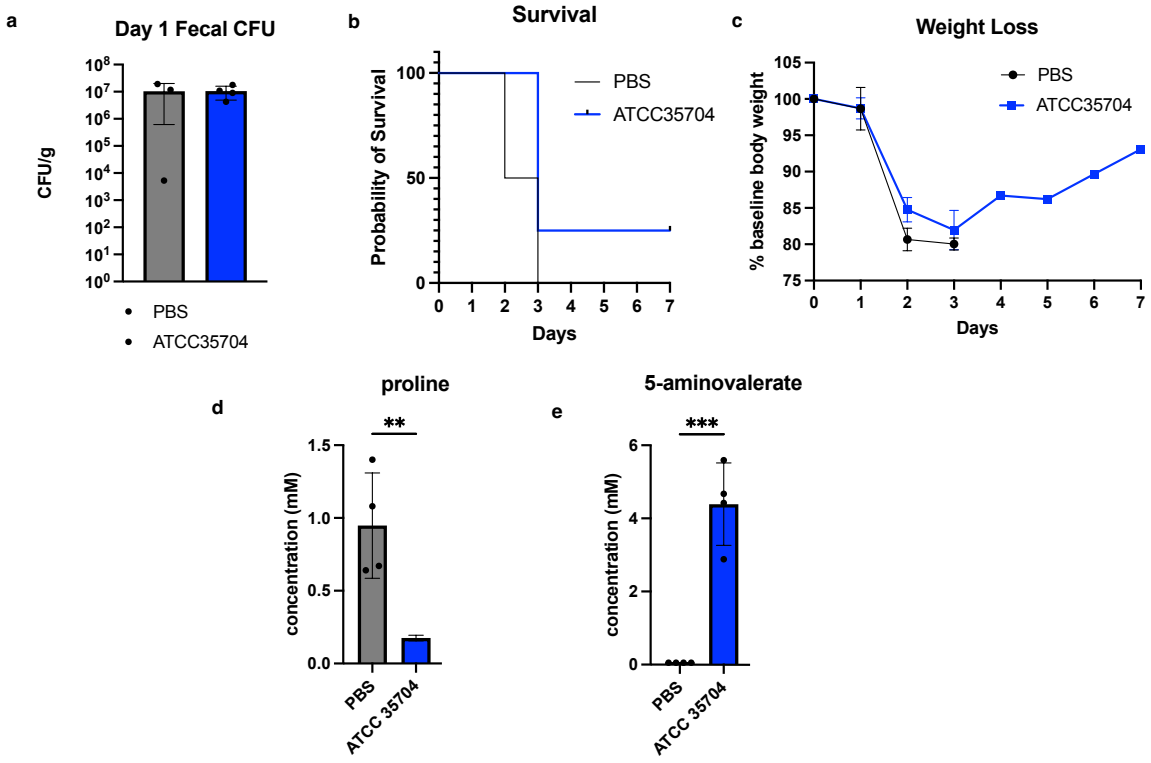

Figure 5.7: *C. scindens* ATCC35704 is insufficient to protect against *C. difficile in vivo*

(a) Quantified *C. difficile* cfu per gram of feces compared between PBS or *C. scindens* ATCC35704 colonized mice.

(b) Survival of between PBS or *C. scindens* ATCC35704 colonized mice post-*C. difficile* infection.

(c) Body weight loss (%) between PBS or *C. scindens* ATCC35704 colonized mice post-*C. difficile* infection.

(d) Fecal concentration of 5-AV and proline (mM) between PBS or *C. scindens* ATCC35704 colonized mice on 1 dpi.

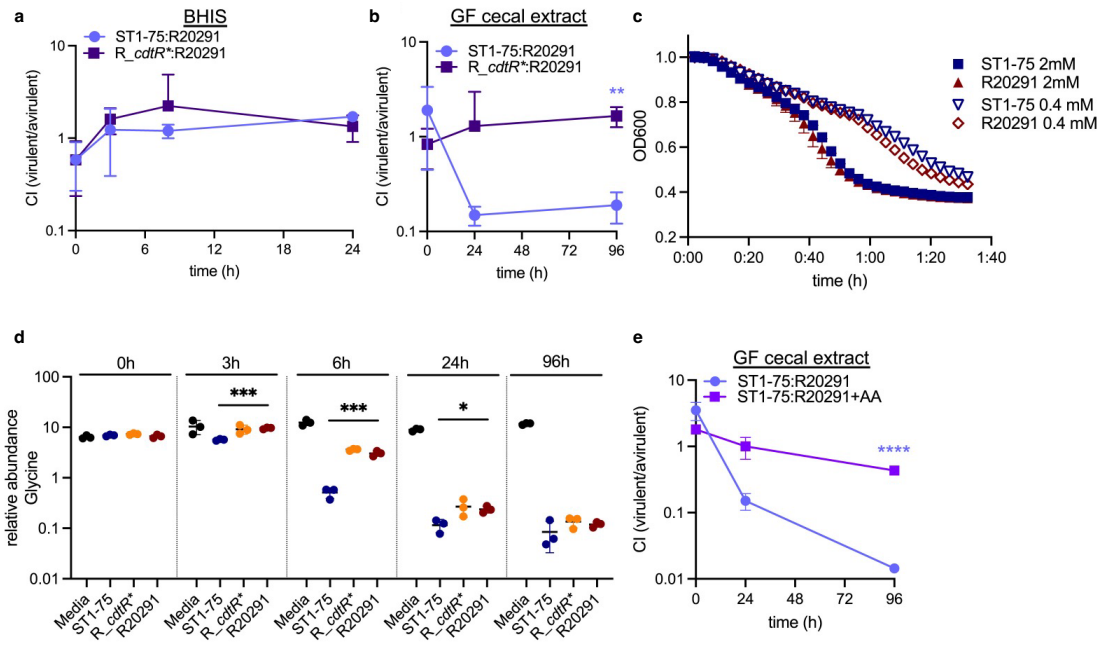

Figure 5.8: ST1-75 protection is attributable to faster amino acids depletion in nutrient-limited environment

(a,b,e) Relative abundance of R20291 in culture was determined by measuring wild-type cdtR copies by qPCR. (a) competitive index (CI) is shown by dividing the RA of R20291 to avirulent strain (ST1-75/R_cdtR∗) in BHIS. (b) CI is shown by dividing the relative abundance of R20291 to avirulent strain (ST1-75/R_cdtR∗) in diluted germ-free cecal extract.

(c) Metabolomic relative abundance of glycine in cecal extracts inoculated with either ST1-75, R_cdtR∗, or R20291.

(d) OD-drop assay measuring germination defects of indicated strains while supplemented with 0.4 and 2 mM glycine.

(e) CI is shown by dividing the RA of R20291 to avirulent strain ST1-75 in diluted germ-free cecal extract (spiked in with or without amino acids). Results represent means SD. Statistical significance was calculated by unpaired t-test or one-way ANOVA, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Experiment designed and data collected by Q.D. and S.S.S. Data analyzed by Q.D. Results available at: https://doi.org/10.1016/j.chom.2024.11.003

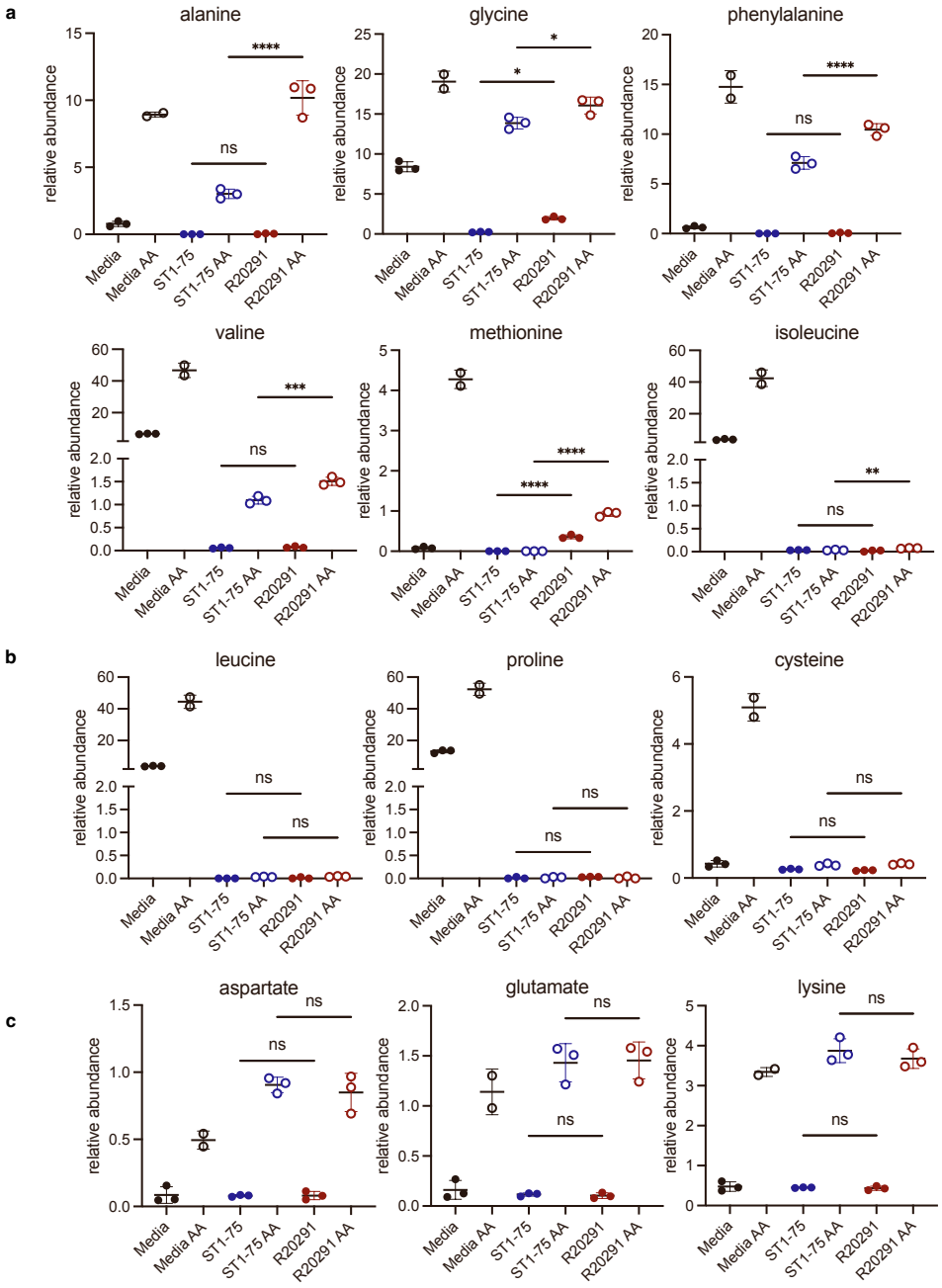

Figure 5.9: ST1-75 depletes specific amino acids faster than R20291

Metabolome analyses of the relative abundance of each amino acid spiked into cecal extracts and 6 hr after inoculation with either ST1-75 or R20291. (a) Amino acids depleted more rapidly by ST1-75 than R20291. (b) Amino acids depleted at similar levels by ST1-75 and R20291. (c) Amino acids that were not depleted by ST1-75 or R20291. Results represent means SD. Statistical significance was calculated by unpaired t-test or one-way ANOVA, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Experiment designed by Q.D. and S.S.S. Data collected by S.S.S. and analyzed by Q.D. Results available at: https://doi.org/10.1016/j.chom.2024.11.003

CHAPTER 6

DISCUSSION

6.1 Applications for strain-level metagenomic tracking in the clinic

The clinical outlook of CDAD is highly variable, ranging from asymptomatic colonization to potentially life-threatening colitis and toxic megacolon. The severity of CDAD has been attributed to a multitude of factors, including pathogen virulence, immune deficiency, host age, and the diversity of the gut microbiota. Despite this, integration of clinical metagenomics has been slow, in part, due to the informatics challenge of interpreting large and rich datasets. Accurate strain-level tracking is essential for deploying effective infection control strategies. Typing strategies for *C. difficile* strains have evolved in the past several decades from ribotyping to MLST to WGS and have revealed a remarkable diversity of clades and strains causing CDAD and this species' unusually large pan-genome. As the cost of sequencing dropped, WGS has quickly gained traction for genomic surveillance and transmission tracking of *C. difficile* in healthcare settings[178,235,59,236,60,175,167,237], and several studies have demonstrated potential cost savings of incorporating WGS surveillance in healthcare facilities[238,239,240,241]. However, these traditional approaches for strain tracking, to date, require the cultivation of *C. difficile* from fecal samples, a tedious and time-consuming process that is not routinely performed in conventional clinical microbiology laboratories.

In chapter 3, we optimized tools for the accurate detection and genomic relatedness of *C. difficile* strains using shotgun-sequenced metagenomes from human fecal samples. In contrast, marker gene-based tools, such as MetaPhlAn4, assess only a fraction of the genetic information in metagenomic samples, reducing sensitivity. This presents a particular challenge for low-abundance species, such as many clinically relevant enteric pathogens. We addressed this challenge with our local sensitivity adjustment and marker gene curation of MetaPhlAn4 allowed for the improved detection accuracy of *C. difficile* from patient sam-

ples without significantly impacting the false positive rate of the overall community. We optimized the use of contemporary metagenomic profiling tools specifically for *C. difficile*, the methods presented in this study can be adapted to enhance the sensitivity for other low-abundance pathogens.

Culture-independent, metagenomic sequencing approaches for bacterial strain determination have been used to measure strain dynamics[242,168,243], track antimicrobial resistance[244], and identify unknown infectious agents[245]. However, the high genome coverage required by most contemporary strain-tracking tools[170], including StrainPhlAn4, has been a barrier to clinical metagenotyping. We found that parameter optimization was necessary to feasibly detect clinically relevant but low-abundance pathogens from complex environments like *C. difficile*. In addition, the rapidly expanding size of reference databases is known to significantly impact the profiling performance of metagenome profiling tools[246,247,248,249,250]. Indeed, we found that the inclusion of evolutionarily distinct environmental cryptic clades[251,117] drastically reduced the number of species-specific marker genes for accurate classification within the genus *Clostridioides*. A custom database including *Clostridioides* genomes associated with human disease, we saw significantly improved accuracy of strain detection by StrainPhlAn4. Thus, careful curation of reference databases and parameter optimization enhances the sensitivity for strain detection. Collectively, these interventions improved the limit of StrainPhlAn4's detection of *C. difficile* at 1x genome coverage, and greater sequencing depth and pathogen abundance can further increase sensitivity[252,253,254]. Samples successfully detected by StrainPhlAn4 clustered accurately and reliably beyond ST level resolution and correlated with genome-wide ANI estimations.

A strength of our study is the pairing of fecal NAAT and microbiome metagenomes with *C. difficile* isolate genome sequences from a panel of patients. Although the application of StrainPhlAn4, after optimization of analytical parameters, demonstrated the feasibility of identifying strains in respiratory samples with high quantities of host DNA[243], our study is

the first to demonstrate strain-level detection of the low-abundance gut pathogen *C. difficile* in a complex fecal microbiome. However, our approach has some limitations. Although the diversity of *C. difficile* isolates in our collection is comparable to the distribution of STs nationally[255], this study is confined to a single hospital in a high-income country, so the results are not broadly generalizable to all settings. Second, inclusion into the study cohort did not consider the diagnosis or symptoms of CDI, therefore, the results do not provide insight into the factors driving the development of CDAD. In addition, the bacterial culture isolation protocol resulted in a single sequenced isolate per fecal sample. Similarly, StrainPhlAn4 reports a single dominant strain per metagenomic sample, so our results potentially overlook instances of co-colonization with different strains.

The clinical utility of any molecular assay is determined by its potential to influence decision-making by the healthcare team. Recent studies demonstrate that in healthcare environments with tight adherence to infection control interventions, such as ICUs[59] and clinical wards serving leukemia and hematopoietic stem cell transplant patients[60,61], *C. difficile* is rarely newly acquired. It is likely, however, that in other healthcare settings, where patients are mobile and contact precautions are not in place, transmission is substantially higher. The rapidity with which shotgun metagenomic sequencing can be performed and analyzed, detection and strain typing of *C. difficile* can inform clinicians and healthcare epidemiologists of potential outbreaks of specific strains, as occurred two decades ago with the introduction of the clade 2, ST-1 strain, better known as the BI/NAP1/027 so-called hyper-virulent strain. An additional benefit of fecal metagenomic sequencing is the characterization of the patient's microbiome, which can increasingly provide insights into the patient's degree of dysbiosis and associated disease susceptibilities, including the potential for CDAD recurrence. A weakness of the metagenomic sequencing approach for diagnosis of *C. difficile* infection is its inability to distinguish toxin-negative from toxin-encoding strains. Therefore, NAAT diagnostics that test for the presence of TcdB continue to have a role in

rapidly and accurately detecting toxigenic *C. difficile*. However, the ability to accurately strain track enables prospective epidemiology and investigations into pathogen dynamics at a fine resolution without the added time and cost of culturing *C. difficile* in the laboratory. To become clinically useful, fecal shotgun metagenomic sequencing will require procedural standardization to reduce turnaround times and costs while improving sensitivity[142]. Our study emphasizes the importance of targeted benchmarking for specific pathogens such as *C. difficile* and suggests that with improvements in molecular techniques, computing power, and application of machine learning platforms, shotgun metagenomic sequencing has the potential to enable tracking of pathogens in close to real-time.

6.2 Multi-omic perspective on *C. difficile*

By leveraging mixed effects regressions and integrative omic analysis, we identified several significant microbial and metabolic features correlated with the presence of *C. difficile* in a retrospective analysis across 4 distinct patient backgrounds and healthy controls. In particular, the metabolites associated with amino acid fermentation were the most discriminatory metabolites for *C. difficile* colonization. Notably, alloisoLCA, a particularly inhibitory SBA[76], was negatively associated with *C. difficile*$^+$ samples and our HD population reports mean and median alloisoLCA levels above the MIC for *C. difficile*. To my knowledge, our study is the second to identify alloisoLCA in the human gut, which has otherwise only been described in Japanese centenarians[76].

However, our study has several limitations. One notable limitation of our study is the scope of our targeted metabolomic panel, which primarily focused on bile acids and SCFAs, metabolites that have largely been associated with CDI and other gastrointestinal diseases. Other multi-omic investigations highlighted the same key biomarkers discriminating CDI, 5-AV, and 4-methylvalerate, but go on to highlight carbohydrates negatively associated with CDI, fructose, rhamnose, and hydroxyproline[126]. In addition, this retrospective analysis

included patients enrolled in other previously reported studies[110,108,109,111] so sampling times often did not coincide with *C. difficile* diagnoses, detection assays, or treatments. The variability in the sample collection methodology likely contributed to additional noise in our models. And as a result, our results are limited to the presence of the *C. difficile* and are not appropriate to draw conclusions about disease severity, recurrence, or other clinical manifestations of CDI.

6.3 Sub-inhibitory SBA production protects against severe CDI

Colonization with our minimal 4-member consortia increased the overall survival against virulent *C. difficile*. However, our murine model (B4mix) was able to restore physiological, yet sub-inhibitory levels of DCA and LCA in GF mice. These results are not directly attributable to the competition for proline and glycine alone, as monocolonization with *C. scindens* strains was ineffective at protecting against CDI. However, we cannot rule out the role of the collective nutrient niche overlap imposed by the entire consortia onto *C. difficile*. In addition, we cannot rule out the role of nutrient competition for metabolites that were not targeted within our panels, such as carbohydrates. Our investigations were limited due to the lack of available genetic tools for *Clostdria spp.*.

Potentially, sub-inhibitory SBAs may contribute sufficient bile stress to induce metabolic reprogramming in *C. difficile*, including the production of isocaproate (4-methylvalerate), a by-product of reductive lysine metabolism. *C. difficile* encodes 2-hydroxyisocaptoyl-CoA dehydratase (HadBC), a highly oxygen-sensitive iron-sulfur cluster protein facilitating the reversible dehydration of (R)-2-hydroxyisocaproyl-CoA to isocaprenoyl-CoA[256]. Typically, proline is considered to be the preferred substrate for Stickland fermentation; however, it was suggested that growth of *C. difficile* in a bile acid stress environment promotes a metabolic shift towards leucine catabolism[80]. In addition, subinhibitory levels of SBAs have been shown to induce biofilm formation[82] and a reduction in flagellation/motility[81,80].

6.4 Strain-specific metabolomic protection against CDI

High genomic plasticity has contributed to the high metabolic adaptability of *C. difficile*[39,36,257,230]. Our investigations into *C. scindens* explored the variability of nutrient competition as the primary driver of *C. difficile* inhibition *in vitro* between various *C. scindens* strains. We also observed strain-level differences, suggesting that the contribution of amino acid competition varied not only between *C. scindens* strains but also between *C. difficile* strains. This was explored further in intrastrain level variability in metabolism, promoting ST1-75 driven protection against a competing toxigenic strain, R20291. Although closely related, ST1-75 consumed more second-string Stickland substrates (glycine, alanine, valine, methionine, isoleucine, and phenylalanine) from a low nutrient environment than R20291 after 6 hours. Proline, leucine, and cysteine were not differentially consumed because these amino acids were completely depleted by both strains. However, these results cannot rule out the possibility that these preferred amino acids were also more rapidly consumed during a shorter time scale ($<$ 6 hrs).

Toxin production from the PaLoc is triggered in hostile, low-nutrient environments, sensed and regulated by CcpA, CodY, and sigma factor H[233]. It is also known that the expression of *prd* is essential for *C. difficile* virulence[258,215] and some studies suggest that *C. difficile* disrupts barrier integrity to scavenge released hydroxyproline, a proline precursor stored in collagen[98,259]. Despite this, low toxin-producing avirulent strains such as ST1-75 had a competitive advantage, possibly due to the metabolic expense of toxin production. The competition between ST1-75 and R20291 was a particularly notable example of nutritional niche variability due to the close genomic relatedness of the strains. Since the R20291 CdtR mutant failed to outcompete R20291 in rich media or low-nutrient environments, the loss of CdtR does not mediate this phenotype. Additional studies analyzing the potential differences in the presence and expression of metabolic genes may elucidate the variability in metabolic adaptability demonstrated by ST1-75.

CHAPTER 7

SUPPLEMENTAL FIGURES

Table S1: Human fecal sample shotgun sequencing read count summary statistics

	Total	**HDs**	**Culture positive patients**
n (samples)	94	21	73
Median (reads)	7.2×10^6	9.2×10^6	6.7×10^6
IQR (reads)	2.9×10^6	3.6×10^6	2.4×10^6
Mean (reads)	8.2×10^6	9.4×10^6	7.8×10^6
SE (reads)	4.6×10^5	5.1×10^5	5.6×10^5

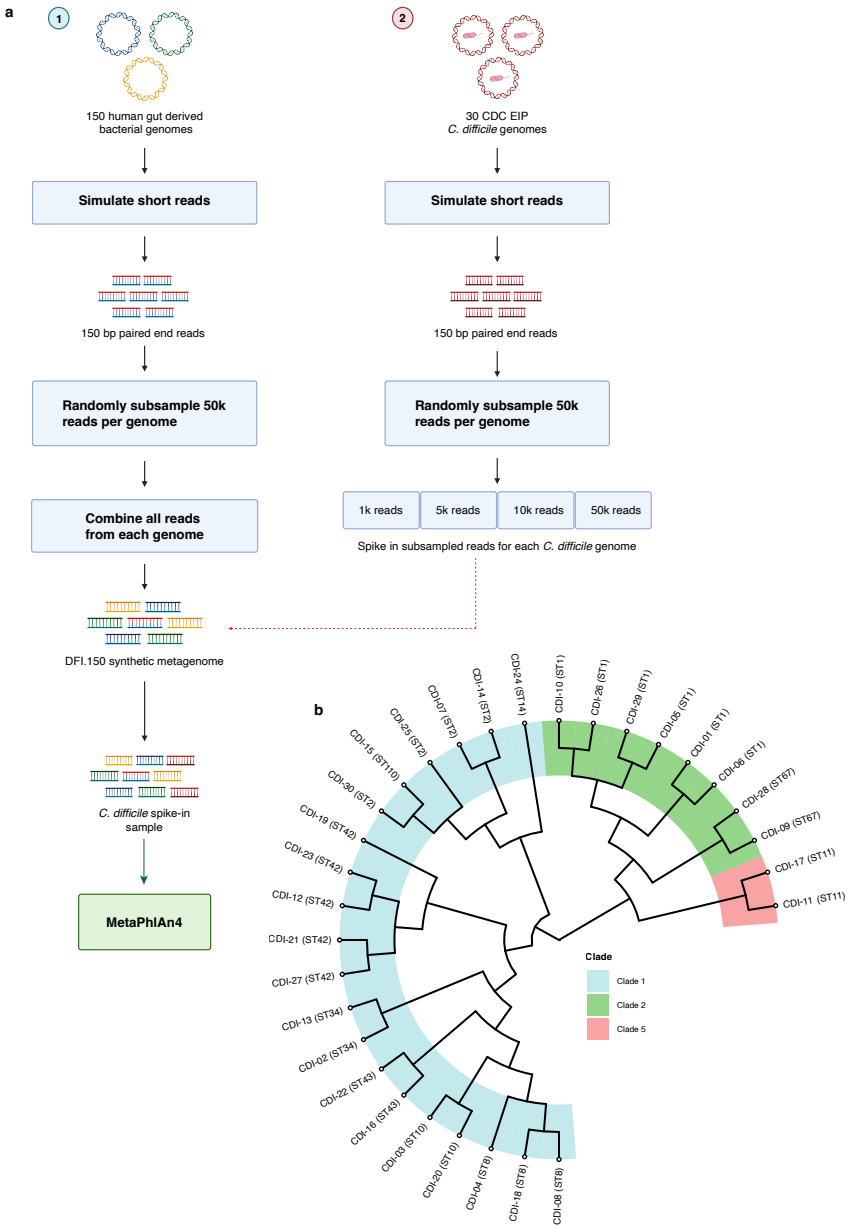

Fig. S1: Model for construction of 150-member synthetic metagenome and *C. difficile* spike-in experiments

(a) 150 bacterial isolates derived from patient fecal samples were whole genome sequenced, These genomes were simulated into 150 bp paired end reads and 50k reads randomly subsampled per genome. These reads were combined into a 150-member synthetic metagenome. Then, *C. difficile* reads were prepared as described above and randomly subsampled at 1k, 5k, 10k, and 50k reads per genome and spiked into the syncom background. Prepared samples were then profiled using MetaPhlAn4.

