Clostridioides difficile beyond the disease-centred perspective: Beneficial properties in healthy infants and over-diagnosis in diseased adults identified by species- and SNV-based metagenomic analysis

Dissertation zur Erlangung des

naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

Pamela Ferretti

Geburtsort: Reggio nell’Emilia, Italy

Würzburg, 2021

Eingereicht am: 08.10.2021

Mitglieder der Promotionskommission:

Vorsitzender: ……………………………………………..

Gutachter: Prof. Dr. Peer Bork

Gutachter: Prof. Dr. Thomas Dandekar

Tag des Promotionskolloquiums: ……………………..…

Doktorurkunde ausgehändigt am: ……………………..…….

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: „***Clostridioides difficile* jenseits der krankheitszentrierten Perspektive: Vorteilhafte Eigenschaften bei gesunden Säuglingen und Überdiagnose bei erkrankten Erwachsenen, identifiziert durch spezies- und SNV-basierte metagenomische Analyse**“, eigenständig, d. h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen, als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Weiterhin erkläre ich, dass bei allen Abbildungen und Texten bei denen die Verwertungsrechte (Copyright) nicht bei mir liegen, diese von den Rechtsinhabern eingeholt wurden und die Textstellen bzw. Abbildungen entsprechend den rechtlichen Vorgaben gekennzeichnet sind sowie bei Abbbildungen, die dem Internet entnommen wurden, der entsprechende Hypertextlink angegeben wurde.

Affidavit

I hereby declare that my thesis entitled: "***Clostridioides difficile* beyond the disease-centred perspective: Beneficial properties in healthy infants and over-diagnosis in diseased adults identified by species- and SNV-based metagenomic analysis**" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore I verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

Besides I declare that if I do not hold the copyright for figures and paragraphs, I obtained it from the rights holder and that paragraphs and figures have been marked according to law or for figures taken from the internet the hyperlink has been added accordingly.

Würzburg, den 27/09/2021

Signature PhD-student

Summary

Clostridioides difficile is a bacterial species well known for its ability to cause *C. difficile* infection (also known as CDI). The investigation of the role of this species in the human gut has been so far dominated by a disease-centred perspective, focused on studying *C. difficile* in relation to its associated disease.

In this context, the first aim of this thesis was to combine publicly available metagenomic data to analyse the microbial composition of stool samples from patients diagnosed with CDI, with a particular focus on identifying a CDI-specific microbial signature.

However, similarly to many other bacterial species inhabiting the human gut, *C. difficile* association with disease is not valid in absolute terms, as *C. difficile* can be found also among healthy subjects. Further aims of this thesis were to 1) identify potential *C. difficile* reservoirs by screening a wide range of habitats, hosts, body sites and age groups, and characterize the biotic context associated with *C. difficile* presence, and 2) investigate *C. difficile* within-species diversity and its toxigenic potential across different age groups.

The first part of the thesis starts with the description of the concepts and definitions used to identify bacterial species and within-species diversity, and then proceeds to provide an overview of the bacterial species at the centre of my investigation, *C. difficile*. The first Chapter includes a detailed description of the discovery, biology and physiology of this clinically relevant species, followed by an overview of the diagnostic protocols used in the clinical setting to diagnose CDI.

The second part of the thesis describes the methodology used to investigate the questions mentioned above, while the third part presents the results of such investigative effort. I first show that *C. difficile* could be found in only a fraction of the CDI samples and that simultaneous colonization of multiple enteropathogenic species able to cause CDI-like clinical manifestations is more common than previously thought, raising concerns about CDI overdiagnosis. I then show that the CDI-associated gut microbiome is characterized by a specific microbial signature, distinguishable from the community composition associated with non-CDI diarrhea. Beyond the nosocomial and CDI context, I show that while rarely found in adults, *C. difficile* is a common member of the infant gut microbiome, where its presence is associated with multiple indicators typical of a desirable healthy microbiome development.

In addition, I describe *C. difficile* extensive carriage among asymptomatic subjects, of all age groups and a potentially novel clade of *C. difficile* identified exclusively among infants.

Finally, I discuss the limitations, challenges and future perspectives of my investigation.

Zusammenfassung

Clostridioides difficile ist eine Bakterienart, die für ihre Fähigkeit bekannt ist, eine *C. difficile*-Infektion (auch bekannt als CDI) zu verursachen. Die Untersuchung der Rolle dieser Spezies im menschlichen Darm wurde bisher von einer krankheitszentrierten Perspektive dominiert, die sich auf die Untersuchung von *C. difficile* in Bezug auf die damit verbundene Erkrankung konzentrierte.

In diesem Zusammenhang war das erste Ziel dieser Arbeit, öffentlich verfügbare metagenomische Daten zu kombinieren, um die mikrobielle Zusammensetzung von Stuhlproben von Patienten mit diagnostizierter CDI zu analysieren, mit besonderem Fokus auf der Identifizierung einer CDI-spezifischen mikrobiellen Signatur.

Wie bei vielen anderen Bakterienarten, die den menschlichen Darm bewohnen, ist jedoch die Assoziation von *C. difficile* mit einer Krankheit nicht absolut gültig, da *C. difficile* auch bei gesunden Probanden gefunden werden kann. Weitere Ziele dieser Arbeit waren 1) die Identifizierung potenzieller *C. difficile*-Reservoirs durch das Screening einer Vielzahl von Habitaten, Wirten, Körperstellen und Altersgruppen und die Charakterisierung des mit der Präsenz von *C. difficile* verbundenen biotischen Kontexts und 2) Untersuchung von *C. difficile* innerhalb der Artenvielfalt und ihr toxigenes Potenzial über verschiedene Altersgruppen hinweg.

Der erste Teil der Dissertation beginnt mit der Beschreibung der Konzepte und Definitionen, die verwendet werden, um Bakterienarten und innerhalb der Artenvielfalt zu identifizieren, und fährt dann fort, einen Überblick über die Bakterienarten zu geben, die im Zentrum meiner Untersuchung, *C. difficile*, stehen. Das erste Kapitel enthält eine detaillierte Beschreibung der Entdeckung, Biologie und Physiologie dieser klinisch relevanten Spezies, gefolgt von einem Überblick über die diagnostischen Protokolle, die im klinischen Umfeld zur Diagnose von CDI verwendet werden.

Der zweite Teil der Arbeit beschreibt die Methodik zur Untersuchung der oben genannten Fragen, während der dritte Teil die Ergebnisse dieser Untersuchungsarbeit präsentiert. Ich zeige zunächst, dass *C. difficile* nur in einem Bruchteil der CDI-Proben gefunden werden konnte und dass die gleichzeitige Besiedlung mehrerer enteropathogener Spezies, die CDI-ähnliche klinische Manifestationen verursachen können, häufiger vorkommt als bisher angenommen, was Bedenken hinsichtlich einer CDI-Überdiagnose aufkommen lässt. Ich zeige dann, dass das CDI-assoziierte Darmmikrobiom durch eine spezifische mikrobielle Signatur gekennzeichnet ist, die sich von der Gemeinschaftszusammensetzung unterscheidet, die mit Nicht-CDI-Diarrhoe verbunden ist. Über den nosokomialen und CDI-Kontext hinaus zeige ich, dass *C. difficile*, obwohl es bei Erwachsenen selten vorkommt, ein häufiges Mitglied des Darmmikrobioms von Säuglingen ist, wo seine Anwesenheit mit mehreren Indikatoren verbunden ist, die typisch für eine wünschenswerte gesunde Mikrobiomentwicklung sind.

Darüber hinaus beschreibe ich die ausgedehnte Beförderung von *C. difficile* bei asymptomatischen Patienten aller Altersgruppen und eine potenziell neue Gruppe von *C. difficile*, die ausschließlich bei Säuglingen identifiziert wurde.

Abschließend diskutiere ich die Grenzen, Herausforderungen und Zukunftsperspektiven meiner Untersuchung.

Acknowledgments

I would like to thank the many people that played an important role over the course of my PhD, and contributed at making these years one of the most exciting and pleasant of my life:

Peer Bork for his continuous supervision, his patience and the many opportunities he gave me to improve as a scientist, visit conferences and meet amazing colleagues around the world.

My TAC members, Thomas Dandekar, Georg Zeller and Nassos Typas for their valuable input and support during my PhD.

Sebastian Schmidt for his supervision which was always spot-on, giving me a lot of freedom to explore (and make mistakes), but always there to help when needed, with a relaxed, optimistic and pragmatic approach. Thanks for your competence and, most of all, for being a great person.

Renato, Ece, Oleks and Alejandro for the absolutely most fantastic time ever in the office (and at parties). I'm sure our friendship will go beyond the PhD and our time at EMBL.

All the members of the Bork group, for the always captivating chats, scientific brainstorming and in general great time.

My friends and colleagues of my batch: Jakob, Tim, Martin, Karen and Matthias for all the parties and get-togethers that made my PhD unforgettable.

A special thanks to Conor, for the great time spent together and for, despite the distance, helping me going through the lockdowns, providing amazing food recipes and for always being up for a chat.

Beyond grateful to my mother Clara, my dad Romano and the rest of my family for their unconditional support, even when my desire for knowledge and exploration led me far away from home. In particular, thanks to my uncle Giulio and my late grandfather Pierino, for pushing me to study hard and get good grades. My grandfather always wanted to become a teacher, but he was forced to abandon school to work in the fields. This PhD is also for him.

Last but not least, I would like to thank the animals that contributed to my happiness during these years:

My 18.5 years old (and counting) cat Lia, for being the most amazing companion, for teaching me always something new about herself and animal communication, for putting up with my (innocuous) behavioral experiments and for letting me train her despite her old age. She will always be in my heart.

The birds that came to my office every day, for making my days so much brighter.

I owe one more acknowledgement: to myself. These PhD years have been the happiest of my life, but were no short of challenges. I challenged myself to leave my comfort zone, to improve myself as a scientist and as a human being, and at times I failed miserably. But I learnt to be kind to myself, to accept my failures, to stand up and try again as a wiser person. One step at a time, without giving up.

Acknowledgements in Italian:

Vorrei ringraziare tutti coloro hanno avuto un ruolo importante nel corso del mio dottorato e che hanno contribuito a rendere questi anni tra i più emozionanti ed indimenticabili della mia vita:

Peer Bork per la sua preziosa supervisione e pazienza, per avermi dato numerose occasioni per migliorare come ricercatrice, partecipare a conferenze e incontrare fantastici colleghi in giro per il mondo.

I membri della mia commissione di valutazione annuale, Thomas Dandekar, Georg Zeller e Nassos Typas, per il loro prezioso contributo e supporto durante il dottorato.

Sebastian Schmidt per la sua impeccabile supervisione, per avermi concesso la libertà di seguire il mio intuito e di fare errori durante durante il percorso, e per esserci nei momenti di necessità, con un approccio rilassato, pragmatico ed ottimista. Grazie per la tua competenza e, sopratutto, per essere una persona fantastica.

Renato, Ece, Oleks e Alejandro per aver reso questi anni (e le tante feste, cene, balli e conferenze) assolutamente indimenticabili. Sono sicura che il nostro legame sarà duraturo.

Tutti i membri del laboratorio di Peer Bork, per le chiacchierate accattivanti ed in generale il bel tempo passato insieme.

Gli amici e colleghi del mio anno: Jakob, Tim, Martin, Karen e Matthias per le numerose feste, ritrovi e per l'indimenticabile predoc course.

Conor, per il fantastico tempo passato assieme, per aver reso, nonostante la distanza, i lockdown e la pandemia molto più leggeri e per essere una continua fonte di ispirazione culinaria.

Un ringraziamento speciale a mia mamma Clara, mio papà Romano e il resto della mia famiglia per il loro supporto incondizionato, anche quando il mio desiderio di conoscenza ed esplorazione mi ha portato lontano da casa. Un grazie particolare a mio zio Giulio e mio nonno Pierino, per avermi sempre spronato a studiare tanto e ad avere buoni voti. Da giovane Pierino voleva diventare un maestro, ma era stato costretto ad abbandonare la scuola per lavorare nei campi. Questo dottorato e' anche per lui.

Per ultimo, ma non in ordine di importanza, vorrei ringraziare gli animali che hanno contribuito alla mia felicità in questi anni: la mia gatta Lia (18.5 anni), per essere la migliore compagna possibile, per avermi insegnato così tanto sulla comunicazione animale e per aver avuto tanta pazienza con i miei (innocui) esperimenti comportamentali e le sessioni di addestramento.

Le cince, per aver reso il mio arrivo in ufficio ogni mattina un vero piacere, e per avermi presentato ai loro piccoli, nidiata dopo nidiata.

Table of Contents

Acknowledgments 9

Table of Contents 13

List of Figures 15

List of Tables 18

List of Publications 19

Abbreviations 21

Chapter 1. Introduction 22

1.1 Species and within-species bacterial diversity 23

1.1.1 Operational definitions for genome-based species boundaries 24

1.1.2 Mechanisms of variation generation and selection within a species 24

1.1.3 Within-species variability definition 26

1.2 Overview of the clinically relevant pathogenic species *Clostridioides difficile* 27

1.2.1 Introduction to *Clostridioides difficile* 27

1.2.2 *C. difficile* in the human gut microbiome 28

1.2.3 Clinical relevance of *C. difficile* 31

1.2.4 *C. difficile* infection (CDI) diagnosis 34

1.3 Thesis outline 37

Chapter 2. Methods 38

Public metagenomes collection 39

Global metagenomic survey of *C. difficile* 39

CDI metagenomic survey of *C. difficile* 39

Data download 39

Metadata collection 39

Preliminary quality control and filtering 40

Taxonomic profiling of the downloaded metagenomic samples 40

Identification of timeseries-representative samples 40

Prevalence and abundance estimations of *C. difficile* 41

Logistic regression and ANOVA 41

Alpha diversity calculation 41

Species co-occurrence analysis 41

Machine learning modelling 42

Linear mixed effect model 42

Evaluation of toxigenic potential in metagenomes 42

Analysis of *C. difficile* within-species diversity from metagenomic samples 42

Chapter 3. Results and discussion 44

3.1 CDI-specific gut microbial signature and potential CDI overdiagnosis revealed by metagenomic meta-analysis 45

3.1.1 *C. difficile* identified in only 30% of CDI metagenomic samples 45

3.1.2 Enrichment of multiple enteropathogens in the gut microbiome of CDI patients 47

3.2 *C. difficile* prevalence, abundance and biotic context across age, geography, health and disease 54

3.2.1 Age-dependent prevalence and abundance of *C. difficile* in human gastrointestinal tract 54

3.2.2 Enrichment of *C. difficile* across diseases 61

3.2.3 Increased diversity and resemblance to the maternal gut associated with *C. difficile* presence in the gut microbiome of healthy infants 62

3.2.4 The unique biotic context associated with *C. difficile* in healthy infants 66

3.2.5 Limited similarity between human and animal gut microbiomes harbouring *C. difficile* 70

3.3 Unexplored *C. difficile* within-species diversity and pathogenicity potential over lifetime 73

3.3.1 Metagenomic-based identification of a potentially novel clade of *C. difficile* exclusively found in infants 73

3.3.2 Extensive asymptomatic carriage of potentially toxigenic *C. difficile* over lifetime 75

3.4 Study limitations and future perspectives 78

Chapter 4. Concluding remarks 80

Appendix 81

References 95

List of Figures

Figure 1. Mechanisms behind variability generation and selection of variants within bacterial species.

Figure 2. Stratification of the terminologies used to identify within-species variability, ordered by resolution, ranging from a single nucleotide variation in the whole genome (100% ANI) to 97% ANI, the recognised species-level boundary.

Figure 3. Growth of *C. difficile*-related research since its first description in 1935, as of June 2021 in Pubmed.

Figure 4. Overview of the processes involved in the gut microbiome transition from healthy to diseased, with focus on factors influencing *C. difficile* sporulation, germination and blooming.

Figure 5. Mechanism of action of *C. difficile* toxins in the human gut.

Figure 6. Divergence of the known *C. difficile* clades.

Figure 7. Overview of 10 public CDI or diarrheal cohorts used for the CDI meta-analysis, and rarefied species richness across the three sample groups.

Figure 8. Prevalence of *C. difficile* and other antibiotic-associated diarrhea (AAD) species, divided by study.

Figure 9. Prevalence and relative abundance of species known to be able to induce antibiotic-associated diarrhea (AAD) among CDI and control samples.

Figure 10. Species significantly enriched or depleted in terms of relative abundance in CDI compared to diseased and healthy controls.

Figure 11. Microbial signature associated with the three sample groups, as seen by LASSO model.

Figure 12. AUC values of the LASSO model for all samples as well as for single study populations, comparing CDI samples with healthy and diseased controls combined (left) and CDI compared with healthy controls only (right).

Figure 13. Prevalence of potentially toxigenic species found among *C. difficile* positive and *C. difficile* negative samples, divided by sample group.

Figure 14. (A) Overview of the 42,814 samples, from 253 publicly available studies, used for the global *C. difficile* meta-analysis. (B) Stratification of the human gut samples, divided by age category and health status. (C) Overview of the *C. difficile* positive samples, as seen by mOTUs v2.0 (Milanese et al. 2019), using 2 marker genes.

Figure 15. *C. difficile* prevalence in human stool samples, divided by health status and age group.

Figure 16. *C. difficile* prevalence in stool samples divided by (A) health status and delivery mode, (B) health status and prematurity, (C) combination of health status, delivery mode and prematurity and by (D) feeding mode.

Figure 17. Time of the first appearance of *C. difficile* in infants and children timeseries.

Figure 18. *C. difficile* relative abundance over life time in human stool samples of (A) healthy and (B) diseased subjects. (C) Differential relative abundance of *C. difficile* divided by health status and age group.

Figure 19. Prevalence of *C. difficile* across diseases and drugs.

Figure 20. Species richness across all human stool samples, age groups and health status in presence or absence of *C. difficile*.

Figure 21. Species richness of human stool samples in presence or absence of *C. difficile*, divided by (A) delivery mode, prematurity and health status. (B) Species richness in healthy adult and infant gut stool samples across continents.

Figure 22. Bray-Curtis dissimilarity of healthy infant-mother pairs in presence or absence of *C. difficile* in stools across the first four years of life.

Figure 23 Positive and negative species associated with *C. difficile* divided by age group and health status. Annotation of those species includes oxygen requirement, prevalence trends over life time in the healthy human gut, enrichment in westernised populations.

Figure 24. Geographical and age group distribution of the samples included in the global survey of *C. difficile* across animal species.

Figure 25. *C. difficile* (A) prevalence and (B) relative abundance in animal stool samples, compared to humans, as seen by mOTUs v2.0, using two marker genes. Species co-occurring (logOR>0) with *C. difficile* across gut stool samples from (C) humans (n=14,095) and (D) animals (n=3,967).

Figure 26. SNV similarity across *C. difficile* positive metagenomic samples from our global survey, and genomes from known *C. difficile* clades.

Figure 27. Prevalence of potentially toxigenic *C. difficile* in the stools of (A) healthy and diseased across age groups, (B) diet during the first year of life, (C) delivery mode and gestational age. Toxigenic potential defined via detection of either one or both of *C. difficile* Toxin A (TcdA) or Toxin B genes (TcdB).

Figure 28. Prevalence of potentially toxigenic species other than *C. difficile*, in infants and adults, divided by health status.

Supplementary figure 1. Overview of the different operational definitions of “strain” in three fields of investigation and correlation between similarity of the core genome (measure via ANI) and similarity of the gene content (measure via Jaccard Index) in conspecific isolate genomes from 155 bacterial species.

Supplementary figure 2. Taxonomic composition at the species level for CDI and control samples.

Supplementary figure 3. Geographical distribution of *C. difficile* prevalence in the stools of healthy infants and adults divided by country and continent.

Supplementary figure 4. Logistic regression adjusted ANOVA p-values, divided by age group.

Supplementary figure 5. Prevalence of species known from the literature to be able to induce antibiotic-associated diarrhea (AAD), across healthy and diseased human gut stool samples over lifetime.

Supplementary figure 6. (A) Species evenness (as the inverse Simpson index divided by Richness) across age groups and health status. (B) Evenness in infant samples divided by delivery mode, prematurity and health status.

Supplementary figure 7. Prevalence of *C. difficile* toxin genes in the stools of diseased patients, divided by disease type or drugs.

List of Tables

Table 1. Diagnostic tests available for CDI diagnosis.

Supplementary table 1. List of publicly available CDI studies used for the *C. difficile* CDI meta-analysis *C. difficile*.

Supplementary Table 2. Most prevalent toxin genes in CDI samples

Supplementary table 3. List of the 253 publicly available studies used in the *C. difficile* global meta-analysis.

Supplementary table 4. List of animal species included in our global *C. difficile* survey, including the number of samples per each species and the country of sampling.

Supplementary table 5. Metadata table for the 42,814 metagenomic samples used for the *C. difficile* global meta-analysis.

Supplementary table 6. mOTU2.0 taxonomic profiles (using 2 marker genes) for the 42,814 metagenomic samples used for the *C. difficile* global meta-analysis.