(b) Neighbor joining phylogenetic tree featuring 30 *C. difficile* isolate genomes, derived from the CDC EIP, constructed from genomic ANI distances.

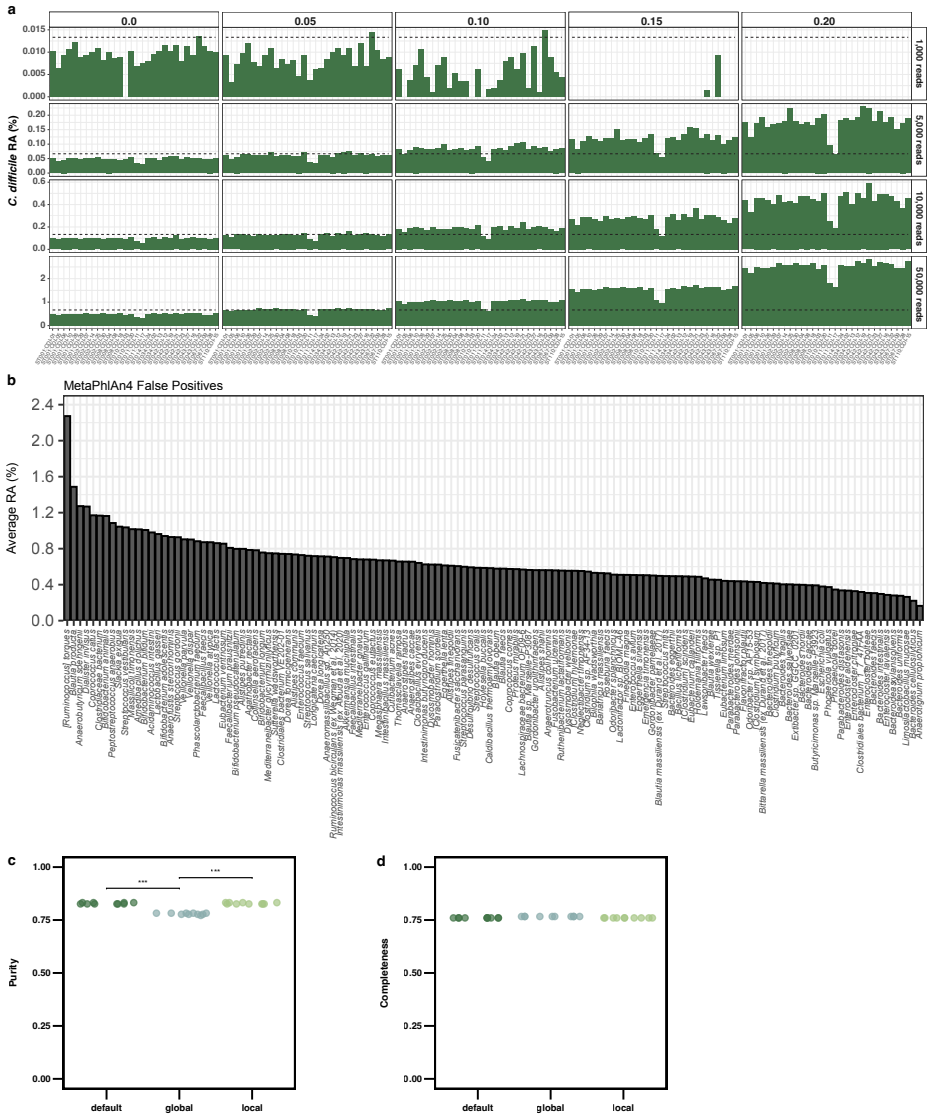

Fig. S2: *C. difficile* classification performance on the 150-member syncom metagenome.

(a) Reported RA of *C. difficile* for 30 isolates of *C. difficile* across various global stat_q parameters (0.0-0.20). Black line represents the target RA of *C. difficile* for 1,000 to 50,000 reads.

(b) False positive species level taxa reported from MetaPhlAn4 using default stat_q (0.20) with the 150-member bacterial synthetic community. Species ordered by decreasing average RA for species with RA over 0.1%.

Binary classification performance of the synthetic metagenomic samples (n=10) between default stat_q (0.20), global stat_q (0.05), and local stat_q (0.05) measured for the purity of the taxonomic profile (c) and completeness of the taxonomic profile (d).

(e) L1 norm error, the total error between the true and predicted abundances of each species, of the synthetic metagenomic samples (n=10) between default stat_q (0.20), global stat_q (0.05), and local stat_q (0.05).

Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest.

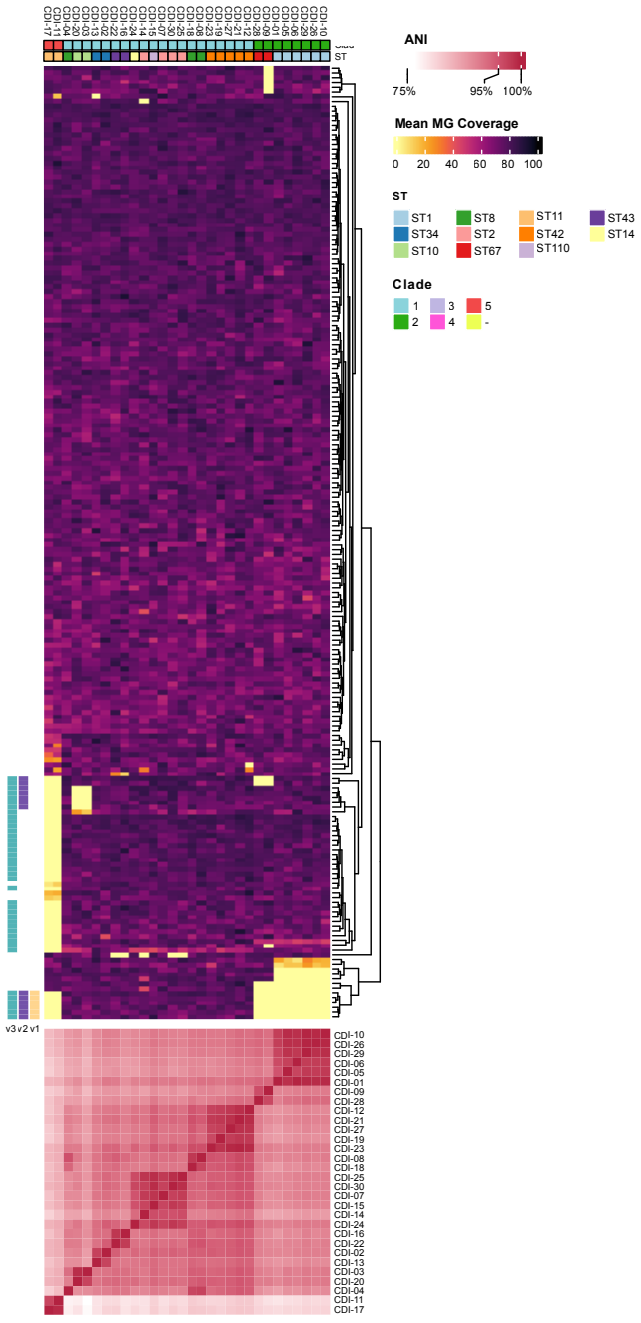

Fig. S3: Marker gene mapping and curation for CDC 30 *C. difficile* genomes.

Heatmap of mean marker gene (n=200) coverage of marker genes mapped to 30 CDC curated *C. difficile* isolates using BowTie2 (top) and heatmap of pairwise genomic ANI distances (bottom). Left annotation displays inclusion into 3 sets of marker genes that are non-uniformly represented across the representative genomes.

Table S2: Curated MetaPhlAn4 *C. difficile* marker genes

	Database		
Marker Genes	**v1**	**v2**	**v3**
UniRef90_A0A069AJC6|1__5|SGB6136	⊗	⊗	⊗
UniRef90_A0A069AN95|1__5|SGB6136	⊗	⊗	⊗
UniRef90_A0A1R2F136|1__9|SGB6136	⊗	⊗	⊗
UniRef90_A0A449PNN4|6__9|SGB6136	⊗	⊗	⊗
UniRef90_A0A449QGZ5|1__4|SGB6136	⊗	⊗	⊗
UniRef90_T4HRY8|1__3|SGB6136	⊗	⊗	⊗
UniRef90_A0A060PHX4|1__5|SGB6136	-	⊗	⊗
UniRef90_A0A069A2R3|6__11|SGB6136	-	⊗	⊗
UniRef90_A0A0N1PAI9|1__5|SGB6136	-	⊗	⊗
UniRef90_A0A381I971|2__14|SGB6136	-	⊗	⊗
UniRef90_A0A449MNF1|1__5|SGB6136	-	⊗	⊗
UniRef90_A0A450KCS6|1__4|SGB6136	-	⊗	⊗
UniRef90_A0A450KCT9|3__12|SGB6136	-	⊗	⊗
UniRef90_A0A031WFH2|2__6|SGB6136	-	-	⊗
UniRef90_A0A068ZZE9|1__7|SGB6136	-	-	⊗
UniRef90_A0A069A2L5|1__3|SGB6136	-	-	⊗
UniRef90_A0A069ASY3|8__11|SGB6136	-	-	⊗
UniRef90_A0A069AT92|1__15|SGB6136	-	-	⊗
UniRef90_A0A069AVB0|1__7|SGB6136	-	-	⊗
UniRef90_A0A069AWD0|1__4|SGB6136	-	-	⊗
UniRef90_A0A069AX49|1__4|SGB6136	-	-	⊗
UniRef90_A0A069B1Z3|1__5|SGB6136	-	-	⊗
UniRef90_A0A0N0P1L3|1__4|SGB6136	-	-	⊗
UniRef90_A0A160WZT9|1__3|SGB6136	-	-	⊗
UniRef90_A0A1R2F4N1|2__5|SGB6136	-	-	⊗
UniRef90_A0A1R5FBT2|1__5|SGB6136	-	-	⊗
UniRef90_A0A1R5FBU2|1__5|SGB6136	-	-	⊗
UniRef90_A0A1R7E666|1__8|SGB6136	-	-	⊗
UniRef90_A0A3T1GUG0|1__4|SGB6136	-	-	⊗
UniRef90_A0A449KG26|1__7|SGB6136	-	-	⊗
UniRef90_A0A449KHH4|1__5|SGB6136	-	-	⊗
UniRef90_A0A449KL93|1__9|SGB6136	-	-	⊗
UniRef90_A0A449KSL1|1__9|SGB6136	-	-	⊗
UniRef90_A0A449KYW6|1__9|SGB6136	-	-	⊗
UniRef90_A0A449KZ44|1__9|SGB6136	-	-	⊗
UniRef90_A0A449LFN9|1__7|SGB6136	-	-	⊗
UniRef90_A0A449LZG3|1__6|SGB6136	-	-	⊗
UniRef90_A0A449T393|3__6|SGB6136	-	-	⊗
UniRef90_A0A449U405|1__5|SGB6136	-	-	⊗
UniRef90_A0A450JKB4|7__11|SGB6136	-	-	⊗

* ⊗ = excluded from database; - = kept in database

Fig. S4: Patient sample inclusion and *C. difficile* isolation protocol

Fig. S4: Patient sample inclusion and *C. difficile* isolation protocol (cont.)

(a) Patient selection criteria for inclusion into study.

(b) Protocol for culture isolation of *C. difficile* from patient fecal samples and whole genome sequencing for *C. difficile* isolates.

(c) Number of isolates by ST from the culture collection.

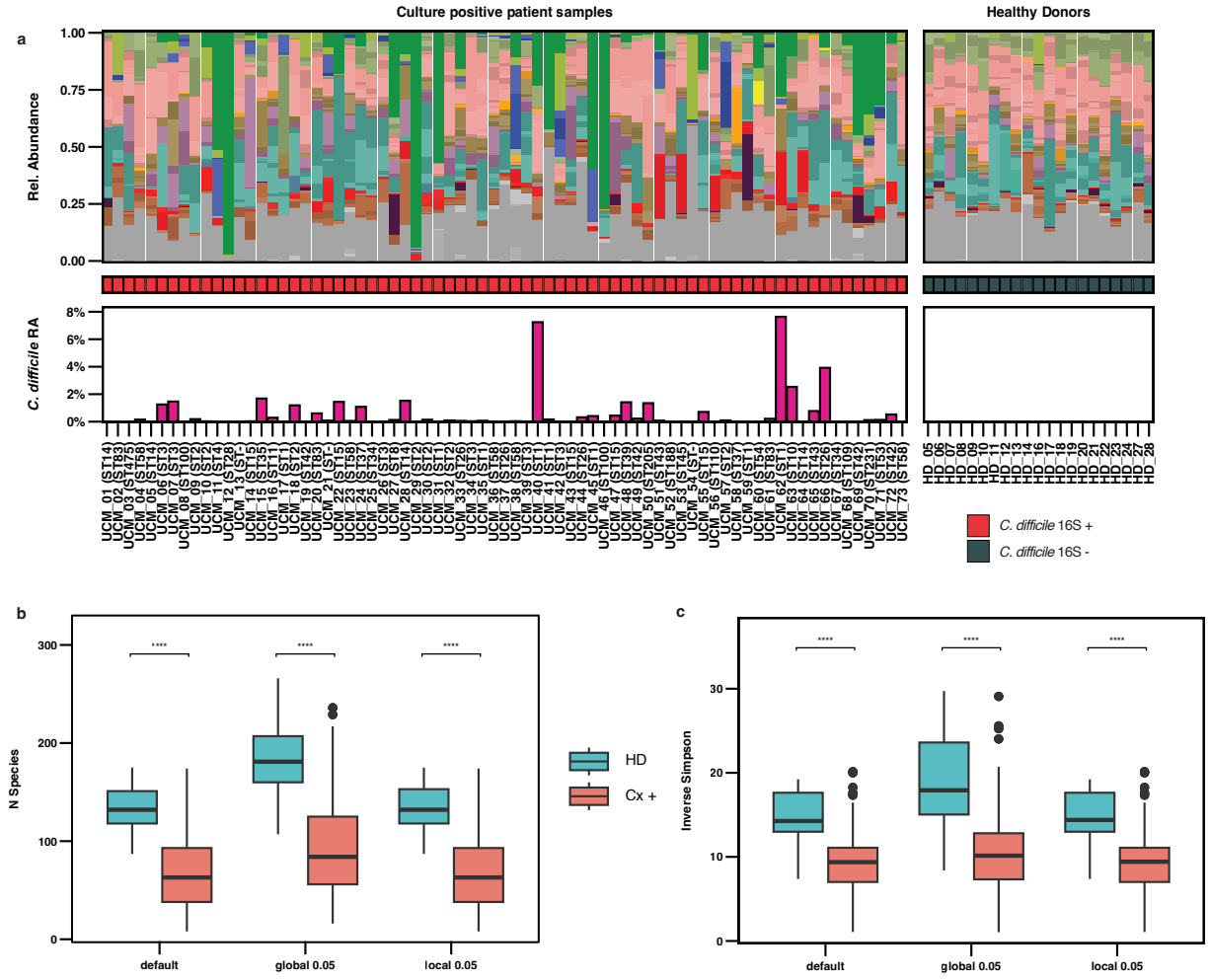

Fig. S5: Taxonomic profiles of healthy donor (HD) and culture positive (Cx pos) fecal metagenomic samples.

(a) Genus-level taxonomy (top), detection of *C. difficile* 16S with PCR (middle), and MetaPhlAn4 estimates of *C. difficile* RA (bottom) of HD (n=21)and Cx positive samples (n=73) using default MetaPhlAn4 (0.20).

(b) Number of reported species of HD and Cx positive fecal samples across compared MetaPhlan4 stat_q settings.

(c) Alpha diversity measured by Inverse Simpson of HD and Cx positive fecal samples across compared MetaPhlan4 stat_q settings.

Statistical comparisons between individual groups were analyzed using a two-tailed Wilcoxon rank sum rest.

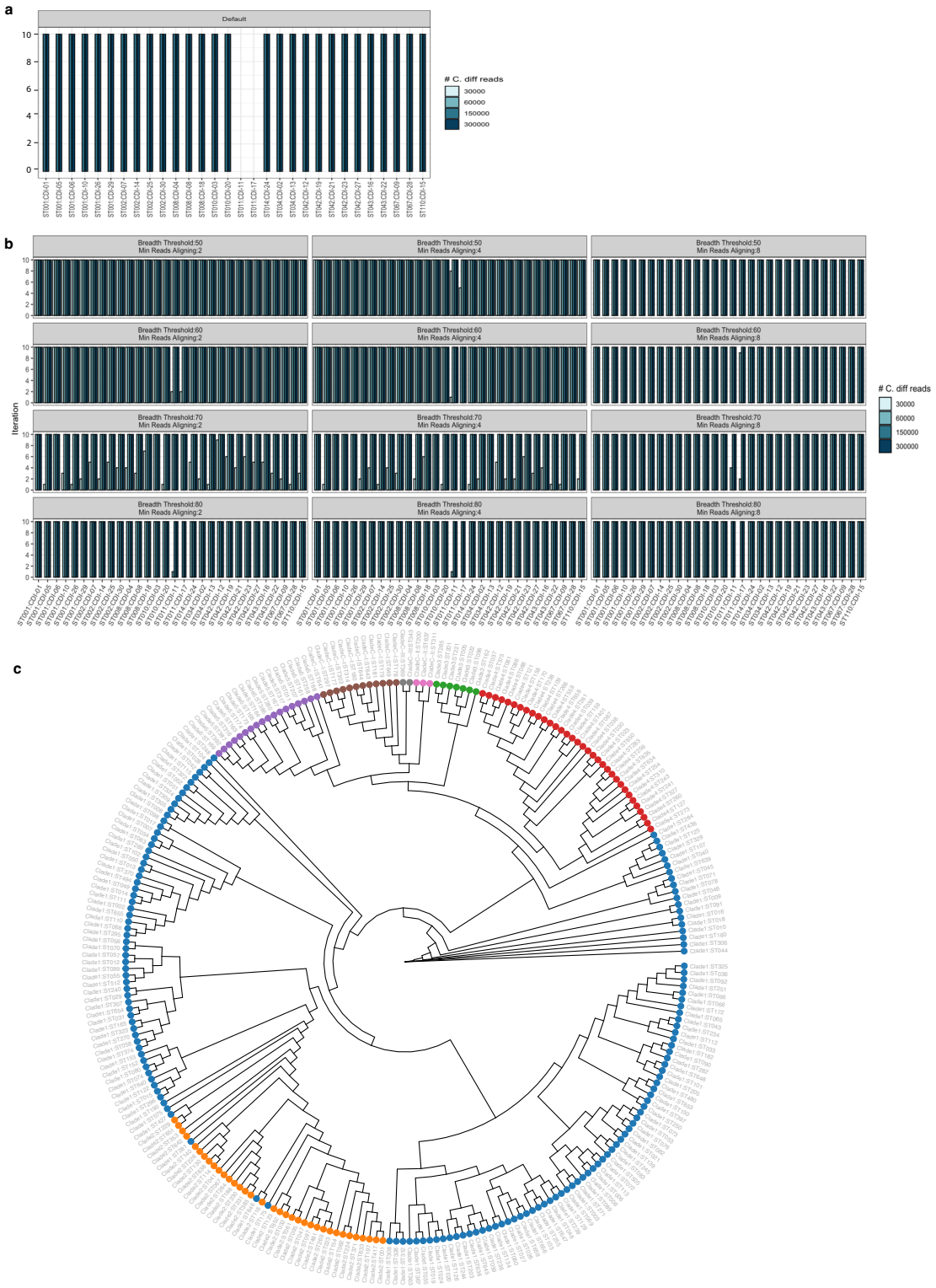

Fig. S6: Optimization of StrainPhlAn4 for the detection of *C. difficile*

Fig. S6: Optimization of StrainPhlAn4 for the detection of *C. difficile* (cont.)

(a) Number of correctly detected iterations (n=10) for each CDC *C. difficile* isolate using default StrainPhlan4 settings

(b) Number of correctly detected iterations (n=10) for each CDC *C. difficile* isolate across varying StrainPhlan4 settings

(c) Maximum-likelihood default StrainPhlAn4 phylogenetic tree output of CDC *C. difficile* isolate spike-ins (red) into patient stool metagenomic samples (n=5) compared to reference genomes (gray).

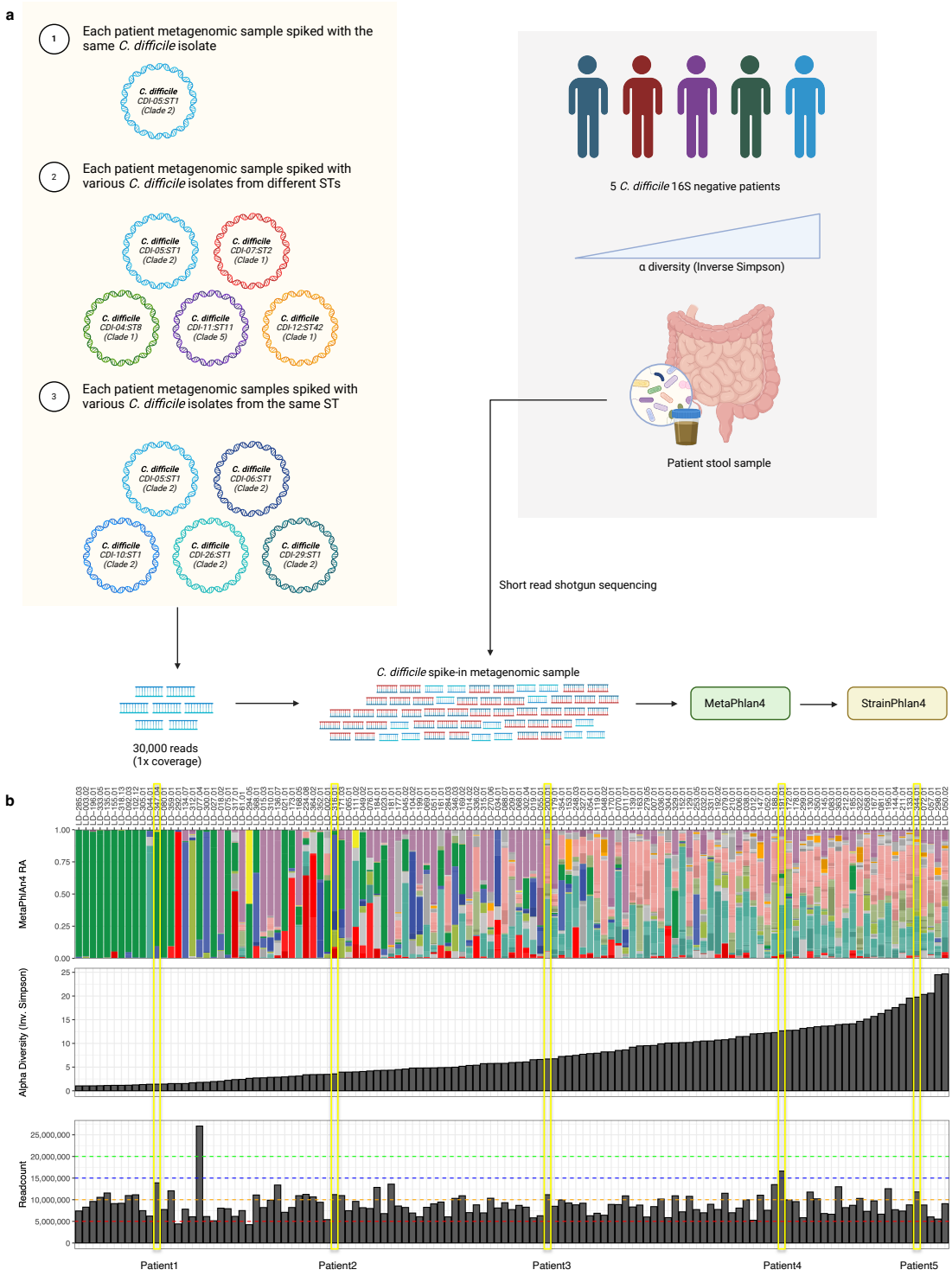

Fig. S7: Model and patient selection for StrainPlAn4 *in silico* experiments (cont.)

Fig. S7: Model and patient selection for StrainPlAn4 *in silico* experiments

(a) Experimental model for *in silico* StrainPhlAn4 experiments.

(b) Genus level taxonomic plots of stool metagenomic samples sourced from Liver Disease patients ranked by increasing α-diversity measured as inverse Simpson indices. 5 samples with over 10 million sequencing reads were selected for inclusion in StrainPhlan4 optimization *in silico* experiments(yellow).

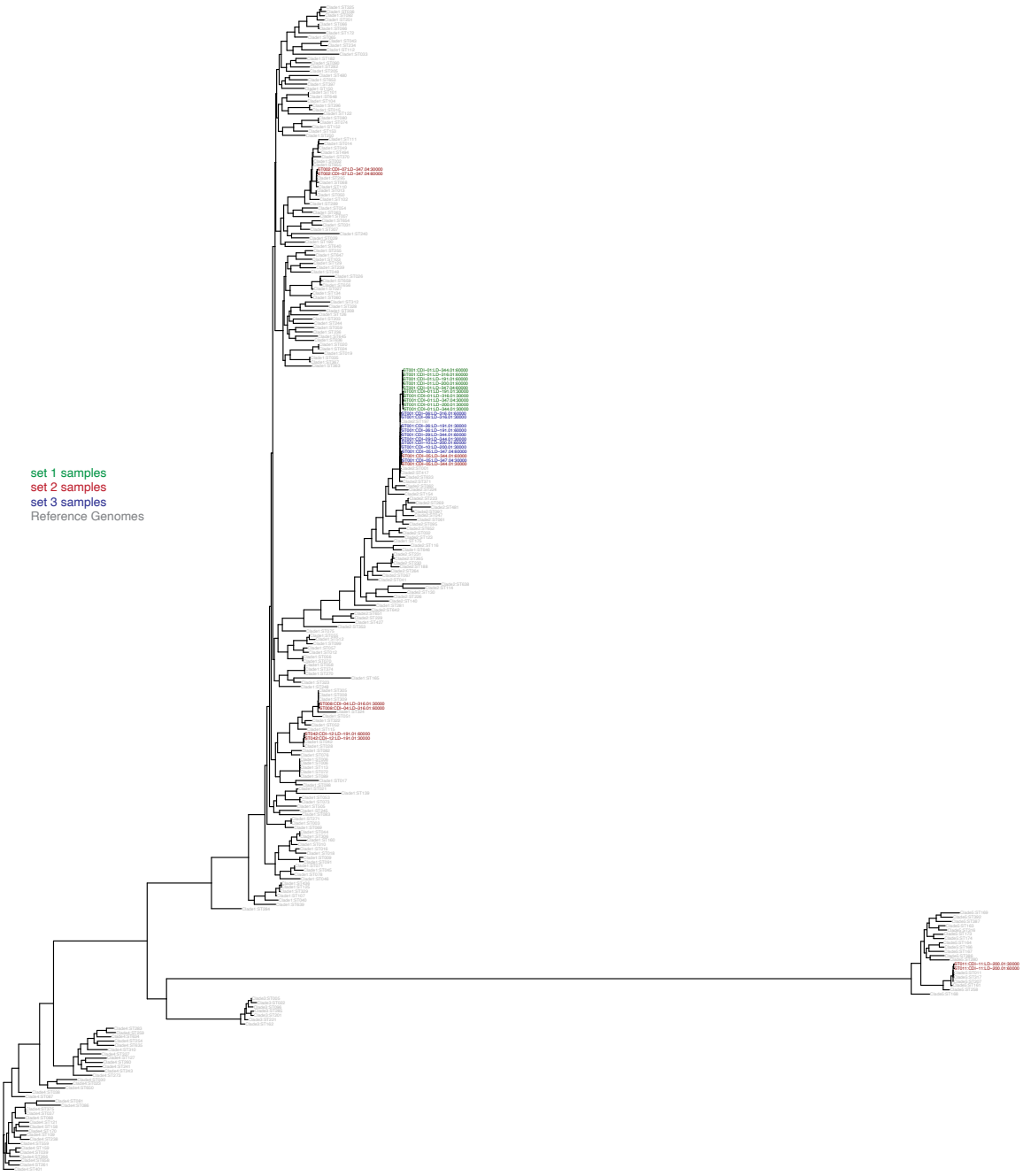

Fig. S8: Maximum-likelihood optimized StrainPhlAn4 phylogenetic tree output of CDC *C. difficile* isolate spike-ins into patient stool metagenomic samples (n=5). Genomes were separated into three sets as previously described (a): set 1 (green), set 2 (red), and set 3 (blue). For each spike-in, the *C. difficile* genome was randomly sampled at 30,000 (1x) or 60,000 (2x) reads while the reference genomes (gray) were sampled at 300,000 reads (10x) using optimized settings.

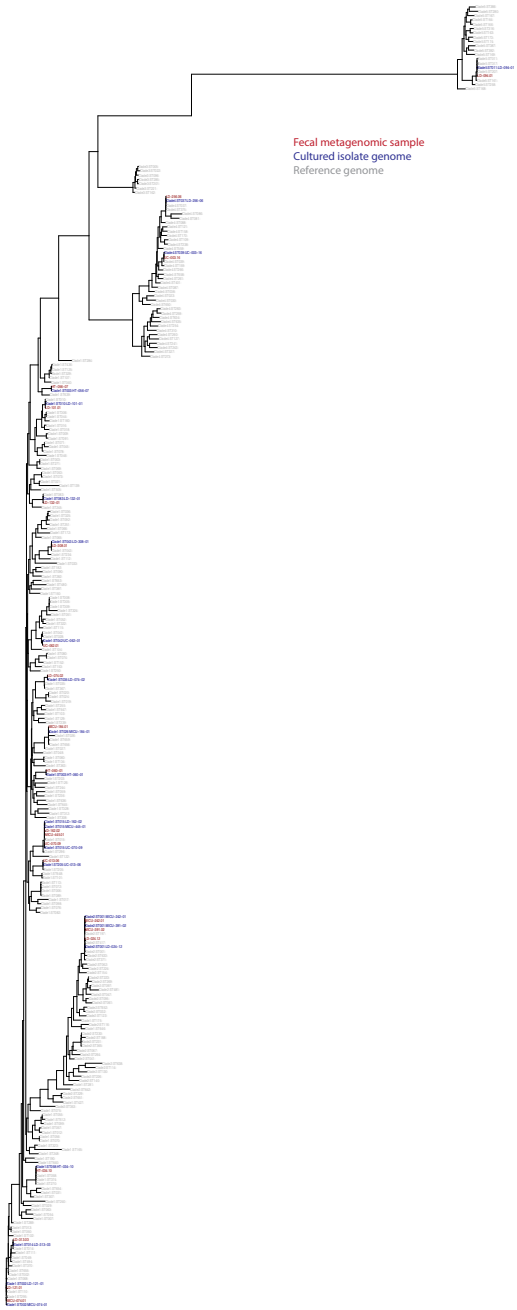

Fig. S9: Maximum-likelihood phylogenetic tree StrainPhlAn4 output of patient-derived *C. difficile* isolate spike-ins (red) into patient stool metagenomic samples (n=5). For each spike-in, the *C. difficile* genome were randomly sampled at 30,000 (1x) or 60,000 (2x) reads while the isolated culture genomes (blue) and the reference tree (gray) were sampled at 300,000 reads (10x) using optimized settings.

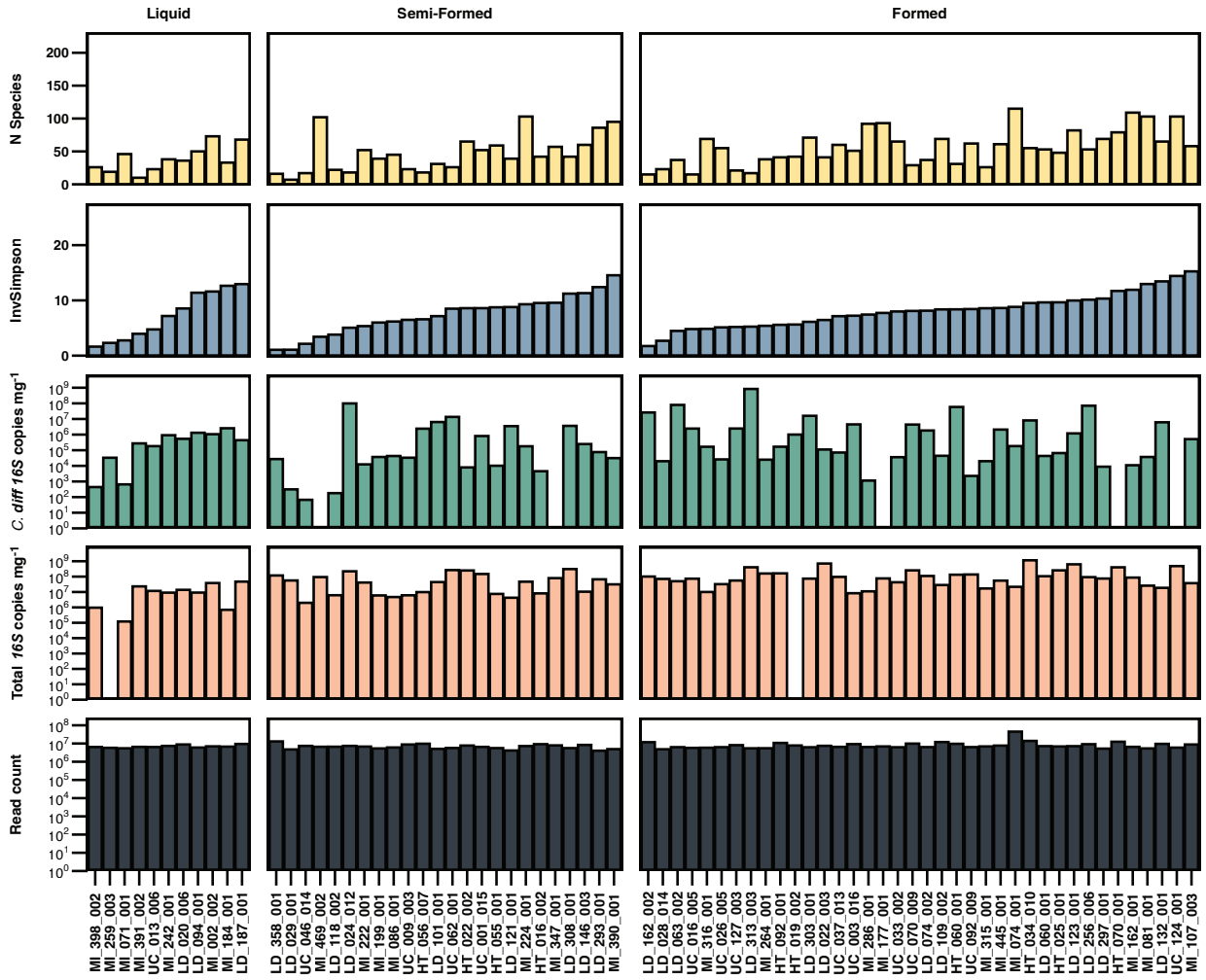

Fig. S10: Quantification of *C. difficile* from patient fecal samples using qPCR

Alpha diversity quantified with species richness (yellow) and inverse Simpson (blue) for *C. difficile* culture+ fecal samples (n=73). Absolute counts of *C. difficile*-specific *16S rRNA* qPCR (green), total*16S rRNA* qPCR (orange), and total read count (navy) for each sample.

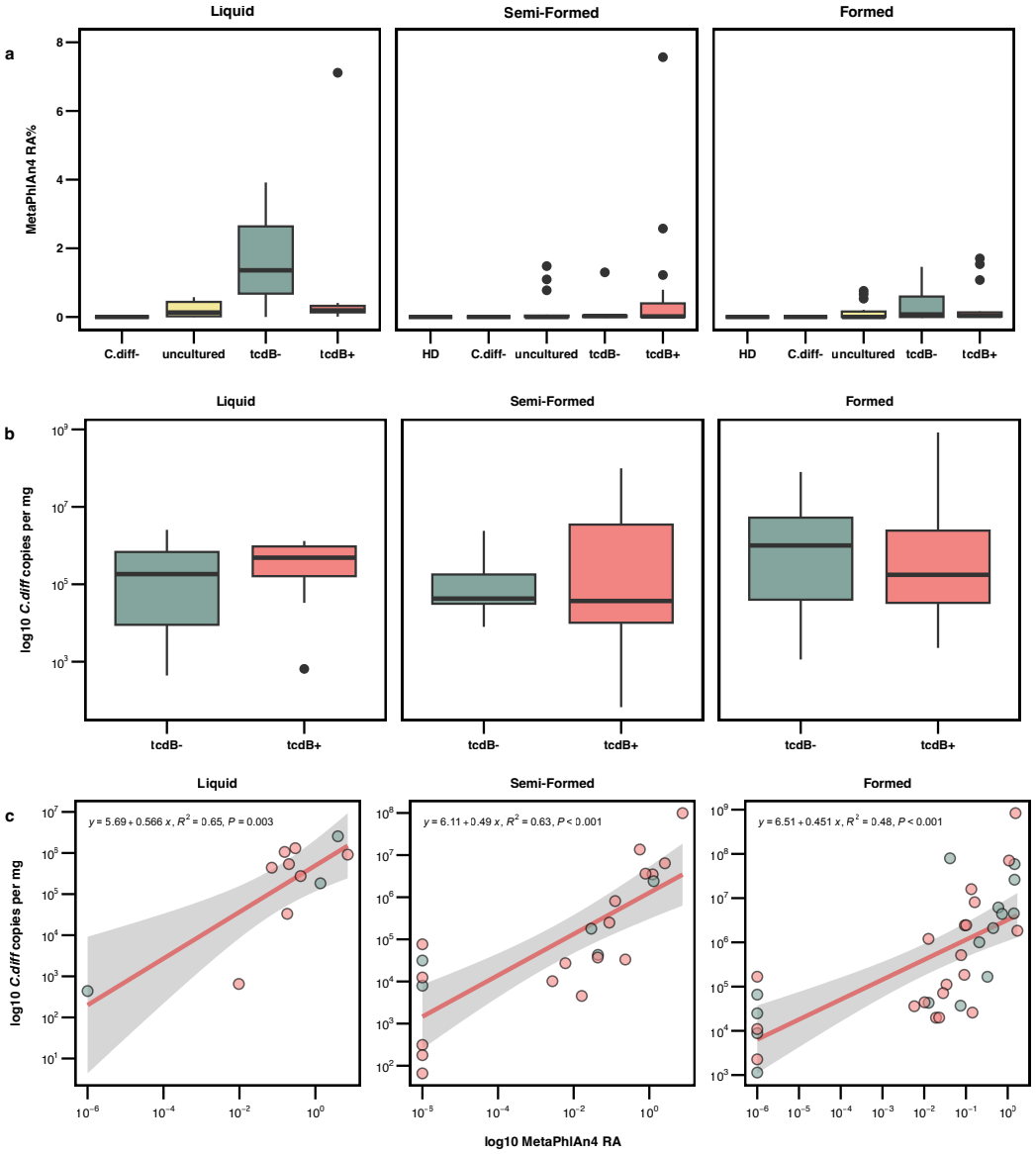

Fig. S11: Quantification of *C. difficile* across *C. difficile* culture-positive patient fecal samples

(a) Comparison of *C. difficile* RA estimated by MetaPhlAn4 with local stat_q 0.05 between all fecal samples (n = 505).

(b) Comparison of *C. difficile* RA calculated from *16S rRNA* qPCR (copies per mg-1) between cultured fecal samples (n = 73).

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests.

(c) Linear regressions of reported MetaPhlAn4 *C. difficile* RA and *C. difficile* RA calculated from *16S rRNA* qPCR.

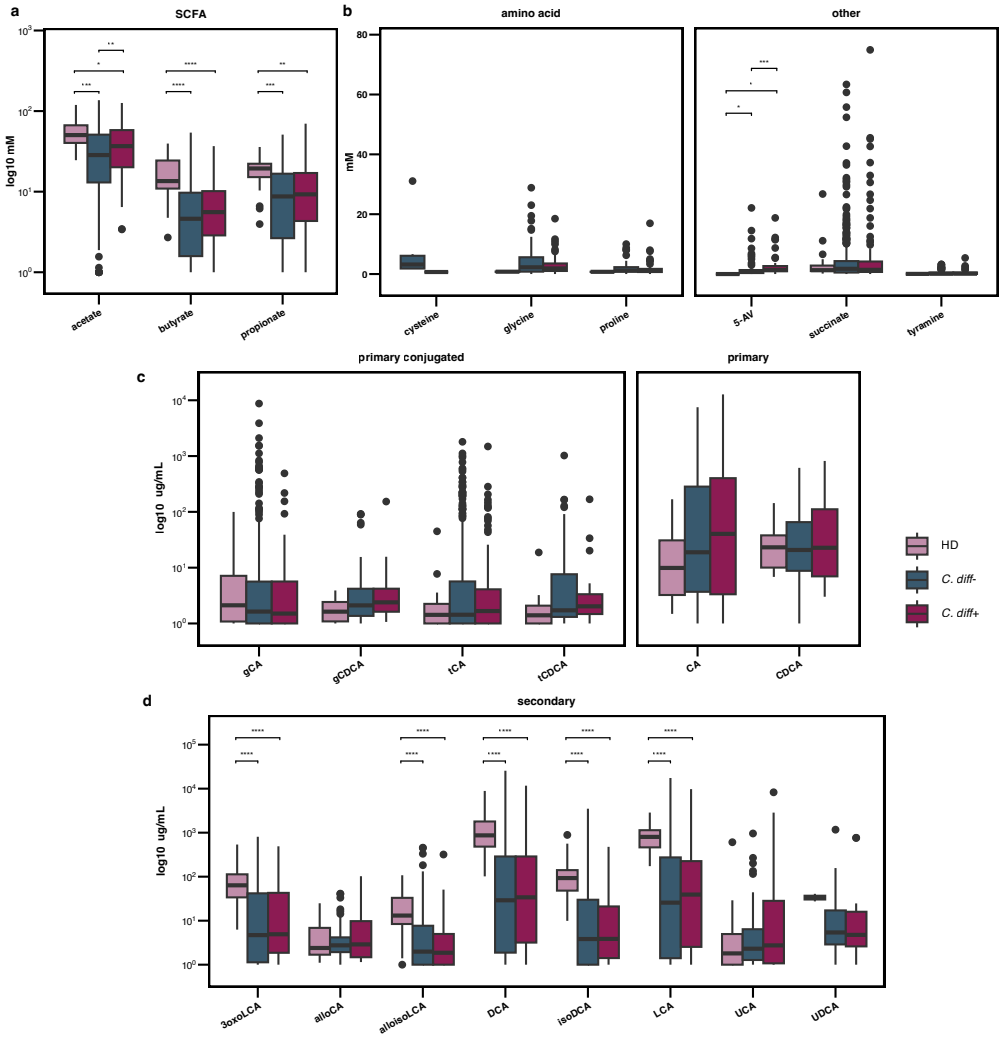

Fig. S12: Quantified fecal metabolites from HDs and patient samples

(a) Quantified amino acids (cysteine, glycine, and proline) and Stickland fermentation product 5-aminovalerate (5-AV) compared between HDs (n=21), *C. difficile 16S-* (n=374), and *C. difficile 16S+* (n=110) patient fecal samples.

(b) Short chain fatty acids (SCFAs), acetate, butyrate, and propionate, quantified from *C. difficile 16S-* or *C. difficile 16S+* patient fecal samples.

(c) Primary bile acids (PBAs) and conjugated PBAs quantified from *C. difficile 16S-* or *C. difficile 16S+* patient fecal samples.

(d) Secondary bile acids (SBAs) quantified from *C. difficile 16S-* or *C. difficile 16S+* patient fecal samples.

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests.

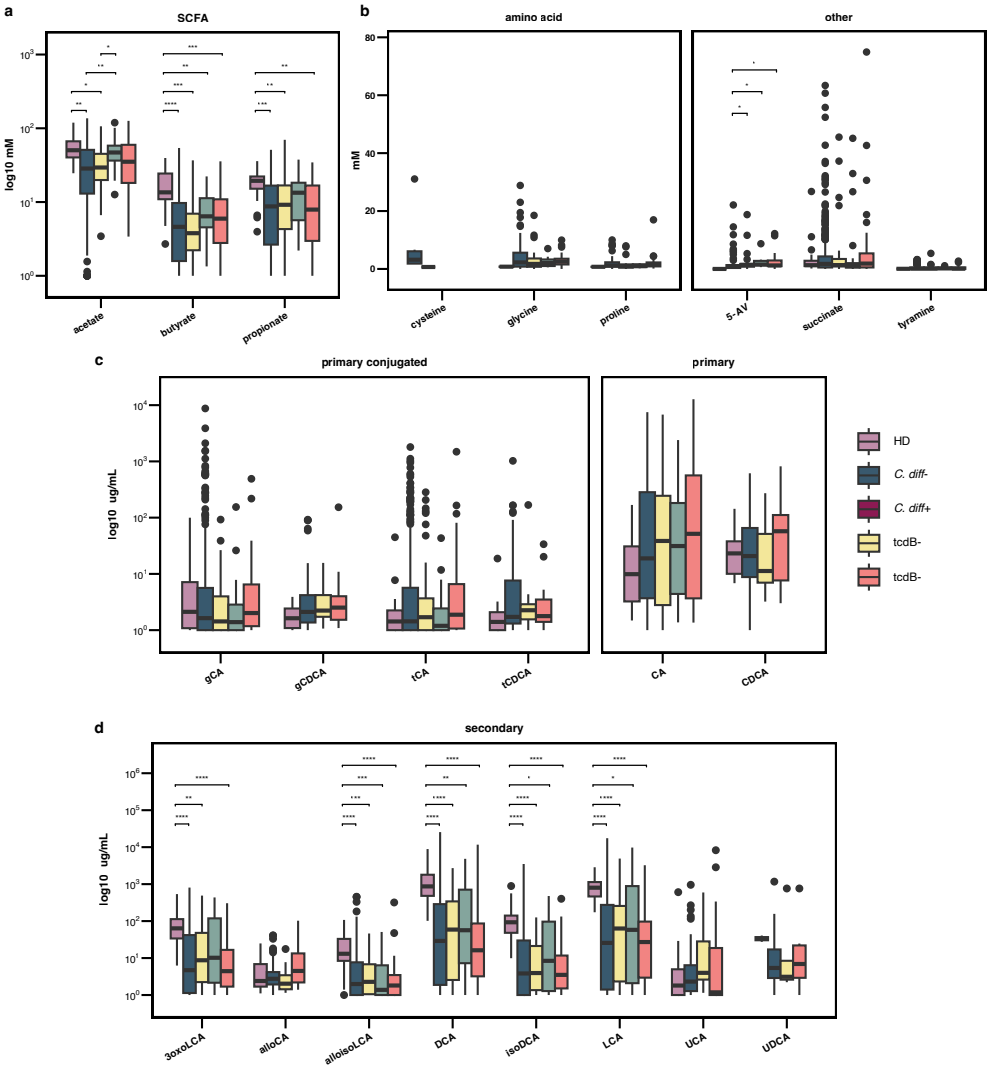

Fig. S13: Quantified fecal metabolites from HDs and Cx positive patient samples

(a) Quantified amino acids (cysteine, glycine, and proline) and Stickland fermentation product 5-aminovalerate (5-AV) compared between HDs (n=21), Cx negative samples (n=374), uncultured Cx positive samples (n=37), *tcdB*- Cx positive fecal samples (n=25), and *tcdB*+ Cx positive fecal samples (n=48)

(b) Short chain fatty acids (SCFAs), acetate, butyrate, and propionate, quantified from *tcdB*- or *tcdB*+ Cx positive patient fecal samples.

(c) Primary bile acids (PBAs) and conjugated PBAs quantified from *tcdB*- or *tcdB*+ Cx positive patient fecal samples.

(d) Secondary bile acids (SBAs) quantified from *tcdB*- or *tcdB*+ Cx positive patient fecal samples.

Statistical comparisons between individual groups were analyzed using unpaired two-tailed Wilcoxon rank sum tests.