List of Publications

Publications resulting directly from Doctoral studies:

***C. difficile* as pathobiont: beneficial properties in infants and overdiagnosis in adults identified by metagenomic analysis**
Pamela Ferretti, Jakob Wirbel, Oleksandr M Maistrenko, Thea Van Rossum, Renato Alves, Anthony Fullam, Wasiu Akanni, Christian Schudoma, Michael Kuhn, Georg Zeller, Thomas SB Schmidt and Peer Bork
Ferretti et al, in preparation

Towards the biogeography of prokaryotic genes
Luis Pedro Coelho, Renato Alves, Álvaro Rodríguez del Río, Pernille Neve Myers, Carlos P. Cantalapiedra, Joaquín Giner-Lamia, Thomas Sebastian Schmidt, Daniel Mende, Askarbek Orakov, Ivica Letunic, Falk Hildebrand, Thea Van Rossum, Sofia K. Forslund, Supriya Khedkar, Oleksandr M. Maistrenko, Longhao Jia, **Pamela Ferretti**, Shinichi Sunagawa, Xing-Ming Zhao, Henrik Bjørn Nielsen, Jaime Huerta-Cepas and Peer Bork
Accepted in principle, Nature

Dispersal strategies shape persistence and evolution of human gut bacteria
Falk Hildebrand, Toni I. Gossmann, Clémence Frioux, Ezgi Özkurt, Pernille Neve Myers, **Pamela Ferretti**, Michael Kuhn, Mohammad Bahram, Henrik Bjørn Nielsen, Peer Bork
Cell Host & Microbe, 2021

Metagenomic assessment of the global diversity and distribution of bacteria and fungi
Mohammad Bahram, Tarquin Netherway, Clémence Frioux, **Pamela Ferretti**, Luis Pedro Coelho, Stefan Geisen, Peer Bork, Falk Hildebrand
Environmental Microbiology, 2021

Diversity within species: interpreting strains in microbiomes
Thea Van Rossum, **Pamela Ferretti**, M. Oleksandr Maistrenko, Peer Bork
Nature Reviews Microbiology, 2020

Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome
Pamela Ferretti, Edoardo Pasolli, Adrian Tett, Francesco Asnicar, Valentina Gorfer, Sabina Fedi, Federica Armanini, Duy Tin Truong, Serena Manara, Moreno Zolfo, Francesco Beghini, Roberto Bertorelli, Veronica De Sanctis, Ilaria Bariletti, Rosarita Canto, Rosanna Clementi, Marina Cologna, Tiziana Crifò, Giuseppina Cusumano, Stefania Gottardi, Claudia Innamorati, Caterina Masè, Daniela Postai, Daniela Savoi, Sabrina Duranti, Andrea Gabriele Lugli, Leonardo Mancabelli, Francesca Turroni, Chiara Ferrario, Christian Milani, Marta Mangifesta, Rosaria Anzalone, Alice Viappiani, Moran Yassour, Hera Vlamakis, Ramnik Xavier, Carmen Maria Collado, Omry Koren, Saverio Tateo, Massimo Soffiati, Anna Pedrotti, Marco Ventura, Curtis Huttenhower, Peer Bork, Nicola Segata
Cell Host & Microbe, 2018

Other publications before and during Doctoral studies:

Strain-level analysis of mother-to-child bacterial transmission during the first few months of life
Moran Yassour, Eeva Jason, Larson J Hogstrom, Timothy D Arthur, Surya Tripathi, Heli Siljander, Jenni Selvenius, Sami Oikarinen, Heikki Hyöty, Suvi M Virtanen, Jorma Ilonen, **Pamela Ferretti**, Edoardo Pasolli, Adrian Tett, Francesco Asnicar, Nicola Segata, Hera Vlamakis, Eric S Lander, Curtis Huttenhower, Mikael Knip, Ramnik J Xavier
Cell Host & Microbe, 2018

The short-term impact of probiotic consumption on the oral cavity microbiome

Erik Dassi, **Pamela Ferretti**, Giuseppina Covello, Roberto Bertorelli, Michela A Denti, Veronica De Sanctis, Adrian Tett, Nicola Segata
Scientific reports, 2018

Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission
Sabrina Duranti, Gabriele Andrea Lugli, Leonardo Mancabelli, Federica Armanini, Francesca Turroni, Kieran James, **Pamela Ferretti**, Valentina Gorfer, Chiara Ferrario, Christian Milani, Marta Mangifesta, Rosaria Anzalone, Moreno Zolfo, Alice Viappiani, Edoardo Pasolli, Ilaria Bariletti, Rosarita Canto, Rosanna Clementi, Marina Cologna, Tiziana Crifò, Giuseppina Cusumano, Sabina Fedi, Stefania Gottardi, Claudia Innamorati, Caterina Masè, Daniela Postai, Daniela Savoi, Massimo Soffiati, Saverio Tateo, Anna Pedrotti, Nicola Segata, Douwe van Sinderen, Marco Ventura
Microbiome, 2017

Experimental metagenomics and ribosomal profiling of the human skin microbiome
Pamela Ferretti, Stefania Farina, Mario Cristofolini, Giampiero Girolomoni, Adrian Tett, Nicola Segata
Experimental dermatology, 2017

Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling
Francesco Asnicar, Serena Manara, Moreno Zolfo, Duy Tin Truong, Matthias Scholz, Federica Armanini, **Pamela Ferretti**, Valentina Gorfer, Anna Pedrotti, Adrian Tett, Nicola Segata
MSystems, 2017

Exploring vertical transmission of bifidobacteria from mother to child
Christian Milani, Leonardo Mancabelli, Gabriele Andrea Lugli, Sabrina Duranti, Francesca Turroni, Chiara Ferrario, Marta Mangifesta, Alice Viappiani, **Pamela Ferretti**, Valentina Gorfer, Adrian Tett, Nicola Segata, Douwe van Sinderen, Marco Ventura
Applied and environmental microbiology, 2015

The new phylogenesis of the genus Mycobacterium
Enrico Tortoli, Tarcisio Fedrizzi, Monica Pecorari, Elisabetta Giacobazzi, Veronica De Sanctis, Roberto Bertorelli, Antonella Grottola, Anna Fabio, **Pamela Ferretti**, Francesca Di Leva, Giulia Fregni Serpini, Sara Tagliazucchi, Fabio Rumpianesi, Olivier Jousson, Nicola Segata
International Journal of Mycobacteriology, 2015

Abbreviations

AAD: antibiotic-associated diarrhea

ANI: average nucleotide identity

ANOVA: analysis of variance

ASCVD: atherosclerotic cardiovascular disease

CDI: *C. difficile* infection

CDT: *C. difficile* binary toxin gene

D-Ctr: diseased controls

DDH: DNA-DNA hybridization

EIA: enzymatic ImmunoAssays

ENA: european nucleotide archive

FDR: false discovery rate

FMT: fecal microbiota transplantation

GDH: glutamate dehydrogenase

H-Ctr: healthy controls

HGT: horizontal gene transfer

HR: homologous recombination

LASSO: least absolute shrinkage and selection operator

MLST: multi-locus sequence typing

NAAT: nucleic acid amplification test

NEC: necrotizing enterocolitis

NICU: neonatal intensive care unit

PaLoc: *C. difficile* pathogenicity locus

PPI: proton pump inhibitor

PMC: pseudomembranous colitis

rCDI: recurrent *C. difficile* infection

SCFA: short chain fatty acids

SNV: single-nucleotide variant

T2D: type II diabetes

TcdA: *C. difficile* toxin A gene

TcdB: *C. difficile* toxin B gene

Chapter 1. Introduction

When observed for the first time in 1676 by the Dutch microscopist Antonie van Leeuwenhoek, microorganisms appeared to be silently present everywhere, from rain water to the human mouth (Lane 2015). While at the time it was not known what these microorganisms were eating or how they were reproducing, it was clear that they could cover a wide range of shapes and sizes. The modern term “bacteria" (latinization from the Greek βακτήριον, meaning “small staff”), now used to refer to these microorganisms, was introduced in 1838 by German naturalist Christian Gottfried Ehrenberg (Hoppe 1983).

The sheer variety of bacteria led to the necessity to label them, and organize the available knowledge into a more comprehensive scheme. The first modern attempt to classify bacteria was done in 1923 with the publication of Bergey’s Manual of Determinative Bacteriology (Hugenholtz et al. 2021). The classification was mostly based on morphology, cultivation conditions and pathogenicity, and was composed of nested hierarchical levels of relatedness, ranging from rank to species-level (similarly to what was already established for the classification of animals and plants). However the bacterial classification on the basis of phenotypic traits is intrinsically subjective and inaccurate, especially at the species level (Hugenholtz et al. 2021).

The introduction of genome sequencing techniques in the past decades offered an alternative and much more reliable and quantitative approach to classify bacteria (Hugenholtz et al. 2021). Genome sequencing also enabled the study of bacteria, not only as single isolates, but also as an integral part of their microbial community.

Bacterial species, in particular in the gut, modulate and are modulated by the surrounding bacteria, via an intricate network of direct and indirect interactions. The investigation of such interactions can help identify potential therapeutic strategies to contrast the overgrowth of pathogenic species, such as *Clostridioides difficile*.

In the first part of this Chapter, I discuss the concept of prokaryotic species and how its operational definition allows the targeted investigation of specific bacterial species using metagenomic data. I then discuss the drivers of diversity within species and how such variability can be stratified.

In the second part of this Chapter, I provide an overview of *Clostridioides difficile,* the pathogenic bacterial species at the centre of our investigation, recently listed among the most urgent threads in hospital-acquired infections worldwide ((u.s.) and Centers for Disease Control and Prevention (U.S.) 2019), and of its associated disease (*C. difficile* infection).

1.1 Species and within-species bacterial diversity

The efforts to identify widely-accepted and biology-driven definitions of species and within species diversity have been fueled by the necessity to classify bacterial genomes, either obtained via cultivation or metagenomic sequencing, and quantify their relatedness. In metagenomics, species-level profiling, based on an operational definition of species, allows to study a specific species of interest, distinguishing it from the rest of the sequenced community, while within-species-level profiling allows a more in-depth analysis of phenotypic traits and sequence variation.

1.1.1 Operational definitions for genome-based species boundaries

For many decades, the operational definition of bacterial species was based on genome similarity, measured using DNA-DNA hybridization (DDH), with genomes belonging to the same species (conspecific genomes) characterized by DDH >70%. More recently, sequencing methods introduced another metric, based on the average nucleotide identity (ANI) (Whitman et al. 2016; Truper and Euzeby 2009), with >70% DDH corresponding approximately to >94% of ANI, if computed on the core genome, and >96% ANI, if computed on universal marker genes (Goris et al. 2007; Konstantinidis and Tiedje 2005; Richter and Rosselló-Móra 2009; Mende et al. 2013).

To date, more than twenty different definitions of species are available in the literature (Mayden 1997; Wilkins 2003; Hey 2001; Bapteste et al. 2009), with some debating the existence of bacterial species altogether (Doolittle 2012). The most widely accepted definition identifies the operational bacterial species boundary at ~95% ANI, when computed on whole genomes, and 96.5% ANI, when computed on marker genes. These specific thresholds, suggested by early studies (Konstantinidis and DeLong 2008; Caro-Quintero and Konstantinidis 2012), have been recently confirmed by large-scale metagenomic analyses (Olm et al. 2020; Rodriguez-R et al. 2021) and are widely used in practice.

However, even within the same species, genomes can substantially differ in terms of mutations and gene content, potentially leading to different phenotypic traits. For example, the *Escherichia coli* species includes a wide variety of environmental and host-associated strains, the latter ranging from commensal to highly pathogenic and even carcinogenic ones (Cuevas-Ramos et al. 2010; Loman et al. 2013). Different strains of *Helicobacter pylori* have been associated with different risks for gastric cancer (Blaser et al. 1995), while some strains of *Eggerthella lenta* have been shown to inactivate the cardiac drug digoxin (Haiser et al. 2013). Other examples include *Clostridioides difficile* and *Bacteroides fragilis*, for which only some strains are pathogenic (Natarajan et al. 2013; Jasemi et al. 2020).

In the following paragraphs, I briefly describe how variation within species is generated and selected, and how it can be quantified and stratified in a meaningful way.

1.1.2 Mechanisms of variation generation and selection within a species

Similarly to the species level, within-species variants are subject to constant pressure from diverging and cohesive forces (**Figure 1**). Diverging forces tend to increase the genetic variability within a species. They include mutations (such as the ones induced by DNA replication errors, DNA repair errors or exposure to mutagenic substances), and within-species gene flow (such as horizontal gene transfer (HGT), homologous recombination (HR) and acquisition of plasmids).

Cohesive forces instead reduce the genetic variability within a species, and include habitat-associated pressure and population size (Charlesworth 2009). The generated within-species variation is then shaped by genetic drift and selective pressure (**Figure 1**).
However, several mechanisms have a double nature, as they can either increase or decrease the genetic variability within a species. An example of such mechanisms is HGT-mediated gene flow. If an HGT event transfers a gene to a cell of a naïve population, it will increase the genetic diversity within the naïve population, while if the HGT event spreads a gene across all cells within a population, it will decrease the genetic diversity within that population. Factors influencing selective pressure can be habitat-specific (such as age, diet and lifestyle in human and animal hosts, or salinity in the ocean) or can be found across multiple habitats (such as temperature, nutrient availability and oxygen concentration) (**Figure 1**).

Diverging and cohesive forces are constantly shaping microbial communities across all hierarchical levels. However, it is at the within-species level that minor changes in the equilibrium between these forces can considerably alter the structure of bacterial species.

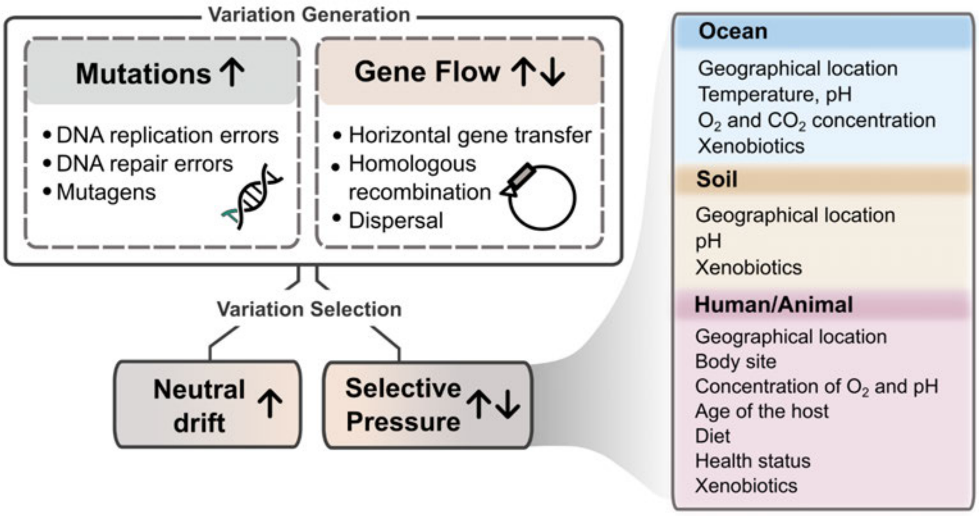


Figure 1. Mechanisms behind variability generation and selection of variants within bacterial species. The arrows indicate if the mechanisms can increase (divergent mechanisms) and/or decrease (cohesive mechanisms) the genetic variability. Figure and caption adapted from (Van Rossum et al. 2020).

When cohesive forces are dominating the diverging ones, the cells within a population are characterized by a very limited genetic variability. This kind of species are called monotypic or “smeared”, and they usually have a limited geographic distribution and/or a limited range of host variability. An example of monotypic species is *Chlamydia trachomatis (Smelov et al. 2017)*. When diverging forces dominate, the resulting population is genetically extremely diverse (polytypic or “clustered”), often observed for bacteria able to adapt to a vast variety of hosts and environments (Bobay and Ochman 2018). An example of polytypic species is *Escherichia coli* (Tenaillon et al. 2010).

The middle ground in between these extremes represents by far the most common case, where a population contains elements of genetic variability but such variability is still not dominating. If an increase in mutation rate is paired with reduced recombination within a species, the population may start diverging and subdividing into subgroups in which the intra-subgroup cohesion is higher than the inter-subgroup one. This process, if sustained in time, can lead to the generation of subspecies, which can be considered groups of strains that are on their way to speciation.

1.1.3 Within-species variability definition

While the 95% to 97% ANI species-level range (for whole genomes) is widely accepted (Jain et al. 2018; Olm et al. 2020), and is accompanied by a sharp reduction in the horizontal gene transfer (HGT) and homologous recombination (HR) (Fraser, Hanage, and Spratt 2007; Bobay and Ochman 2017; Moldovan and Gelfand 2018), the area above 97% ANI remains loosely defined. Usually the term “strain” is used to refer to higher than species-level resolution, but no universally accepted definition of this term exists. The existing definitions are contrasting and dependent on the field of study (Van Rossum et al. 2020) (**Supplementary figure 1A**).

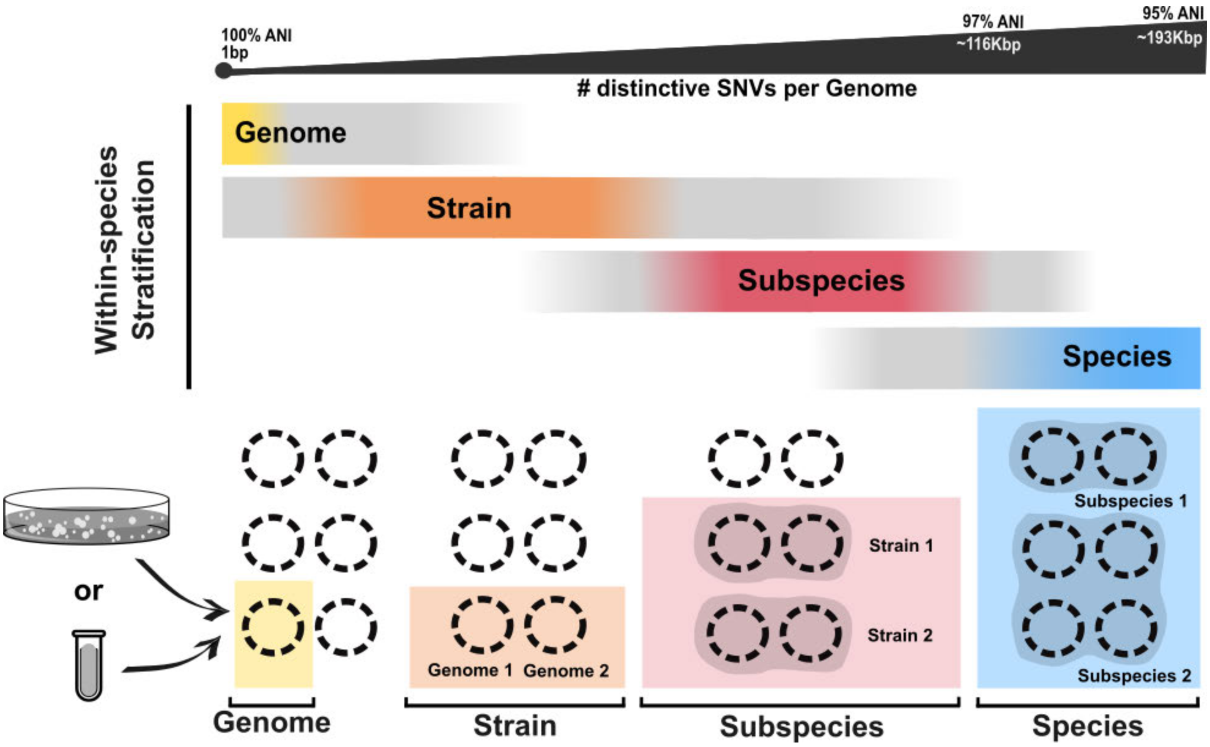


Figure 2. Stratification of the terminologies used to identify within-species variability, ordered by resolution, ranging from a single nucleotide variation in the whole genome (100% ANI) to 97% ANI, the recognised species-level boundary. For each bar, the colored portion refers to the recommended scope of use, while the grey portions refer to nonspecific common use. Figure and caption adapted from (Van Rossum et al. 2020).

To quantify, and ultimately stratify, within-species variability, genome similarity is commonly evaluated via SNV- or gene content-based profiling. While not all SNVs are equally capable of influencing the phenotype, a correlation exists between the core genome identity of within-species variants and their gene content overlap (Van Rossum et al. 2020; Maistrenko et al. 2020; Andreani, Hesse, and Vos 2017) (**Supplementary figure 1B**). In (Van Rossum et al. 2020), me and co-authors describe the range of genetic variation below the species level using three terms, officially recognised by the International Code of Nomenclature of Prokaryotes (Truper and Euzeby 2009), and with increasing more divergent SNV profile: genome, strain and subspecies (**Figure 2**).

Since there is no clear biology-driven threshold to identify how many SNVs are needed to differentiate between strains, the operational definitions are usually driven by the choice of tool used. In the Results discussed in **Chapter 3.3**, the tool metaSNV (Paul Igor Costea et al. 2017) was used to investigate the SNV profiles associated with within-species variability in *Clostridioides difficile* genomes and metagenomes.

1.2 Overview of the clinically relevant pathogenic species *Clostridioides difficile*

1.2.1 Introduction to *Clostridioides difficile*

Clostridioides difficile is an obligately anaerobic, motile, Gram positive bacterial species, isolated for the first time by Ivan Hall and Elizabeth O'Toole in 1935 from the stools of four out of ten healthy breast-fed newborns sampled within the first ten days of life (Hall and O'toole 1935). *C. difficile* was initially named "*Bacillus difficilis* because of the unusual difficulty encountered in its isolation and study (Latin: difficilis, meaning difficult)" (Hall and O'toole 1935). In the following decades researchers identified *C. difficile* in the stools of guinea pigs and rodents that died after the administration of large amounts of penicillin, suggesting that the widely used antibiotic could have serious side effects and trigger *C. difficile* overgrowth (Hamre et al. 1943; Green 1974). However, it is only in 1978 that John Bartlett and colleagues discovered the toxins produced by *C. difficile* and could characterize their role in causing antibiotic-associated pseudomembranous colitis (J. G. Bartlett et al. 1978). Together with the better understanding of the disease-causing mechanisms of *C. difficile*, new questions arose about its correct taxonomic classification.

After its initial naming, *C. difficile* was placed among the *Clostridium* genus, being therefore named *Clostridium difficile*. However this genus was at the time used to classify any anaerobic Gram positive bacteria able to produce spores, and was therefore extremely heterogeneous in its composition. Due to its significant differences with the representative species of the *Clostridium* genus, *C. butyricum*, a tentative proposal to rename *Clostridium difficile* as *Peptoclostridium difficile* was published in 2013 (Yutin and Galperin 2013). However, due to the considerable number of commercially available diagnostic tests for *C. difficile*, as well as ongoing clinical trials, renaming *C. difficile* to *P. difficile* would have been extremely expensive from a financial point of view. For this and other reasons, another study in 2016 (Lawson and Rainey 2016) suggested to solve the taxonomic conundrum by creating a new genus, named *Clostridioides*, and renaming *Clostridium difficile* as *Clostridioides difficile*, therefore keeping the abbreviated form unchanged.

As *C. difficile*-related research grew over the decades (**Figure 3**), *C. difficile* was found not only in the human gut but also in the gastrointestinal tract of farmed as well as wild animals, in fresh and wastewaters, in soil, and on hospital and long term care home surfaces (al Saif and Brazier 1996; Diaz, Seyboldt, and Rupnik 2018; J. S. Weese 2010). In soil, transient presence of *C. difficile* can be considered a hallmark of fecal contamination (Dharmasena and Jiang 2018).

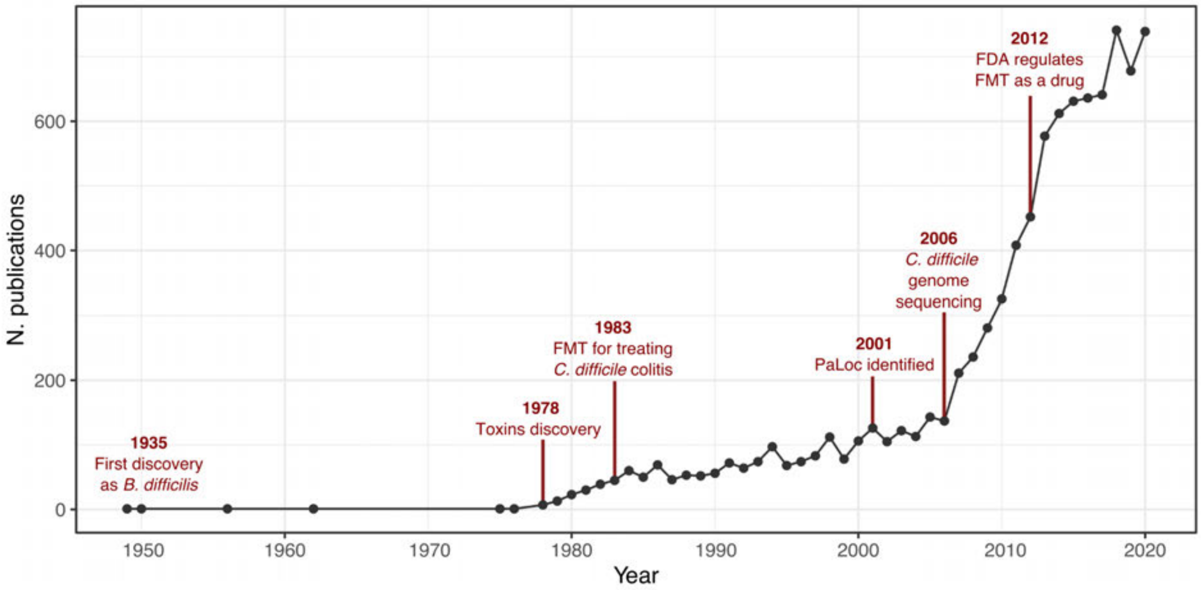


Figure 3. Growth of *C. difficile*-related research since its first description in 1935, as of June 2021 in Pubmed. The major milestones in *C. difficile* research are highlighted in red.