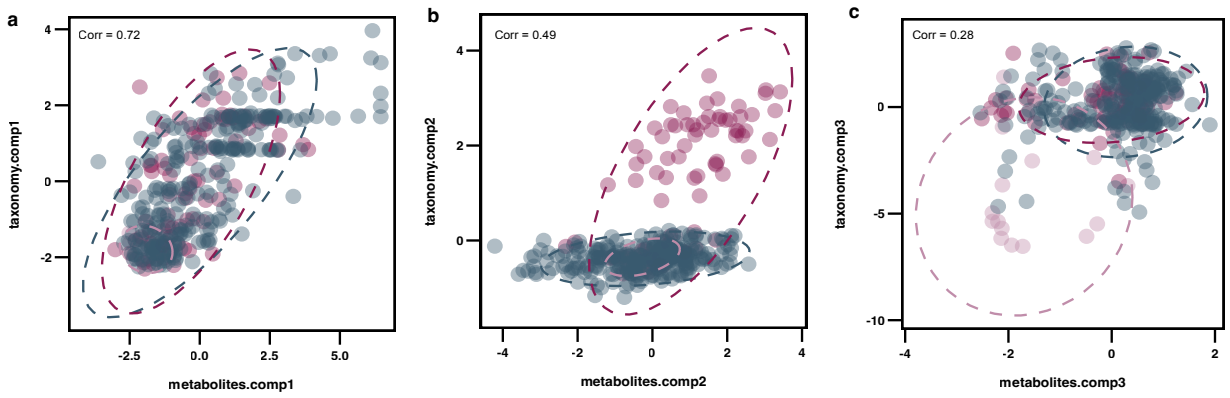

Fig. S14: Spread of training sample data set by *C. difficile* group inclusion on latent omic block component axes

Variance attributable to taxonomy and metabolite features on component 1.

Variance attributable to taxonomy and metabolite features on component 2.

Variance attributable to taxonomy and metabolite features on component 3.

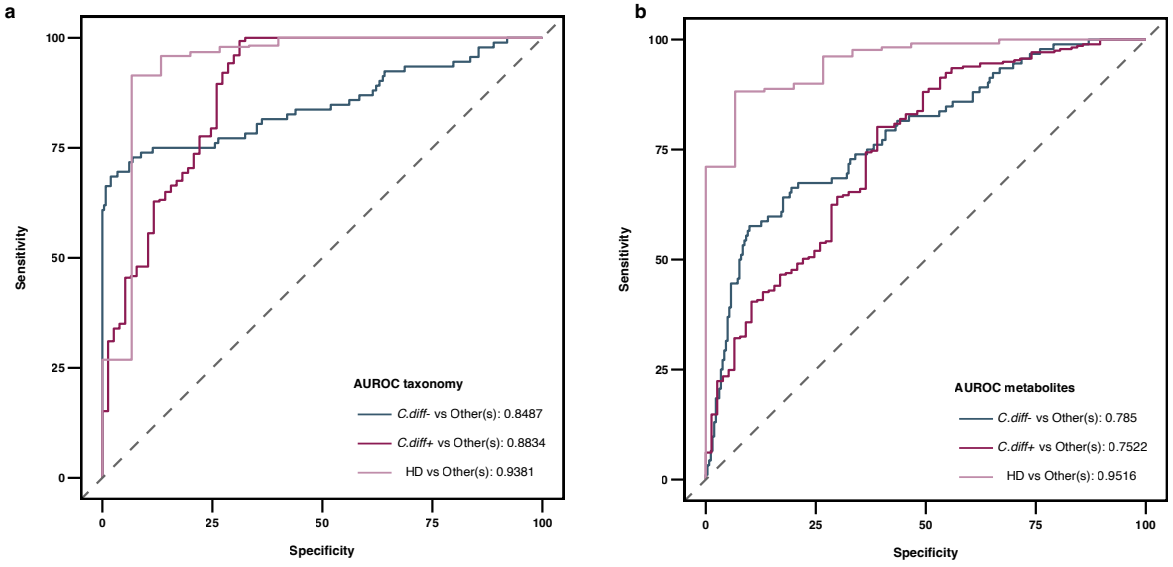

Fig. S15: Quantification of *C. difficile* from patient fecal samples using qPCR

Cumulative AUROC of discrimination *C. difficile*-, *C. difficile* +, or HD samples from the multi-block sPLSDA model using the taxonomic (**??**) or metabolomic (**??**) datasets on the test set (n = 140) provided by the mixOmics::auroc() function.

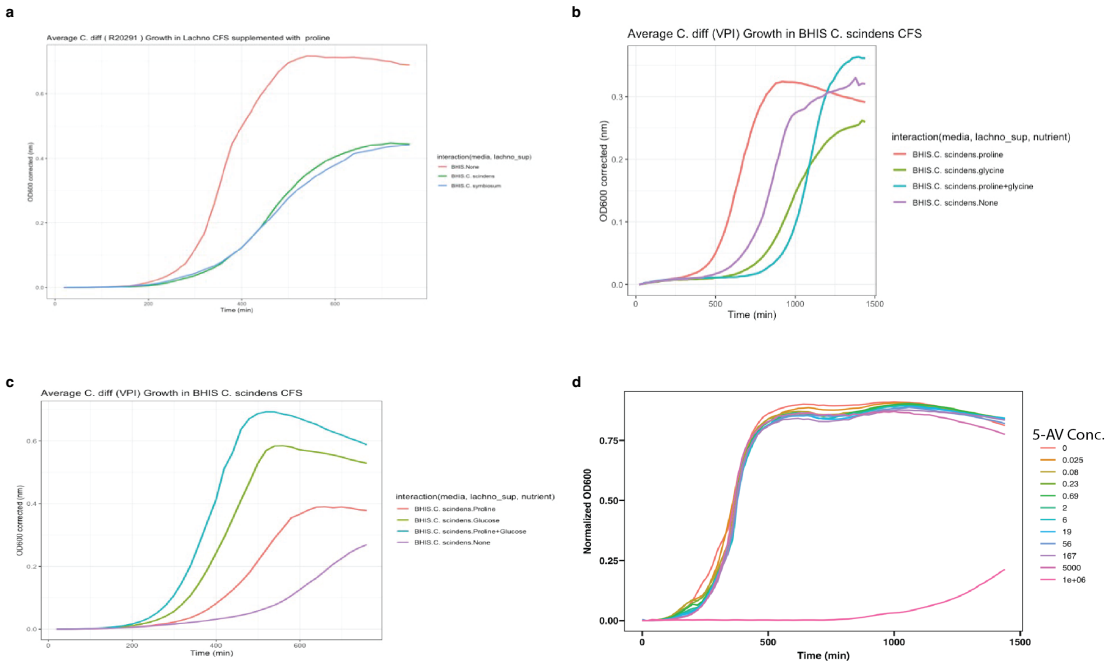

Fig. S15: Competition for proline and carbon sources impact *C. difficile* vegetative growth *in vitro*

Growth of *C. difficile* R20291 in 24 hr *C. scindens* or *C. symbiosum* CFS supplemented with 2 mM L-proline.

Growth of *C. difficile* R20291 in 24 hr *C. scindens* or *C. symbiosum* CFS supplemented with 2 mM L-proline and 3mM L-glycine.

Growth of *C. difficile* R20291 in 24 hr *C. scindens* or *C. symbiosum* CFS supplemented with 2 mM L-proline and 11 mM glucose.

Growth of *C. difficile* R20291 in BHIS supplemented with 5-AV 0-1M.

REFERENCES

[1] Natasha Bagdasarian, Krishna Rao, and Preeti N. Malani. Diagnosis and Treatment of *Clostridium difficile* in Adults: A Systematic Review. *JAMA*, 313(4):398, January 2015. ISSN 0098-7484. doi:10.1001/jama.2014.17103.

[2] John G. Bartlett, Nancy Moon, Te Wen Chang, Nancy Taylor, and Andrew B. Onderdonk. Role of Clostridium difficile in antibiotic-associated pseudomembranous colitis. *Gastroenterology*, 75(5):778–782, November 1978. ISSN 00165085. doi:10.1016/0016-5085(78)90457-2.

[3] P H Gilligan, L R McCarthy, and V M Genta. Relative frequency of Clostridium difficile in patients with diarrheal disease. *Journal of Clinical Microbiology*, 14(1): 26–31, July 1981. ISSN 0095-1137, 1098-660X. doi:10.1128/jcm.14.1.26-31.1981.

[4] L. R. Hirschhorn, Y. Trnka, A. Onderdonk, M.-L. T. Lee, and R. Platt. Epidemiology of Community-Acquired Clostridium difficile-Associated Diarrhea. *Journal of Infectious Diseases*, 169(1):127–133, January 1994. ISSN 0022-1899, 1537-6613. doi:10.1093/infdis/169.1.127.

[5] Casey M. Theriot, Mark J. Koenigsknecht, Paul E. Carlson, Gabrielle E. Hatton, Adam M. Nelson, Bo Li, Gary B. Huffnagle, Jun Z. Li, and Vincent B. Young. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature Communications*, 5(1):3114, January 2014. ISSN 2041-1723. doi:10.1038/ncomms4114.

[6] IVAN C. HALL and ELIZABETH O'TOOLE. INTESTINAL FLORA IN NEWBORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, BACILLUS DIFFICILIS. *American Journal of Diseases of Children*, 49(2): 390–402, February 1935. ISSN 0096-8994. doi:10.1001/archpedi.1935.01970020105010.

[7] RH George. Identification of Clostridium difficile as a cause of pseudomembranous colitis. *Br. Med. J.*, 1:695, 1978.

[8] Paul A. Lawson, Diane M. Citron, Kerin L. Tyrrell, and Sydney M. Finegold. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe*, 40:95–99, August 2016. ISSN 1075-9964. doi:10.1016/j.anaerobe.2016.06.008.

[9] Matthew T. Sorbara and Eric G. Pamer. Microbiome-based therapeutics. *Nature Reviews Microbiology*, 20(6):365–380, June 2022. ISSN 1740-1526, 1740-1534. doi:10.1038/s41579-021-00667-9.

[10] van Nood Els, Vrieze Anne, Nieuwdorp Max, Fuentes Susana, Zoetendal Erwin G., de Vos Willem M., Visser Caroline E., Kuijper Ed J., Bartelsman Joep F.W.M., Tijssen Jan G.P., Speelman Peter, Dijkgraaf Marcel G.W., and Keller Josbert J. Duodenal Infusion of Donor Feces for Recurrent Clostridium difficile. *New England Journal of Medicine*, 368(5):407–415, 2013. doi:10.1056/NEJMoa1205037.

[11] Christian Lodberg Hvas, Simon Mark Dahl Jørgensen, Søren Peter Jørgensen, Merete Storgaard, Lars Lemming, Mette Mejlby Hansen, Christian Erikstrup, and Jens Frederik Dahlerup. Fecal Microbiota Transplantation Is Superior to Fidaxomicin for Treatment of Recurrent Clostridium difficile Infection. *Gastroenterology*, 156(5):1324–1332.e3, April 2019. ISSN 0016-5085, 1528-0012. doi:10.1053/j.gastro.2018.12.019.

[12] Anna M. Seekatz, Casey M. Theriot, Krishna Rao, Yu-Ming Chang, Alison E. Freeman, John Y. Kao, and Vincent B. Young. 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, October 2018. ISSN 10759964. doi:10.1016/j.anaerobe.2018.04.001.

[13] Alexa R. Weingarden, Peter I. Dosa, Erin DeWinter, Clifford J. Steer, Megan K. Shaughnessy, James R. Johnson, Alexander Khoruts, and Michael J. Sadowsky. Changes in Colonic Bile Acid Composition following Fecal Microbiota Transplantation Are Sufficient to Control Clostridium difficile Germination and Growth. *PLOS ONE*, 11(1):e0147210, January 2016. ISSN 1932-6203. doi:10.1371/journal.pone.0147210.

[14] Marc Monot, Catherine Eckert, Astrid Lemire, Audrey Hamiot, Thomas Dubois, Carine Tessier, Bruno Dumoulard, Benjamin Hamel, Amandine Petit, Valérie Lalande, Laurence Ma, Christiane Bouchier, Frédéric Barbut, and Bruno Dupuy. Clostridium difficile: New Insights into the Evolution of the Pathogenicity Locus. *Scientific Reports*, 5(1):15023, October 2015. ISSN 2045-2322. doi:10.1038/srep15023.

[15] Thomas Jank, Torsten Giesemann, and Klaus Aktories. Rho-glucosylating Clostridium difficile toxins A and B: New insights into structure and function. *Glycobiology*, 17(4): 15R–22R, April 2007. ISSN 0959-6658. doi:10.1093/glycob/cwm004.

[16] Daniel E. Voth and Jimmy D. Ballard. Clostridium difficile Toxins: Mechanism of Action and Role in Disease. *Clinical Microbiology Reviews*, 18(2):247–263, April 2005. doi:10.1128/cmr.18.2.247-263.2005.

[17] Nagraj Mani and Bruno Dupuy. Regulation of toxin synthesis in Clostridium difficile by an alternative RNA polymerase sigma factor. *Proceedings of the National Academy of Sciences*, 98(10):5844–5849, May 2001. doi:10.1073/pnas.101126598.

[18] Susana Matamouros, Patrick England, and Bruno Dupuy. Clostridium difficile toxin expression is inhibited by the novel regulator TcdC. *Molecular Microbiology*, 64(5): 1274–1288, 2007. ISSN 1365-2958. doi:10.1111/j.1365-2958.2007.05739.x.

[19] Revathi Govind and Bruno Dupuy. Secretion of Clostridium difficile Toxins A and B Requires the Holin-like Protein TcdE. *PLOS Pathogens*, 8(6):e1002727, June 2012. ISSN 1553-7374. doi:10.1371/journal.ppat.1002727.

[20] Maja Rupnik, Mark H. Wilcox, and Dale N. Gerding. Clostridium difficile infection: New developments in epidemiology and pathogenesis. *Nature Reviews Microbiology*, 7 (7):526–536, July 2009. ISSN 1740-1526, 1740-1534. doi:10.1038/nrmicro2164.

[21] Veit Braun, Thomas Hundsberger, Petra Leukel, Markus Sauerborn, and Christoph Von Eichel-Streiber. Definition of the single integration site of the pathogenicity locus in Clostridium difficile. *Gene*, 181(1-2):29–38, November 1996. ISSN 03781119. doi:10.1016/S0378-1119(96)00398-8.

[22] Kate E. Dingle, Briony Elliott, Esther Robinson, David Griffiths, David W. Eyre, Nicole Stoesser, Alison Vaughan, Tanya Golubchik, Warren N. Fawley, Mark H. Wilcox, Timothy E. Peto, A. Sarah Walker, Thomas V. Riley, Derrick W. Crook, and Xavier Didelot. Evolutionary History of the Clostridium difficile Pathogenicity Locus. *Genome Biology and Evolution*, 6(1):36–52, January 2014. ISSN 1759-6653. doi:10.1093/gbe/evt204.

[23] M. Riegler, R. Sedivy, C. Pothoulakis, G. Hamilton, J. Zacherl, G. Bischof, E. Cosentini, W. Feil, R. Schiessel, and J. T. LaMont. Clostridium difficile toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. https://www.jci.org/articles/view/117885/pdf, May 1995.

[24] Dena Lyras, Jennifer R. O'Connor, Pauline M. Howarth, Susan P. Sambol, Glen P. Carter, Tongted Phumoonna, Rachael Poon, Vicki Adams, Gayatri Vedantam, Stuart Johnson, Dale N. Gerding, and Julian I. Rood. Toxin B is essential for virulence of Clostridium difficile. *Nature*, 458(7242):1176–1179, April 2009. ISSN 1476-4687. doi:10.1038/nature07822.

[25] Glen P. Carter, Dena Lyras, David L. Allen, Kate E. Mackin, Pauline M. Howarth, Jennifer R. O'Connor, and Julian I. Rood. Binary Toxin Production in Clostridium difficile Is Regulated by CdtR, a LytTR Family Response Regulator. *Journal of Bacteriology*, 189(20):7290–7301, October 2007. doi:10.1128/jb.00731-07.

[26] C. Eckert, A. Emirian, A. Le Monnier, L. Cathala, H. De Montclos, J. Goret, P. Berger, A. Petit, A. De Chevigny, H. Jean-Pierre, B. Nebbad, S. Camiade, R. Meckenstock, V. Lalande, H. Marchandin, and F. Barbut. Prevalence and pathogenicity of binary toxin–positive *Clostridium difficile* strains that do not produce toxins A and B. *New Microbes and New Infections*, 3:12–17, January 2015. ISSN 2052-2975. doi:10.1016/j.nmni.2014.10.003.

[27] M R Popoff, E J Rubin, D M Gill, and P Boquet. Actin-specific ADP-ribosyltransferase produced by a Clostridium difficile strain. *Infection and Immunity*, 56(9):2299–2306, September 1988. doi:10.1128/iai.56.9.2299-2306.1988.

[28] Qiwen Dong, Huaiying Lin, Marie-Maude Allen, Julian R. Garneau, Jonathan K. Sia, Rita C. Smith, Fidel Haro, Tracy McMillen, Rosemary L. Pope, Carolyn Metcalfe, Victoria Burgo, Che Woodson, Nicholas Dylla, Claire Kohout, Anitha Sundararajan, Evan S. Snitkin, Vincent B. Young, Louis-Charles Fortier, Mini Kamboj, and Eric G. Pamer. Virulence and genomic diversity among clinical isolates of ST1 (BI/NAP1/027) Clostridioides difficile. *Cell Reports*, 42(8):112861, August 2023. ISSN 22111247. doi:10.1016/j.celrep.2023.112861.

[29] Ellie J.C. Goldstein, Diane M. Citron, Pamela J. Goldman, and Ronald J. Goldman. National hospital survey of anaerobic culture and susceptibility methods: III. *Anaerobe*, 14(2):68–72, April 2008. ISSN 10759964. doi:10.1016/j.anaerobe.2008.01.001.

[30] Lana Dbeibo, Christine W. Lucky, William F. Fadel, Joshua Sadowski, Cole Beeler, Kristen Kelley, Joy Williams, Douglas Webb, and Areeba Kara. Two-step algorithm-based Clostridioides difficile testing as a tool for antibiotic stewardship. *Clinical Microbiology and Infection*, 29(6):798.e1–798.e4, June 2023. ISSN 1198743X. doi:10.1016/j.cmi.2023.02.008.

[31] Daniel R Knight, Korakrit Imwattana, Brian Kullin, Enzo Guerrero-Araya, Daniel Paredes-Sabja, Xavier Didelot, Kate E Dingle, David W Eyre, César Rodríguez, and Thomas V Riley. Major genetic discontinuity and novel toxigenic species in Clostridioides difficile taxonomy. *eLife*, 10:e64325, June 2021. ISSN 2050-084X. doi:10.7554/eLife.64325.

[32] Miao He, Mohammed Sebaihia, Trevor D. Lawley, Richard A. Stabler, Lisa F. Dawson, Melissa J. Martin, Kathryn E. Holt, Helena M.B. Seth-Smith, Michael A. Quail, Richard Rance, Karen Brooks, Carol Churcher, David Harris, Stephen D. Bentley, Christine Burrows, Louise Clark, Craig Corton, Vicky Murray, Graham Rose, Scott Thurston, Andries Van Tonder, Danielle Walker, Brendan W. Wren, Gordon Dougan, and Julian Parkhill. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences*, 107(16):7527–7532, April 2010. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.0914322107.

[33] Xavier Didelot, David W Eyre, Madeleine Cule, Camilla Lc Ip, M Azim Ansari, David Griffiths, Alison Vaughan, Lily O'Connor, Tanya Golubchik, Elizabeth M Batty, Paolo Piazza, Daniel J Wilson, Rory Bowden, Peter J Donnelly, Kate E Dingle, Mark Wilcox, A Sarah Walker, Derrick W Crook, Tim E A Peto, and Rosalind M Harding. Microevolutionary analysis of Clostridium difficile genomes to investigate transmission. *Genome Biology*, 13(12):R118, December 2012. ISSN 1474-760X. doi:10.1186/gb-2012-13-12-r118.

[34] Brittany B. Lewis, Rebecca A. Carter, Lilan Ling, Ingrid Leiner, Ying Taur, Mini Kamboj, Erik R. Dubberke, Joao Xavier, and Eric G. Pamer. Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to Clostridium difficile Virulence. *mBio*, 8(4):10.1128/mbio.00885–17, August 2017. doi:10.1128/mbio.00885-17.

[35] Skye R.S. Fishbein, Anna L. DeVeaux, Sakshi Khanna, Aura L. Ferreiro, James Liao, Wesley Agee, Jie Ning, Bejan Mahmud, Miranda J. Wallace, Tiffany Hink, Kimberly A. Reske, Candice Cass, Janaki Guruge, Sidh Leekha, Sunaina Rengarajan, Erik R. Dubberke, and Gautam Dantas. Commensal-pathogen dynamics structure disease outcomes during Clostridioides difficile colonization. *Cell Host & Microbe*, page S1931312824004475, December 2024. ISSN 19313128. doi:10.1016/j.chom.2024.12.002.

[36] Mohammed Sebaihia, Brendan W Wren, Peter Mullany, Neil F Fairweather, Nigel Minton, Richard Stabler, Nicholas R Thomson, Adam P Roberts, Ana M Cerdeño-Tárraga, Hongmei Wang, Matthew Tg Holden, Anne Wright, Carol Churcher, Michael A Quail, Stephen Baker, Nathalie Bason, Karen Brooks, Tracey Chillingworth, Ann Cronin, Paul Davis, Linda Dowd, Audrey Fraser, Theresa Feltwell, Zahra Hance, Simon Holroyd, Kay Jagels, Sharon Moule, Karen Mungall, Claire Price, Ester Rabbinowitsch, Sarah Sharp, Mark Simmonds, Kim Stevens, Louise Unwin, Sally Whithead, Bruno Dupuy, Gordon Dougan, Bart Barrell, and Julian Parkhill. The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature Genetics*, 38(7):779–786, July 2006. ISSN 1061-4036, 1546-1718. doi:10.1038/ng1830.

[37] Michael S. M. Brouwer, Philip J. Warburton, Adam P. Roberts, Peter Mullany, and Elaine Allan. Genetic Organisation, Mobility and Predicted Functions of Genes on Integrated, Mobile Genetic Elements in Sequenced Strains of Clostridium difficile. *PLoS ONE*, 6(8):e23014, August 2011. ISSN 1932-6203. doi:10.1371/journal.pone.0023014.

[38] R. A. Stabler, D. N. Gerding, J. G. Songer, D. Drudy, J. S. Brazier, H. T. Trinh, A. A. Witney, J. Hinds, and B. W. Wren. Comparative Phylogenomics of Clostridium difficile Reveals Clade Specificity and Microevolution of Hypervirulent Strains. *Journal of Bacteriology*, 188(20):7297–7305, October 2006. doi:10.1128/jb.00664-06.

[39] Maria Kulecka, Edyta Waker, Filip Ambrozkiewicz, Agnieszka Paziewska, Karolina Skubisz, Patrycja Cybula, Łukasz Targoński, Michał Mikula, Jan Walewski, and Jerzy Ostrowski. Higher genome variability within metabolism genes associates with recurrent Clostridium difficile infection. *BMC Microbiology*, 21(1):36, January 2021. ISSN 1471-2180. doi:10.1186/s12866-021-02090-9.

[40] Charles J. Norsigian, Heather A. Danhof, Colleen K. Brand, Numan Oezguen, Firas S. Midani, Bernhard O. Palsson, Tor C. Savidge, Robert A. Britton, Jennifer K. Spinler, and Jonathan M. Monk. Systems biology analysis of the Clostridioides difficile core-genome contextualizes microenvironmental evolutionary pressures leading to genotypic and phenotypic divergence. *npj Systems Biology and Applications*, 6(1):31, October 2020. ISSN 2056-7189. doi:10.1038/s41540-020-00151-9.

[41] Hailong Zhao, David C. Nickle, Zhen Zeng, Pierra Y. T. Law, Mark H. Wilcox, Lan Chen, Ye Peng, Jie Meng, Ziqing Deng, Andrew Albright, Huanzi Zhong, Xun Xu, Shida Zhu, Judong Shen, Rebecca L. Blanchard, Mary Beth Dorr, Peter M. Shaw, and Junhua Li. Global Landscape of Clostridioides Difficile Phylogeography, Antibiotic Susceptibility, and Toxin Polymorphisms by Post-Hoc Whole-Genome Sequencing from the MODIFY I/II Studies. *Infectious Diseases and Therapy*, 10(2):853–870, June 2021. ISSN 2193-8229, 2193-6382. doi:10.1007/s40121-021-00426-6.

[42] Charles J. Norsigian, Heather A. Danhof, Colleen K. Brand, Firas S. Midani, Jared T. Broddrick, Tor C. Savidge, Robert A. Britton, Bernhard O. Palsson, Jennifer K. Spinler, and Jonathan M. Monk. Systems biology approach to functionally assess the

Clostridioides difficile pangenome reveals genetic diversity with discriminatory power. *Proceedings of the National Academy of Sciences*, 119(18):e2119396119, May 2022. doi:10.1073/pnas.2119396119.

[43] M Delmee, M Homel, and G Wauters. Serogrouping of Clostridium difficile strains by slide agglutination. *Journal of Clinical Microbiology*, 21(3):323–327, March 1985. doi:10.1128/jcm.21.3.323-327.1985.

[44] David C. Schwartz and Charles R. Cantor. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1):67–75, May 1984. ISSN 00928674. doi:10.1016/0092-8674(84)90301-5.

[45] A. E. Simor, S. L. Yake, and K. Tsimidis. Infection Due to Clostridium difficile Among Elderly Residents of a Long-Term-Care Facility. *Clinical Infectious Diseases*, 17(4): 672–678, October 1993. ISSN 1058-4838, 1537-6591. doi:10.1093/clinids/17.4.672.

[46] H Kato, N Kato, K Watanabe, K Ueno, H Ushijima, S Hashira, and T Abe. Application of typing by pulsed-field gel electrophoresis to the study of Clostridium difficile in a neonatal intensive care unit. *Journal of Clinical Microbiology*, 32(9):2067–2070, September 1994. ISSN 0095-1137, 1098-660X. doi:10.1128/jcm.32.9.2067-2070.1994.

[47] Haru Kato, Naoki Kato, Kunitomo Watanabe, Kazue Ueno, Yoko Sakata, and Kozo Fujita. Relapses or reinfections: Analysis of a case of Clostridium difficile-associated colitis by two typing systems. *Current Microbiology*, 33(4):220–223, October 1996. ISSN 0343-8651, 1432-0991. doi:10.1007/s002849900103.

[48] Vivian G Loo, Matthew Oughton, Anne-Marie Bourgault, Mirabelle Kelly, Ken Dewar, and Yury Monczak. A Predominantly Clonal Multi-Institutional Outbreak of Clostridium difficile–Associated Diarrhea with High Morbidity and Mortality. *n engl j med*, 2005.

[49] Charlotte A. Huber, Niki F. Foster, Thomas V. Riley, and David L. Paterson. Challenges for Standardization of Clostridium difficile Typing Methods. *Journal of Clinical Microbiology*, 51(9):2810–2814, September 2013. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.00143-13.

[50] Nourkhoda Sadeghifard, Volker Gürtler, Michael Beer, and Robert J. Seviour. The Mosaic Nature of Intergenic 16S-23S rRNA Spacer Regions Suggests rRNA Operon Copy Number Variation in Clostridium difficile Strains. *Applied and Environmental Microbiology*, 72(11):7311–7323, November 2006. doi:10.1128/AEM.01179-06.

[51] Volker Gütler. Typing of Clostridium difficile strains by PCR-amplification of variable length 16S-23S rDNA spacer regions. *Microbiology*, 139(12):3089–3097, 1993. ISSN 1465-2080. doi:10.1099/00221287-139-12-3089.

[52] Ed J. Kuijper, Renate J. van den Berg, and Jon S. Brazier. Comparison of Molecular Typing Methods Applied to Clostridium difficile. In Dominique A. Caugant, editor,

Molecular Epidemiology of Microorganisms: Methods and Protocols, pages 159–171. Humana Press, Totowa, NJ, 2009. ISBN 978-1-60327-999-4. doi:10.1007/978-1-60327-999-4_13.

[53] David Griffiths, Warren Fawley, Melina Kachrimanidou, Rory Bowden, Derrick W. Crook, Rowena Fung, Tanya Golubchik, Rosalind M. Harding, Katie J. M. Jeffery, Keith A. Jolley, Richard Kirton, Tim E. Peto, Gareth Rees, Nicole Stoesser, Alison Vaughan, A. Sarah Walker, Bernadette C. Young, Mark Wilcox, and Kate E. Dingle. Multilocus Sequence Typing of *Clostridium difficile*. *Journal of Clinical Microbiology*, 48(3):770–778, March 2010. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.01796-09.

[54] Daniel R. Knight, Briony Elliott, Barbara J. Chang, Timothy T. Perkins, and Thomas V. Riley. Diversity and Evolution in the Genome of Clostridium difficile. *Clinical Microbiology Reviews*, 28(3):721–741, June 2015. doi:10.1128/cmr.00127-14.

[55] Mini Kamboj, Tracy McMillen, Mustafa Syed, Hoi Yan Chow, Krupa Jani, Anoshe Aslam, Jennifer Brite, Brian Fanelli, Nur A. Hasan, Manoj Dadlani, Lars Westblade, Ahmet Zehir, Matthew Simon, and N. Esther Babady. Evaluation of a Combined Multilocus Sequence Typing and Whole-Genome Sequencing Two-Step Algorithm for Routine Typing of Clostridioides difficile. *Journal of Clinical Microbiology*, 59(2): 10.1128/jcm.01955–20, January 2021. doi:10.1128/jcm.01955-20.

[56] Keith A. Jolley, James E. Bray, and Martin C. J. Maiden. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Research*, 3:124, 2018. ISSN 2398-502X. doi:10.12688/wellcomeopenres.14826.1.

[57] Lars F. Westblade, Alex Van Belkum, Adam Grundhoff, George M. Weinstock, Eric G. Pamer, Mark J. Pallen, and W. Michael Dunne. Role of Clinicogenomics in Infectious Disease Diagnostics and Public Health Microbiology. *Journal of Clinical Microbiology*, 54(7):1686–1693, July 2016. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.02664-15.

[58] Arianna Miles-Jay, Vincent B. Young, Eric G. Pamer, Tor C. Savidge, Mini Kamboj, Kevin W. Garey, and Evan S. Snitkin. A multisite genomic epidemiology study of Clostridioides difficile infections in the USA supports differential roles of healthcare versus community spread for two common strains. *Microbial Genomics*, 7(6), March 2021. ISSN 2057-5858. doi:10.1099/mgen.0.000590.

[59] Arianna Miles-Jay, Evan S. Snitkin, Michael Y. Lin, Teppei Shimasaki, Michael Schoeny, Christine Fukuda, Thelma Dangana, Nicholas Moore, Sarah E. Sansom, Rachel D. Yelin, Pamela Bell, Krishna Rao, Micah Keidan, Alexandra Standke, Christine Bassis, Mary K. Hayden, and Vincent B. Young. Longitudinal genomic surveillance of carriage and transmission of Clostridioides difficile in an intensive care unit. *Nature Medicine*, 29(10):2526–2534, October 2023. ISSN 1078-8956, 1546-170X. doi:10.1038/s41591-023-02549-4.

[60] Erin P. Newcomer, Skye R. S. Fishbein, Kailun Zhang, Tiffany Hink, Kimberly A. Reske, Candice Cass, Zainab H. Iqbal, Emily L. Struttmann, Carey-Ann D. Burnham, Erik R. Dubberke, and Gautam Dantas. Genomic surveillance of *Clostridioides difficile* transmission and virulence in a healthcare setting. *mBio*, 15(3):e03300–23, March 2024. ISSN 2150-7511. doi:10.1128/mbio.03300-23.

[61] Melissa A. Kinnebrew, Yeon Joo Lee, Robert R. Jenq, Lauren Lipuma, Eric R. Littmann, Asia Gobourne, Daniel No, Marcel Van Den Brink, Eric G. Pamer, and Ying Taur. Early Clostridium difficile Infection during Allogeneic Hematopoietic Stem Cell Transplantation. *PLoS ONE*, 9(3):e90158, March 2014. ISSN 1932-6203. doi:10.1371/journal.pone.0090158.

[62] Francesco Beghini, Lauren J McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, Paolo Manghi, Matthias Scholz, Andrew Maltez Thomas, Mireia Valles-Colomer, George Weingart, Yancong Zhang, Moreno Zolfo, Curtis Huttenhower, Eric A Franzosa, and Nicola Segata. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*, 10:e65088, May 2021. ISSN 2050-084X. doi:10.7554/eLife.65088.

[63] Aitor Blanco-Míguez, Francesco Beghini, Fabio Cumbo, Lauren J. McIver, Kelsey N. Thompson, Moreno Zolfo, Paolo Manghi, Leonard Dubois, Kun D. Huang, Andrew Maltez Thomas, William A. Nickols, Gianmarco Piccinno, Elisa Piperni, Michal Punčochář, Mireia Valles-Colomer, Adrian Tett, Francesca Giordano, Richard Davies, Jonathan Wolf, Sarah E. Berry, Tim D. Spector, Eric A. Franzosa, Edoardo Pasolli, Francesco Asnicar, Curtis Huttenhower, and Nicola Segata. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, 41(11):1633–1644, November 2023. ISSN 1087-0156, 1546-1696. doi:10.1038/s41587-023-01688-w.

[64] Ben Langmead and Steven L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359, April 2012. ISSN 1548-7105. doi:10.1038/nmeth.1923.

[65] Francesco Asnicar, Andrew Maltez Thomas, Francesco Beghini, Claudia Mengoni, Serena Manara, Paolo Manghi, Qiyun Zhu, Mattia Bolzan, Fabio Cumbo, Uyen May, Jon G. Sanders, Moreno Zolfo, Evguenia Kopylova, Edoardo Pasolli, Rob Knight, Siavash Mirarab, Curtis Huttenhower, and Nicola Segata. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications*, 11(1):2500, May 2020. ISSN 2041-1723. doi:10.1038/s41467-020-16366-7.

[66] K. Katoh. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066, July 2002. ISSN 13624962. doi:10.1093/nar/gkf436.

[67] Alexandros Stamatakis. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, May 2014. ISSN 1367-4803. doi:10.1093/bioinformatics/btu033.

[68] Alan F. Hofmann. The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Archives of Internal Medicine*, 159(22):2647, December 1999. ISSN 0003-9926. doi:10.1001/archinte.159.22.2647.

[69] Heidi L. Doden and Jason M. Ridlon. Microbial Hydroxysteroid Dehydrogenases: From Alpha to Omega. *Microorganisms*, 9(3):469, March 2021. ISSN 2076-2607. doi:10.3390/microorganisms9030469.

[70] Maria Chiara Di Gregorio, Jacopo Cautela, and Luciano Galantini. Physiology and Physical Chemistry of Bile Acids. *International Journal of Molecular Sciences*, 22(4): 1780, February 2021. ISSN 1422-0067. doi:10.3390/ijms22041780.

[71] Jason M. Ridlon and H. Rex Gaskins. Another renaissance for bile acid gastrointestinal microbiology. *Nature Reviews Gastroenterology & Hepatology*, February 2024. ISSN 1759-5045, 1759-5053. doi:10.1038/s41575-024-00896-2.

[72] Alan F. Hofmann and Lee R. Hagey. Key discoveries in bile acid chemistry and biology and their clinical applications: History of the last eight decades. *Journal of Lipid Research*, 55(8):1553–1595, August 2014. ISSN 0022-2275. doi:10.1194/jlr.R049437.

[73] Robert A. Quinn, Alexey V. Melnik, Alison Vrbanac, Ting Fu, Kathryn A. Patras, Mitchell P. Christy, Zsolt Bodai, Pedro Belda-Ferre, Anupriya Tripathi, Lawton K. Chung, Michael Downes, Ryan D. Welch, Melissa Quinn, Greg Humphrey, Morgan Panitchpakdi, Kelly C. Weldon, Alexander Aksenov, Ricardo da Silva, Julian Avila-Pacheco, Clary Clish, Sena Bae, Himel Mallick, Eric A. Franzosa, Jason Lloyd-Price, Robert Bussell, Taren Thron, Andrew T. Nelson, Mingxun Wang, Eric Leszczynski, Fernando Vargas, Julia M. Gauglitz, Michael J. Meehan, Emily Gentry, Timothy D. Arthur, Alexis C. Komor, Orit Poulsen, Brigid S. Boland, John T. Chang, William J. Sandborn, Meerana Lim, Neha Garg, Julie C. Lumeng, Ramnik J. Xavier, Barbara I. Kazmierczak, Ruchi Jain, Marie Egan, Kyung E. Rhee, David Ferguson, Manuela Raffatellu, Hera Vlamakis, Gabriel G. Haddad, Dionicio Siegel, Curtis Huttenhower, Sarkis K. Mazmanian, Ronald M. Evans, Victor Nizet, Rob Knight, and Pieter C. Dorrestein. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature*, 579(7797):123–129, March 2020. ISSN 1476-4687. doi:10.1038/s41586-020-2047-9.

[74] Bipin Rimal, Stephanie L. Collins, Ceylan E. Tanes, Edson R. Rocha, Megan A. Granda, Sumeet Solanki, Nushrat J. Hoque, Emily C. Gentry, Imhoi Koo, Erin R. Reilly, Fuhua Hao, Devendra Paudel, Vishal Singh, Tingting Yan, Min Soo Kim, Kyle Bittinger, Joseph P. Zackular, Kristopher W. Krausz, Dhimant Desai, Shantu Amin, James P. Coleman, Yatrik M. Shah, Jordan E. Bisanz, Frank J. Gonzalez, John P. Vanden Heuvel, Gary D. Wu, Babette S. Zemel, Pieter C. Dorrestein, Emily E. Weinert,

and Andrew D. Patterson. Bile salt hydrolase catalyses formation of amine-conjugated bile acids. *Nature*, February 2024. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-023-06990-w.

[75] Jae Won Lee, Elise S. Cowley, Patricia G. Wolf, Heidi L. Doden, Tsuyoshi Murai, Kelly Yovani Olivos Caicedo, Lindsey K. Ly, Furong Sun, Hajime Takei, Hiroshi Nittono, Steven L. Daniel, Isaac Cann, H. Rex Gaskins, Karthik Anantharaman, João M. P. Alves, and Jason M. Ridlon. Formation of secondary allo-bile acids by novel enzymes from gut Firmicutes. *Gut Microbes*, 14(1):2132903, December 2022. ISSN 1949-0976. doi:10.1080/19490976.2022.2132903.

[76] Yuko Sato, Koji Atarashi, Damian R. Plichta, Yasumichi Arai, Satoshi Sasajima, Sean M. Kearney, Wataru Suda, Kozue Takeshita, Takahiro Sasaki, Shoki Okamoto, Ashwin N. Skelly, Yuki Okamura, Hera Vlamakis, Youxian Li, Takeshi Tanoue, Hajime Takei, Hiroshi Nittono, Seiko Narushima, Junichiro Irie, Hiroshi Itoh, Kyoji Moriya, Yuki Sugiura, Makoto Suematsu, Nobuko Moritoki, Shinsuke Shibata, Dan R. Littman, Michael A. Fischbach, Yoshifumi Uwamino, Takashi Inoue, Akira Honda, Masahira Hattori, Tsuyoshi Murai, Ramnik J. Xavier, Nobuyoshi Hirose, and Kenya Honda. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature*, 599(7885):458–464, November 2021. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-021-03832-5.

[77] S M Huijghebaert, J A Mertens, and H J Eyssen. Isolation of a bile salt sulfatase-producing Clostridium strain from rat intestinal microflora. *Applied and Environmental Microbiology*, 43(1):185–192, January 1982. ISSN 0099-2240, 1098-5336. doi:10.1128/aem.43.1.185-192.1982.

[78] J Robben, G Parmentier, and H Eyssen. Isolation of a rat intestinal Clostridium strain producing 5 alpha- and 5 beta-bile salt 3 alpha-sulfatase activity. *Applied and Environmental Microbiology*, 51(1):32–38, January 1986. ISSN 0099-2240, 1098-5336. doi:10.1128/aem.51.1.32-38.1986.

[79] Joseph A. Sorg and Abraham L. Sonenshein. Bile Salts and Glycine as Cogerminants for *Clostridium difficile* Spores. *Journal of Bacteriology*, 190(7):2505–2512, April 2008. ISSN 0021-9193, 1098-5530. doi:10.1128/JB.01765-07.

[80] Susanne Sievers, Nicole G. Metzendorf, Silvia Dittmann, Daniel Troitzsch, Viola Gast, Sophie Marlen Tröger, Christian Wolff, Daniela Zühlke, Claudia Hirschfeld, Rabea Schlüter, and Katharina Riedel. Differential View on the Bile Acid Stress Response of Clostridioides difficile. *Frontiers in Microbiology*, 10:258, February 2019. ISSN 1664-302X. doi:10.3389/fmicb.2019.00258.

[81] Nicole G. Metzendorf, Lena Melanie Lange, Nina Lainer, Rabea Schlüter, Silvia Dittmann, Lena-Sophie Paul, Daniel Troitzsch, and Susanne Sievers. Destination

and Specific Impact of Different Bile Acids in the Intestinal Pathogen Clostridioides difficile. *Frontiers in Microbiology*, 13:814692, March 2022. ISSN 1664-302X. doi:10.3389/fmicb.2022.814692.

[82] Andrea Martinez Aguirre, Adegoke Oyeleye Adegbite, and Joseph A. Sorg. Clostridioides difficile bile salt hydrolase activity has substrate specificity and affects biofilm formation. *npj Biofilms and Microbiomes*, 8(1):94, November 2022. ISSN 2055-5008. doi:10.1038/s41522-022-00358-0.

[83] Yusuke Yoshino, Takatoshi Kitazawa, Mahoko Ikeda, Keita Tatsuno, Shintaro Yanagimoto, Shu Okugawa, Hiroshi Yotsuyanagi, and Yasuo Ota. Clostridium difficile flagellin stimulates toll-like receptor 5, and toxin B promotes flagellin-induced chemokine production via TLR5. *Life Sciences*, 92(3):211–217, February 2013. ISSN 00243205. doi:10.1016/j.lfs.2012.11.017.

[84] Anna L. Gregory, Daniel A. Pensinger, and Andrew J. Hryckowian. A short chain fatty acid–centric view of Clostridioides difficile pathogenesis. *PLOS Pathogens*, 17 (10):e1009959, October 2021. ISSN 1553-7374. doi:10.1371/journal.ppat.1009959.

[85] Simonida Gencic and David A. Grahame. Diverse Energy-Conserving Pathways in Clostridium difficile: Growth in the Absence of Amino Acid Stickland Acceptors and the Role of the Wood-Ljungdahl Pathway. *Journal of Bacteriology*, 202(20): 10.1128/jb.00233–20, September 2020. doi:10.1128/jb.00233-20.

[86] José Luís Fachi, Cristiane Sécca, Patrícia Brito Rodrigues, Felipe Cézar Pinheiro De Mato, Blanda Di Luccia, Jaqueline De Souza Felipe, Laís Passariello Pral, Marcella Rungue, Victor De Melo Rocha, Fabio Takeo Sato, Ulliana Sampaio, Maria Teresa Pedrosa Silva Clerici, Hosana Gomes Rodrigues, Niels Olsen Saraiva Câmara, Sílvio Roberto Consonni, Angélica Thomaz Vieira, Sergio Costa Oliveira, Charles Reay Mackay, Brian T. Layden, Karina Ramalho Bortoluci, Marco Colonna, and Marco Aurélio Ramirez Vinolo. Acetate coordinates neutrophil and ILC3 responses against *C. difficile* through FFAR2. *Journal of Experimental Medicine*, 217(3):e20190489, March 2020. ISSN 0022-1007, 1540-9538. doi:10.1084/jem.20190489.

[87] Li Fang Lu, Dae Hong Kim, Ik Hwan Lee, Ji Hong, Peng Zhang, I. Na Yoon, Jae Sam Hwang, and Ho Kim. Potassium Acetate Blocks Clostridium difficile Toxin A-Induced Microtubule Disassembly by Directly Inhibiting Histone Deacetylase 6, Thereby Ameliorating Inflammatory Responses in the Gut. 26(4):693–699, April 2016. doi:10.4014/jmb.1511.11063.

[88] Petra Louis, Sheila I. McCrae, Cédric Charrier, and Harry J. Flint. Organization of butyrate synthetic genes in human colonic bacteria: Phylogenetic conservation and horizontal gene transfer. *FEMS microbiology letters*, 269(2):240–247, April 2007. ISSN 0378-1097. doi:10.1111/j.1574-6968.2006.00629.x.

[89] Yael Litvak, Mariana X. Byndloss, and Andreas J. Bäumler. Colonocyte metabolism shapes the gut microbiota. *Science*, 362(6418):eaat9076, November 2018. doi:10.1126/science.aat9076.

[90] José Luís Fachi, Jaqueline De Souza Felipe, Laís Passariello Pral, Bruna Karadi Da Silva, Renan Oliveira Corrêa, Mirella Cristiny Pereira De Andrade, Denise Morais Da Fonseca, Paulo José Basso, Niels Olsen Saraiva Câmara, Éricka Lorenna De Sales E Souza, Flaviano Dos Santos Martins, Suzana Eiko Sato Guima, Andrew Maltez Thomas, João Carlos Setubal, Yuli Thamires Magalhães, Fábio Luis Forti, Thamiris Candreva, Hosana Gomes Rodrigues, Marcelo Bispo De Jesus, Sílvio Roberto Consonni, Alessandro Dos Santos Farias, Patrick Varga-Weisz, and Marco Aurélio Ramirez Vinolo. Butyrate Protects Mice from Clostridium difficile-Induced Colitis through an HIF-1-Dependent Mechanism. *Cell Reports*, 27(3):750–761.e7, April 2019. ISSN 22111247. doi:10.1016/j.celrep.2019.03.054.

[91] Atsushi Hayashi, Hiroko Nagao-Kitamoto, Sho Kitamoto, Chang H Kim, and Nobuhiko Kamada. The Butyrate-Producing Bacterium Clostridium butyricum Suppresses Clostridioides difficile Infection via Neutrophil- and Antimicrobial Cytokine–Dependent but GPR43/109a-Independent Mechanisms. *The Journal of Immunology*, 206(7):1576–1585, April 2021. ISSN 0022-1767. doi:10.4049/jimmunol.2000353.

[92] Jessica A. Ferreyra, Katherine J. Wu, Andrew J. Hryckowian, Donna M. Bouley, Bart C. Weimer, and Justin L. Sonnenburg. Gut Microbiota-Produced Succinate Promotes C. difficile Infection after Antibiotic Treatment or Motility Disturbance. *Cell Host & Microbe*, 16(6):770–777, December 2014. ISSN 19313128. doi:10.1016/j.chom.2014.11.003.

[93] Swadha Anand, Harrisham Kaur, and Sharmila S. Mande. Comparative In silico Analysis of Butyrate Production Pathways in Gut Commensals and Pathogens. *Frontiers in Microbiology*, 7, December 2016. ISSN 1664-302X. doi:10.3389/fmicb.2016.01945.

[94] Susan Hromada, Yili Qian, Tyler B Jacobson, Ryan L Clark, Lauren Watson, Nasia Safdar, Daniel Amador-Noguez, and Ophelia S Venturelli. Negative interactions determine Clostridioides difficile growth in synthetic human gut communities. *Molecular Systems Biology*, 17(10):e10355, October 2021. ISSN 1744-4292. doi:10.15252/msb.202110355.

[95] Christopher A. Lopez, Tess P. McNeely, Kamila Nurmakova, William N. Beavers, and Eric P. Skaar. Clostridioides difficile proline fermentation in response to commensal clostridia. *Anaerobe*, 63, June 2020. ISSN 10958274. doi:10.1016/j.anaerobe.2020.102210. Publisher: Academic Press.

[96] Andrea Martinez Aguirre and Joseph A. Sorg. Gut associated metabolites and their roles in Clostridioides difficile pathogenesis. *https://doi.org/10.1080/19490976.2022.2094672*, 14(1), December 2022. ISSN

1949-0976. doi:10.1080/19490976.2022.2094672. URL https://www.tandfonline.com/doi/abs/10.1080/19490976.2022.2094672. Publisher: Taylor & Francis.

[97] Aidan Pavao, Madeline Graham, Mario Arrieta-Ortiz, Selva Rupa Christinal Immanuel, Nitin S. Baliga, and Lynn Bry. Reconsidering the in vivo functions of Clostridial Stickland amino acid fermentations. *Anaerobe*, 76:102600, August 2022. ISSN 1075-9964. doi:10.1016/j.anaerobe.2022.102600.

[98] Eric J. Battaglioli, Vanessa L. Hale, Jun Chen, Patricio Jeraldo, Coral Ruiz-Mojica, Bradley A. Schmidt, Vayu M. Rekdal, Lisa M. Till, Lutfi Huq, Samuel A. Smits, William J. Moor, Yava Jones-Hall, Thomas Smyrk, Sahil Khanna, Darrell S. Pardi, Madhusudan Grover, Robin Patel, Nicholas Chia, Heidi Nelson, Justin L. Sonnenburg, Gianrico Farrugia, and Purna C. Kashyap. *Clostridioides Difficile* uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. *Science Translational Medicine*, 10(464):eaam7019, October 2018. ISSN 1946-6234, 1946-6242. doi:10.1126/scitranslmed.aam7019.

[99] Julia D. Hofmann, Andreas Otto, Mareike Berges, Rebekka Biedendieck, Annika-Marisa Michel, Dörte Becher, Dieter Jahn, and Meina Neumann-Schaal. Metabolic Reprogramming of Clostridioides difficile During the Stationary Phase With the Induction of Toxin Production. *Frontiers in Microbiology*, 9, August 2018. ISSN 1664-302X. doi:10.3389/fmicb.2018.01970.

[100] Laurent Bouillaut, William T. Self, and Abraham L. Sonenshein. Proline-Dependent Regulation of Clostridium difficile Stickland Metabolism. *Journal of Bacteriology*, 195 (4):844–854, January 2013. doi:10.1128/jb.01492-12.