Due to its obligate anaerobic nature, *C. difficile* in its vegetative form can not survive for extended periods of time in oxygen-rich environments such as surfaces, freshwater and wastewater. However, in the form of spores, *C. difficile* can survive extreme conditions, including desiccation, elevated temperatures and prolonged exposure to an oxygenated environment (Lawler, Lambert, and Worthington 2020).

C. difficile sporulation, believed to be initiated by quorum sensing and lack of nutrients, is regulated by the transcription regulator Spo0A. Spo0A is highly conserved and is key in the sporulation initiation not only in *C. difficile* but in all *Bacillus* and *Clostridium* species (Deakin et al. 2012). Once in a favourable environment, such as the gastrointestinal tract of a host, *C. difficile* spores germinate and enter the vegetative growth phase.

1.2.2 *C. difficile* in the human gut microbiome

C. difficile is one of the several hundreds of bacterial species that can inhabit the gastrointestinal tract of a human host. As for any other member of the gut microbiome, *C. difficile* survival and metabolism is influenced by a variety of factors, such as the host immune system, host lifestyle (including diet and usage of antibiotics), and neighboring bacterial species.

A healthy gut microbiome is able to inhibit *C. difficile* germination and vegetative growth in multiple ways, the most important ones being via direct competition, production of secondary bile acids and production of short chain fatty acids (SCFAs). While the precise definition of a healthy microbiome remains elusive (Eisenstein 2020), a healthy gut microbiome usually has a high species diversity (Lozupone et al. 2012). In the colon, where the bacterial density is the highest (Donaldson, Lee, and Mazmanian 2016), the competition for nutrients is fierce and some species can outcompete *C. difficile* in harvesting mucus-derived sugars (Pereira et al. 2020). Some species, including *Bacteroides thetaiotaomicron*, *Akkermansia muciniphila*, *Ruminococcus gnavus* and *Bifidobacterium longum (Juge, Tailford, and Owen 2016)*, are able to cleave and release sialic acids from the mucosal layer, which are then metabolized by other members of the gut microbiome, leaving not enough nutrients for *C. difficile* to grow and expand (Ng et al. 2013; Ley 2014).

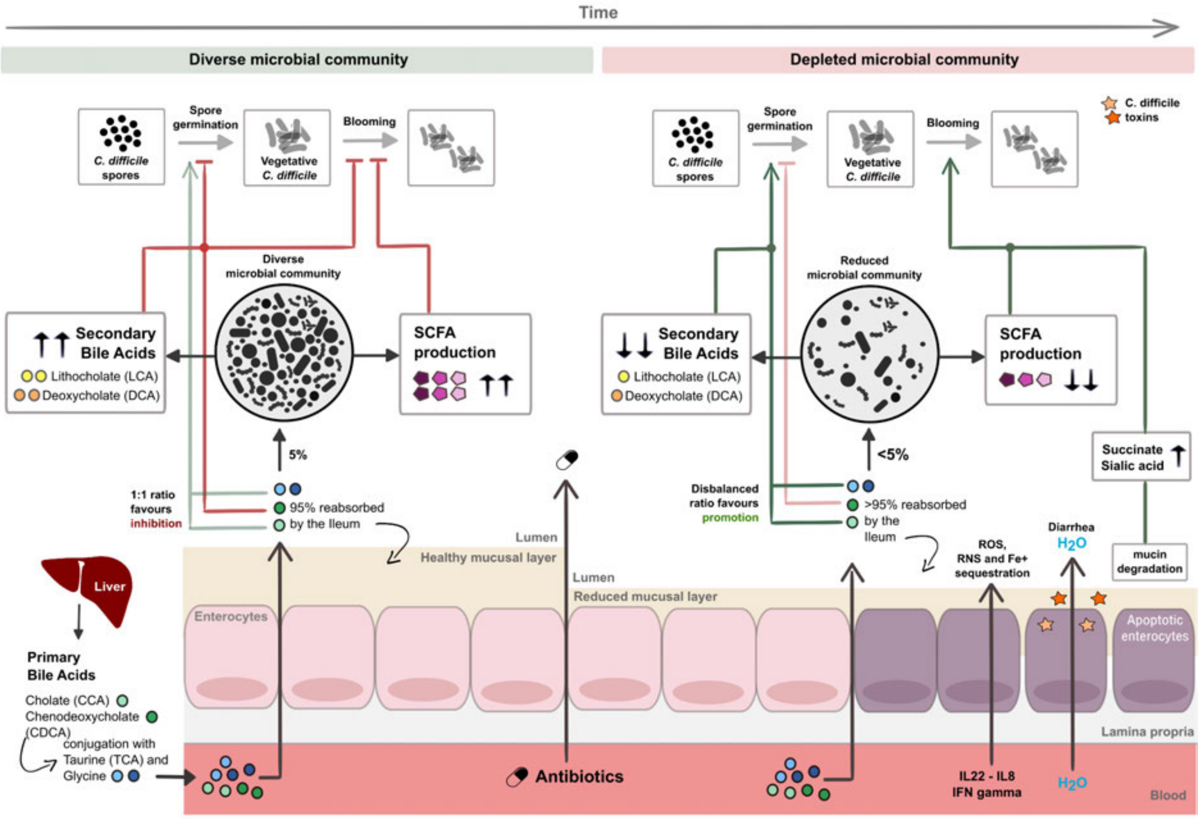


Figure 4. Overview of the community-wide changes that are associated with the transition from a healthy to a diseased gut microbiome, upon antibiotic intake, and the processes that allow the members of the gut microbiome to influence *C. difficile* sporulation, germination and blooming. Abbreviations: reactive oxygen species (ROS), reactive nitrogen species (ROS), interferon (IFN).

In the large intestine, around 5% of the primary bile acids produced by the liver is converted by anaerobic bacteria into secondary bile acids (Theriot, Bowman, and Young 2016; Theriot and Young 2015) (**Figure 4**). In vitro studies showed that low concentrations of primary bile acids and high concentrations of secondary bile acids inhibit *C. difficile* germination and growth (Theriot, Bowman, and Young 2016; Theriot and Young 2015).

In addition to producing secondary bile acids, a diverse microbial community also produces high quantities of SCFAs, such as butyrate, acetate and propionate, as an end product of the anaerobic fermentation of carbohydrates ((Theriot, Bowman, and Young 2016; Theriot and Young 2015) (**Figure 4**). In vitro assays showed that butyrate can reduce the pH in the gastrointestinal lumen and inhibit *C. difficile* growth (Theriot and Young 2015; Theriot, Bowman, and Young 2016; Rivière et al. 2016). Butyrate is able to promote mucosal layer production (Cornick, Tawiah, and Chadee 2015; Gaudier et al. 2004; Willemsen et al. 2003), to modulate the epithelial immune response (Ghimire et al. 2020) as well as cell-to-cell tight junctions (Abt, McKenney, and Pamer 2016; Rivière et al. 2016) and has been also associated with a protective effect against *C. difficile* toxins (Fachi et al. 2019).

Antibiotic treatment, particularly with broad spectrum antibiotics, reduces the microbial diversity in the gut, resulting in reduced secondary bile acids concentration and loss of key species for the production of SCFAs, such as *Faecalibacterium prausnitzii, Eubacterium rectale, Eubacterium hallii, Anaerostipes caccae, Anaerostipes hadrus, Roseburia faecis, Roseburia inulinivorans, Roseburia intestinalis* and *Roseburia hominis (Rivière et al. 2016).* Loss of butyrogenic species was found in the gut of patients diagnosed with CDI and nosocomial diarrhea (Antharam et al. 2013).

In addition, in the reduced microbial community following an antibiotic treatment, sialic acids are still released in abundance in the gastrointestinal lumen, but they are not readily metabolized by the few surviving gut bacteria, leaving high concentrations of sialic acid available for *C. difficile* consumption (Ng et al. 2013; Ley 2014). Degradation of the mucosal layer via the release of sialic acids, combined with the lack of mucosal growth promotion (due to the low levels of butyrate), lead to an overall reduction of the mucosal layer thickness (**Figure 4**), which has been associated with increased probability of bacterial tissue invasion, immune response instigation and release of reactive oxygen species (ROS) (Cornick, Tawiah, and Chadee 2015; Aviello and Knaus 2017). As butyrogenic species are particularly sensitive to ROS (Devaux, Million, and Raoult 2020), their release further reduces the microbial diversity, fuelling the vicious cycle (**Figure 4**). Antibiotic treatment has also been associated with intraluminal increase of reactive nitrogen species (RNS) (Faber et al. 2016).

Together, these changes drastically alter the niche and nutrient availability in the lumen and in the proximity of the mucosal layer, favouring *C. difficile* expansion and the progressive shift to a more dysbiotic state (**Figure 4**).

While the community-wide changes described in **Figure 4** are generally well understood, their interpretation is complicated by the incredible versatility of *C. difficile*, which is able to metabolise a vast array of nutrients depending on the situation (Fletcher et al. 2021; Jenior et al. 2017; Theriot and Young 2015) and the surrounding microbial species (Lopez et al. 2020; Melinda Anne Engevik et al. 2020), and to adapt and react to different types of stress (Knippel et al. 2020).

1.2.3 Clinical relevance of *C. difficile*

C. difficile represents a serious burden for the healthcare system, with over 223,900 estimated hospitalizations and 12,800 estimated deaths in 2017 in the US alone ((u.s.) and Centers for Disease Control and Prevention (U.S.) 2019). However, not all *C. difficile* strains are toxigenic and can therefore cause disease (*C. difficile* infection also known as CDI). The prevalence of non-toxigenic *C. difficile* is highly variable, depending on the testing method, health status and host age (Natarajan et al. 2013). Non-toxigenic *C. difficile* isolates have been found in human hosts as well as from animal and environmental sources, suggesting that non-toxigenic *C. difficile* might be widespread (Natarajan et al. 2013).

Toxigenic *C. difficile* isolates possess the pathogenicity locus (PaLoc), a 19.6kb long stretch of sequences that includes the two toxin genes, *tcdA* and *tcdB*, and several other regulatory genes (such as *tcdD*, *tcdE* and *tcdC*) (Voth and Ballard 2005) (**Figure 5A**). Several truncated versions of PaLoc are known (Rupnik 2008). TcdA substantially increases the intestinal permeability, while TcdB leads to intense inflammation of the colon. Despite TcdB being about 1000 times more potent than TcdA, they both have cytotoxic effects (causing cell death by apoptosis and necrosis) and proinflammatory activity (Voth and Ballard 2005).

Both TcdA and TcdB have an A/B-type structure: composed of an A-subunit, the enzymatically active subunit, and a B-subunit, which is responsible for binding with the host cell’s receptor and the intracellular intake (Gerhard 2017). Both toxins are able to bind to a wide range of cell types in a variety of host species (Gerhard 2017).

Once the toxins contact the receptors on the intestinal cells, they inactivate the Rho-GTPase proteins in the host cell's cytosol via glucosylation (Sun, Savidge, and Feng 2010), resulting in dysregulation of the actin cytoskeleton, microtubule dynamics, cell-to-cell contact, epithelial barrier functions and cytokine production (**Figure 5B**). *C. difficile* toxins damage cell-to-cell junctions and favour bacterial invasion and increase the passage of water into the lumen, causing diarrhea. Another effect mediated by the endocytosis of TcdA and TcdB is the production of cytokines, responsible for the inflammation response. However, the link between the Rho glucosylation pathway and the host immune response activation remains unclear (Sun, Savidge and Feng, 2010).

Furthermore, upon the activation of the host inflammatory response, *C. difficile* is able to alter its metabolism and exclude members of gut microbiome (in particular belonging to the *Bacteroides* genus) via nutrient competition (Fletcher et al. 2021), exacerbating the vicious cycle described above.

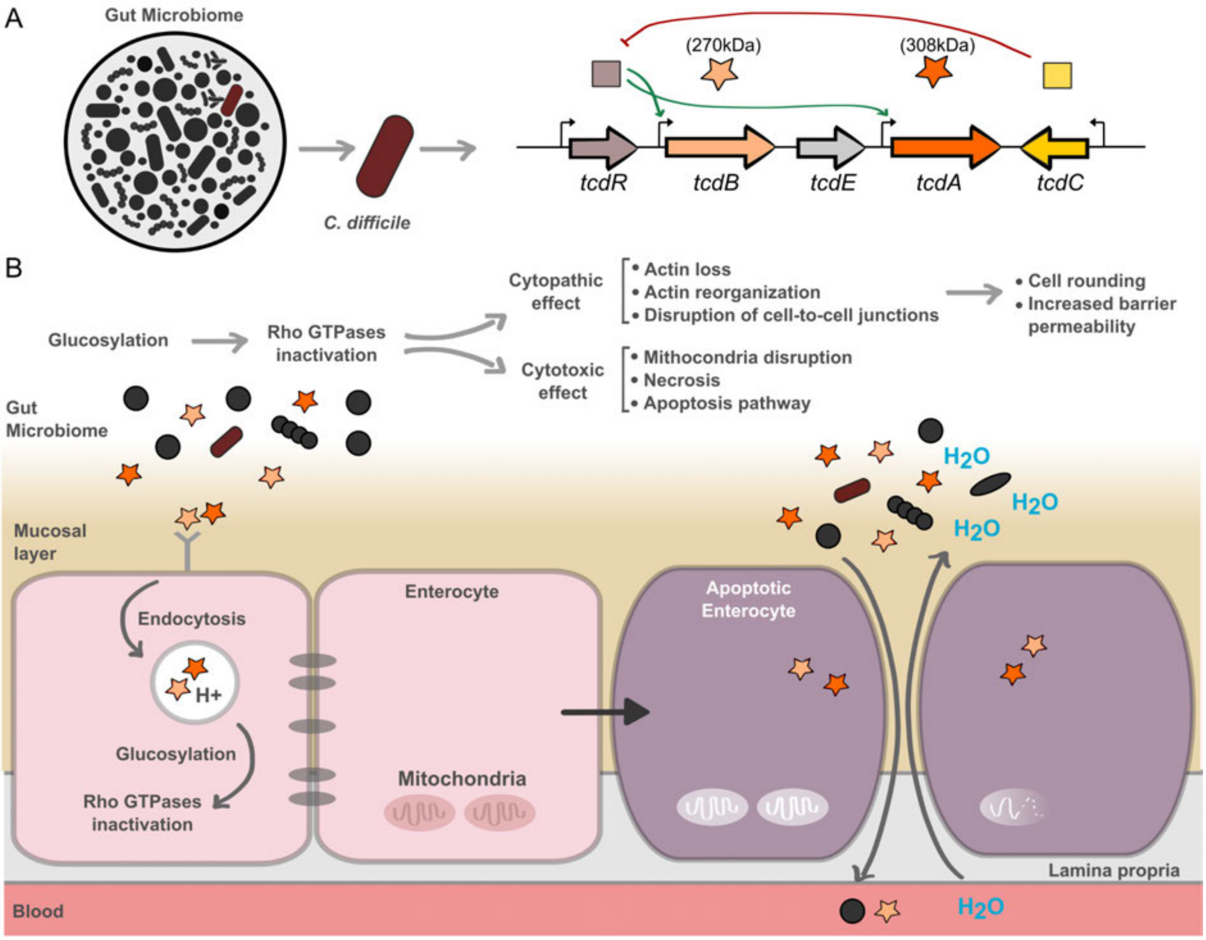


Figure 5. Mechanism of action of *C. difficile* toxins in the human gut. (A) Structure of *C. difficile* pathogenicity locus (PaLoc), including two toxin genes TcdA and TcdB and three additional regulatory ORFs. *tcdD* and *tcdC* are (positive and negative, respectively) regulators of the toxin gene expression, while *tcdE* is a putative holin protein, suspected to mediate toxin release through its ability to disrupt the bacterial cytoplasmic membrane. (B) Overview of the cascade induced by the endocytosis of *C. difficile* TcdA and TcdB in the host cell. Cytopathic and cytotoxic effects are mediated directly by the toxins or by Rho-dependent mechanisms. Toxins also induce the activation of pro-inflammatory pathways, inducing the production of cytokines. Figure and caption adapted from (Voth and Ballard 2005; Chen et al. 2015; Fortier and Sekulovic 2013).

Besides TcdA and TcdB, another toxin, called binary toxin (CDT), can be present. CDT can be found in 17-23% of cases ((Eckert et al. 2015), but estimations are complicated by the fact that most of the toxin tests used in the diagnostic and clinical setting are targeting TcdA and/or TcdB and not CDT (see **Chapter 1.2.4** for more detailed discussion on diagnostic methods).

The binary toxin is encoded by two genes (*cdtA* and *cdtB*), both located in the CDT locus (CdtLoc), for which so far only a single truncated version is known (Stare, Delmée, and Rupnik 2007). When found together with TcdA and TcdB, CDT has been associated with increased disease severity. Sole presence of CDT (A-/B-/CT+) is rare and usually identified in animal hosts (Schneeberg et al. 2013; Knight, Squire, and Riley 2015; Eckert et al. 2015).

Toxigenic *C. difficile* strains can be commonly found among asymptomatic patients (Eyre et al. 2013; Alasmari et al. 2014). It remains unclear whether non-toxigenic *C. difficile* originated from toxigenic *C. difficile* isolates or vice versa (Natarajan et al. 2013). Carriage and expression of toxin genes is highly energy-consuming. In fact, *C. difficile* toxin production was found to be negatively correlated with growth rate, suggesting that toxin production might be a much more energetically demanding process than previously thought (Tschudin-Sutter et al. 2016).

As toxigenic and non-toxigenic *C. difficile* are likely to compete for nutrients and niche occupancy, non-toxigenic *C. difficile* can be protective against the colonization of toxigenic *C. difficile* (Natarajan et al. 2013). However, HGT of the PaLoc from toxigenic to non-toxigenic *C. difficile* is possible, raising safety concerns on the preventive colonization with non-toxigenic isolates as protective procedure (Brouwer et al. 2013, 2016).

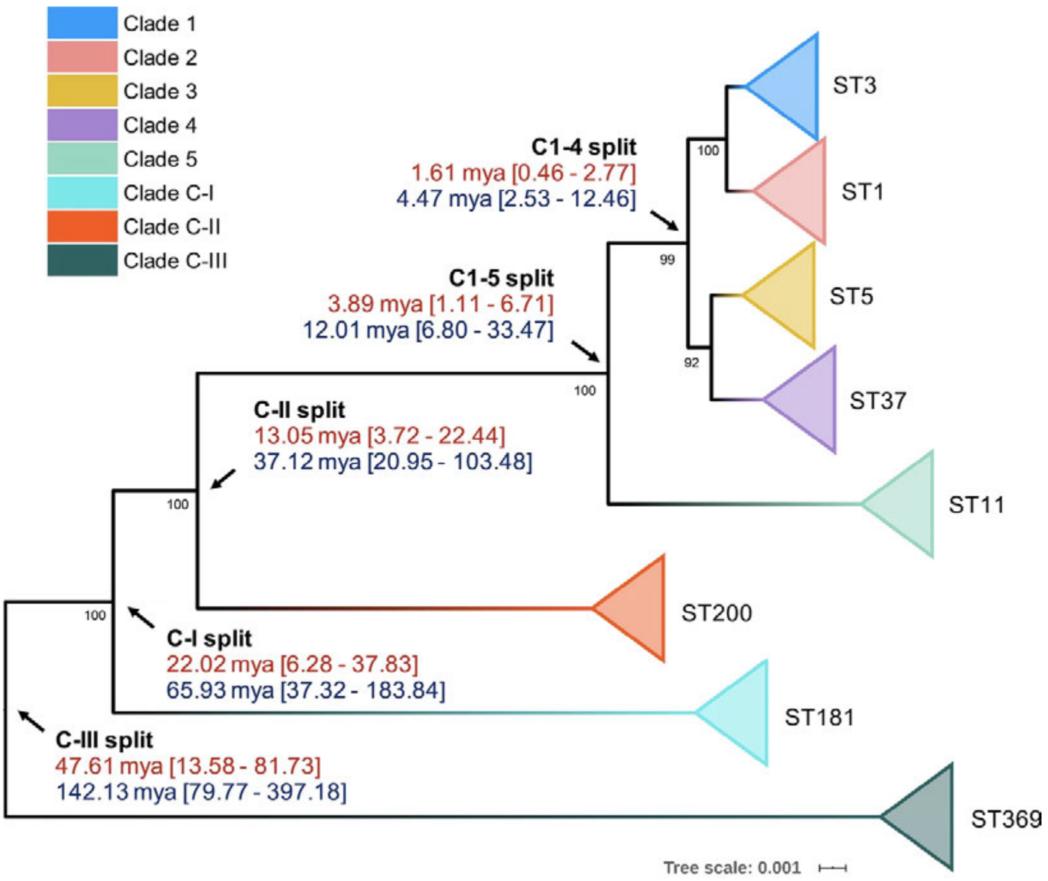


Figure 6. Updated *C. difficile* clades divergence, using BacDating and BEAST. Figure from (Knight et al. 2021).

The phylogenetic diversity of *C. difficile* isolates is usually investigated via multilocus sequence typing (MLST). MLST typing steps involve i) PCR amplification and DNA sequencing of the internal fragment (usually 400-500bp long) of up to seven housekeeping genes, ii) assignment of allele numbers to the unique sequence variants within species for each housekeeping gene, and iii) combination of the allelic numbers to obtain the sequence type (ST) (Maiden et al. 1998; Ruppitsch 2016). In other words, sequence types are assigned based on the allelic variants of seven highly conserved housekeeping genes (Stabler et al. 2012).

There are 8 known phylogenetic clades (or monophyletic groups) of *C. difficile*, including both toxigenic and non-toxigenic sequence types (Knight et al. 2021) (**Figure 6**). Clades 1-5 include many of the most prevalent CDI-causing sequence types worldwide. Clades C-I, C-II and C-III, called cryptic clades, are less characterized and recent data suggests that they have emerged before the other clades (Knight et al. 2021) (**Figure 6**).

Overall, as non-toxigenic sequence types are not associated with disease, they have been so far overlooked, with the vast majority of the research effort focused on the study of toxigenic sequence types and their clinical implications.

1.2.4 *C. difficil*e infection (CDI) diagnosis

Risk factors associated with CDI include antibiotic treatment for an extended period of time, advanced age, previous hospitalization, use of proton pump inhibitors, and presence of other comorbidities (Bagdasarian, Rao, and Malani 2015). CDI can present itself with varying degrees of severity, ranging from mild (diarrhea <3 times per day) to life threatening (systemic infection, megacolon or ileus) (Allen et al. 2013).

The first step in CDI diagnosis is the identification of the symptoms, which can include fever, abdominal pain, nausea, vomiting, diarrhea and, occasionally, pseudomembranous colitis (John G. Bartlett and Gerding 2008). Diarrhea is by far considered the most common symptom of CDI and presence of 3 or more unformed stools in 24 hours usually grants further investigation (Bagdasarian, Rao, and Malani 2015). Pseudomembranous colitis (PMC) prevalence varies from 10% (Sartelli et al. 2019) to 25% (Karanika et al. 2016) of CDI cases, however this number is likely an underestimation as detection requires invasive procedures and is therefore performed in a limited number of cases. Following the presence of symptoms, a wide range of laboratory tests can be used to verify the presence of toxigenic *C. difficile* in the patient’s stools (**Table 1**). These tests vary considerably in what they target, in their sensitivity, specificity, cost and execution time.

Test Type	**Detected substance**	**Mayor advantages**	**Mayor limitations**
Toxigenic culture	*C. difficile* bacterium or spores	gold standard	time consuming (24-48h)
CCNA	Presence of active *C. difficile* toxin production	gold standard	time consuming (24-48h); fresh stools are needed
GDH EIA	GDH (*C. difficile* common enzyme)	rapid and cheap	unable to distinguish toxigenic and non-toxigenic isolates; wide range of sensitivity and specificity
Toxin EIA	Presence of active *C. difficile* toxin production	rapid and cheap	wide range of sensitivity and specificity
NAAT	*C. difficile* toxin genes	viable cells not needed, rapid	more expensive than EIA

Table 1. Overview of the diagnostic tests available for CDI diagnosis. Adapted from (Lee et al. 2021; Monique J. T. Crobach et al. 2018; Martínez-Meléndez et al. 2017). Abbreviations: CCNA, cell cytotoxic neutralization assay; GDH, glutamate dehydrogenase; EIA, enzyme immunoassay; NAAT, nucleic acid amplification test;

The gold standard definition of CDI includes presence of symptoms, evidence of pseudomembranous colitis and positive toxigenic test for *C. difficile* (Keighley et al. 1978). The diagnostic tests available for diagnosing CDI can be divided into two major categories, based on their rapidity (and therefore applicability in the nosocomial context). Rapid tests include NAAT and EIA tests (targeting GDH or toxins). Non-rapid tests include toxigenic culture and CCNA (**Table 1)**. Below, I briefly describe the sub-categories of diagnostic tests available to aid in the CDI diagnosis:

- In the toxigenic cultivation, *C. difficile* spores are recovered from the patient's stools, and germination is promoted by the use of cycloserine-cefoxitin-fructose-agar (CCFA) with the addition of biliary salts (Sorg and Sonenshein 2008; George et al. 1979). The ability of *C. difficile* strains to produce toxins is then tested on a variety of cell lines or in combination with EIA tests (Martínez-Meléndez et al. 2017). Despite being a reference standard, toxigenic culture is rarely performed in the clinical setting, as they take time (up to 48 hours) and require dedicated equipment and trained personnel (Bagdasarian, Rao, and Malani 2015).
- Cell Cytotoxic Neutralization Assay (CCNA) is the gold standard for toxin detection. The patient's stools filtrate is directly tested on selected cell lines. The test is positive if there is a visible cytopathic effect (cell rounding) and if it can be successfully reversed

by an antitoxin (Planche and Wilcox 2011). For this test, fresh stools are needed to have reliable results (Freeman and Wilcox 2003).

- GDH assays target glutamate dehydrogenase (GDH), produced by all *C. difficile* strains, with the important function of protecting *C. difficile* from oxidative stress. However, GDH is produced in both toxigenic and non-toxigenic *C. difficile* strains, and for this reason GDH tests should always be paired with more specific, toxin-targeting tests (Shetty, Wren, and Coen 2011).

- Enzyme immunoassays (EIA) tests aim at detecting *C. difficile* toxins TcdA and/or TcdB. Tests targeting both toxins are now more common, as symptomatic cases associated with single toxin production have been reported (M. J. T. Crobach et al. 2016). EIA are very frequently used in the hospital context, as they can be rapidly performed and are relatively cheap. However, these tests are also the most inconsistent ones, with sensitivity ranging from 51% to 94% and specificity ranging from 75% to 100% (Lee et al. 2021; Martínez-Meléndez et al. 2017). In addition, considerable variation of these parameters can be found across different manufacturers (Martínez-Meléndez et al. 2017).

- Nucleic Acid Amplification Tests (NAAT) target the toxin genes. Compared to EIA, NAAT are more sensitive (>62.1%) and more specific (>87.5%), even though considerable differences persist between manufacturers (Martínez-Meléndez et al. 2017; M. J. T. Crobach et al. 2016).

Recommended international guidelines strongly discourage the use of stand-alone tests, and advise a two-step algorithm (M. J. T. Crobach et al. 2016): a first test characterized by high sensitivity, followed in case of a positive result by a second test with high specificity. One of the recommended combinations of tests is the use of NAAT or GDH EIA, followed by toxin EIA (targeting both TcdA and TcdB). If the second test is positive, CDI is considered likely, while if negative, a third optional testing (toxigenic culture or NAAT, if the first test was GDH EIA) is recommended to help the clinical diagnosis (M. J. T. Crobach et al. 2016). In alternative, another recommended combination is the use of GDH EIA and toxin EIA (targeting both TcdA and TcdB). If both tests are positive, CDI is considered likely, while if only one is positive a second optional test (toxigenic culture or NAAT) is recommended (M. J. T. Crobach et al. 2016).

While CDI cases are for the vast majority affecting adult and elderly subjects, pediatric and neonatal CDI cases are possible (Lees et al. 2016). However, the definition and assessment of disease severity in pediatric CDI patients is complicated by the lack of a standardized scoring system (Lees et al. 2016). As both toxigenic and non-toxigenic *C. difficile* are known to be common in asymptomatic newborns and children below the age of 3, testing for toxigenic *C. difficile* or for its toxins is discouraged by the recommended guidelines (M. J. T. Crobach et al. 2016). Irrespective of the age, a “test of cure” is also discouraged (McDonald et al. 2018), due to the elevated asymptomatic carriage of *C. difficile*.

The first line of treatment against CDI is antibiotic administration (in particular metronidazole, vancomycin or fidaxomicin) (Wu et al. 2020; Gateau et al. 2018), combined with the discontinuation of other unnecessary antibiotic treatments (Martínez-Meléndez et al. 2017). The efficacy of *C. difficile*-targeting antibiotic treatment can be reduced by the presence of biofilms (organized bacterial communities composed by single or multiple species) at the interface with the mucosal layer of the gastrointestinal tract of the host.

The majority of relapses are in fact traceable to the same ribotype responsible for the initial episode (Figueroa et al. 2012), suggesting that mechanisms such as biofilms might play an important role in *C. difficile* ability to escape the action of antibiotics. A recent study showed that *C. difficile*-containing biofilms can act as reservoir for recurrent CDI (rCDI), and that specific members of the gut microbiome can modulate the biofilm formation rate (Normington et al. 2021; Melinda A. Engevik et al. 2021). In case of rCDI, fecal microbiota transplantation (FMT) is considered. Despite its very high success rate (Kim et al. 2019), FMT has been associated with long-term adverse events, such as obesity and immune-mediated disorders (Park and Seo 2021). In addition, the complex composition of donor's stools poses important challenges in the complete characterization of its components (Gupta, Allen-Vercoe, and Petrof 2016). These reasons pushed the US Food and Drug Administration (FDA) to consider the FMT donor stools comparable to drugs (Gupta, Allen-Vercoe, and Petrof 2016).

1.3 Thesis outline

In **Chapter 2**, I describe in detail the methods behind the data collection, analysis and interpretation.

In **Chapter 3.1**, I discuss the results of the meta-analysis aiming at characterizing the microbial signature associated with *C. difficile* infection (CDI), at the species level.

In **Chapter 3.2**, I describe the results of a broader survey for *C. difficile*, including samples from a wide variety of habitats, hosts, body sites, health status and age groups. In particular, I compare *C. difficile* prevalence, relative abundance and biotic context in infant and adult subjects.

In **Chapter 3.3**, I discuss *C. difficile* within-species diversity and the toxigenic potential of *C. difficile,* as well as other enteropathogenic species, across age groups.

In **Chapter 3.4**, I provide an overview of the technical limitations of this study and the future perspectives.

Final remarks are discussed in **Chapter 4**.

Chapter 2. Methods

This Chapter describes in detail how the raw data were collected and how they were subsequently analyzed.

Public metagenomes collection

The investigation of *C. difficile* carriage and its association with other members of the gut microbiome was carried out in two groups of samples: an extended collection of samples from all over the world, including a wide range of habitats, hosts, age ranges and diseases (global *C. difficile* survey), and a smaller subset of samples from cohorts including patients diagnosed with CDI (CDI meta-analysis). As described in **Chapter 1.2**, the gut microbiome in CDI is characterized by drastic changes, compared to the healthy counterparts, therefore deserving a dedicated meta-analysis.

Global metagenomic survey of *C. difficile*

A total of 42,814 publicly available metagenomic samples from 253 different study populations were downloaded (**Supplementary Table 2**). The collection includes samples from 84 countries, 25 animal species and 6 different human body sites: gastrointestinal tract (stools, rectal swabs and biopsies), vagina, skin, oral cavity, respiratory tract and milk.

CDI metagenomic survey of *C. difficile*

A subset of 10 CDI or diarrhea-associated publicly available cohorts were used for the downstream analysis of CDI samples (**Supplementary Table 1**). Out of 294 samples in total, 100 were identified as CDI and 194 as controls. Samples were divided in three groups: (i) CDI: including samples from subjects diagnosed with CDI, (ii) D-Ctr, including samples from subjects with diarrhea but no CDI diagnosis, or subjects without diarrhea but diagnosed for a disease other than CDI and (iii) H-Ctr, including samples from subjects identified as “control” in the study’s metadata.

Data download

Metagenomes publicly available on the 1st of January 2020 were downloaded using fetch-data (Coelho et al, *in revision*). Only shotgun metagenomic samples sequenced using Illumina platforms have been included. No minimum number of samples per study or minimum sequencing depth threshold was applied during data download.

Metadata collection

The metadata collected for each dataset include: bodysite, sample type (stools, rectal swabs or biopsy), health status (healthy or diseased), age (in months), age group, geography (country and continent), westernised lifestyle or not, diagnosis for *C. difficile* infection (CDI), use of antibiotics, delivery mode, gestational age (pre-term or full-term), sex and diet (for infants: exclusive formula feeding, exclusive breastfeeding and mixed, for the other age groups: omnivorous, vegetarian, vegan or other).

Age group was defined as follows: infants from birth to 1 year of life (included), children from 1 year to 10 years (included), adolescents from 10 to 18 years (included), adults from 18 to 65 years (included) and elderly above 65 years of age. In the cases lacking specific age, a range of age was provided when possible.

Given the broad range of perturbations in the gut microbiome associated with antibiotics intake, in particular in the context of *C. difficile* carriage, an extended definition of “diseased” subjects was used, including (i) subjects diagnosed with any kind of disease and/or (ii) subjects taking antibiotics at the time of the sampling. Metadata about antibiotic intake prior to sampling was scarce, therefore I could not take this aspect into consideration in our analysis. However, the majority of studies had “exposure to antibiotics in the 6 months prior sampling” among the exclusion criteria for healthy controls.

For animal samples, the following metadata were collected: host species, bodysite, sample type (stool or rectal swab), health status, age group (juvenile or adult), geography (country and continent).

Preliminary quality control and filtering

Two consecutive filtering steps were performed on the total number of reads mapping to mOTU marker genes for each sample: (i) samples with zero reads were discarded (n=1,091) and (ii) samples with less than 59 reads mapping to mOTU marker genes, corresponding to 95%ile calculated on the remaining samples, were discarded (n= 2,050). An additional third read count filtering was performed only on human gut stool samples, corresponding to 99%ile, removing samples with less than 100 reads mapping to mOTU marker genes (n=132 samples). The resulting file included 39,106 samples, of which 26,421 were human gut stools.

Taxonomic profiling of the downloaded metagenomic samples

The downloaded datasets were taxonomically profiled at the species level with mOTUs v2.0[5], using two marker genes. In brief, taxonomic profiling of microbial communities can be achieved using specific genes. Previous studies identified 40 universal marker genes (Sorek et al. 2007; Ciccarelli et al. 2006) that occur in single copy in the majority of known bacterial species and that can be used to delineate prokaryotic organisms at the species level (Mende et al. 2013). mOTUs v2.0 utilizes 10 of the 40 marker genes described above, and clusters them to generate a database of marker genes-based operational taxonomic units (mOTUs). Marker genes are extracted from both prokaryotic reference genomes (ref-mOTUs) and publicly available metagenomes (meta-mOTUs). Alignments against this database are then used to taxonomically classify metagenomic reads. All data analyses were conducted in the R Statistical Computing framework v3.5 or higher.

Identification of timeseries-representative samples

Out of 26,421 samples, 24,331 had subject metadata and were associated with 12,012 unique subjects. For 3,545 subjects multiple timepoints were available. Mean number of timepoints per subject was 2.02 (2.98 for infants, 2.30 for children, 2.01 for adolescents, 1.60 for adults and 1.56 for elderly), with a maximum of 205 timepoints per subject. In order to avoid under- or over-estimating the prevalence of *C. difficile*, only one sample per timeseries was used in the downstream analyses. In a timeseries, three cases were possible:

- All samples had *C. difficile*. In this case, the sample with the highest *C. difficile* read count was selected as representative and the corresponding subject was considered *C. difficile* positive.
- None of the samples had *C. difficile*. In this case, the sample of the first timepoint was selected as representative and the corresponding subject was considered *C. difficile* negative.

- Some samples in the timeseries had *C. difficile,* while some didn't.
    - In case all samples of the timeseries had the same health status (all healthy or all diseased), the sample with the highest *C. difficile* read count was selected as representative and the corresponding subject was considered *C. difficile* positive.
    - In case the subject over the course of the timeseries changed health status (i.e. from healthy to diseased): the timeseries was split into portions with consistent health status, and the representative sample for each portion was identified based on the cases described above.

Additionally, for 2,090 samples no subject metadata were available. In this case, we assumed one sample per patient. One representative sample for each time series was used for all downstream analyses, if not specified otherwise.

Prevalence and abundance estimations of *C. difficile*

Prevalence estimations of *C. difficile* over life time were based on human gut samples (stools, rectal swabs and biopsies) for which the precise age in months was available (samples with "na" or age ranges too big to fit into one of the plotted age ranges were discarded). *C. difficile* mOTUs ("ref_mOTU_0051" and "ref_mOTU_0052") were considered as cumulative sums in the downstream analyses. For *C. difficile* relative abundance estimates in humans, only *C. difficile* positive samples from stool samples with precise age metadata available were included. For *C. difficile* relative abundance estimates in animals, all *C. difficile* positive samples were considered, independently of the age group.

Logistic regression and ANOVA

In order to identify the association of factors such as antibiotics usage, age, sex, health status, delivery mode or prematurity, with the presence of *C. difficile* I used logistic regression combined with ANOVA (R packages: car v3.0-6, rsq v2.0). For this analysis, only human stool samples with complete metadata regarding age, prematurity, delivery mode, sex, antibiotics use, westernisation, geography, healthy status and presence of *C. difficile* a diagnosied infection were included. In total, 4,096 metagenomic samples passed these selection criteria.

Alpha diversity calculation

Alpha diversity calculation was performed on human gut stool samples only, with at least 100 reads mapping to mOTU marker genes per sample, with known age group and health status and from studies with at least two *C. difficile* positive samples. Unclassified fraction ("-1") was preemptively removed. Rarefied per-sample taxa diversity ('alpha diversity', averaged over 100 rarefaction iterations) was calculated as effective number of taxa with Hill coefficients of q=0 (i.e., taxa richness), q=1 (exponential of Shannon entropy) and q=2 (inverse Simpson index), and evenness measures as ratios thereof. Unless otherwise stated, results in the main text refer to taxa richness. Differences in alpha diversity were tested using ANOVA followed by post hoc tests and Benjamini-Hochberg correction for multiple hypothesis tests, as specified in the main text.

Species co-occurrence analysis

Human gut stool samples, with at least 100 reads per sample and with known age group and health status were considered for this analysis. One representative sample per time series was considered. From the mOTUs v2.0 profiles (Milanese et al. 2019), the unclassified fraction and taxa with prevalence <1% were removed. Fisher's exact test, followed by Benjamini-Hochberg correction, were applied to the contingency table.

Machine learning modelling

L1-regularized LASSO logistic regression models to predict CDI status were built using the SIAMCAT R package with 10-fold cross-validation. For this analysis, I focused on the subset of 10 CDI or diarrhea-associated datasets (**Supplementary Table 1**) and then trained two different sets of models: one set of models to distinguish CDI samples and samples from healthy controls (excluding controls from diseased subjects) and another set of models to distinguish CDI samples and any type of control samples. In order to minimize overfitting and to counter batch effects (Wirbel et al. 2021), datasets across studies were pooled in a leave-one-study-out approach. In short, all except one study were jointly processed to train a LASSO model that was then used to predict the left-out study. Additionally, to check if the microbial signature for CDI was independent of *C. difficile*, another set of models was trained with the same cross-validation splits but excluded *C. difficile* from the feature table. Feature weights were extracted from the models, normalized by the absolute sum of feature weights, and averaged across cross-validation folds. For the heatmap in **Figure 11**, all microbial species that were assigned non-zero weights in at least 80% of cross-validation folds were included.

Linear mixed effect model

In order to test for differential abundance of microbial species between CDI and non-CDI samples while taking into account possible confounding factors, linear mixed effect models were implemented via the lmerTest package (Kuznetsova, Brockhoff, and Christensen 2017). After filtering for prevalence (prevalence of at least 5% in three or more studies), the log-transformed abundance of each microbial species was tested using a linear mixed effect model with "CDI status" as fixed and "Study" and "Age group" as random effects. Effect size and p-values were extracted from the model and p-values were corrected for multiple hypotheses using the Benjamini-Hochberg procedure.

Evaluation of toxigenic potential in metagenomes

Identification of *tcdA* and *tcdB* toxin-coding genes was performed using the Virulence Factor DataBase (VFDB (B. Liu et al. 2019), as downloaded in March 2021. Evaluation of *C. difficile* binary toxin gene was not included, since its presence has been not associated with disease severity (Goldenberg and French 2011).

Analysis of *C. difficile* within-species diversity from metagenomic samples

C. difficile subspecies detection was performed on all *C. difficile* positive samples with MetaSNV (Paul Igor Costea et al. 2017). Briefly, MetaSNV is a tool for single nucleotide variant (SNV) analysis in metagenomic samples that uses nucleotide sequence alignments to reference genomes to perform SNV calling for individual samples as well as for the whole set of samples. Output included allele frequencies and nucleotide diversity per sample as well as distances across samples.

Reads were mapped against the progenomes v1 (Mende et al. 2017) species representatives genomes for 3 species in the Clostridioides genus:

- specI_v2_0051 (NCBI taxonomy ID 272563, PRJNA78) *Clostridioides difficile*
- specI_v2_0052 (NCBI taxonomy ID 1151292, PRJNA85757) *Clostridioides difficile*

- specI_v2_1125 (NCBI taxonomy ID 1408823, PRJNA223331) *Clostridioides mangenotii*

By mapping to multiple genomes and only using the uniquely mapped reads, we focused on the species-specific core. Mappings that had at least 97% identity across at least 45bp were kept. Mapping and filtering was performed using BWA (Li and Durbin 2009) and ngless (Coelho et al. 2019). Reads that mapped uniquely across the 3 reference genomes were used to call SNVs using metaSNV with default parameters. SNVs were detected at 48,482 positions in 307 samples, after filtering for prevalence across the sample population. Substructure within the population was assessed in the resultant SNV profiles according to a previously reported approach (Paul I. Costea et al. 2017).

Briefly, dissimilarities between samples were calculated based on SNV abundance profiles (produced by metaSNV) and the resultant distance matrix was tested for clusters using the Prediction Strength algorithm (Tibshirani and Walther 2005). No clusters passed the cluster detection threshold (threshold was 0.8, max value found was 0.63). Distance matrix is plotted using R and the pheatmap package with average clustering. Six samples with extreme dissimilarity to all other samples were removed from the distance matrix for illustrative purposes (SAMN08918181, SAMN09980608, SAMN10722477, SAMN13091313, SAMN13091317, SAMN13091322). *C. difficile* reference genomes for each of the known *C. difficile* clades were randomly selected from the Supplementary table 1 of (Knight et al. 2021): ERR1024380, ERR125919 and ERR125977 for Clade 1, ERR029530, ERR031688 and ERR026854 for Clade 2, ERR232393, SRR3630175 and SRR3938313 for Clade 3, ERR1015455, ERR125966 and ERR232390 for Clade 4, ERR1854834, ERR1854840 and ERR232396 for Clade 5, ERR2216002, ERR232401 and ERR789085 for Clade C-I, ERR3296451, SRR3629287 and SRR3654506 for Clade C-II, ERR2215981 and ERR2216003 for Clade C-III. Reference genomes from Clade C-I and C-III did not pass the initial filtering steps (mapping to reference genome lower than 97% identity threshold), and are therefore not shown in **Figure 26**.

Chapter 3. Results and discussion

Over the decades, *C. difficile* has been predominantly investigated from the disease-centred perspective, mostly studying its burden in subjects diagnosed with *C. difficile* infection (CDI). However the metagenomic characterization of the gut microbial communities associated with *C. difficile* presence in CDI patients is limited by low sample size and narrow geographical span. To overcome these limitations, I performed a meta-analysis combining 10 publicly available CDI cohorts, including a total of 534 samples.

In the first part of this Chapter (**Chapter 3.1**), I discuss the microbial diversity and community composition associated with *C. difficile* presence in CDI patients, and the prevalence and abundance of other potential enteropathogens able to cause CDI-like symptomatology.

However, as *C. difficile* can be found not only in CDI patients but also in healthy subjects of all ages, I extended the metagenomic survey of *C. difficile* beyond the disease-centred perspective, including 42,814 public metagenomic samples, covering a wide range of habitats, hosts, health status, host age ranges and geographic locations.

The second part of this Chapter (**Chapter 3.2**) is dedicated to present the results of this extended meta-analysis. I discuss *C. difficile* prevalence, abundance and *C. difficile*-associated microbial community composition first in the human gut over lifetime, then in other hosts and environments.

In the last part of this Chapter, I provide an overview of *C. difficile* diversity below the species level, discussing a potentially novel clade of *C. difficile* found exclusively in infants. I also discuss *C. difficile* toxigenic potential over lifetime, via the identification of toxin-producing genes from metagenomic reads (**Chapter 3.3**). I then conclude this Chapter presenting the limitations of my meta-analysis and its future developments (**Chapter 3.4**).

3.1 CDI-specific gut microbial signature and potential CDI overdiagnosis revealed by metagenomic meta-analysis

3.1.1 *C. difficile* identified in only 30% of CDI metagenomic samples

In the survey I included 534 gut metagenomic samples, from 10 publicly available cohorts (**Supplementary table 1**). 43% of the samples were obtained from CDI patients, 21% from diseased individuals (hospitalized, with diarrhea but no CDI diagnosis, or hospitalized and without diarrhea) and 36% from healthy subjects. The species-level composition of the microbial community associated with *C. difficile* presence was investigated in all three sample groups.

I detected *C. difficile* among 30.8% of CDI samples, 2.6% of diseased controls and 1.1% of healthy controls (**Figure 7A**). About half of the *C. difficile* positive samples, or 15.4% of the total number of CDI samples, were carrying toxin genes (**Figure 7A**). No case of *C. difficile* toxin gene detection was found among *C. difficile* negative samples.