[101] T. Karasawa, S. Ikoma, K. Yamakawa, and S. Nakamura. A defined growth medium for Clostridium difficile. *Microbiology*, 141(2):371–375, February 1995. ISSN 1350-0872, 1465-2080. doi:10.1099/13500872-141-2-371.

[102] Caitlin C. Murdoch and Eric P. Skaar. Nutritional immunity: The battle for nutrient metals at the host–pathogen interface. *Nature Reviews Microbiology*, 20(11):657–670, November 2022. ISSN 1740-1534. doi:10.1038/s41579-022-00745-6.

[103] Theresa D. Ho and Craig D. Ellermeier. Ferric Uptake Regulator Fur Control of Putative Iron Acquisition Systems in Clostridium difficile. *Journal of Bacteriology*, 197(18):2930–2940, August 2015. doi:10.1128/jb.00098-15.

[104] Mareike Berges, Annika-Marisa Michel, Christian Lassek, Aaron M. Nuss, Michael Beckstette, Petra Dersch, Katharina Riedel, Susanne Sievers, Dörte Becher, Andreas Otto, Sandra Maaß, Manfred Rohde, Denitsa Eckweiler, Jose M. Borrero-de Acuña, Martina Jahn, Meina Neumann-Schaal, and Dieter Jahn. Iron Regulation in Clostridioides difficile. *Frontiers in Microbiology*, 9, 2018. ISSN 1664-302X.

[105] Elaine R. Frawley and Ferric C. Fang. The Ins and Outs of Bacterial Iron Metabolism. *Molecular microbiology*, 93(4):609–616, August 2014. ISSN 0950-382X. doi:10.1111/mmi.12709.

[106] Christopher A. Lopez, William N. Beavers, Andy Weiss, Reece J. Knippel, Joseph P. Zackular, Walter Chazin, and Eric P. Skaar. The Immune Protein Calprotectin Impacts Clostridioides difficile Metabolism through Zinc Limitation. *mBio*, 10(6):e02289–19, December 2019. ISSN 2161-2129, 2150-7511. doi:10.1128/mBio.02289-19.

[107] Verónica Urdaneta and Josep Casadesús. Interactions between Bacteria and Bile Salts in the Gastrointestinal and Hepatobiliary Tracts. *Frontiers in Medicine*, 4:163, October 2017. ISSN 2296-858X. doi:10.3389/fmed.2017.00163.

[108] Matthew R. Stutz, Nicholas P. Dylla, Steven D. Pearson, Paola Lecompte-Osorio, Ravi Nayak, Maryam Khalid, Emerald Adler, Jaye Boissiere, Huaiying Lin, William Leiter, Jessica Little, Amber Rose, David Moran, Michael W. Mullowney, Krysta S. Wolfe, Christopher Lehmann, Matthew Odenwald, Mark De La Cruz, Mihai Giurcanu, Anne S. Pohlman, Jesse B. Hall, Jean-Luc Chaubard, Anitha Sundararajan, Ashley Sidebottom, John P. Kress, Eric G. Pamer, and Bhakti K. Patel. Immunomodulatory fecal metabolites are associated with mortality in COVID-19 patients with respiratory failure. *Nature Communications*, 13(1):6615, November 2022. ISSN 2041-1723. doi:10.1038/s41467-022-34260-2.

[109] Christopher J. Lehmann, Nicholas P. Dylla, Matthew Odenwald, Ravi Nayak, Maryam Khalid, Jaye Boissiere, Jackelyn Cantoral, Emerald Adler, Matthew R. Stutz, Mark Dela Cruz, Angelica Moran, Huaiying Lin, Ramanujam Ramaswamy, Anitha Sundararajan, Ashley M. Sidebottom, Jessica Little, Eric G. Pamer, Andrew Aronsohn, John Fung, Talia B. Baker, and Aalok Kacha. Fecal metabolite profiling identifies liver transplant recipients at risk for postoperative infection. *Cell Host & Microbe*, 32(1): 117–130.e4, January 2024. ISSN 19313128. doi:10.1016/j.chom.2023.11.016.

[110] Matthew A. Odenwald, Huaiying Lin, Christopher Lehmann, Nicholas P. Dylla, Cody G. Cole, Jake D. Mostad, Téa E. Pappas, Ramanujam Ramaswamy, Angelica Moran, Alan L. Hutchison, Matthew R. Stutz, Mark Dela Cruz, Emerald Adler, Jaye Boissiere, Maryam Khalid, Jackelyn Cantoral, Fidel Haro, Rita A. Oliveira, Emily Waligurski, Thomas G. Cotter, Samuel H. Light, Kathleen G. Beavis, Anitha Sundararajan, Ashley M. Sidebottom, K. Gautham Reddy, Sonali Paul, Anjana Pillai, Helen S. Te, Mary E. Rinella, Michael R. Charlton, Eric G. Pamer, and Andrew I. Aronsohn. Bifidobacteria metabolize lactulose to optimize gut metabolites and prevent systemic infection in patients with liver disease. *Nature Microbiology*, 8(11):2033–2049, October 2023. ISSN 2058-5276. doi:10.1038/s41564-023-01493-w.

[111] Mark Dela Cruz, Huaiying Lin, Jiho Han, Emerald Adler, Jaye Boissiere, Maryam Khalid, Ashley Sidebottom, Anitha Sundararajan, Christopher Lehmann, Angelica

Moran, Matthew Odenwald, Matthew Stutz, Gene Kim, Sean Pinney, Valluvan Jeevanandam, Maria-Luisa Alegre, Eric Pamer, and Ann B. Nguyen. Reduced immunomodulatory metabolite concentrations in peri-transplant fecal samples from heart allograft recipients. *Frontiers in Transplantation*, 2:1182534, July 2023. ISSN 2813-2440. doi:10.3389/frtra.2023.1182534.

[112] Ashley Paulick, Michelle Adamczyk, Karen Anderson, Nick Vlachos, María-José Machado, Gillian McAllister, Lauren Korhonen, Alice Y. Guh, Alison Laufer Halpin, J. Kamile Rasheed, Maria Karlsson, Joseph D. Lutgring, Amy S. Gargis, and the Emerging Infections Program Clostridioides difficile Pathogen Group. Characterization of Clostridioides difficile Isolates Available through the CDC & FDA Antibiotic Resistance Isolate Bank. *Microbiology Resource Announcements*, 10(1):e01011–20, January 2021. ISSN 2576-098X. doi:10.1128/MRA.01011-20.

[113] Wan Yu. readsimulatorv0.0.1. `https://github.com/wanyuac/readSimulator`, 2019.

[114] Heng Li. seqtkv1.4. `https://github.com/lh3/seqtk`, 2023.

[115] Tony Bolger. trimmomatic.0.39. `https://github.com/usadellab/Trimmomatic`, 2021.

[116] Curtis Huttenhower, Robert D. Finn, and Alice Carolyn McHardy. Challenges and opportunities in sharing microbiome data and analyses. *Nature Microbiology*, 8(11): 1960–1970, October 2023. ISSN 2058-5276. doi:10.1038/s41564-023-01484-x.

[117] Daniel R. Knight, Brian Kullin, Grace O. Androga, Frederic Barbut, Catherine Eckert, Stuart Johnson, Patrizia Spigaglia, Kazuhiro Tateda, Pei-Jane Tsai, and Thomas V. Riley. Evolutionary and Genomic Insights into Clostridioides difficile Sequence Type 11: A Diverse Zoonotic and Antimicrobial-Resistant Lineage of Global One Health Importance. *mBio*, 10(2):e00446–19, April 2019. ISSN 2150-7511. doi:10.1128/mBio.00446-19.

[118] Fernando Meyer, Andreas Bremges, Peter Belmann, Stefan Janssen, Alice C. McHardy, and David Koslicki. Assessing taxonomic metagenome profilers with OPAL. *Genome Biology*, 20(1):51, March 2019. ISSN 1474-760X. doi:10.1186/s13059-019-1646-y.

[119] Fernando Meyer, Adrian Fritz, Zhi-Luo Deng, David Koslicki, Till Robin Lesker, Alexey Gurevich, Gary Robertson, Mohammed Alser, Dmitry Antipov, Francesco Beghini, Denis Bertrand, Jaqueline J. Brito, C. Titus Brown, Jan Buchmann, Aydin Buluç, Bo Chen, Rayan Chikhi, Philip T. L. C. Clausen, Alexandru Cristian, Piotr Wojciech Dabrowski, Aaron E. Darling, Rob Egan, Eleazar Eskin, Evangelos Georganas, Eugene Goltsman, Melissa A. Gray, Lars Hestbjerg Hansen, Steven Hofmeyr, Pingqin Huang, Luiz Irber, Huijue Jia, Tue Sparholt Jørgensen, Silas D. Kieser, Terje Klemetsen, Axel Kola, Mikhail Kolmogorov, Anton Korobeynikov, Jason Kwan, Nathan LaPierre, Claire Lemaitre, Chenhao Li, Antoine Limasset, Fabio Malcher-Miranda, Serghei Mangul, Vanessa R. Marcelino, Camille Marchet, Pierre Marijon, Dmitry

Meleshko, Daniel R. Mende, Alessio Milanese, Niranjan Nagarajan, Jakob Nissen, Sergey Nurk, Leonid Oliker, Lucas Paoli, Pierre Peterlongo, Vitor C. Piro, Jacob S. Porter, Simon Rasmussen, Evan R. Rees, Knut Reinert, Bernhard Renard, Espen Mikal Robertsen, Gail L. Rosen, Hans-Joachim Ruscheweyh, Varuni Sarwal, Nicola Segata, Enrico Seiler, Lizhen Shi, Fengzhu Sun, Shinichi Sunagawa, Søren Johannes Sørensen, Ashleigh Thomas, Chengxuan Tong, Mirko Trajkovski, Julien Tremblay, Gherman Uritskiy, Riccardo Vicedomini, Zhengyang Wang, Ziye Wang, Zhong Wang, Andrew Warren, Nils Peder Willassen, Katherine Yelick, Ronghui You, Georg Zeller, Zhengqiao Zhao, Shanfeng Zhu, Jie Zhu, Ruben Garrido-Oter, Petra Gastmeier, Stephane Hacquard, Susanne Häußler, Ariane Khaledi, Friederike Maechler, Fantin Mesny, Simona Radutoiu, Paul Schulze-Lefert, Nathiana Smit, Till Strowig, Andreas Bremges, Alexander Sczyrba, and Alice Carolyn McHardy. Critical Assessment of Metagenome Interpretation: The second round of challenges. *Nature Methods*, 19(4):429–440, April 2022. ISSN 1548-7091, 1548-7105. doi:10.1038/s41592-022-01431-4.

[120] National Institutes of Health Bethesda MD 20894 USA National Center for Biotechnology Information, National Library of Medicine. The ncbi pathogen detection project. `https://www.ncbi.nlm.nih.gov/pathogens/`, 2016.

[121] Weizhong Li and Adam Godzik. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, July 2006. ISSN 1367-4811, 1367-4803. doi:10.1093/bioinformatics/btl158.

[122] Fabiano Sviatopolk-Mirsky Pais, Patrícia de Cássia Ruy, Guilherme Oliveira, and Roney Santos Coimbra. Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology : AMB*, 9:4, March 2014. ISSN 1748-7188. doi:10.1186/1748-7188-9-4.

[123] Robert D. Finn, Jody Clements, William Arndt, Benjamin L. Miller, Travis J. Wheeler, Fabian Schreiber, Alex Bateman, and Sean R. Eddy. HMMER web server: 2015 update. *Nucleic Acids Research*, 43(W1):W30–W38, July 2015. ISSN 0305-1048, 1362-4962. doi:10.1093/nar/gkv397.

[124] T. Rinttila, A. Kassinen, E. Malinen, L. Krogius, and A. Palva. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *Journal of Applied Microbiology*, 97(6):1166–1177, December 2004. ISSN 1364-5072, 1365-2672. doi:10.1111/j.1365-2672.2004.02409.x.

[125] William A. Nickols, Thomas Kuntz, Jiaxian Shen, Sagun Maharjan, Himel Mallick, Eric A. Franzosa, Kelsey N. Thompson, Jacob T. Nearing, and Curtis Huttenhower. MaAsLin 3: Refining and extending generalized multivariable linear models for metaomic association discovery, December 2024.

[126] Skye Rs Fishbein, John I Robinson, Tiffany Hink, Kimberly A Reske, Erin P Newcomer, Carey-Ann D Burnham, Jeffrey P Henderson, Erik R Dubberke, and Gautam

Dantas. Multi-omics investigation of Clostridioides difficile-colonized patients reveals pathogen and commensal correlates of C. difficile pathogenesis. *eLife*, 11:e72801, January 2022. ISSN 2050-084X. doi:10.7554/eLife.72801.

[127] Laura Sisk-Hackworth and Scott T. Kelley. An application of compositional data analysis to multiomic time-series data. *NAR Genomics and Bioinformatics*, 2(4):lqaa079, October 2020. doi:10.1093/nargab/lqaa079.

[128] Florian Rohart, Benoît Gautier, Amrit Singh, and Kim-Anh Lê Cao. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, 13(11):e1005752, November 2017. ISSN 1553-7358. doi:10.1371/journal.pcbi.1005752.

[129] Melissa Weldy, Clayton Evert, Peter I. Dosa, Alexander Khoruts, and Michael J. Sadowsky. Convenient Protocol for Production and Purification of Clostridioides difficile Spores for Germination Studies. *STAR Protocols*, 1(2):100071, September 2020. ISSN 26661667. doi:10.1016/j.xpro.2020.100071.

[130] Joseph A Sorg and Sean S Dineen. Laboratory Maintenance of Clostridium difficile. doi:10.1002/9780471729259.mc09a01s12. URL `www.interscience.wiley.com`.

[131] Alexandre R. Marra, Eli N. Perencevich, Richard E. Nelson, Matthew Samore, Karim Khader, Hsiu-Yin Chiang, Margaret L. Chorazy, Loreen A. Herwaldt, Daniel J. Diekema, Michelle F. Kuxhausen, Amy Blevins, Melissa A. Ward, Jennifer S. McDanel, Rajeshwari Nair, Erin Balkenende, and Marin L. Schweizer. Incidence and Outcomes Associated With *Clostridium difficile* Infections: A Systematic Review and Meta-analysis. *JAMA Network Open*, 3(1):e1917597, January 2020. ISSN 2574-3805. doi:10.1001/jamanetworkopen.2019.17597.

[132] Erik R. Dubberke and Margaret A. Olsen. Burden of Clostridium difficile on the Healthcare System. *Clinical Infectious Diseases*, 55(suppl_2):S88–S92, August 2012. ISSN 1537-6591, 1058-4838. doi:10.1093/cid/cis335.

[133] Casey M. Theriot, Mark J. Koenigsknecht, Paul E. Carlson, Gabrielle E. Hatton, Adam M. Nelson, Bo Li, Gary B. Huffnagle, Jun Z. Li, and Vincent B. Young. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature Communications*, 5(1), May 2014. ISSN 2041-1723. doi:10.1038/ncomms4114.

[134] Florian P Breitwieser, Jennifer Lu, and Steven L Salzberg. A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4):1125–1136, July 2019. ISSN 1467-5463, 1477-4054. doi:10.1093/bib/bbx120.

[135] Simon H. Ye, Katherine J. Siddle, Daniel J. Park, and Pardis C. Sabeti. Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell*, 178(4):779–794, August 2019. ISSN 00928674. doi:10.1016/j.cell.2019.07.010.

[136] Qing Miao, Yuyan Ma, Qingqing Wang, Jue Pan, Yao Zhang, Wenting Jin, Yumeng Yao, Yi Su, Yingnan Huang, Mengran Wang, Bing Li, Huaying Li, Chunmei Zhou, Chun Li, Maosong Ye, Xiaoling Xu, Yongjun Li, and Bijie Hu. Microbiological Diagnostic Performance of Metagenomic Next-generation Sequencing When Applied to Clinical Practice. *Clinical Infectious Diseases*, 67(suppl_2):S231–S240, November 2018. ISSN 1058-4838, 1537-6591. doi:10.1093/cid/ciy693.

[137] Steve Miller, Samia N. Naccache, Erik Samayoa, Kevin Messacar, Shaun Arevalo, Scot Federman, Doug Stryke, Elizabeth Pham, Becky Fung, William J. Bolosky, Danielle Ingebrigtsen, Walter Lorizio, Sandra M. Paff, John A. Leake, Rick Pesano, Roberta DeBiasi, Samuel Dominguez, and Charles Y. Chiu. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Research*, 29(5):831–842, May 2019. ISSN 1088-9051, 1549-5469. doi:10.1101/gr.238170.118.

[138] Kumeren N. Govender, Teresa L. Street, Nicholas D. Sanderson, and David W. Eyre. Metagenomic Sequencing as a Pathogen-Agnostic Clinical Diagnostic Tool for Infectious Diseases: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies. *Journal of Clinical Microbiology*, 59(9):e02916–20, August 2021. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.02916-20.

[139] Min Kong, Wei Li, Qingsheng Kong, Haixin Dong, Aizhong Han, and Liqing Jiang. Application of metagenomic next-generation sequencing in cutaneous tuberculosis. *Frontiers in Cellular and Infection Microbiology*, 12:942073, September 2022. ISSN 2235-2988. doi:10.3389/fcimb.2022.942073.

[140] Yanjiao Zhou, Kristine M. Wylie, Rana E. El Feghaly, Kathie A. Mihindukulasuriya, Alexis Elward, David B. Haslam, Gregory A. Storch, and George M. Weinstock. Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens. *Journal of Clinical Microbiology*, 54(2):368–375, February 2016. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.01965-15.

[141] Jieun Kim, Youna Cho, Mi-Ran Seo, Mi Hyun Bae, Bongyoung Kim, Mina Rho, and Hyunjoo Pai. Quantitative characterization of Clostridioides difficile population in the gut microbiome of patients with C. difficile infection and their association with clinical factors. *Scientific Reports*, 10(1):17608, October 2020. ISSN 2045-2322. doi:10.1038/s41598-020-74090-0.

[142] Charles Y. Chiu and Steven A. Miller. Clinical metagenomics. *Nature Reviews Genetics*, 20(6):341–355, June 2019. ISSN 1471-0056, 1471-0064. doi:10.1038/s41576-019-0113-7.

[143] L Clifford McDonald, Dale N Gerding, Stuart Johnson, Johan S Bakken, Karen C Carroll, Susan E Coffin, Erik R Dubberke, Kevin W Garey, Carolyn V Gould, Ciaran Kelly, Vivian Loo, Julia Shaklee Sammons, Thomas J Sandora, and Mark H Wilcox. Clinical Practice Guidelines for Clostridium difficile Infection in Adults and Children:

2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clinical Infectious Diseases*, 66(7):e1–e48, March 2018. ISSN 1058-4838, 1537-6591. doi:10.1093/cid/cix1085.

[144] L. Clifford McDonald, Bruno Coignard, Erik Dubberke, Xiaoyan Song, Teresa Horan, Preeta K. Kutty, and the Ad Hoc Clostridium difficile Surveillance Working Group. Recommendations for Surveillance of Clostridium difficile–Associated Disease. *Infection Control & Hospital Epidemiology*, 28(2):140–145, February 2007. ISSN 0899-823X, 1559-6834. doi:10.1086/511798.

[145] Silin Tang, Renato H. Orsi, Hao Luo, Chongtao Ge, Guangtao Zhang, Robert C. Baker, Abigail Stevenson, and Martin Wiedmann. Assessment and Comparison of Molecular Subtyping and Characterization Methods for Salmonella. *Frontiers in Microbiology*, 10, 2019. ISSN 1664-302X.

[146] Hans-Joachim Ruscheweyh, Alessio Milanese, Lucas Paoli, Nicolai Karcher, Quentin Clayssen, Marisa Isabell Keller, Jakob Wirbel, Peer Bork, Daniel R. Mende, Georg Zeller, and Shinichi Sunagawa. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome*, 10(1):212, December 2022. ISSN 2049-2618. doi:10.1186/s40168-022-01410-z.

[147] Vadim Puller, Florian Plaza Oñate, Edi Prifti, and Raynald De Lahondès. Impact of simulation and reference catalogues on the evaluation of taxonomic profiling pipelines. Preprint, In Review, October 2023.

[148] Dieter M. Tourlousse, Koji Narita, Takamasa Miura, Akiko Ohashi, Masami Matsuda, Yoshifumi Ohyama, Mamiko Shimamura, Masataka Furukawa, Ken Kasahara, Keishi Kameyama, Sakae Saito, Maki Goto, Ritsuko Shimizu, Riko Mishima, Jiro Nakayama, Koji Hosomi, Jun Kunisawa, Jun Terauchi, Yuji Sekiguchi, and Hiroko Kawasaki. Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements. *Microbiology Spectrum*, 10 (2):e01915–21, March 2022. doi:10.1128/spectrum.01915-21.

[149] Kyung Hyun Kim, Dongbin Park, Baolei Jia, Ju Hye Baek, Yoonsoo Hahn, and Che Ok Jeon. Identification and Characterization of Major Bile Acid 7α-Dehydroxylating Bacteria in the Human Gut. *mSystems*, 7(4):e00455–22, August 2022. ISSN 2379-5077. doi:10.1128/msystems.00455-22.

[150] Gianluca Ianiro, Michal Punčochář, Nicolai Karcher, Serena Porcari, Federica Armanini, Francesco Asnicar, Francesco Beghini, Aitor Blanco-Míguez, Fabio Cumbo, Paolo Manghi, Federica Pinto, Luca Masucci, Gianluca Quaranta, Silvia De Giorgi, Giusi Desirè Sciumè, Stefano Bibbò, Federica Del Chierico, Lorenza Putignani, Maurizio Sanguinetti, Antonio Gasbarrini, Mireia Valles-Colomer, Giovanni Cammarota, and Nicola Segata. Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases.

Nature Medicine, 28(9):1913–1923, September 2022. ISSN 1078-8956, 1546-170X. doi:10.1038/s41591-022-01964-3.

[151] Caroline Vincent, Mark A. Miller, Thaddeus J. Edens, Sudeep Mehrotra, Ken Dewar, and Amee R. Manges. Bloom and bust: Intestinal microbiota dynamics in response to hospital exposures and Clostridium difficile colonization or infection. *Microbiome*, 4(1):12, December 2016. ISSN 2049-2618. doi:10.1186/s40168-016-0156-3.

[152] Ranjit Kumar, Nengjun Yi, Degui Zhi, Peter Eipers, Kelly T. Goldsmith, Paula Dixon, David K. Crossman, Michael R. Crowley, Elliot J. Lefkowitz, J. Martin Rodriguez, and Casey D. Morrow. Identification of donor microbe species that colonize and persist long term in the recipient after fecal transplant for recurrent Clostridium difficile. *npj Biofilms and Microbiomes*, 3(1):12, June 2017. ISSN 2055-5008. doi:10.1038/s41522-017-0020-7.

[153] David B. Stewart, Justin R. Wright, Maria Fowler, Christopher J. McLimans, Vasily Tokarev, Isabella Amaniera, Owen Baker, Hoi-Tong Wong, Jeff Brabec, Rebecca Drucker, and Regina Lamendella. Integrated Meta-omics Reveals a Fungus-Associated Bacteriome and Distinct Functional Pathways in Clostridioides difficile Infection. *mSphere*, 4(4):e00454–19, August 2019. ISSN 2379-5042. doi:10.1128/mSphere.00454-19.

[154] Tanya M. Monaghan, Tim J. Sloan, Stephen R. Stockdale, Adam M. Blanchard, Richard D. Emes, Mark Wilcox, Rima Biswas, Rupam Nashine, Sonali Manke, Jinal Gandhi, Pratishtha Jain, Shrejal Bhotmange, Shrikant Ambalkar, Ashish Satav, Lorraine A. Draper, Colin Hill, and Rajpal Singh Kashyap. Metagenomics reveals impact of geography and acute diarrheal disease on the Central Indian human gut microbiome. *Gut Microbes*, 12(1):1752605, November 2020. ISSN 1949-0976, 1949-0984. doi:10.1080/19490976.2020.1752605.

[155] Daniel Podlesny, Marija Durdevic, Sudarshan Paramsothy, Nadeem O. Kaakoush, Christoph Högenauer, Gregor Gorkiewicz, Jens Walter, and W. Florian Fricke. Identification of clinical and ecological determinants of strain engraftment after fecal microbiota transplantation using metagenomics. *Cell Reports Medicine*, 3(8):100711, August 2022. ISSN 26663791. doi:10.1016/j.xcrm.2022.100711.

[156] Ashley M. Rooney, Kyla Cochrane, Stephanie Fedsin, Samantha Yao, Shaista Anwer, Satyender Dehmiwal, Susy Hota, Susan Poutanen, Emma Allen-Vercoe, Bryan Coburn, MTOP Investigators, Johane Allard, Kenneth Croitoru, Herbert Gaisano, David Guttman, Valerie Taylor, Dana Philpott, and Dan Winer. A microbial consortium alters intestinal *Pseudomonadota* and antimicrobial resistance genes in individuals with recurrent *Clostridioides difficile* infection. *mBio*, pages e03482–22, July 2023. ISSN 2150-7511. doi:10.1128/mbio.03482-22.