In line with previous studies (Berkell et al. 2021; Schubert et al. 2014), CDI samples were characterized by a significantly reduced species richness compared to diseased and healthy controls, independently of the presence of *C. difficile* (**Figure 7B**). Diseased controls had a less diverse community than healthy controls (**Figure 7B**). No clear clustering by species composition was found when comparing sample groups (**Supplementary figure 2**).

C. difficile detection rate varied considerably across different studies (9.2%-92.3%) (**Figure 8A**). While counterintuitive, absence of *C. difficile* in samples from CDI diagnosed subjects was previously reported in other shotgun metagenomic-based studies (Vincent et al. 2016; Zhou et al. 2016), as well as 16S rRNA- and laboratory assay-based studies (Daquigan et al. 2017; Seekatz et al. 2016). It has been hypothesized that low sequencing depth could be related to the lack of *C. difficile* in CDI samples (Vincent et al. 2016). However in our meta-analysis, I found that our *C. difficile* detection approach based on marker genes (Milanese et al. 2019) was not impacted by sequencing depth (ANOVA, adjusted p=0.4533; R^2=0.01).

This apparent over-diagnosis of *C. difficile* could be due to differences in the diagnostic approaches used between different studies. Correct CDI diagnosis is challenging. Diarrheal symptoms can be caused by other entopathogenic species (Polage, Solnick, and Cohen 2012; Larcombe et al. 2018; Chia et al. 2017; Zollner-Schwetz et al. 2008; Kiu and Hall 2018), besides *C. difficile*, such as *Klebsiella oxytoca*, *Clostridium innocuum, Citrobacter amalonaticus, Clostridium perfringens, Staphylococcus aureus, Enterococcus faecalis, Enterobacter cloacae* and *Pseudomonas aeruginosa*. Even pseudomembranous colitis (PMC), once considered a hallmark of CDI, can be the result of an *E. coli* or *C. innocuum* infection (Tang, Urrunaga, and von Rosenvinge 2016; Chia et al. 2018). Despite the wide variety of laboratory tests available for CDI diagnosis (see **Chapter 1**), none have sufficient specificity and sensitivity to be used as a stand-alone test (Gateau et al. 2018).

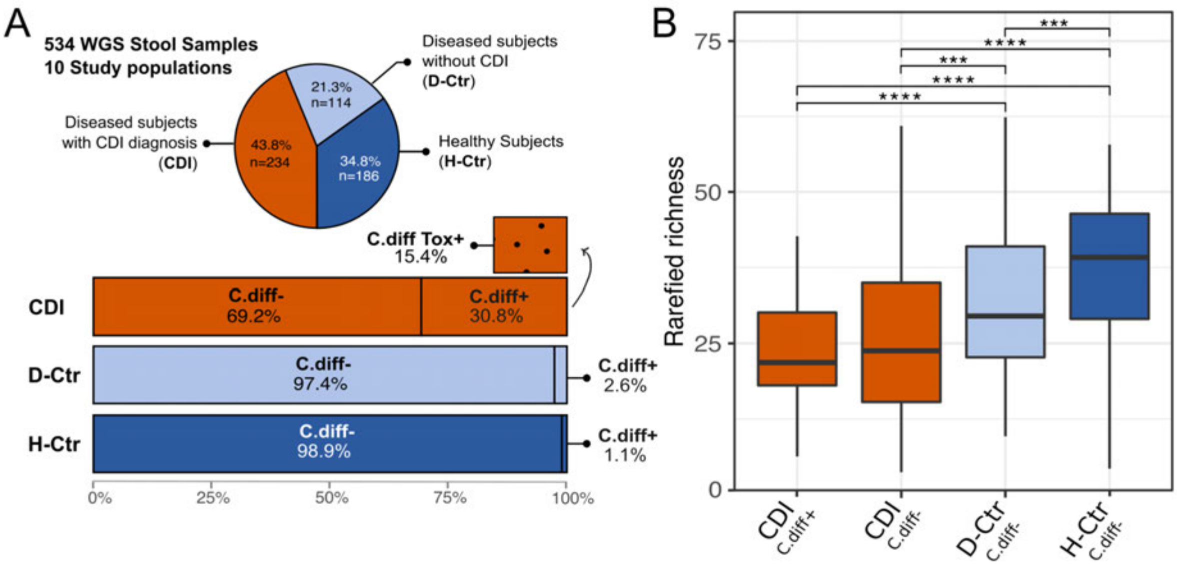


Figure 7. (A) Samples used in the meta-analysis from 10 public CDI or diarrheal cohorts, divided by groups: CDI (n=234), diseased controls (n=114), and healthy control (n=186) samples. (B) Rarefied species richness across the three sample groups. Mean comparison p-values calculated using t-test.

Reliance on a single test, even when in presence of symptoms, has been associated with elevated rate of CDI over- and mis-diagnosis (Polage et al. 2015; Lee et al. 2021). For two of the 10 studies included in our meta-analysis no information on the used CDI diagnostic protocol was available. Out of the remaining studies, 62.5% (5 out of 8) did not comply with the recommended guidelines on CDI diagnostic procedure (see Chapter 1 and (Gateau et al. 2018; M. J. T. Crobach et al. 2016)). Reasons for full or partial non compliance with the guidelines may include costs, limited testing capacity or test availability, and turnaround times of each hospital (Tenover et al. 2011). In the report of a recently failed phase 2 clinical trial of a Microbiome Therapeutic for CDI, CDI over-diagnosis due to reliance on a single test is listed among the potential underlying reasons for the failure (Vincent Bensan Young 2021).

In addition to the diagnostic challenges mentioned above, even when considering the combination of tests recommended in the guidelines (M. J. T. Crobach et al. 2016), there is a significant degree of variability in both sensitivity (>77%) and specificity (>91%) across different manufacturers, with more sensitive tests not necessarily being the most specific ones and vice versa (M. J. T. Crobach et al. 2016; Martínez-Meléndez et al. 2017; Lee et al. 2021). Moreover, different studies assessing specificity and sensitivity of the same diagnostic test from the same manufacturer produced contrasting results, with estimates at times differing by more than 40% (M. J. T. Crobach et al. 2016). To help put these variability ranges in perspective, a study found that a reduction of only 1.2% in the specificity of *C. difficile* toxins-targeting ELISA batch resulted in 32% increase in CDI diagnosed cases in a hospital in Missouri, in a pseudo-outbreak of *C. difficile (Litvin et al. 2009)*.

Unfortunately, the description of the CDI diagnostic procedure was not available for all cohorts, and several studies adopted multiple alternative diagnostic protocols, confirming how little standardized the diagnostic procedures are, even within the same hospital and study population. This aspect, combined with the lack of per-sample detailed information on how the CDI diagnosis was carried out, prevented us from assessing which diagnostic protocols were associated with the highest detection rate of *C. difficile*.

3.1.2 Enrichment of multiple enteropathogens in the gut microbiome of CDI patients

Considering the procedural pitfalls described in the previous paragraph, I hypothesized that when *C. difficile* was not present in CDI samples, other enteropathogenic bacterial species could explain the clinical symptomatology (diarrhea and/or PMC). All studies included in our meta-analysis had at least 3 species, besides *C. difficile*, known to be able to induce antibiotic-associated diarrhea (AAD) (**Figure 8A**).

90.1% of CDI samples lacking *C. difficile* had at least one AAD-species (excluding *C. difficile*), and 11.7% had more than 3 (**Figure 8B**). In *C. difficile* positive CDI samples, 98.6% had at least another AAD-species, and 25% had more than 3, suggesting that simultaneous colonization of multiple enteropathogenic species is significantly more common than previously estimated (Hensgens et al. 2014) (**Figure 8B**).

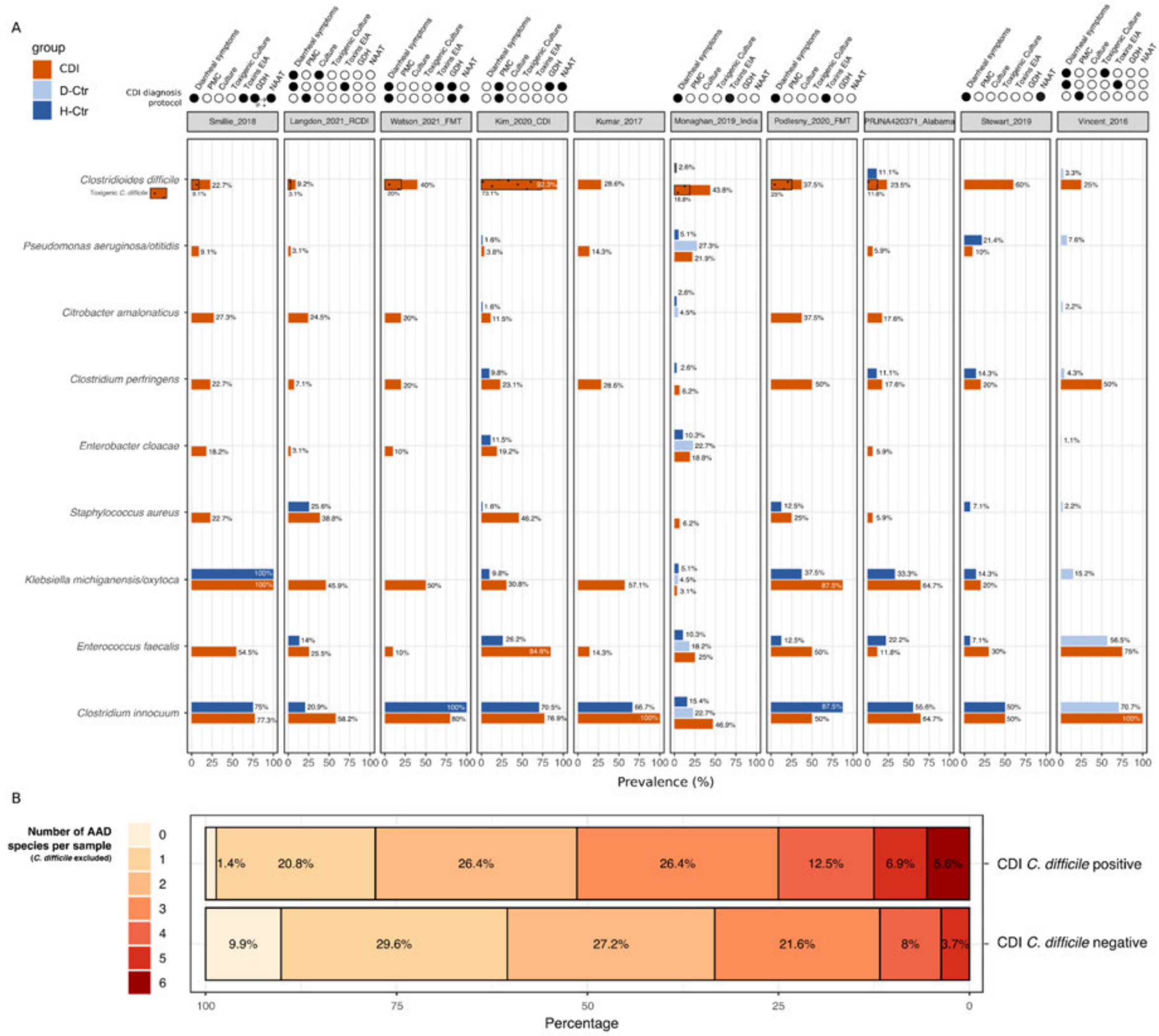


Figure 8. (A) Prevalence of *C. difficile* and other antibiotic-associated diarrhea (AAD) species, divided by study. For *C. difficile* only, the portion of samples with potentially toxigenic *C. difficile* is shown (dotted segments). CDI diagnosis procedure is shown on top of each study. Each row represents the diagnostic algorithm (combination of tests) used to diagnose CDI. Tests performed are indicated with black dots (white if not). For example, in "Langdon et al. 2021", CDI was diagnosed if symptoms were present and toxigenic culture for *C. difficile* was positive, or if symptoms were present and the enzymatic immunoassays for *C. difficile* was positive, or if pseudomembranous colitis was identified. Diagnostic protocols legend: "symptoms" refer to diarrheal stools, "EIA": Enzymatic ImmunoAssays, "GDH": glutamate dehydrogenase, "NAAT" nucleic acid amplification test (including PCR), "PMC": pseudomembranous colitis. (B) Prevalence of the number of AAD species (*C. difficile* not included) identified in each CDI sample, divided by *C. difficile* positivity.

All the AAD-species mentioned above were more prevalent in CDI samples, with the only exceptions of *P. aeruginaosa* and *E. faecalis*, more prevalent in diseased controls (**Figure 9**). *C. innocuum*, until recently considered innocuous (Chia et al. 2017, 2018), was found in CDI as much as in diseased controls. However when looking at the relative abundance, *C.*

innocuum was the only one associated with a significantly higher relative abundance in CDI samples compared to both diseased and healthy controls (**Figure 9**).

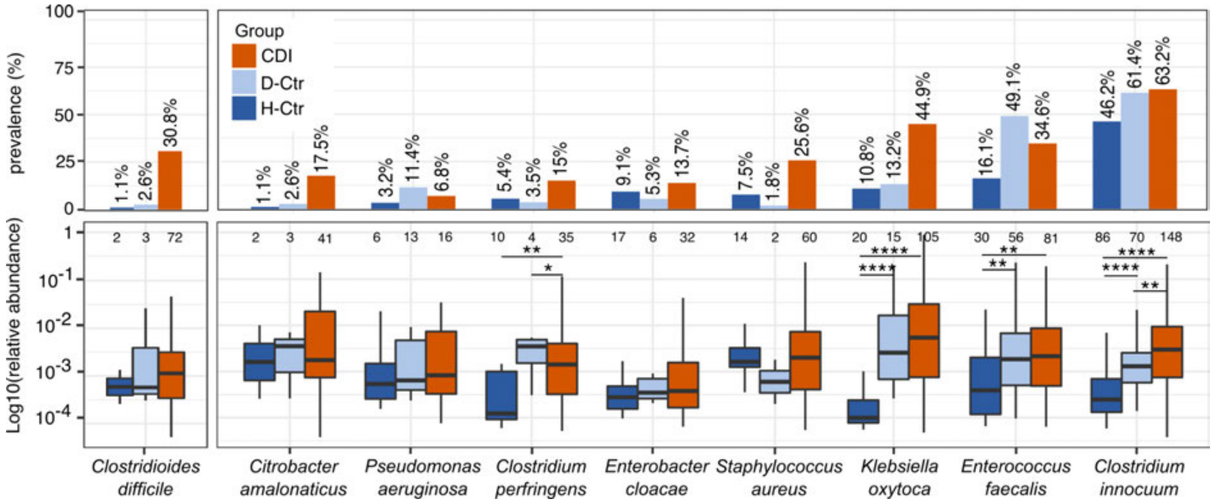


Figure 9. Prevalence and relative abundance of species known to be able to induce antibiotic-associated diarrhea (AAD) among CDI and control samples. Mean comparison p-values calculated using t-test.

Species enrichment analysis, comparing CDI samples against all controls altogether, showed that *K. oxytoca*, *C. amalonaticus*, *S. aureus* and *C. perfringens* were significantly enriched in CDI samples (**Figure 10**). Enterobacteriaceae sp., by far the most enriched mOTU cluster in CDI samples, includes *Shigella flexneri* and *E. coli*.

Other enriched species included species typically found in the oral cavity, such as *Veillonella parvula* and *Veillonella atypica* and probiotic species, such as *Lactobacillus rhamnosus* and *Lactobacillus salivarius* (**Figure 10**).

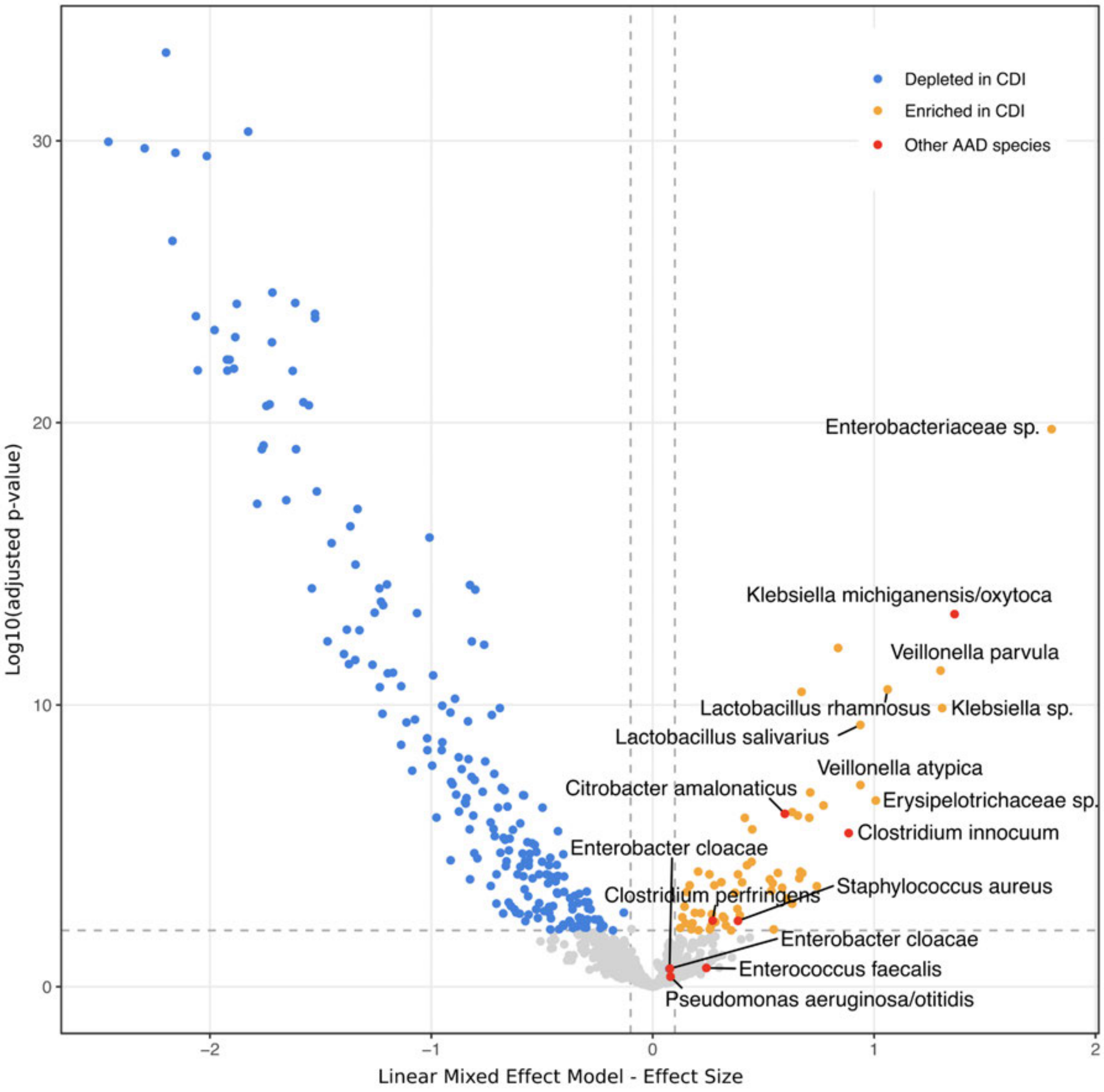


Figure 10. Species significantly enriched (in yellow) or depleted (in blue) in terms of relative abundance in CDI compared to diseased and healthy controls, as seen by linear mixed effect model analysis. Species known to cause antibiotic-associated diarrhea (AAD) are highlighted in red. *C. difficile* is not included in the analysis. Study and age group were considered as random effects. To be noted that Enterobacter sp. (ref_mOTU_0036) includes *Shigella flexneri* and *Escherichia coli*. Analysis performed in collaboration with Jakob Wirbel.

To identify the microbial signature associated with CDI samples, LASSO-regularised logistic regression models were trained using a cross-validated leave-one-study-out approach (see Methods in **Chapter 2**). The most predictive species for CDI samples included *C. difficile*, as expected, and multiple AAD enteropathogens (such as *S. aureus* or *C. perfringens*). AAD enteropathogens predictive signature was particularly pronounced in CDI samples lacking *C. difficile* (**Figure 11**). Several oral species (*Veillonella parvula, Veillonella atypica* and *Rothia dentocariosa*) and common probiotic species (*Lactobacillus casei*, *L. plantarum* and *L. fermentum*) were found among the most predictive species of CDI. These results, in line with what described in **Figure 10**, suggest that CDI patients might be characterized by an increased oral-gut microbial transmission, similarly to what is found in other diseases (Schmidt et al. 2019). The enrichment in Lactobacilli species among CDI patients might be associated with probiotic intake, possible during CDI hospitalization (Golić et al. 2017; Na and Kelly 2011). However, an alternative explanation could involve the extended antibiotic resistance typical of Lactobacilli (Tynkkynen, Singh, and Varmanen 1998; Klare et al. 2007), which could therefore be better equipped to survive the antibiotic treatments typically associated with CDI.

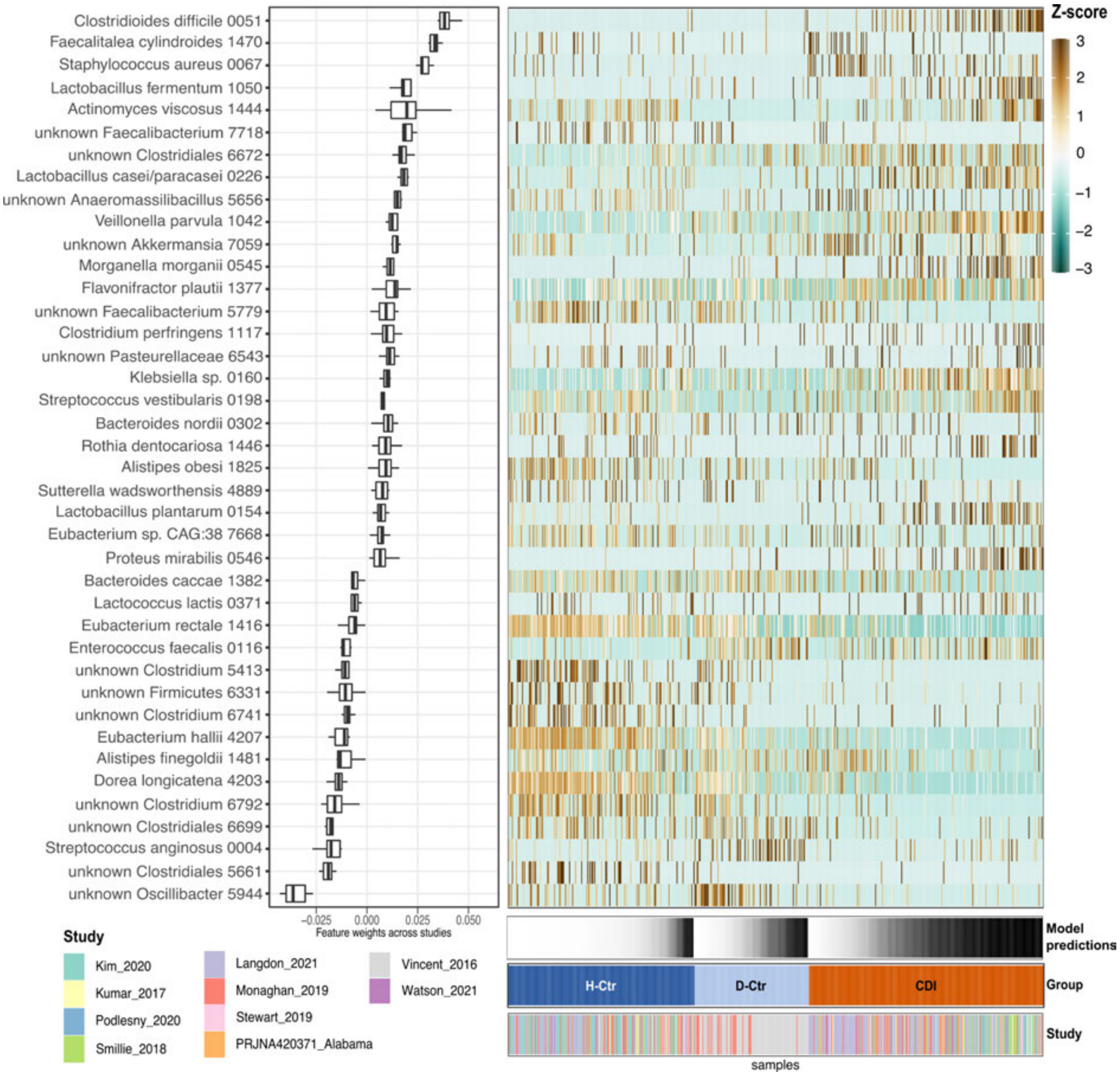


Figure 11. Microbial signature associated with the three sample groups, as seen by the LASSO model. List of the study population included in the model can be found in the **Supplementary Table 1.** Analysis performed in collaboration with Jakob Wirbel.

CDI samples were characterized by a specific microbial signature, distinguishable from the signature associated with diseased and healthy controls (**Figure 11**). CDI signature was not dominated by a single species, but was instead community-wide, as demonstrated by the high prediction accuracy obtained when excluding *C. difficile* from the community composition (**Figure 12**). This indicates that the gut microbiome of CDI patients is overall profoundly different from the one of a healthy individual or an individual diagnosed with other diseases (including non-CDI diarrhea), even in absence of *C. difficile*.

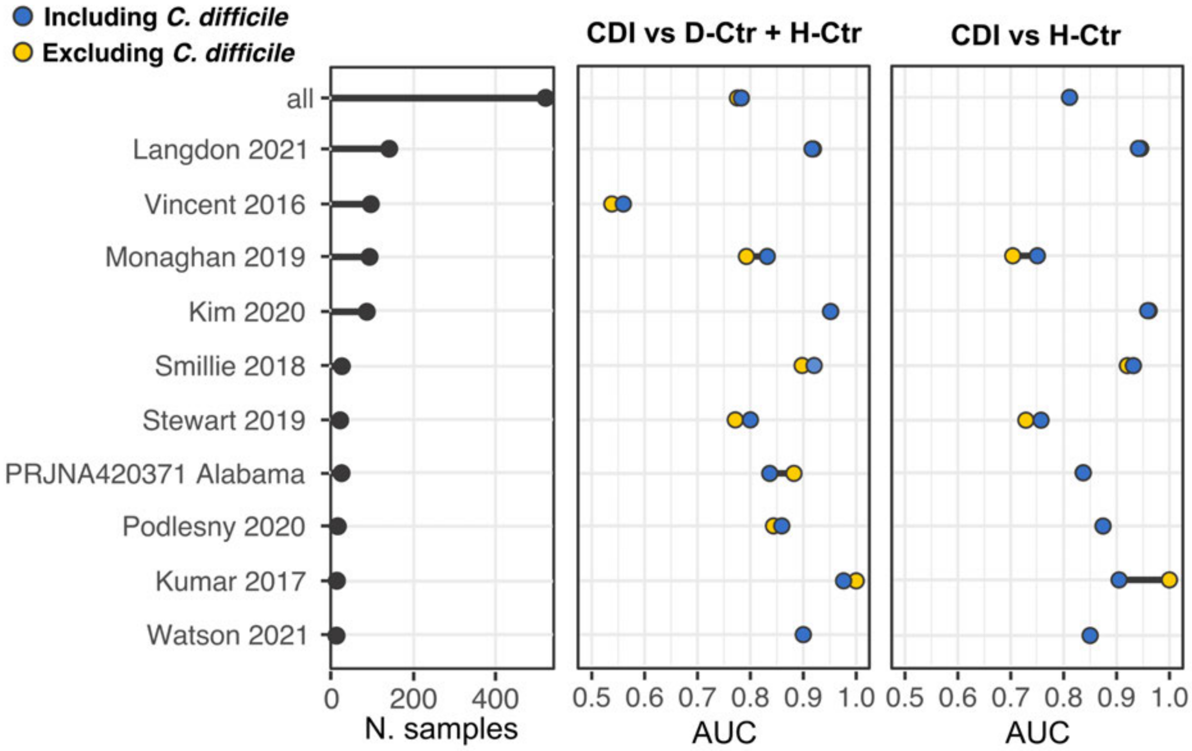


Figure 12. AUC values of the LASSO model for all samples as well as for single study populations. AUC values on the left refer to the comparison between CDI samples and healthy and diseased controls combined, while the AUC values on the right refer to CDI compared with healthy controls only. The AUC value of "Vincent 2016" when comparing CDI to healthy control sample is left intentionally blank as only diseased controls (D-Ctr) are available for this study population (see **Supplementary Table 1**). Analysis performed in collaboration with Jakob Wirbel.

Overall, the LASSO models identified the CDI-signature with high accuracy, with an average area under the receiver operating characteristics curve (AUROC) of 0.78. AUROC varied significantly across study populations, ranging from 0.56, for (Vincent et al. 2016), to 0.98, for (Kumar et al. 2017) (**Figure 12**). The predictive accuracy was higher when comparing CDI samples with healthy controls only (average AUROC = 0.81) (**Figure 12**), compared to CDI with both healthy and diseased controls.

C. difficile is not the only bacterial species inhabiting the gut that is able to cause diarrhea and to potentially carry toxigenic genes. We found that, compared to diseased controls, CDI samples were significantly enriched in toxin gene-carrying *C. difficile*, *S. flexneri*, *C. perfringens* and *S. aureus* species and that 58% of CDI samples lacking *C. difficile* had at least one strain of *E. coli*, *S. flexneri*, *S. aureus* or *C. perfringens* harbouring toxin genes (**Figure 13**).

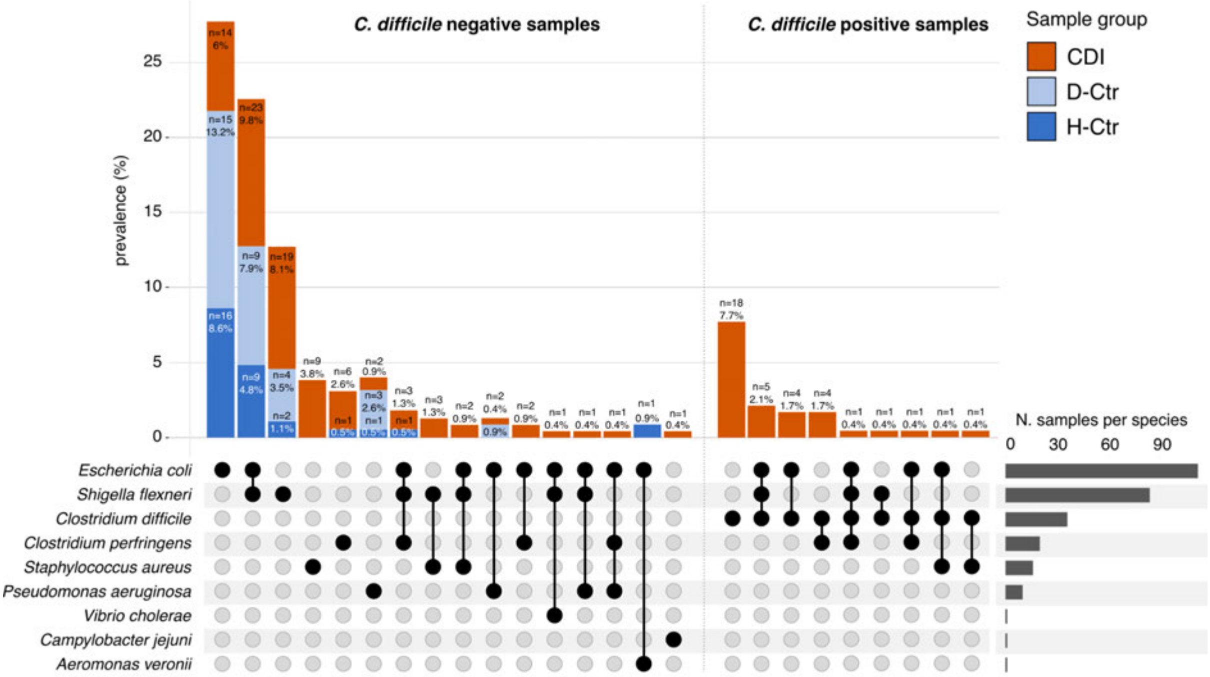


Figure 13. Prevalence of toxin genes-carrying species found among *C. difficile* positive and *C. difficile* negative samples, divided by sample group. Besides AAD species, also *E. coli* and *S. flexneri* can cause diarrhea, but not in an antibiotic-dependent manner. Cumulative prevalence across multiple strains is shown: *C. difficile* 630, *E. coli* CFT073, O44:H18 042 and O157:H7 str. EDL933, *S. flexneri* 2a str.301, *P. aeruginosa* PAO1, *C. perfringens* str.13 and SM101, *S. aureus* RN4220, subsp. aureus MW2 and N315.

Toxin gene-carrying *S. flexneri* can produce *Shigella* enterotoxin 1 (ShET1) and 2 (ShET2), that can cause diarrhea by altering electrolyte and water transport (Fasano 2002), while *S. aureus* and *C. perfringens* can produce a wide variety of toxins with different mechanism of action (Navarro, McClane, and Uzal 2018; Otto 2014). In our survey, the most commonly found toxin genes among CDI samples included *senB* from S. flexneri; *toxA* and *toxB* for *C. difficile*; *sat* and *astA* for *E. coli*; *tsst-1*, *esaD* and *lukD* for *S. aureus*; *cdtA* and *cdtB* from *C. jejuni*; *nagH*, *nagK* and *nagJ* for *C. perfringens* (**Supplementary Table 2**). Multiple species able to produce toxins were also found in *C. difficile* positive CDI samples, suggesting that even in presence of toxin gene-carrying *C. difficile*, other species might be contributing to or even leading the toxin-mediated inflammatory process associated with the clinical manifestations.

Together, these results suggest that potential mis- and over-diagnosis of CDI are serious concerns and that further studies are needed to discern the real burden of *C. difficile* from the burden of other enteropathogens capable of producing toxins, especially in case of simultaneous infection with multiple enteropathogenic species. Our results indicate that CDI mis- and over- diagnosis might be rampant, which besides inflating the numbers of infections and deaths associated with *C. difficile*, might also lead to the systematic under-estimation of the burden of other enteropathogenic species. CDI mis- and over-diagnosis have also serious financial implications. Hospital-acquired CDI is associated with a length of hospital stay up to 21.3 days, costing up to $29,000 per patient (Gabriel and Beriot-Mathiot 2014). Unnecessary treatment due to an incorrect CDI diagnosis can damage the health of the patient, increase antimicrobial resistance and rapidly drain the resources of a hospital (Lee et al. 2021).

3.2 *C. difficile* prevalence, abundance and biotic context across age, geography, health and disease

3.2.1 Age-dependent prevalence and abundance of *C. difficile* in human gastrointestinal tract

In order to investigate the global prevalence of *C. difficile* in a more exhaustive manner, 42,814 publicly available shotgun metagenomic samples were analysed, in what, to our knowledge, is the largest cross-habitat survey of *C. difficile* to date (**Figure 14A** and **Supplementary table 3**). Samples from 253 public study populations were included, covering a wide range of human body sites (gastrointestinal tract, oral cavity, skin, vagina, respiratory tract and human milk) and well as a broad geographical area (84 countries). The collection also included several thousand samples from the gastrointestinal tract of 25 animal species and from environmental sources such as indoor surfaces, soil, freshwater, wastewater and harbours (**Figure 14A**). The 28,347 human gut samples included all age groups, spanning from 1 day to 107 years of age, and both healthy and diseased subjects. Healthy adults represented the biggest category (7,538 samples), followed by healthy infants (4,972) and diseased adults (4,305). 19.7% of the human gut samples could not be assigned to a specific age group and/or health status (**Figure 14B**). Here, I considered as diseased those subjects who either had a medical diagnosis of any kind or that were under antibiotic treatment at the time of the sampling (see **Chapter 2** for details).

Species-level taxonomic assignment was performed via mOTU v2.0 (Milanese et al. 2019). In the entirety of our sample collection, 2,441 samples (5.7%) were *C. difficile* positive (**Figure 14C**). *C. difficile* was predominantly found in the human gut, in particular in the stools of healthy infants (15.9%) and diseased children (7.8%). Besides in stools, I identified *C. difficile* also in biopsy samples from the colon, cecum and terminal ileum lumen, and from the colon (all segments) and rectal mucosal tissue of healthy adult subjects. While typically found in the large intestine (cecum, colon, rectum and anal canal), previous reports confirm that *C. difficile* can also be sporadically found in the small intestine (duodenum, jejunum and ileum) and that incidence of small intestine *C. difficile* carriage might be underestimated (Navaneethan and Giannella 2009; Schubl et al. 2016).

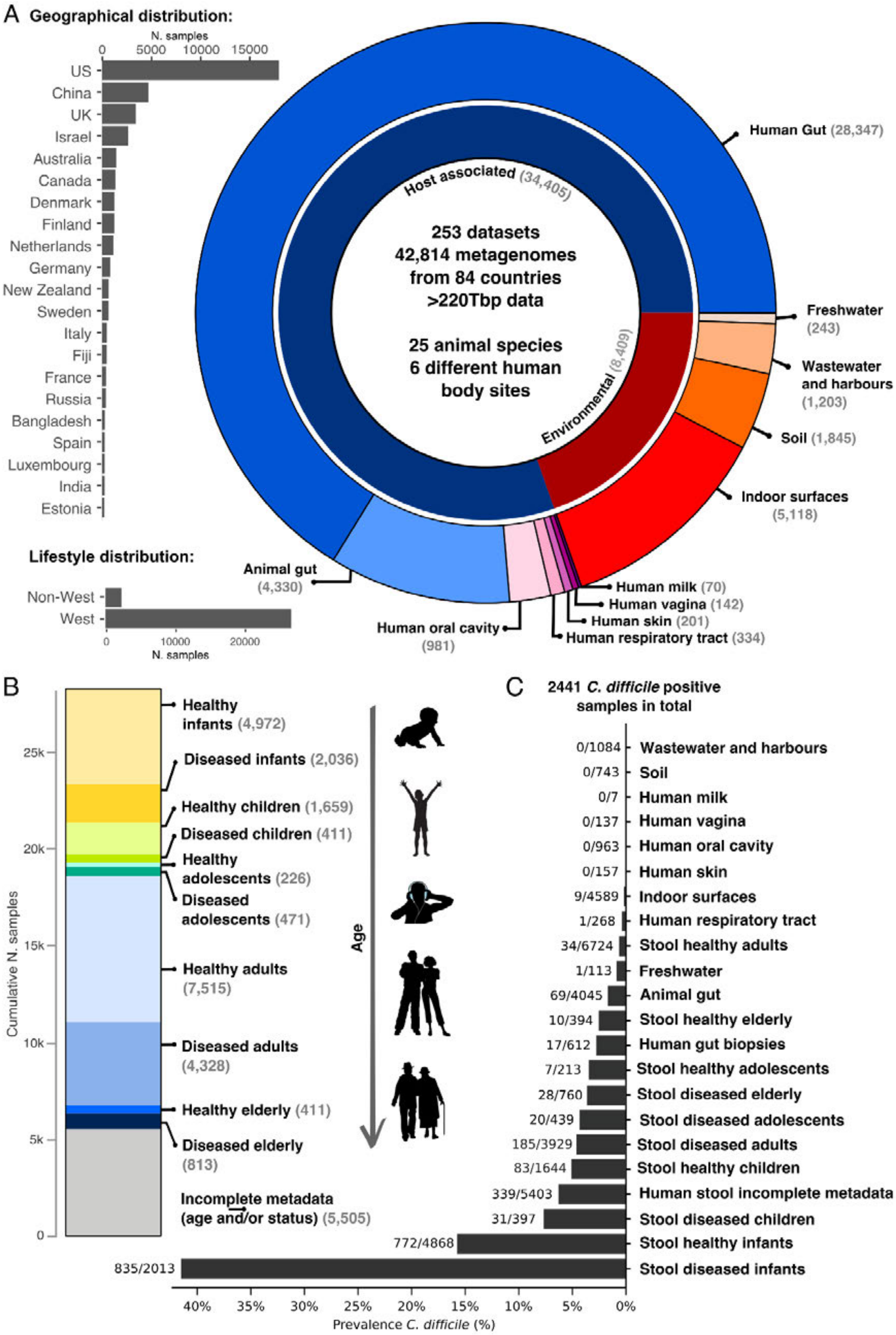


Figure 14. (A) Overview of the dataset collection composed of 42,814 samples, from 253 publicly available studies. Our dataset collection, mainly composed of samples from developed Westernized countries (left), includes both host-associated and environmental

samples (right, internal ring). These two categories can be further divided into human and animal body sites, and different types of environmental habitats (right, external ring). Only countries with >200 samples are shown. (B) Stratification of the human gut samples, divided by age category and health status. Numbers refer to the number of samples prior to read filtering. (C) Overview of the *C. difficile* positive samples, as seen by mOTUs v2.0 (Milanese et al. 2019), using 2 marker genes. Total values refer to the number of samples after the initial read filtering and before time series dereplication (see Methods).

C. difficile was identified in 1.7% of animal gut samples (including stools and rectal swabs) and in a sporadic manner (0.19%) also in the hospital surfaces of a neonatal intensive care unit (NICU), in an urban freshwater sample from Singapore (0.88%), and in the respiratory tract of a diseased child affected with pneumonia (0.37%). The child did not receive any antibiotic treatment in the month prior to sample collection (nasopharyngeal swab), and was diagnosed with *Mycoplasma pneumoniae* infection (Dai et al. 2019). It is possible to speculate that in this case the increased mucus production in the airways (Zhang et al. 2021) might have contributed to create an anaerobic environment suitable for the survival of *C. difficile*. Persistence of *C. difficile* on hospital surfaces has been previously recorded (Claro, Daniels, and Humphreys 2014; Brown et al. 2018), but, to the best of our knowledge, there is no previous report on *C. difficile* presence (without CDI diagnosis) in the human airways. No other body site resulted positive for *C. difficile*, indicating that this species is particularly well adapted at surviving in the human and animal gastrointestinal tract and only occasionally found elsewhere.

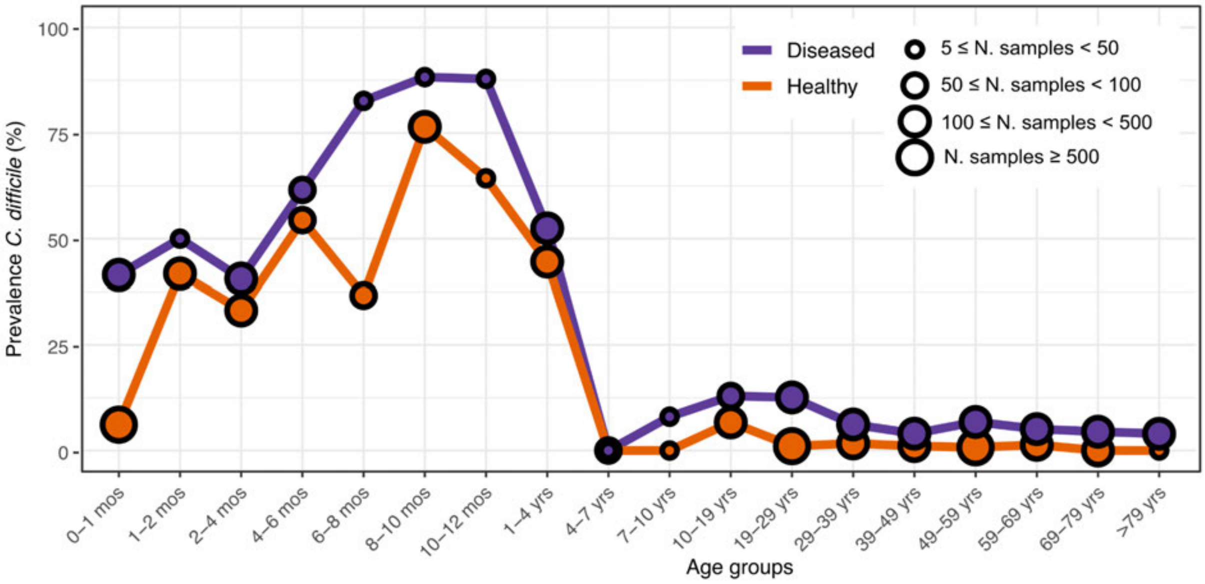


Figure 15. *C. difficile* prevalence in human stool samples. Diseased subjects are defined by either presence of a medically diagnosed condition and/or antibiotic intake at the time of the sampling. Prevalence was calculated based on presence/absence of *C. difficile* from mOTUs v2.0 taxonomic profiles.

This survey provided, for the first time, an in-depth survey of *C. difficile* prevalence, with bimestral resolution within the first year of life, for a large pool of healthy as well as diseased infants. While *C. difficile* was extremely common among healthy infants (25.5%) and children (7%), with significant fluctuations during the first year of life, it was rarely found in the gut of older subjects, independently of their health status (**Figure 15**).

Despite the significant reduction after the first year of life, *C. difficile* prevalence remained above 40% until the fourth year of life (**Figure 15**), in contrast with previous studies reporting an average prevalence ranging from 5 to 10% between the first and the second year of life (Lees et al. 2016; Jangi and Lamont 2010). However, it is important to highlight that, due to the metagenomic nature of the survey, these might be conservative estimations.

Doubling the window of persistent elevated asymptomatic carriage of *C. difficile* in infancy and early childhood has important implications for surveillance purposes, as healthy infants colonized with *C. difficile* are believed to be an important reservoir for community-acquired cases of *C. difficile* infection in adult and elderly subjects (Rousseau et al. 2012).

C. difficile prevalence was elevated in infants born via C-section (**Figure 16A**), and in premature infants (**Figure 16B**), in line with previous smaller scale studies (Stoesser et al. 2017; Ferraris et al. 2019). Indeed, premature birth and C-section were more strongly associated with *C. difficile* prevalence than health status (**Figure 16C**). In fact, when looking at the average prevalence of *C. difficile* across the first four years of life, the highest carriage was found among diseased premature infants born via C-section, while the lowest was in diseased full-term vaginally delivered infants (**Figure 16C**).