[157] Matthew T. Sorbara, Eric R. Littmann, Emily Fontana, Thomas U. Moody, Claire E. Kohout, Mergim Gjonbalaj, Vincent Eaton, Ruth Seok, Ingrid M. Leiner, and Eric G.

Pamer. Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity. *Cell Host & Microbe*, 28(1): 134–146.e4, July 2020. ISSN 19313128. doi:10.1016/j.chom.2020.05.005.

[158] Zhenrun J. Zhang, Cody G. Cole, Michael J. Coyne, Huaiying Lin, Nicholas Dylla, Rita C. Smith, Téa E. Pappas, Shannon A. Townson, Nina Laliwala, Emily Waligurski, Ramanujam Ramaswamy, Che Woodson, Victoria Burgo, Jessica C. Little, David Moran, Amber Rose, Mary McMillin, Emma McSpadden, Anitha Sundararajan, Ashley M. Sidebottom, Eric G. Pamer, and Laurie E. Comstock. Comprehensive analyses of a large human gut Bacteroidales culture collection reveal species- and strain-level diversity and evolution. *Cell Host & Microbe*, page S1931312824003238, September 2024. ISSN 19313128. doi:10.1016/j.chom.2024.08.016.

[159] Kumeren N. Govender and David W. Eyre. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. *Microbial Genomics*, 8(10), October 2022. ISSN 2057-5858. doi:10.1099/mgen.0.000886.

[160] C. Royer, N. V. Patin, K. J. Jesser, A. Peña-Gonzalez, J. K. Hatt, G. Trueba, K. Levy, and K. T. Konstantinidis. Comparison of metagenomic and traditional methods for diagnosis of E. coli enteric infections. *mBio*, 15(4):e03422–23, March 2024. doi:10.1128/mbio.03422-23.

[161] Yuan Wu, Lin Yang, Wen-Ge Li, Wen Zhu Zhang, Zheng Jie Liu, and Jin-Xing Lu. Microevolution within ST11 group Clostridioides difficile isolates through mobile genetic elements based on complete genome sequencing. *BMC Genomics*, 20(1):796, December 2019. ISSN 1471-2164. doi:10.1186/s12864-019-6184-1.

[162] Nicola Segata, Levi Waldron, Annalisa Ballarini, Vagheesh Narasimhan, Olivier Jousson, and Curtis Huttenhower. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 9(8):811–814, August 2012. ISSN 1548-7091, 1548-7105. doi:10.1038/nmeth.2066.

[163] Martha Zepeda-Rivera, Samuel S. Minot, Heather Bouzek, Hanrui Wu, Aitor Blanco-Míguez, Paolo Manghi, Dakota S. Jones, Kaitlyn D. LaCourse, Ying Wu, Elsa F. McMahon, Soon-Nang Park, Yun K. Lim, Andrew G. Kempchinsky, Amy D. Willis, Sean L. Cotton, Susan C. Yost, Ewa Sicinska, Joong-Ki Kook, Floyd E. Dewhirst, Nicola Segata, Susan Bullman, and Christopher D. Johnston. A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature*, March 2024. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-024-07182-w.

[164] David W Eyre, Tanya Golubchik, N Claire Gordon, Rory Bowden, Paolo Piazza, Elizabeth M Batty, Camilla L C Ip, Daniel J Wilson, Xavier Didelot, Lily O'Connor, Rochelle Lay, David Buck, Angela M Kearns, Angela Shaw, John Paul, Mark H Wilcox, Peter J Donnelly, Tim E A Peto, A Sarah Walker, and Derrick W Crook. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium*

difficile for outbreak detection and surveillance. *BMJ Open*, 2(3):e001124, 2012. ISSN 2044-6055, 2044-6055. doi:10.1136/bmjopen-2012-001124.

[165] David W. Eyre, Madeleine L. Cule, Daniel J. Wilson, David Griffiths, Alison Vaughan, Lily O'Connor, Camilla L.C. Ip, Tanya Golubchik, Elizabeth M. Batty, John M. Finney, David H. Wyllie, Xavier Didelot, Paolo Piazza, Rory Bowden, Kate E. Dingle, Rosalind M. Harding, Derrick W. Crook, Mark H. Wilcox, Tim E.A. Peto, and A. Sarah Walker. Diverse Sources of *C. difficile* Infection Identified on Whole-Genome Sequencing. *New England Journal of Medicine*, 369(13):1195–1205, September 2013. ISSN 0028-4793, 1533-4406. doi:10.1056/NEJMoa1216064.

[166] Nicole Pecora, Stacy Holzbauer, Xiong Wang, Yu Gu, Samantha Taffner, Trupti Hatwar, Dwight Hardy, Michelle Dziejman, Paige D'Heilly, Kelly Pung, Alice Guh, Xing Qiu, Steven Gill, and Ghinwa Dumyati. Genomic Analysis of *Clostridioides difficile* in 2 Regions of the United States Reveals a Diversity of Strains and Limited Transmission. *The Journal of Infectious Diseases*, 225(1):121–129, January 2022. ISSN 0022-1899, 1537-6613. doi:10.1093/infdis/jiab294.

[167] Taito Miyazaki, Kotaro Aoki, Tadashi Maeda, Kohji Komori, Sadako Yoshizawa, Yoshikazu Ishii, Yoshihisa Urita, and Kazuhiro Tateda. A molecular epidemiological and transmission analysis of Clostridioides difficile using draft whole-genome sequencing in a single hospital. *BMC Infectious Diseases*, 24(1):989, September 2024. ISSN 1471-2334. doi:10.1186/s12879-024-09841-9.

[168] Mireia Valles-Colomer, Aitor Blanco-Míguez, Paolo Manghi, Francesco Asnicar, Leonard Dubois, Davide Golzato, Federica Armanini, Fabio Cumbo, Kun D. Huang, Serena Manara, Giulia Masetti, Federica Pinto, Elisa Piperni, Michal Punčochář, Liviana Ricci, Moreno Zolfo, Olivia Farrant, Adriana Goncalves, Marta Selma-Royo, Ana G. Binetti, Jimmy E. Becerra, Bei Han, John Lusingu, John Amuasi, Loredana Amoroso, Alessia Visconti, Claire M. Steves, Mario Falchi, Michele Filosi, Adrian Tett, Anna Last, Qian Xu, Nan Qin, Huanlong Qin, Jürgen May, Daniel Eibach, Maria Valeria Corrias, Mirco Ponzoni, Edoardo Pasolli, Tim D. Spector, Enrico Domenici, Maria Carmen Collado, and Nicola Segata. The person-to-person transmission landscape of the gut and oral microbiomes. *Nature*, 614(7946):125–135, February 2023. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-022-05620-1.

[169] Duy Tin Truong, Adrian Tett, Edoardo Pasolli, Curtis Huttenhower, and Nicola Segata. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Research*, 27(4):626–638, April 2017. ISSN 1088-9051, 1549-5469. doi:10.1101/gr.216242.116.

[170] Lucas R. Van Dijk, Bruce J. Walker, Timothy J. Straub, Colin J. Worby, Alexandra Grote, Henry L. Schreiber, Christine Anyansi, Amy J. Pickering, Scott J. Hultgren, Abigail L. Manson, Thomas Abeel, and Ashlee M. Earl. StrainGE: A toolkit to track and characterize low-abundance strains in complex microbial communities. *Genome Biology*, 23(1):74, December 2022. ISSN 1474-760X. doi:10.1186/s13059-022-02630-0.

[171] Barry G. Hall, Garth D. Ehrlich, and Fen Z. Hu. Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. *Microbiology*, 156(Pt 4):1060, April 2010. doi:10.1099/mic.0.035188-0.

[172] Martin C. J. Maiden, Jane A. Bygraves, Edward Feil, Giovanna Morelli, Joanne E. Russell, Rachel Urwin, Qing Zhang, Jiaji Zhou, Kerstin Zurth, Dominique A. Caugant, Ian M. Feavers, Mark Achtman, and Brian G. Spratt. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, 95(6): 3140–3145, March 1998. doi:10.1073/pnas.95.6.3140.

[173] Monique J. T. Crobach, Bastian V. H. Hornung, Cees Verduin, Margreet C. Vos, Joost Hopman, Nitin Kumar, Celine Harmanus, Ingrid Sanders, Elisabeth M. Terveer, Mark D. Stares, Trevor D. Lawley, and Ed J. Kuijper. Screening for *Clostridioides difficile* colonization at admission to the hospital: A multi-centre study. *Clinical Microbiology and Infection*, 29(7):891–896, July 2023. ISSN 1198-743X. doi:10.1016/j.cmi.2023.02.022.

[174] Mustapha M. Mustapha, Vatsala R. Srinivasa, Marissa P. Griffith, Shu-Ting Cho, Daniel R. Evans, Kady Waggle, Chinelo Ezeonwuka, Daniel J. Snyder, Jane W. Marsh, Lee H. Harrison, Vaughn S. Cooper, and Daria Van Tyne. Genomic Diversity of Hospital-Acquired Infections Revealed through Prospective Whole-Genome Sequencing-Based Surveillance. *mSystems*, 7(3):e01384–21, June 2022. doi:10.1128/msystems.01384-21.

[175] Ling Yuan Kong, David W Eyre, Jacques Corbeil, Frederic Raymond, A Sarah Walker, Mark H Wilcox, Derrick W Crook, Sophie Michaud, Baldwin Toye, Eric Frost, Nandini Dendukuri, Ian Schiller, Anne-Marie Bourgault, Andrew Dascal, Matthew Oughton, Yves Longtin, Louise Poirier, Paul Brassard, Nathalie Turgeon, Rodica Gilca, and Vivian G Loo. Clostridium difficile: Investigating Transmission Patterns Between Infected and Colonized Patients Using Whole Genome Sequencing. *Clinical Infectious Diseases*, 68(2):204–209, January 2019. ISSN 1058-4838. doi:10.1093/cid/ciy457.

[176] Jonathan J. Hunt and Jimmy D. Ballard. Variations in Virulence and Molecular Biology among Emerging Strains of Clostridium difficile. *Microbiology and Molecular Biology Reviews*, 77(4):567–581, December 2013. ISSN 1092-2172, 1098-5557. doi:10.1128/MMBR.00017-13.

[177] Esther R Robinson, Timothy M Walker, and Mark J Pallen. Genomics and outbreak investigation: From sequence to consequence. *Genome Medicine*, 5(4):36, December 2013. ISSN 1756-994X. doi:10.1186/gm440.

[178] Jon E. Edman-Wallér, Michael Toepfer, Johan Karp, Kristina Rizzardi, Gunnar Jacobsson, and Maria Werner. Clostridioides difficile outbreak detection: Evaluation by

ribotyping and whole-genome sequencing of a surveillance algorithm based on ward-specific cutoffs. *Infection Control & Hospital Epidemiology*, 44(12):1948–1952, December 2023. ISSN 0899-823X, 1559-6834. doi:10.1017/ice.2023.113.

[179] Luis M. Rodriguez-R, Roth E. Conrad, Tomeu Viver, Dorian J. Feistel, Blake G. Lindner, Stephanus N. Venter, Luis H. Orellana, Rudolf Amann, Ramon Rossello-Mora, and Konstantinos T. Konstantinidis. An ANI gap within bacterial species that advances the definitions of intra-species units. *mBio*, 15(1):e02696–23, January 2024. ISSN 2150-7511. doi:10.1128/mbio.02696-23.

[180] Skye R. S. Fishbein, Bejan Mahmud, and Gautam Dantas. Antibiotic perturbations to the gut microbiome. *Nature Reviews Microbiology*, July 2023. ISSN 1740-1526, 1740-1534. doi:10.1038/s41579-023-00933-y.

[181] David VanInsberghe, Joseph A. Elsherbini, Bernard Varian, Theofilos Poutahidis, Susan Erdman, and Martin F. Polz. Diarrhoeal events can trigger long-term Clostridium difficile colonization with recurrent blooms. *Nature Microbiology*, 5(4):642–650, April 2020. ISSN 2058-5276. doi:10.1038/s41564-020-0668-2.

[182] Frederic D. Bushman, Maire Conrad, Yue Ren, Chunyu Zhao, Christopher Gu, Christopher Petucci, Min-Soo Kim, Arwa Abbas, Kevin J. Downes, Nina Devas, Lisa M. Mattei, Jessica Breton, Judith Kelsen, Sarah Marakos, Alissa Galgano, Kelly Kachelries, Jessi Erlichman, Jessica L. Hart, Michael Moraskie, Dorothy Kim, Huanjia Zhang, Casey E. Hofstaedter, Gary D. Wu, James D. Lewis, Joseph P. Zackular, Hongzhe Li, Kyle Bittinger, and Robert Baldassano. Multi-omic Analysis of the Interaction between Clostridioides difficile Infection and Pediatric Inflammatory Bowel Disease. *Cell Host & Microbe*, 28(3):422–433.e7, September 2020. ISSN 19313128. doi:10.1016/j.chom.2020.07.020.

[183] Charlie G. Buffie, Vanni Bucci, Richard R. Stein, Peter T. McKenney, Lilan Ling, Asia Gobourne, Daniel No, Hui Liu, Melissa Kinnebrew, Agnes Viale, Eric Littmann, Marcel R.M. Van Den Brink, Robert R. Jenq, Ying Taur, Chris Sander, Justin R. Cross, Nora C. Toussaint, Joao B. Xavier, and Eric G. Pamer. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature*, 517 (7533):205–208, January 2015. ISSN 14764687. doi:10.1038/nature13828. Publisher: Nature Publishing Group.

[184] Himel Mallick, Eric A. Franzosa, Lauren J. Mclver, Soumya Banerjee, Alexandra Sirota-Madi, Aleksandar D. Kostic, Clary B. Clish, Hera Vlamakis, Ramnik J. Xavier, and Curtis Huttenhower. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nature Communications*, 10(1):3136, July 2019. ISSN 2041-1723. doi:10.1038/s41467-019-10927-1.

[185] Xiaoxuan Xia, William Ka Kei Wu, Sunny Hei Wong, Dabin Liu, Thomas Ngai Yeung Kwong, Geicho Nakatsu, Pearlly S. Yan, Yu-Ming Chuang, Michael Wing-Yan Chan, Olabisi Oluwabukola Coker, Zigui Chen, Yun Kit Yeoh, Liuyang Zhao, Xiansong Wang,

Wing Yin Cheng, Matthew Tak Vai Chan, Paul Kay Sheung Chan, Joseph Jao Yiu Sung, Maggie Haitian Wang, and Jun Yu. Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. *Microbiome*, 8(1):108, July 2020. ISSN 2049-2618. doi:10.1186/s40168-020-00847-4. URL https://doi.org/10.1186/s40168-020-00847-4.

[186] for the CDC Prevention Epicenter Program, Amy Langdon, Drew J. Schwartz, Christopher Bulow, Xiaoqing Sun, Tiffany Hink, Kimberly A. Reske, Courtney Jones, Carey-Ann D. Burnham, Erik R. Dubberke, and Gautam Dantas. Microbiota restoration reduces antibiotic-resistant bacteria gut colonization in patients with recurrent Clostridioides difficile infection from the open-label PUNCH CD study. *Genome Medicine*, 13 (1):28, December 2021. ISSN 1756-994X. doi:10.1186/s13073-021-00843-9.

[187] Alexander B. Smith, Matthew L. Jenior, Orlaith Keenan, Jessica L. Hart, Jonathan Specker, Arwa Abbas, Paula C. Rangel, Chao Di, Jamal Green, Katelyn A. Bustin, Jennifer A. Gaddy, Maribeth R. Nicholson, Clare Laut, Brendan J. Kelly, Megan L. Matthews, Daniel R. Evans, Daria Van Tyne, Emma E. Furth, Jason A. Papin, Frederic D. Bushman, Jessi Erlichman, Robert N. Baldassano, Michael A. Silverman, Gary M. Dunny, Boone M. Prentice, Eric P. Skaar, and Joseph P. Zackular. Enterococci enhance Clostridioides difficile pathogenesis. *Nature*, 611(7937):780–786, November 2022. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-022-05438-x.

[188] Alexander B. Smith, Matthew L. Jenior, Orlaith Keenan, Jessica L. Hart, Jonathan Specker, Arwa Abbas, Paula C. Rangel, Chao Di, Jamal Green, Katelyn A. Bustin, Jennifer A. Gaddy, Maribeth R. Nicholson, Clare Laut, Brendan J. Kelly, Megan L. Matthews, Daniel R. Evans, Daria Van Tyne, Emma E. Furth, Jason A. Papin, Frederic D. Bushman, Jessi Erlichman, Robert N. Baldassano, Michael A. Silverman, Gary M. Dunny, Boone M. Prentice, Eric P. Skaar, and Joseph P. Zackular. Enterococci enhance Clostridioides difficile pathogenesis. *Nature*, 611(7937):780–786, November 2022. ISSN 0028-0836, 1476-4687. doi:10.1038/s41586-022-05438-x. URL https://www.nature.com/articles/s41586-022-05438-x.

[189] Catabolic Network of the Fermentative Gut Bacterium Phocaeicola vulgatus (Phylum Bacteroidota) from a Physiologic-Proteomic Perspective | Microbial Physiology | Karger Publishers. https://karger.com/mip/article/34/1/88/894550/Catabolic-Network-of-the-Fermentative-Gut.

[190] Eiko Hayase, Tomo Hayase, Akash Mukherjee, Stuart C. Stinson, Mohamed A. Jamal, Miriam R. Ortega, Christopher A. Sanchez, Saira S. Ahmed, Jennifer L. Karmouch, Chia-Chi Chang, Ivonne I. Flores, Lauren K. McDaniel, Alexandria N. Brown, Rawan K. El-Himri, Valerie A. Chapa, Lin Tan, Bao Q. Tran, Yao Xiao, Christopher Fan, Dung Pham, Taylor M. Halsey, Yimei Jin, Wen-Bin Tsai, Rishika Prasad, Israel K. Glover, Altai Enkhbayar, Aqsa Mohammed, Maren Schmiester, Katherine Y. King, Robert A. Britton, Pavan Reddy, Matthew C. Wong, Nadim J. Ajami, Jennifer A. Wargo, Samuel Shelburne, Pablo C. Okhuysen, Chen Liu, Stephanie W.

Fowler, Margaret E. Conner, Zoe Katsamakis, Natalie Smith, Marina Burgos da Silva, Doris M. Ponce, Jonathan U. Peled, Marcel R. M. van den Brink, Christine B. Peterson, Gabriela Rondon, Jeffrey J. Molldrem, Richard E. Champlin, Elizabeth J. Shpall, Philip L. Lorenzi, Rohtesh S. Mehta, Eric C. Martens, Amin M. Alousi, and Robert R. Jenq. Bacteroides ovatus alleviates dysbiotic microbiota-induced graft-versus-host disease. *Cell Host & Microbe*, 32(9):1621–1636.e6, September 2024. ISSN 1931-3128. doi:10.1016/j.chom.2024.08.004.

[191] YiTing Yan, Yu Lei, Ying Qu, Zhen Fan, Ting Zhang, Yangbin Xu, Qian Du, Daniel Brugger, Yulin Chen, Ke Zhang, and Enping Zhang. Bacteroides uniformis-induced perturbations in colonic microbiota and bile acid levels inhibit TH17 differentiation and ameliorate colitis developments. *npj Biofilms and Microbiomes*, 9(1):1–14, August 2023. ISSN 2055-5008. doi:10.1038/s41522-023-00420-5.

[192] Jordy Evan Sulaiman, Jaron Thompson, Pak Lun Kevin Cheung, Yili Qian, Jericha Mill, Isabella James, Hanhyeok Im, Eugenio I. Vivas, Judith Simcox, and Ophelia S. Venturelli. Phocaeicola vulgatus shapes the long-term growth dynamics and evolutionary adaptations of Clostridioides difficile. *Cell Host & Microbe*, 33(1):42–58.e10, January 2025. ISSN 1931-3128. doi:10.1016/j.chom.2024.12.001.

[193] Daoming Wang, Marwah Doestzada, Lianmin Chen, Sergio Andreu-Sánchez, Inge C.L. van den Munckhof, Hannah E. Augustijn, Martijn Koehorst, Angel J. Ruiz-Moreno, Vincent W. Bloks, Niels P. Riksen, Joost H.W. Rutten, Leo A.B. Joosten, Mihai G. Netea, Cisca Wijmenga, Alexandra Zhernakova, Folkert Kuipers, and Jingyuan Fu. Characterization of gut microbial structural variations as determinants of human bile acid metabolism. *Cell Host & Microbe*, 29(12):1802–1814.e5, December 2021. ISSN 19313128. doi:10.1016/j.chom.2021.11.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S1931312821005096`.

[194] Fátima C. Pereira, Kenneth Wasmund, Iva Cobankovic, Nico Jehmlich, Craig W. Herbold, Kang Soo Lee, Barbara Sziranyi, Cornelia Vesely, Thomas Decker, Roman Stocker, Benedikt Warth, Martin Von Bergen, Michael Wagner, and David Berry. Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. *Nature Communications*, 11(1):5104, October 2020. ISSN 2041-1723. doi:10.1038/s41467-020-18928-1.

[195] Charlie G. Buffie and Eric G. Pamer. Microbiota-mediated colonization resistance against intestinal pathogens. *Nature Reviews Immunology 2013 13:11*, 13(11):790–801, October 2013. ISSN 1474-1741. doi:10.1038/nri3535. URL `https://www.nature.com/articles/nri3535`. Publisher: Nature Publishing Group.

[196] Hubertus Wietholtz, Jan Sjovall', and Siegfried Matem. Stimulation of bile acid 6a-hydroxylation by rifampin.

[197] Sidney R. Elsden and Martin G. Hilton. Volatile acid production from threonine,

valine, leucine and isoleucine by clostridia. *Archives of Microbiology*, 117(2):165–172, 1978. ISSN 0302-8933, 1432-072X. doi:10.1007/BF00402304.

[198] Skye RS Fishbein, John I Robinson, Tiffany Hink, Kimberly A Reske, Erin P Newcomer, Carey-Ann D Burnham, Jeffrey P Henderson, Erik R Dubberke, and Gautam Dantas. Multi-omics investigation of Clostridioides difficile-colonized patients reveals pathogen and commensal correlates of C. difficile pathogenesis. *eLife*, 11: e72801, January 2022. ISSN 2050-084X. doi:10.7554/eLife.72801. URL `https://elifesciences.org/articles/72801`.

[199] Andrea Martinez Aguirre, Nazli Yalcinkaya, Qinglong Wu, Alton Swennes, Mary Elizabeth Tessier, Paul Roberts, Fabio Miyajima, Tor Savidge, and Joseph A. Sorg. Bile acid-independent protection against Clostridioides difficile infection. *PLOS Pathogens*, 17(10):e1010015, October 2021. ISSN 1553-7374. doi:10.1371/journal.ppat.1010015.

[200] Kim-Anh Lê Cao, Simon Boitard, and Philippe Besse. Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, 12(1):253, June 2011. ISSN 1471-2105. doi:10.1186/1471-2105-12-253.

[201] Kim-Anh Lê Cao, Debra Rossouw, Christèle Robert-Granié, and Philippe Besse. A Sparse PLS for Variable Selection when Integrating Omics Data. *Statistical Applications in Genetics and Molecular Biology*, 7(1), November 2008. ISSN 1544-6115. doi:10.2202/1544-6115.1390.

[202] Andrew Bell, Jason Brunt, Emmanuelle Crost, Laura Vaux, Ridvan Nepravishta, C. David Owen, Dimitrios Latousakis, An Xiao, Wanqing Li, Xi Chen, Martin A. Walsh, Jan Claesen, Jesus Angulo, Gavin H. Thomas, and Nathalie Juge. Elucidation of a sialic acid metabolism pathway in mucus-foraging Ruminococcus gnavus unravels mechanisms of bacterial adaptation to the gut. *Nature Microbiology*, 4(12):2393–2404, December 2019. ISSN 2058-5276. doi:10.1038/s41564-019-0590-7.

[203] Matthew T. Henke, Douglas J. Kenny, Chelsi D. Cassilly, Hera Vlamakis, Ramnik J. Xavier, and Jon Clardy. Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proceedings of the National Academy of Sciences*, 116(26):12672–12677, June 2019. doi:10.1073/pnas.1904099116.

[204] Chin Wen Png, Sara K. Lindén, Kristen S. Gilshenan, Erwin G. Zoetendal, Chris S. McSweeney, Lindsay I. Sly, Michael A. McGuckin, and Timothy H. J. Florin. Mucolytic Bacteria With Increased Prevalence in IBD Mucosa AugmentIn VitroUtilization of Mucin by Other Bacteria. *Official journal of the American College of Gastroenterology / ACG*, 105(11):2420, November 2010. ISSN 0002-9270. doi:10.1038/ajg.2010.281.