C. difficile prevalence varied considerably across these groups within the first semester of life (from 5.9% to 39.9% in healthy subjects, from 40.6% to 52.4% in diseased subjects), suggesting that high-resolution stratification, considering precise age, health status, prematurity and delivery mode altogether, is needed in this age group to ensure appropriate comparative meta-studies. Higher *C. difficile* prevalence in preterm and C-section born infants could be due to an increased exposure to environmental sources of bacteria, such as prolonged hospital stay, admission to neonatal intensive care unit (NICU), higher exposure to (also preventive) antibiotic treatments, and higher influx of maternal non-gut bacterial strains (Lees et al. 2016). *C. difficile* prevalence was also differentially associated with dietary intake in infants. In particular, exclusively formula-fed infants were characterized by a significant increase in *C. difficile* carriage, compared to exclusively breast-fed infants or infants fed with a mixture of formula and breast milk (**Figure 16D**). This result, in line with a previous study (Drall et al. 2019) suggests that, at least for *C. difficile* carriage, breastfeeding is considered protective, independently of the dosage.

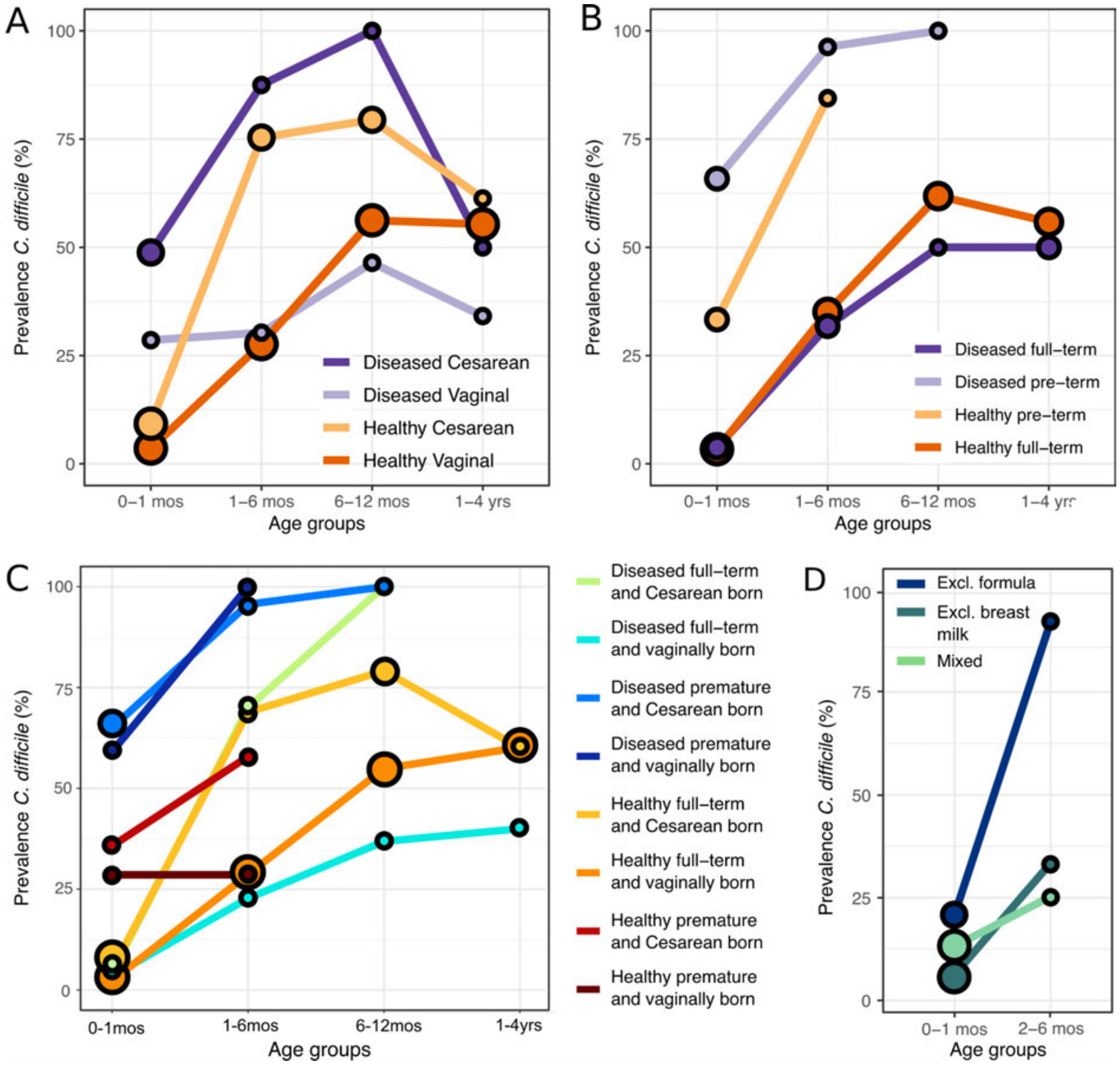


Figure 16. *C. difficile* prevalence in stool samples divided by (A) health status and delivery mode, (B) health status and prematurity, (C) combination of health status, delivery mode and prematurity and by (D) feeding mode.

The availability of time series in our meta-analysis enabled us to investigate *C. difficile* first appearance in the infant and child gut microbiome, during the first four years of life (**Figure 17**). It was previously suggested that *C. difficile* first colonization predominantly takes place in two stages: immediately after birth, as a result of hospital exposure, and in the second semester of life (Rousseau et al. 2012). Here, I confirm that indeed *C. difficile* first colonization is not equally distributed over the first year of life. However, despite early (within the first month of age) colonization events, I identified two peaks in the *C. difficile* first appearance distribution: one between 2-4 months of age, and the second between 8-10 months.

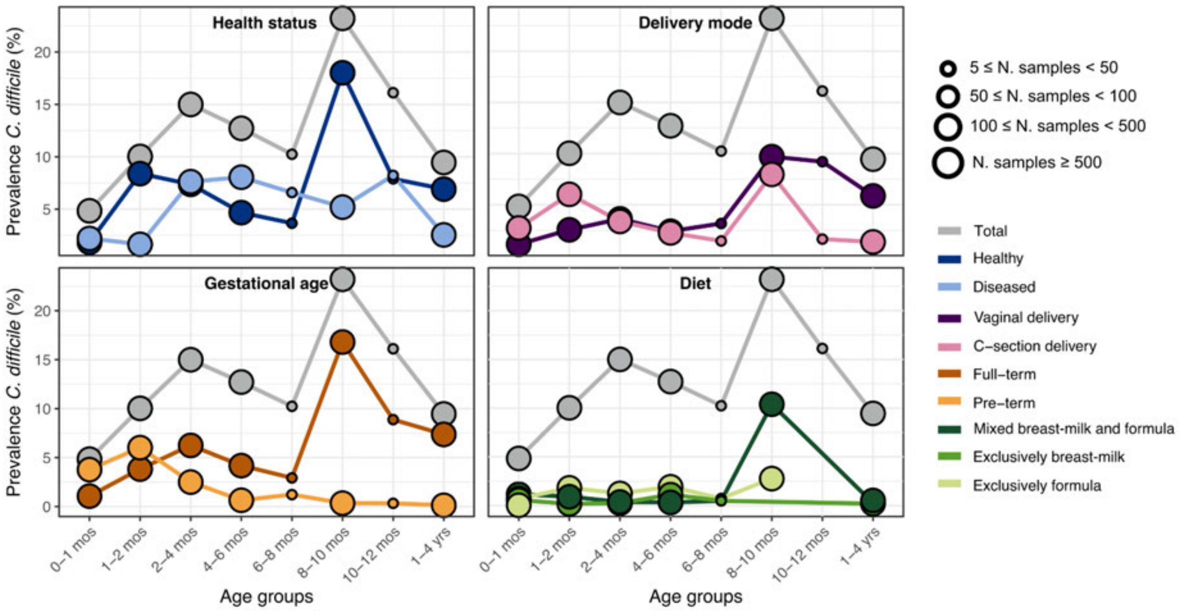


Figure 17. Time of the first appearance of *C. difficile* in infants and children timeseries. The grey line indicates the total number of samples per age interval.

I hypothesize that these two intervals coincide with increased exposure of the infant gut microbiome to environmental species. In particular, the interval between 2-4 months coincides with the beginning of mouthing, the practice typical of infants of this age to start exploring the surrounding environment and objects by putting them in the mouth. In my previous study on infant microbiome development over the first 4 months of life, this time interval was associated with an overall increase in non maternal strains influx (Ferretti et al. 2018). The peak between 8-10 months of first appearance of *C. difficile* was prominent in healthy full-term babies fed with a mixture of formula and breast milk (**Figure 17**). 66% of the samples in our survey in this age interval are from UK infants, and interestingly this interval coincides with the maximal duration of maternal paid leave in that country (Thevenon, Adema, and Clarke 2016), suggesting that at this age infants might start day care and therefore be exposed to a variety of new bacterial sources from the environments as well as from peers and day care staff. Introduction of solid food (starting from the sixth month of age) is likely another contributing factor, but the available solid diet metadata was not enough to investigate this aspect.

Infants born via C-section and premature infants were characterized by a higher prevalence of *C. difficile* acquisition during the first month of life (**Figure 17**) compared to vaginally born or at-term infants, suggesting that early acquisition of *C. difficile* might be related to increased exposure to the hospital environment.

The abundance of *C. difficile* relative to the rest of the bacterial community was highest in the first two months of life, then gradually decreasing later in life, with the lowest values found in elder subjects, independently of health status (1.1x10^{-2} and 3.8x10^{-2} for healthy and diseased infants, 3.1x10^{-4} and 3.4x10^{-3} for healthy and diseased adults) (**Figure 18A and 18B**). The highest *C. difficile* relative abundance (50%) was found in a healthy two weeks old premature infant (data not shown). *C. difficile* gradual reduction in terms of relative abundance was less linear in diseased subjects, compared to healthy ones, probably a result of the less stable gut microbiome that can characterize diseased subjects. When comparing health status within age groups, diseased infants and adults had significantly higher relative abundance of *C. difficile* compared to age-matched healthy subjects and the gap was more pronounced in adults than in infants (**Figure 18C**).

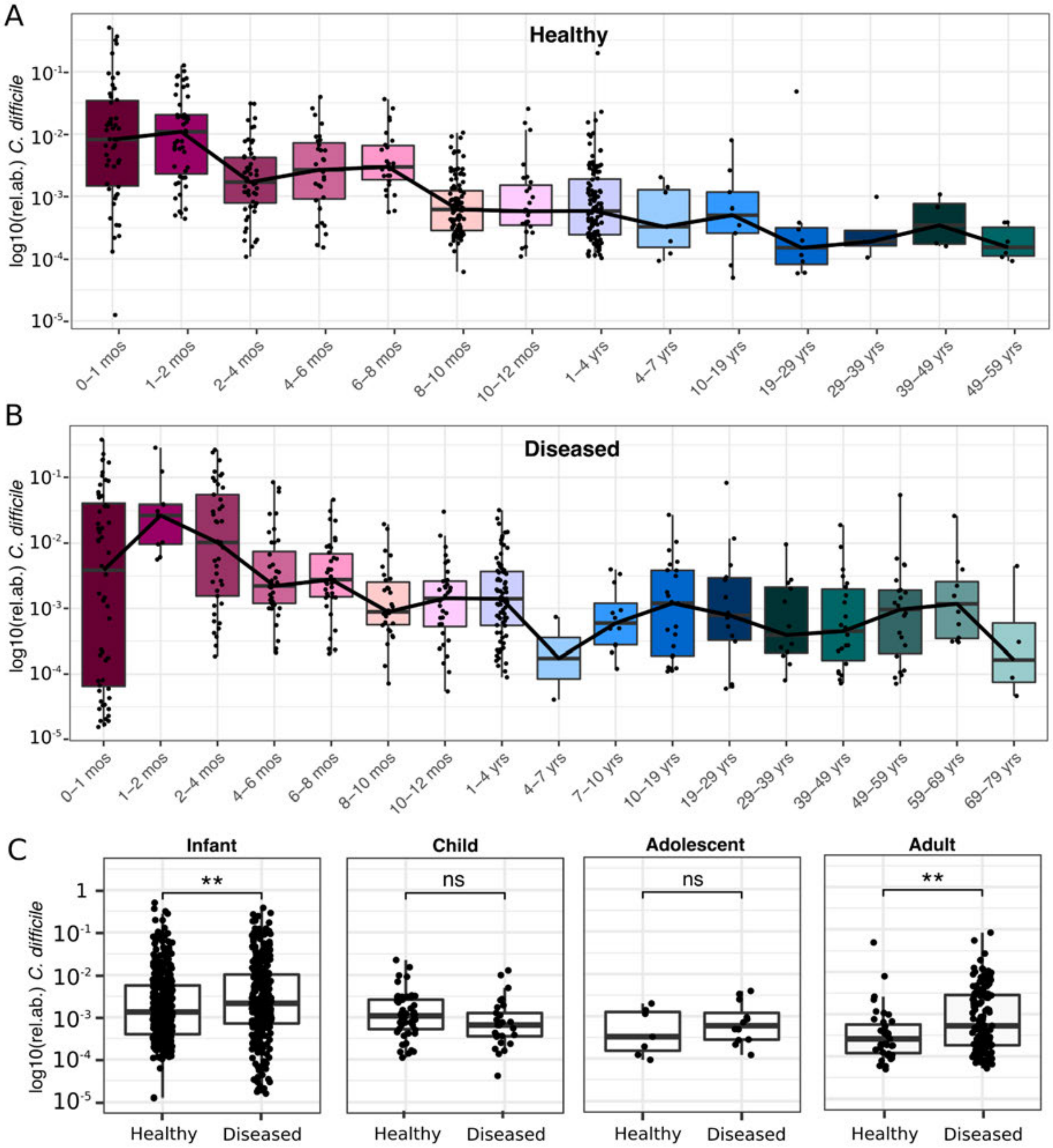


Figure 18. *C. difficile* relative abundance over life time in human stool samples of (A) healthy and (B) diseased subjects. (C) Differential relative abundance of *C. difficile* divided by health status and age group. Mean comparison p-values calculated using t-test.

Besides providing age-specific estimations on *C. difficile* prevalence and relative abundance, at times with monthly resolution, our collection of 42,814 samples allowed us to investigate *C. difficile* carriage in over 84 countries, representing over 40% of the countries in the world. In healthy populations, *C. difficile* prevalence variability within continents was as pronounced as between continents (**Supplementary figure 3**). In healthy adults, the only countries with prevalence above 1% were Canada (4.2%), China (2.5%) and UK (1.3%). *C. difficile* is recognized as a globally distributed species (K. E. Burke and Lamont 2014), with symptomatic carriage rate estimations varying considerably between countries, year of investigation, studies and detection method used (Stabler et al. 2012; Argamany et al. 2015; Zhao et al. 2021). For this reason, combined with the lack of metagenomics-based extensive *C. difficile* surveillance and the scarcity of *C. difficile* carriage in asymptomatic subjects, it is difficult to put our results in perspective of the currently available literature.

3.2.2 Enrichment of *C. difficile* across diseases

C. difficile was found in a wide range of diseases, including conditions not directly related to the gastrointestinal tract, such as cystic fibrosis, breast cancer, tuberculosis, liver cirrhosis and atherosclerotic cardiovascular disease (**Figure 19**).

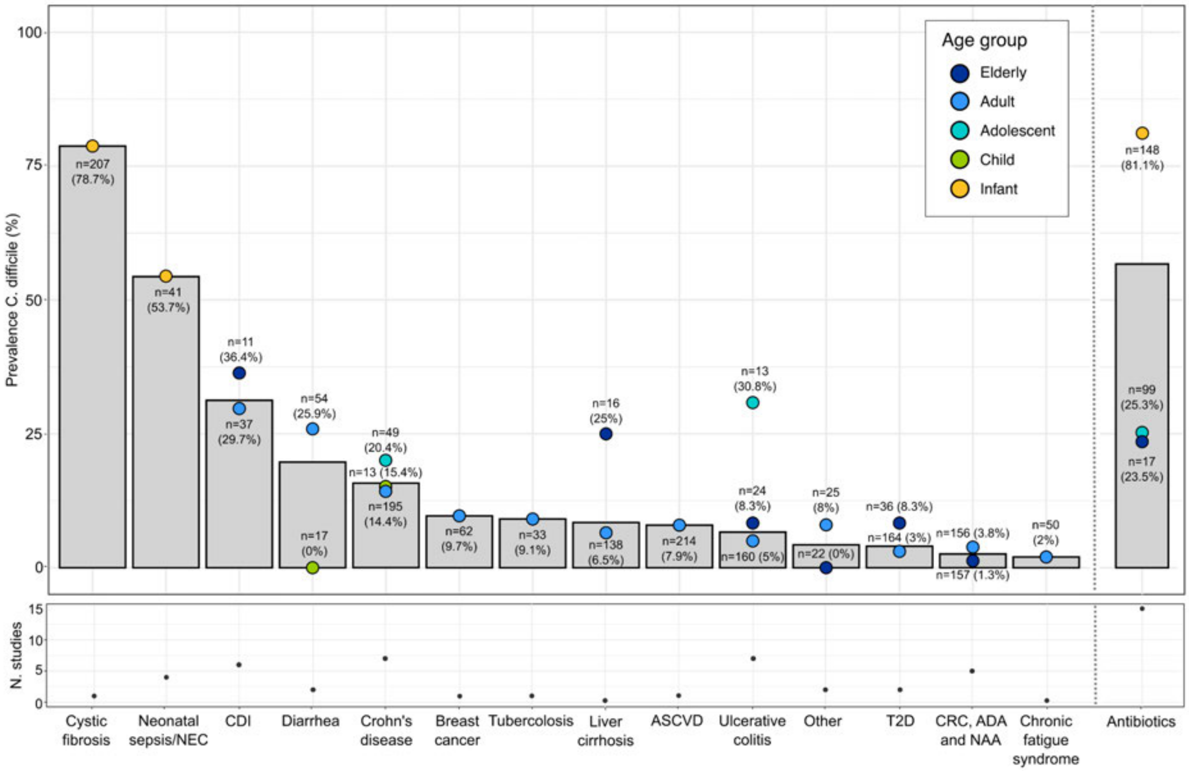


Figure 19. Prevalence of *C. difficile* across diseases (left) and drugs (right), across all age groups (grey bar) and divided by age group (colored dots). Number of cohorts used for prevalence estimations in each disease or drug is reported in the lower part. Abbreviations:

NEC: “Necrotizing enterocolitis”; ASCVD: “atherosclerotic cardiovascular disease”; T2D: “Type 2 diabetes”; CRC: “Colorectal cancer”; ADA: “advanced adenoma”; NAA: “non-advanced adenoma”.

Infants taking antibiotics had the highest *C. difficile* prevalence (81.1%), followed by infants diagnosed with cystic fibrosis (78.8%) and neonatal sepsis or NEC (53.7%). Later in life, *C. difficile* prevalence associated with antibiotic intake was significantly lower (25.3% in adolescents and 23.5% in elderly subjects). In diseased adults, *C. difficile* was more commonly found in CDI patients (29.7%), to an extent comparable to what found in patients suffering from generic unspecified diarrhea (25.9%), and Crohn’s disease patients (14.4%).

High *C. difficile* carriage rate in cystic fibrosis patients has been previously reported (Deane et al. 2021; Bauer et al. 2014), and is thought to be associated with the increased exposure to the nosocomial context and the frequent use of antibiotics for prolonged periods of time (Bauer et al. 2014). Cystic fibrosis patients are frequently colonized with toxigenic *C. difficile,* independently of age*,* but rarely develop CDI (D. G. Burke et al. 2017; Chaun 2001; Welkon et al. 1985; Tamma and Sandora 2012). At the moment there are only speculations as to why, despite the elevated prevalence of toxigenic *C. difficile*, the rate of symptomatic manifestations remains low.

I used logistic regression to identify which parameters could be predictive of *C. difficile* presence in the infant and adult gut microbiome (**Supplementary figure 4**). In univarite models, parameters such as health status in adult subjects, and geography, sex, gestational age, delivery mode and feeding practice in infants were significantly associated with *C. difficile* presence. However, almost none of these associations remained significant when considered in combined models. In combined models, geography (p=2.1x10^{-15}) and feeding pattern (p=7.8x10^{-5}) were predictive of *C. difficile* carriage in infants; geography (p=2.1x10^{-4}) and BMI (p=2.2x10^{-3}) in adults. As previously mentioned in **Chapter 3.1**, sequencing depth was not predictive of *C. difficile* presence in any of the combined models (p=0.5).

3.2.3 Increased diversity and resemblance to the maternal gut associated with *C. difficile* presence in the gut microbiome of healthy infants

Our collection of 42,814 samples allowed us to investigate, besides the prevalence of *C. difficile* over life, also the prevalence of other clinically relevant species, such as the AAD-species mentioned in **Chapter 3.1**. Overall, prevalence of AAD-species was higher in diseased subjects compared to healthy ones, but trends over lifetime varied significantly among species within the same health status (**Supplementary figure 5**). In healthy infants, *E. faecalis* and *C. innoccuum* were the most prevalent species, with the latter resembling the trend seen for *C. difficile* in the first four years of life, with over 80% average prevalence between 10 and 12 months of age (**Supplementary figure 5A**). *C. innocuum* was present at high prevalence also in elderly subjects, independently of the health status. Diseased infants were characterized by an increased prevalence of *C. perfringens*, *K. oxytoca* and *E. faecalis* (**Supplementary figure 5B**).

These results show that first, even among healthy infants carriage of opportunistic enteropathogens different from *C. difficile* is common; second, that in diseased infants, diarrheal manifestations could be ascribed to species other than *C. difficile*, similarly to what has been hypothesized for CDI patients (see **Figure 8B**).

Precise estimates of *C. difficile* prevalence and abundance over life time and across healthy as well as diseased populations are essential for tailoring surveillance programs and guide future studies. However in order to investigate *C. difficile* as a member of the gut microbiome, it's important to look at the broader picture and evaluate community-wide changes associated with *C. difficile* presence. Species richness is one of the parameters that can be used to evaluate microbial community composition, by counting the number of microbial species present in a sample.

While species richness over the first year of life has been extensively investigated (Ferretti et al. 2018; Stewart et al. 2018; Roswall et al. 2021; Guittar, Shade, and Litchman 2019), much less is known about the overall trend over lifetime. Leveraging our extensive collection of samples, I calculated species richness for healthy and diseased individuals, from birth up to 107 years of age (**Figure 20**). Species richness gradually increases over lifetime, with the biggest increase taking place between infancy and childhood, as shown also in (Yatsunenko et al. 2012; Roswall et al. 2021). Species richness did not significantly increase from adulthood to elder age in healthy subjects, but it did in diseased ones. When taking into account *C. difficile* presence, *C. difficile* positive samples had a significantly lower species richness than samples without *C. difficile*, across almost all age groups. The sole exception were infants, where presence of *C. difficile* was significantly associated with higher species richness. While postulated in a previous study (Drall et al. 2019), this is the first time this trend is confirmed empirically, with sufficient statistical power.