[205] Maximilian Baumgartner, Michaela Lang, Hunter Holley, Daniel Crepaz, Bela Hausmann, Petra Pjevac, Doris Moser, Felix Haller, Fabian Hof, Andrea Beer, Elisabeth

Orgler, Adrian Frick, Vineeta Khare, Rayko Evstatiev, Susanne Strohmaier, Christian Primas, Werner Dolak, Thomas Köcher, Kristaps Klavins, Timo Rath, Markus F. Neurath, David Berry, Athanasios Makristathis, Markus Muttenthaler, and Christoph Gasche. Mucosal Biofilms Are an Endoscopic Feature of Irritable Bowel Syndrome and Ulcerative Colitis. *Gastroenterology*, 161(4):1245–1256.e20, October 2021. ISSN 0016-5085. doi:10.1053/j.gastro.2021.06.024.

[206] Si Yu, Xiaomeng Ge, Hui Xu, Bei Tan, Bowen Tian, Yujie Shi, Yimin Dai, Yue Li, Songnian Hu, and Jiaming Qian. Gut microbiome and mycobiome in inflammatory bowel disease patients with Clostridioides difficile infection. *Frontiers in Cellular and Infection Microbiology*, 13, February 2023. ISSN 2235-2988. doi:10.3389/fcimb.2023.1129043.

[207] Jeremy R. Chen See, Jillian Leister, Justin R. Wright, Peter I. Kruse, Mohini V. Khedekar, Catharine E. Besch, Carol A. Kumamoto, Gregory R. Madden, David B. Stewart, and Regina Lamendella. Clostridioides difficile infection is associated with differences in transcriptionally active microbial communities. *Frontiers in Microbiology*, 15, April 2024. ISSN 1664-302X. doi:10.3389/fmicb.2024.1398018.

[208] Manabu Nagayama, Yano , Tomonori, Atarashi , Koji, Tanoue , Takeshi, Sekiya , Mariko, Kobayashi , Yasutoshi, Sakamoto , Hirotsugu, Miura , Kouichi, Sunada , Keijiro, Kawaguchi , Takaaki, Morita , Satoru, Sugita , Kayoko, Narushima , Seiko, Barnich , Nicolas, Isayama , Jun, Kiridooshi , Yuko, Shiota , Atsushi, Suda , Wataru, Hattori , Masahira, Yamamoto , Hironori, and Kenya and Honda. TH1 cell-inducing Escherichia coli strain identified from the small intestinal mucosa of patients with Crohn's disease. *Gut Microbes*, 12(1):1788898, November 2020. ISSN 1949-0976. doi:10.1080/19490976.2020.1788898.

[209] Wei Li, Saiyu Hang, Yuan Fang, Sena Bae, Yancong Zhang, Minghao Zhang, Gang Wang, Megan D. McCurry, Munhyung Bae, Donggi Paik, Eric A. Franzosa, Fraydoon Rastinejad, Curtis Huttenhower, Lina Yao, A. Sloan Devlin, and Jun R. Huh. A bacterial bile acid metabolite modulates Treg activity through the nuclear hormone receptor NR4A1. *Cell Host & Microbe*, 29(9):1366–1377.e9, September 2021. ISSN 19313128. doi:10.1016/j.chom.2021.07.013.

[210] Jiani Wang, Sally Ghali, Chunlan Xu, Caroline C. Mussatto, Christina Ortiz, Elaine C. Lee, Diana H. Tran, Jonathan P. Jacobs, Venu Lagishetty, Kym F. Faull, Travis Moller, Maura Rossetti, Xinhua Chen, and Hon Wai Koon. Ceragenin CSA13 Reduces *Clostridium difficile* Infection in Mice by Modulating the Intestinal Microbiome and Metabolites. *Gastroenterology*, 154(6):1737–1750, May 2018. ISSN 0016-5085. doi:10.1053/j.gastro.2018.01.026.

[211] Wolfgang Buckel, Peter H. Janssen, Alexandra Schuhmann, Ulrich Eikmanns, Paul Messner, Uwe Sleytr, and Werner Liesack. Clostridium viride sp. nov., a strictly anaerobic bacterium using 5-aminovalerate as growth substrate, previously assigned

to Clostridium aminovalericum. *Archives of Microbiology*, 162(6):387–394, December 1994. ISSN 1432-072X. doi:10.1007/BF00282102.

[212] Vincent B. Young and Thomas M. Schmidt. Antibiotic-Associated Diarrhea Accompanied by Large-Scale Alterations in the Composition of the Fecal Microbiota. *Journal of Clinical Microbiology*, 42(3):1203–1206, March 2004. ISSN 0095-1137, 1098-660X. doi:10.1128/JCM.42.3.1203-1206.2004.

[213] Joshua R. Fletcher, Samantha Erwin, Cristina Lanzas, and Casey M. Theriot. Shifts in the Gut Metabolome and Clostridium difficile Transcriptome throughout Colonization and Infection in a Mouse Model. *mSphere*, 3(2):e00089–18, April 2018. ISSN 2379-5042. doi:10.1128/mSphere.00089-18.

[214] Charlie G. Buffie, Vanni Bucci, Richard R. Stein, Peter T. McKenney, Lilan Ling, Asia Gobourne, Daniel No, Hui Liu, Melissa Kinnebrew, Agnes Viale, Eric Littmann, Marcel R. M. Van Den Brink, Robert R. Jenq, Ying Taur, Chris Sander, Justin R. Cross, Nora C. Toussaint, Joao B. Xavier, and Eric G. Pamer. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature*, 517 (7533):205–208, January 2015. ISSN 0028-0836, 1476-4687. doi:10.1038/nature13828.

[215] Xiaoyun Huang, April E. Johnson, Joshua N. Brehm, Thi Van Thanh Do, Thomas A. Auchtung, Hugh C. McCullough, Armando I. Lerma, Sigmund J. Haidacher, Kathleen M. Hoch, Thomas D. Horvath, Joseph A. Sorg, Anthony M. Haag, and Jennifer M. Auchtung. *Clostridioides Difficile* colonization is not mediated by bile salts and utilizes Stickland fermentation of proline in an *in vitro* model. *mSphere*, pages e01049–24, January 2025. ISSN 2379-5042. doi:10.1128/msphere.01049-24.

[216] Pingting Guo, Ke Zhang, Xi Ma, and Pingli He. Clostridium species as probiotics: Potentials and challenges. *Journal of Animal Science and Biotechnology*, 11(1):24, February 2020. ISSN 2049-1891. doi:10.1186/s40104-019-0402-1.

[217] Shogo Takahashi, Tatsuki Fukami, Yusuke Masuo, Chad N. Brocker, Cen Xie, Kristopher W. Krausz, C. Roland Wolf, Colin J. Henderson, and Frank J. Gonzalez. Cyp2c70 is responsible for the species difference in bile acid metabolism between mice and humans. *Journal of Lipid Research*, 57(12):2130–2137, December 2016. ISSN 00222275. doi:10.1194/jlr.M071183.

[218] Akira Honda, Teruo Miyazaki, Junichi Iwamoto, Takeshi Hirayama, Yukio Morishita, Tadakuni Monma, Hajime Ueda, Seiya Mizuno, Fumihiro Sugiyama, Satoru Takahashi, and Tadashi Ikegami. Regulation of bile acid metabolism in mouse models with hydrophobic bile acid composition. *Journal of Lipid Research*, 61(1):54–69, January 2020. ISSN 0022-2275. doi:10.1194/jlr.RA119000395.

[219] Kinchel C. Doerner, Fusae Takamine, Crystal P. LaVoie, Darrell H. Mallonee, and Phillip B. Hylemon. Assessment of fecal bacteria with bile acid 7-dehydroxylating activity for the presence of bai-like genes. *Applied and Environmental Microbiology*, 63(3):

1185–1188, 1997. ISSN 00992240. doi:10.1128/AEM.63.3.1185-1188.1997. Publisher: American Society for Microbiology.

[220] Saiyu Hang, Donggi Paik, Lina Yao, Eunha Kim, Jamma Trinath, Jingping Lu, Soyoung Ha, Brandon N. Nelson, Samantha P. Kelly, Lin Wu, Ye Zheng, Randy S. Longman, Fraydoon Rastinejad, A. Sloan Devlin, Michael R. Krout, Michael A. Fischbach, Dan R. Littman, and Jun R. Huh. Bile acid metabolites control TH17 and Treg cell differentiation. *Nature*, 576(7785):143–148, December 2019. ISSN 0028-0836. doi:10.1038/s41586-019-1785-z. URL `http://www.nature.com/articles/s41586-019-1785-z`.

[221] Donggi Paik, Lina Yao, Yancong Zhang, Sena Bae, Gabriel D. D'Agostino, Minghao Zhang, Eunha Kim, Eric A. Franzosa, Julian Avila-Pacheco, Jordan E. Bisanz, Christopher K. Rakowski, Hera Vlamakis, Ramnik J. Xavier, Peter J. Turnbaugh, Randy S. Longman, Michael R. Krout, Clary B. Clish, Fraydoon Rastinejad, Curtis Huttenhower, Jun R. Huh, and A. Sloan Devlin. Human gut bacteria produce 17-modulating bile acid metabolites. *Nature 2022 603:7903*, 603(7903):907–912, March 2022. ISSN 1476-4687. doi:10.1038/s41586-022-04480-z. URL `https://www.nature.com/articles/s41586-022-04480-z`. Publisher: Nature Publishing Group ISBN: 4158602204480.

[222] Xinyang Song, Ximei Sun, Sungwhan F. Oh, Meng Wu, Yanbo Zhang, Wen Zheng, Naama Geva-Zatorsky, Ray Jupp, Diane Mathis, Christophe Benoist, and Dennis L. Kasper. Microbial bile acid metabolites modulate gut ROR+ regulatory T cell homeostasis. *Nature*, 577(7790):410–415, January 2020. ISSN 0028-0836. doi:10.1038/s41586-019-1865-0. URL `http://www.nature.com/articles/s41586-019-1865-0`.

[223] Clarissa Campbell, Peter T McKenney, Daniel Konstantinovsky, Olga I Isaeva, Michail Schizas, Jacob Verter, Cheryl Mai, Wen-Bing Jin, Chun-Jun Guo, Sara Violante, Ruben J Ramos, Justin R Cross, Krishna Kadaveru, John Hambor, and Alexander Y Rudensky. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature*, 581:475, 2020. doi:10.1038/s41586-020-2193-0. URL `https://doi.org/10.1038/s41586-020-2193-0`.

[224] Solenne Marion, Lyne Desharnais, Nicolas Studer, Yuan Dong, Matheus D. Notter, Suresh Poudel, Laure Menin, Andrew Janowczyk, Robert L. Hettich, Siegfried Hapfelmeier, and Rizlan Bernier-Latmani. Biogeography of microbial bile acid transformations along the murine gut. *Journal of Lipid Research*, 61(11):1450–1463, November 2020. ISSN 00222275. doi:10.1194/jlr.RA120001021.

[225] Amber D. Reed and Casey M. Theriot. Contribution of Inhibitory Metabolites and Competition for Nutrients to Colonization Resistance against Clostridioides difficile by Commensal Clostridium. *Microorganisms*, 9(2):371, February 2021. ISSN 2076-2607. doi:10.3390/microorganisms9020371.

[226] Grace L. Guo and John Y.L. Chiang. Is CYP2C70 the key to new mouse models to understand bile acids in humans? *Journal of Lipid Research*, 61(3):269–271, March 2020. ISSN 00222275. doi:10.1194/jlr.C120000621.

[227] Jianing Li and Paul A. Dawson. Animal models to study bile acid metabolism. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1865(5):895–911, May 2019. ISSN 09254439. doi:10.1016/j.bbadis.2018.05.011.

[228] Thomas Dubois, Yannick D. N. Tremblay, Audrey Hamiot, Isabelle Martin-Verstraete, Julien Deschamps, Marc Monot, Romain Briandet, and Bruno Dupuy. A microbiota-generated bile salt induces biofilm formation in Clostridium difficile. *npj Biofilms and Microbiomes*, 5(1):14, May 2019. ISSN 2055-5008. doi:10.1038/s41522-019-0087-4.

[229] Samantha C. Kisthardt, Rajani Thanissery, Colleen M. Pike, Matthew H. Foley, and Casey M. Theriot. The microbial-derived bile acid lithocholate and its epimers inhibit *Clostridioides difficile* growth and pathogenicity while sparing members of the gut microbiota. *Journal of Bacteriology*, 205(9):e00180–23, September 2023. ISSN 0021-9193, 1098-5530. doi:10.1128/jb.00180-23.

[230] Jordy Evan Sulaiman, Jaron Thompson, Yili Qian, Eugenio I. Vivas, Christian Diener, Sean M. Gibbons, Nasia Safdar, and Ophelia S. Venturelli. Elucidating human gut microbiota interactions that robustly inhibit diverse Clostridioides difficile strains across different nutrient landscapes. *Nature Communications*, 15(1):7416, August 2024. ISSN 2041-1723. doi:10.1038/s41467-024-51062-w.

[231] Jhansi L. Leslie, Matthew L. Jenior, Kimberly C. Vendrov, Alexandra K. Standke, Madeline R. Barron, Tricia J. O'Brien, Lavinia Unverdorben, Pariyamon Thaprawat, Ingrid L. Bergin, Patrick D. Schloss, and Vincent B. Young. Protection from Lethal Clostridioides difficile Infection via Intraspecies Competition for Cogerminant. *mBio*, 12(2):e00522–21, April 2021. ISSN 2150-7511. doi:10.1128/mBio.00522-21.

[232] Kristin J. Nagaro, S. Tyler Phillips, Adam K. Cheknis, Susan P. Sambol, Walter E. Zukowski, Stuart Johnson, and Dale N. Gerding. Nontoxigenic Clostridium difficile Protects Hamsters against Challenge with Historic and Epidemic Strains of Toxigenic BI/NAP1/027 C. difficile. *Antimicrobial Agents and Chemotherapy*, 57(11):5266–5270, October 2013. doi:10.1128/aac.00580-13.

[233] Meina Neumann-Schaal, Dieter Jahn, and Kerstin Schmidt-Hohagen. Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile. *Frontiers in Microbiology*, 10:219, February 2019. ISSN 1664-302X. doi:10.3389/fmicb.2019.00219.

[234] Marijana Bosnjak, Avinash V. Karpe, Thi Thu Hao Van, Despina Kotsanas, Grant A. Jenkin, Samuel P. Costello, Priscilla Johanesen, Robert J. Moore, David J. Beale,

Yogitha N. Srikhanta, Enzo A. Palombo, Sarah Larcombe, and Dena Lyras. Multi-omics analysis of hospital-acquired diarrhoeal patients reveals biomarkers of enterococcal proliferation and Clostridioides difficile infection. *Nature Communications*, 14 (1):7737, November 2023. ISSN 2041-1723. doi:10.1038/s41467-023-43671-8.

[235] Helena M. B. Seth-Smith, Michael Biggel, Tim Roloff, Vladimira Hinic, Thomas Bodmer, Martin Risch, Carlo Casanova, Andreas Widmer, Rami Sommerstein, Jonas Marschall, Sarah Tschudin-Sutter, and Adrian Egli. Transition From PCR-Ribotyping to Whole Genome Sequencing Based Typing of Clostridioides difficile. *Frontiers in Cellular and Infection Microbiology*, 11, 2021. ISSN 2235-2988.

[236] Xingxing Xu, Qiao Bian, Yun Luo, Xiaojun Song, Shan Lin, Huan Chen, Qian Liang, Meixia Wang, Guangyong Ye, Bo Zhu, Liang Chen, Yi-Wei Tang, Xianjun Wang, and Dazhi Jin. Comparative Whole Genome Sequence Analysis and Biological Features of Clostridioides difficile Sequence Type 2‡. *Frontiers in Microbiology*, 12, 2021. ISSN 1664-302X.

[237] Ruba Barbar, Jessica N Brazelton, Karen C Carroll, Shawna Lewis, Dimitrios Bourdas, Anita Tembo, Linda Gluck, Hana Hakim, and Randall T Hayden. Molecular Epidemiology and Genetic Relatedness of *Clostridioides difficile* Isolates in Pediatric Oncology and Transplant Patients Using Whole Genome Sequencing. *Clinical Infectious Diseases*, 76(3):e1071–e1078, February 2023. ISSN 1058-4838, 1537-6591. doi:10.1093/cid/ciac459.

[238] Brad Brown, Marc Allard, Michael C. Bazaco, Joseph Blankenship, and Travis Minor. An economic evaluation of the Whole Genome Sequencing source tracking program in the U.S. *PLOS ONE*, 16(10):e0258262, October 2021. ISSN 1932-6203. doi:10.1371/journal.pone.0258262.

[239] Travis K. Price, Susan Realegeno, Ruel Mirasol, Allison Tsan, Sukantha Chandrasekaran, Omai B. Garner, and Shangxin Yang. Validation, Implementation, and Clinical Utility of Whole Genome Sequence-Based Bacterial Identification in the Clinical Microbiology Laboratory. *The Journal of Molecular Diagnostics*, 23(11):1468–1477, November 2021. ISSN 1525-1578. doi:10.1016/j.jmoldx.2021.07.020.

[240] Thomas M. Elliott, Patrick N. Harris, Leah W. Roberts, Michelle Doidge, Trish Hurst, Krispin Hajkowicz, Brian Forde, David L. Paterson, and Louisa G. Gordon. Cost-effectiveness analysis of whole-genome sequencing during an outbreak of carbapenem-resistant *Acinetobacter baumannii*. *Antimicrobial Stewardship & Healthcare Epidemiology*, 1(1):e62, 2021. ISSN 2732-494X. doi:10.1017/ash.2021.233.

[241] Praveen Kumar, Alexander J Sundermann, Elise M Martin, Graham M Snyder, Jane W Marsh, Lee H Harrison, and Mark S Roberts. Method for Economic Evaluation of Bacterial Whole Genome Sequencing Surveillance Compared to Standard of Care in Detecting Hospital Outbreaks. *Clinical Infectious Diseases*, 73(1):e9–e18, July 2021. ISSN 1058-4838, 1537-6591. doi:10.1093/cid/ciaa512.

[242] Claudio Donati, Moreno Zolfo, Davide Albanese, Duy Tin Truong, Francesco Asnicar, Valerio Iebba, Duccio Cavalieri, Olivier Jousson, Carlotta De Filippo, Curtis Huttenhower, and Nicola Segata. Uncovering oral Neisseria tropism and persistence using metagenomic sequencing. *Nature Microbiology*, 1(7):16070, May 2016. ISSN 2058-5276. doi:10.1038/nmicrobiol.2016.70.

[243] Nadja Mostacci, Tsering Monika Wüthrich, Léa Siegwald, Silas Kieser, Ruth Steinberg, Olga Sakwinska, Philipp Latzin, Insa Korten, and Markus Hilty. Informed interpretation of metagenomic data by StrainPhlAn enables strain retention analyses of the upper airway microbiome. *mSystems*, 8(6):e00724–23, December 2023. ISSN 2379-5077. doi:10.1128/msystems.00724-23.

[244] Gustavo Monasterio, Rodrigo A. Morales, David A. Bejarano, Xesús M. Abalo, Jennifer Fransson, Ludvig Larsson, Andreas Schlitzer, Joakim Lundeberg, Srustidhar Das, and Eduardo J. Villablanca. A versatile tissue-rolling technique for spatial-omics analyses of the entire murine gastrointestinal tract. *Nature Protocols*, June 2024. ISSN 1754-2189, 1750-2799. doi:10.1038/s41596-024-01001-2.

[245] Regina Z. Cer, Logan J. Voegtly, Bishwo N. Adhikari, Brian L. Pike, Matthew R. Lueder, Lindsay A. Glang, Francisco Malagon, Ernesto Santa Ana, James M. Regeimbal, Maria F. Potts-Szoke, Kevin L. Schully, Darci R. Smith, and Kimberly A. Bishop-Lilly. Genomic and virologic characterization of samples from a shipboard outbreak of COVID-19 reveals distinct variants within limited temporospatial parameters. *Frontiers in Microbiology*, 13, 2022. ISSN 1664-302X.

[246] Yunlong Liu, Morteza H. Ghaffari, Tao Ma, and Yan Tu. Impact of database choice and confidence score on the performance of taxonomic classification using Kraken2. *aBIOTECH*, 5(4):465, July 2024. doi:10.1007/s42994-024-00178-0.

[247] Daniel J. Nasko, Sergey Koren, Adam M. Phillippy, and Todd J. Treangen. RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. *Genome Biology*, 19(1):165, December 2018. ISSN 1474-760X. doi:10.1186/s13059-018-1554-6.

[248] Chunyu Zhao, Zhou Jason Shi, and Katherine S. Pollard. Pitfalls of genotyping microbial communities with rapidly growing genome collections. *Cell Systems*, 14(2): 160–176.e3, February 2023. ISSN 24054712. doi:10.1016/j.cels.2022.12.007.

[249] Robyn J. Wright, Andrè M. Comeau, and Morgan G. I. Langille. From defaults to databases: Parameter and database choice dramatically impact the performance of metagenomic taxonomic classification tools. *Microbial Genomics*, 9(3):mgen000949, March 2023. ISSN 2057-5858. doi:10.1099/mgen.0.000949.

[250] Qinzhong Tian, Pinglu Zhang, Yixiao Zhai, Yansu Wang, and Quan Zou. Application and Comparison of Machine Learning and Database-Based Methods in Taxonomic Classification of High-Throughput Sequencing Data. *Genome Biology and Evolution*, 16(5):evae102, May 2024. ISSN 1759-6653. doi:10.1093/gbe/evae102.

[251] Charles Hall Davis Williamson, Nathan E. Stone, Amalee E. Nunnally, Chandler C. Roe, Adam J. Vazquez, Samantha A. Lucero, Heidie Hornstra, David M. Wagner, Paul Keim, Maja Rupnik, Sandra Janezic, and Jason William Sahl. Identification of novel, cryptic Clostridioides species isolates from environmental samples collected from diverse geographical locations. *Microbial Genomics*, 8(2), February 2022. ISSN 2057-5858. doi:10.1099/mgen.0.000742.

[252] Benjamin Hillmann, Gabriel A. Al-Ghalith, Robin R. Shields-Cutler, Qiyun Zhu, Daryl M. Gohl, Kenneth B. Beckman, Rob Knight, and Dan Knights. Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems*, 3(6): 10.1128/msystems.00069–18, November 2018. doi:10.1128/msystems.00069-18.

[253] Donglai Liu, Haiwei Zhou, Teng Xu, Qiwen Yang, Xi Mo, Dawei Shi, Jingwen Ai, Jingjia Zhang, Yue Tao, Donghua Wen, Yigang Tong, Lili Ren, Wen Zhang, Shumei Xie, Weijun Chen, Wanli Xing, Jinyin Zhao, Yilan Wu, Xianfa Meng, Chuan Ouyang, Zhi Jiang, Zhikun Liang, Haiqin Tan, Yuan Fang, Nan Qin, Yuanlin Guan, Wei Gai, Sihong Xu, Wenjuan Wu, Wenhong Zhang, Chuntao Zhang, and Youchun Wang. Multicenter assessment of shotgun metagenomics for pathogen detection. *eBioMedicine*, 74:103649, December 2021. ISSN 23523964. doi:10.1016/j.ebiom.2021.103649.

[254] Jin Liu, Xiaokai Wang, Hailiang Xie, Qinghua Zhong, and Yan Xia. Analysis and evaluation of different sequencing depths from 5 to 20 million reads in shotgun metagenomic sequencing, with optimal minimum depth being recommended. *Genome*, 65(9): 491–504, September 2022. ISSN 0831-2796. doi:10.1139/gen-2021-0120.

[255] Marina Muñoz, Dora Inés Ríos-Chaparro, Manuel Alfonso Patarroyo, and Juan David Ramírez. Determining Clostridium difficile intra-taxa diversity by mining multilocus sequence typing databases. *BMC Microbiology*, 17(1):62, December 2017. ISSN 1471-2180. doi:10.1186/s12866-017-0969-7.

[256] Jihoe Kim, Daniel Darley, and Wolfgang Buckel. 2-Hydroxyisocaproyl-CoA dehydratase and its activator from *Clostridium difficile*. *The FEBS Journal*, 272(2):550–561, January 2005. ISSN 1742-464X, 1742-4658. doi:10.1111/j.1742-4658.2004.04498.x.

[257] Matthew L. Jenior, Jhansi L. Leslie, Vincent B. Young, and Patrick D. Schloss. *Clostridium Difficile* Alters the Structure and Metabolism of Distinct Cecal Microbiomes during Initial Infection To Promote Sustained Colonization. *mSphere*, 3(3): e00261–18, June 2018. ISSN 2379-5042. doi:10.1128/mSphere.00261-18.

[258] Laurent Bouillaut, William T. Self, and Abraham L. Sonenshein. Proline-dependent regulation of Clostridium difficile stickland metabolism. *Journal of Bacteriology*, 195 (4):844–854, February 2013. ISSN 00219193. doi:10.1128/JB.01492-12.

[259] Joshua R. Fletcher, Colleen M. Pike, Ruth J. Parsons, Alissa J. Rivera, Matthew H. Foley, Michael R. McLaren, Stephanie A. Montgomery, and Casey M. Theriot. Clostridioides difficile exploits toxin-mediated inflammation to alter the host nutritional land-

scape and exclude competitors from the gut microbiota. *Nature Communications*, 12 (1):462, January 2021. ISSN 2041-1723. doi:10.1038/s41467-020-20746-4.