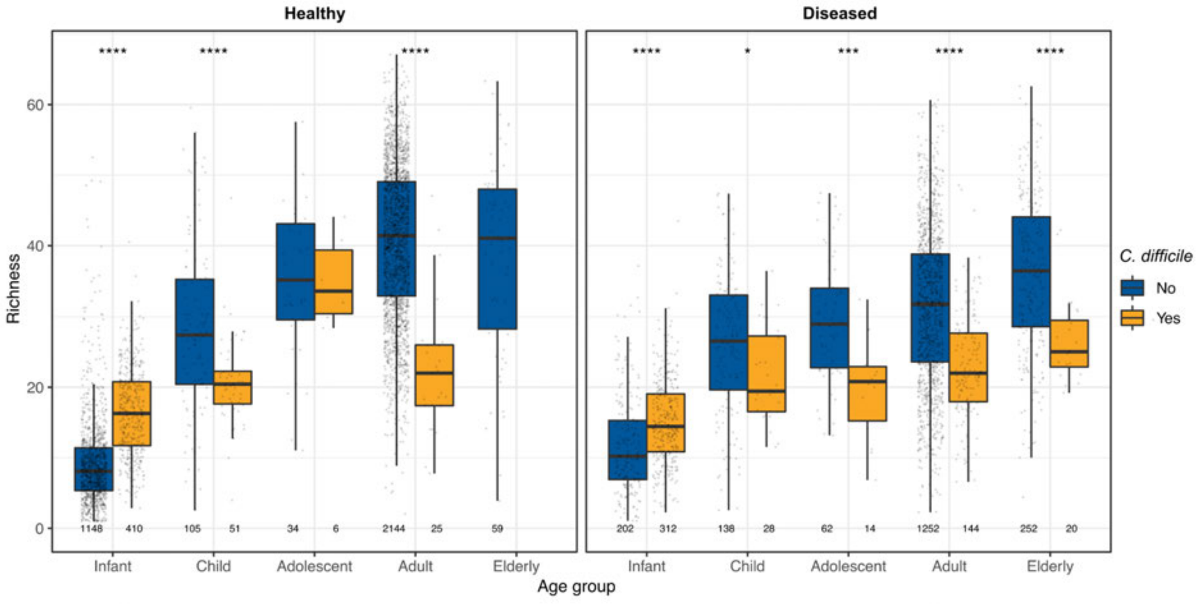


Figure 20. Species richness across all human stool samples, age groups and health status in presence or absence of *C. difficile*. The number of samples per group is shown under each boxplot. Mean comparison p-values calculated using t-test.

Moreover, I found that the increase in species richness associated with *C. difficile* positive samples was significant independently of health status (**Figure 21**), delivery mode, gestational age and geography (**Figure 21A-B**).

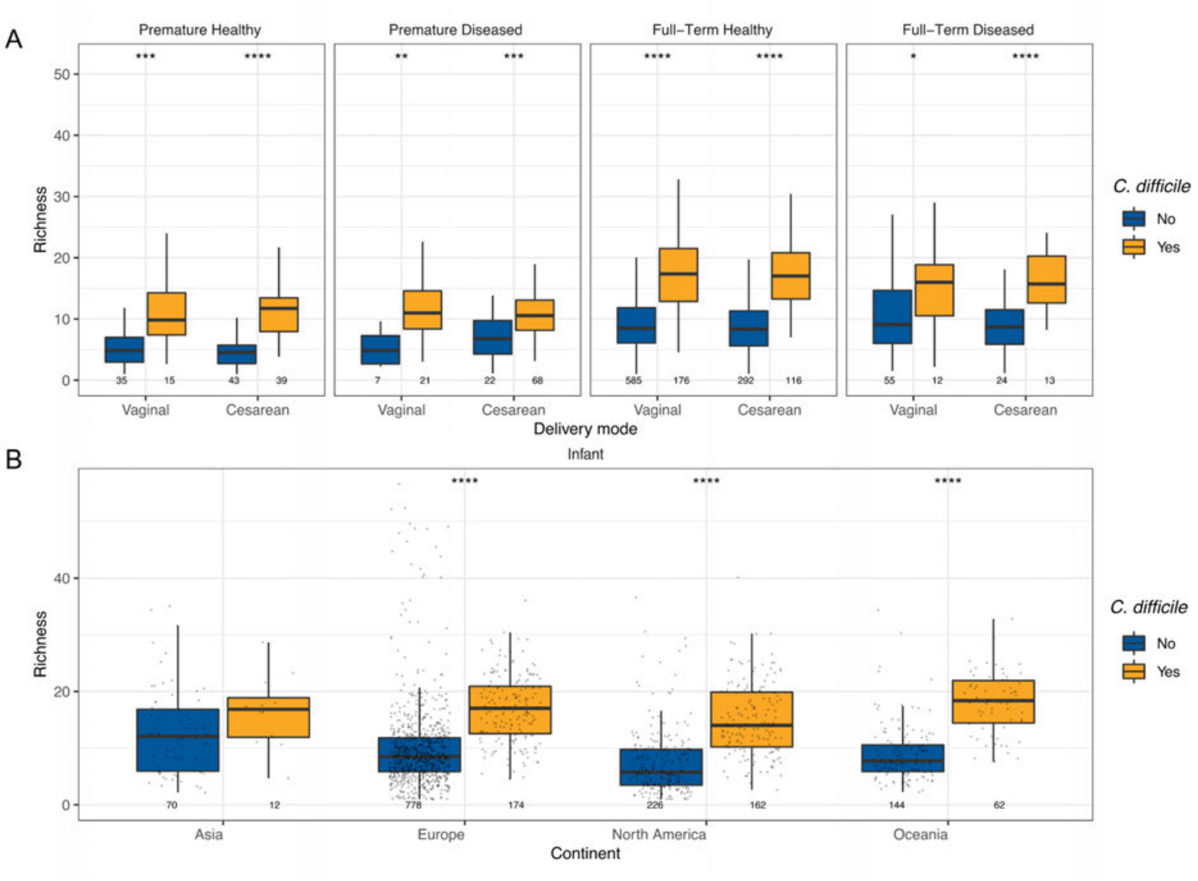


Figure 21. Species richness of human stool samples in presence or absence of *C. difficile*, divided by (A) delivery mode, prematurity and health status. (B) Species richness in healthy adult and infant gut stool samples across continents. Continents with at least five *C. difficile* positive samples are shown. The number of samples per group is shown under each boxplot. Mean comparison p-values calculated using t-test.

Health-related implications of an increased species richness have to be carefully evaluated. Reduced species richness has been identified in subjects taking antibiotics, subjects diagnosed with IBD (Lozupone et al. 2013), CDI (as I showed in **Chapter 3.1**), liver disease (Qin et al. 2014), chronic fatigue syndrome (Giloteaux et al. 2016) and cancer patients (C. Liu et al. 2017), compared to healthy counterparts. In particular in infants, reduced species richness is associated with cesarean delivery and premature birth (Rutayisire et al. 2016; Chernikova et al. 2018). In healthy infants, exclusive breastfeeding is associated with a reduced Bifidobacteria-dominated community, compared to infants exclusively or partially fed with formula (Roger et al. 2010; Roger and McCartney 2010; Bridgman et al. 2017). In the infant gut, bifidobacteria are able to metabolise human milk oligosaccharides, known as HMOs, and dominate the community (Garrido, Barile, and Mills 2012; Taft et al. 2018). Therefore, in the context of the infant gut microbiome, an increased species richness *per se* can be associated with both beneficial (i.e. vaginal at-term birth) as well as with sub-optimal practices (i.e. formula feeding). However, there are multiple reasons to consider the increased species richness observed in infants in presence of *C. difficile* beneficial, rather than detrimental. First, overall, the gut microbiome of every infant experiences an expansion in terms of species diversity (species richness) over time. Such expansion might be delayed or reduced in amplitude due to antibiotic treatment, hospitalisation, C-section or premature birth, but will take place sooner or later as it is a necessary step towards the maturation to an adult-like state. Second, introduction of solid food is associated with an increase in microbial species richness, in both formula- and breast-fed infants (Moore and Townsend 2019). For these reasons, the increase in species richness in infants can be considered, overall, as beneficial and as part of the normal infant gut development towards an adult-like state, as suggested also by (Bäckhed et al. 2015; Ferretti et al. 2018).

Species evenness represents another method to evaluate community-wide changes in the microbiome, by describing the relative differences in the abundance of different species (Vincent B. Young and Schmidt 2008). In other words, this parameter describes how evenly represented different species are in the community.

I found that *C. difficile* presence was associated with significantly elevated species evenness in healthy full-term infants, independently of delivery mode (**Supplementary figure 6A-B**). In adults, *C. difficile* was associated with a significant reduction in evenness (**Supplementary figure 6**). While the evenness reduction was found in both healthy and diseased subjects, the difference between samples with and without *C. difficile* was more pronounced in healthy subjects. Therefore, while in healthy full-term infants *C. difficile* is associated with a more even community (less dominated by few microbial species), in adults *C. difficile* is associated with a community dominated by fewer bacterial species. Together these results suggest that *C. difficile* presence is associated with age-specific community structure, with a significant and opposite trend in infants compared to adults.

Over the first year of life, the maturing infant gut microbiome becomes more diverse and more homogeneous in terms of abundance, as well as more similar to the maternal gut microbiome composition (Ferretti et al. 2018). Investigation of gut microbiome similarity in healthy infant-mother pairs revealed that the gut microbiome of healthy infants colonized with *C. difficile* was significantly more similar to the gut microbiome of their mothers, in terms of species composition, compared to the infants that did not harbour *C. difficile* (**Figure 22**). The similarity was more pronounced between 4 and 10 months, compared to the first months of life. However, the opposite trend was found starting from the tenth month of age, as infants above this age and children colonized with *C. difficile* had a significantly lower similarity with their mothers' microbiome, compared to infants that were not colonized with *C. difficile* (**Figure 22**).

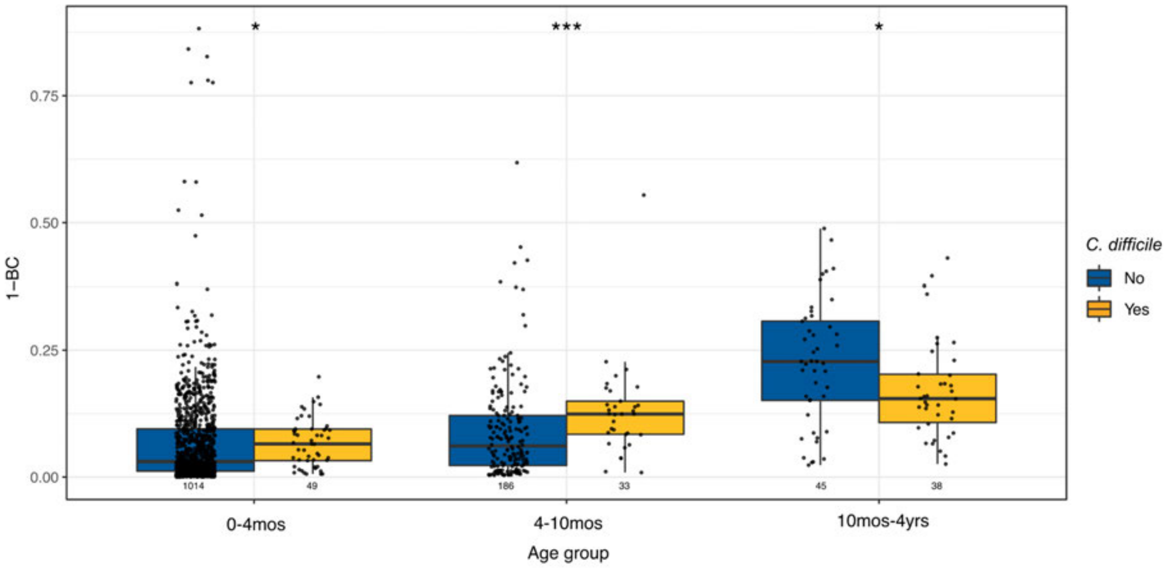


Figure 22. Bray-Curtis dissimilarity of healthy infant-mother pairs in presence or absence of *C. difficile* in stools across the first four years of life. The number of samples per group is shown under each boxplot. Mean comparison p-values calculated using t-test.

An increased similarity of the infant gut microbiome to the maternal microbiome during the first year of life is, similarly to the increase in species richness previously discussed, usually considered beneficial and highlighted as part of the normal development of the infant gut microbiome (Ferretti et al. 2018; Bäckhed et al. 2015).

3.2.4 The unique biotic context associated with *C. difficile* in healthy infants

To further characterize the microbial community in terms of species composition, I calculated pairwise associations between each bacterial species and *C. difficile* using Fisher tests, identifying co-occurrence or avoidance patterns (see Methods in **Chapter 2**).

Two clusters of species emerged, hereafter called Group 1 and Group 2 (**Figure 23A**). Group 1 was composed of species consistently co-occurring with a *C. difficile* across all age groups, independently of the health status. Despite this overall trend, *Enterobacter sp*., *Streptococcus mitis*, *Streptococcus pseudopneumoniae* and *Streptococcus pneumoniae* were co-occurring with *C. difficile* only in diseased infants. In Group 1, not a single species was found to be consistently co-occurring with *C. difficile* across all age groups in healthy subjects but not in diseased, or vice versa.

Group 1 included several species typically found in the oral cavity, such as *Veillonella parvula*, *Veillonella atypica* and *Veillonella dispar*. Interestingly, while an enrichment in oral species was previously found among adult and elderly CDI patients, here these species co-occur with *C. difficile* in healthy subjects and diseased infants, but show neutral association in diseased adults and elderly. This might indicate that the mechanism underlying the presence of oral species in CDI patients (i.e increased oral-gut leakage) is different from the co-occurrence of oral species in the *C. difficile*-colonized gut of infants and healthy subjects.

Additionally, Group 1 was characterized by several opportunistic enteropathogens, such as *Klebsiella oxytoca*, *Clostridium perfringens*, *Clostridium innocuum* and *Citrobacter amalonaticus.*

To further characterize the species in this group, I included, for each species, the species oxygen requirement, its prevalence over life time and its enrichment in Westernized populations (**Figure 23B**). Only 64% of species in Group 1 were obligate anaerobic species. An increase in oxygen tolerant species in the gut microbiome of adult subjects has been associated with IBD and Salmonella proliferation (Rigottier-Gois 2013; Rivera-Chávez et al. 2016). Furthermore, the increased intra-luminal concentration of oxygen is hypothesized to be the consequence of a compromised mucosal barrier, ongoing inflammation and subsequent release of ROS and NOS (Daniel, Lécuyer, and Chassaing 2021; Swidsinski et al. 2007).

Contrary to Group 1, species in Group 2 co-occurred with *C. difficile* almost exclusively in healthy infants, but not in diseased infants nor in later stages of life (**Figure 23A**). The sole exceptions were “unknown Peptostreptococcaceae (meta_mOTU_7331)”, co-occurring with *C. difficile* in both healthy and diseased infants; *Eubacterium sulci*, positively associated with *C. difficile* in healthy infants and healthy adults; and *Clostridium sp*. and *Clostridium sporosphaeroides*, found with *C. difficile* in healthy infants and diseased adults. Group 2 also included several butyrate producing species, such as *Roseburia sp.*, *Faecalibacterium prausnitzii* or *Anaerostipes hadrus.* Butyrate producers are important members of the gut microbiome, as butyrate production is involved in a myriad of processes that ultimately constitute the first line of defence against *C. difficile* toxins (Pruitt and Lacy 2012)), such as i) production of mucin, an important component of the gastrointestinal mucosal layer, ii) regulation of cell-to-cell tight junctions, essential to the integrity of physical barrier against pathogen invasion, and iii) inflammation inhibition (Cornick, Tawiah, and Chadee 2015) among other functions (Zheng, Kelly, and Colgan 2015). While a previous 16S rRNA-based study on a small cohort (Rousseau et al. 2011) identified relevant associations between *C. difficile* and species such as *K. pneumonia* and *R. gnavus*, in our study neither of these species was positively or negatively associated with *C. difficile* in any category.

Among the species co-occurring with *C. difficile* in healthy infants (Group 2), there were also several species known for their ability to degrade the intestinal mucus, such as *Akkermansia municiphila, Ruminococcus bromii* and *Ruminococcus torques (Yang et al. 2017; Tailford et al. 2015)*. As previously discussed in **Chapter 1.2.2**, the thickness of the mucosal layer is directly and indirectly influenced by the gut microbiome species, and it represents an important line of defence against the action of *C. difficile* toxins. However, in breastfed infants mucin degrading species might not be degrading the mucosal layer after all. A recent study showed that HMOs found in breastmilk are structurally similar to the host’s mucin, and suggested that dietary mucin intake could prevent mucin degraders from consuming the host mucosal layer (Pruss et al. 2021).

It is possible to speculate that the combination of inflammation inhibition and increased mucin production (both associated with an enrichment in butyrate producers) and lack or reduced rate of mucin degradation, may help explain the low rate of CDI symptomatology among healthy infants, especially if partially or exclusively breast-fed. However, as the gut microbiome of breast-fed infants is usually dominated by Bifidobacteria , which are particularly efficient at metabolising HMOs (Garrido, Barile, and Mills 2012; Taft et al. 2018), it is likely that nutrient competition and niche exclusion play an important role in the low rate of CDI among this age group.

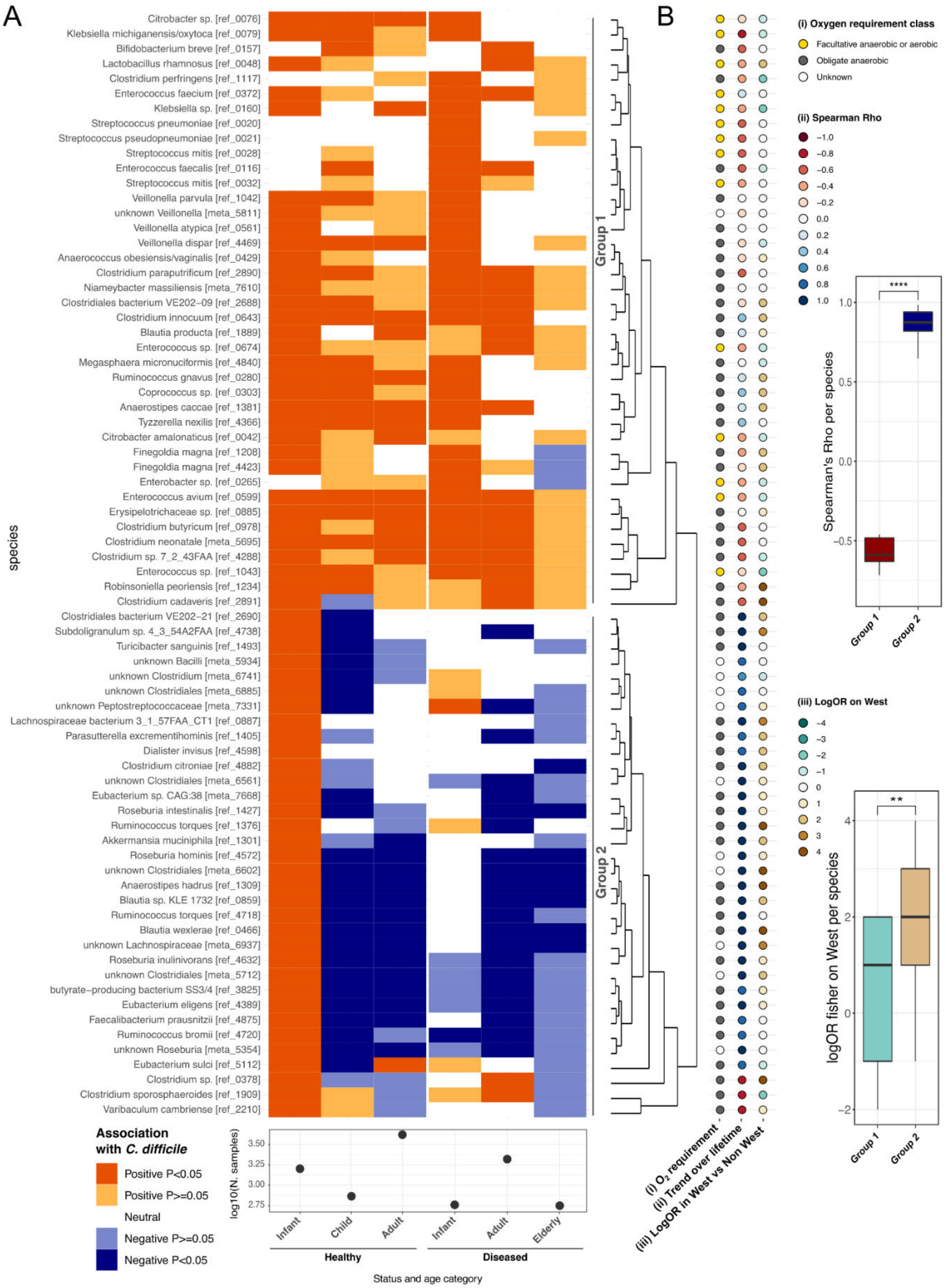


Figure 23 (A) Positive (orange) and negative (blue) species associated with *C. difficile* divided by age group and health status. Number of samples per each category shown in the lower part. (B) Annotation of those species includes oxygen requirement, prevalence trends over life

time in the healthy human gut, and enrichment in westernised vs non-westernised populations. Negative Spearman Rho values indicate that the species is more commonly found in the gut of healthy infants or children, rather than in the gut of adults or elderly. Positive Spearman Rho values indicate the opposite trend. Only species significantly associated either positively or negatively, with *C. difficile* in at least one age group are shown. Mean comparison p-values calculated using t-test.

In terms of oxygen tolerance, all the species in Group 2 with known oxygen tolerance were obligate anaerobes. While presence of oxygen tolerant species could be considered deleterious in the adult gut, anaerobes and facultative anaerobic species are commonly found in the infant gut microbiome after birth (Rodríguez et al. 2015; Houghteling and Walker 2015). However, within months, the community shifts from being mildly aerotolerant to being dominated by obligate anaerobic bacteria. This transition, confirmed in several studies (Ferretti et al. 2018; Bäckhed et al. 2015; Rodríguez et al. 2015; Houghteling and Walker 2015), can be considered part of the normal infant gut microbiome development towards the adult-like state, and therefore beneficial. Therefore, while Group 1 was enriched in oxygen tolerant species, considered detrimental in adulthood, Group 2 was almost exclusively composed of obligate anaerobes, considered beneficial in infancy.

In terms of prevalence over lifetime, Group 2 was significantly enriched in species typically found in the gut of healthy adult subjects (**Figure 23B**). This analysis was performed looking at the prevalence of each species in the stools of healthy individuals, from infancy to elderhood, and assigning a trend score (Spearman Rho). Positive values are associated with higher prevalence of the target species in adulthood and lower in infancy, with negative values associated with the opposite trend.

The species consistently co-occurring with *C. difficile* across (almost) all age groups (Group 1), were more commonly found in infancy than in adulthood, compared to Group 2. This is consistent with the observations on oxygen tolerance, as many of the species more commonly found in infancy were indeed oxygen tolerant species. I also found that species in Group 2 were significantly enriched in species more commonly found in the gut of healthy Westernised individuals, compared to healthy non-Westernised ones (**Figure 23B**).

Together these results suggest that not all the microbial communities harbouring *C. difficile* are equal, and that significant age-related differences exist when comparing the *C. difficile*-positive infant microbial communities with the adult ones, both in terms of community richness, evenness, composition and oxygen tolerance.

In particular, in healthy infants, *C. difficile* was found to be associated with higher resemblance to the maternal gut microbiome, enrichment in obligate anaerobes and in species typically found in the gut of healthy adults. As all of these parameters are considered beneficial and are considered important milestones in the healthy infant gut microbiome development towards an adult-like state (Bäckhed et al. 2015; Yassour et al. 2016; Ferretti et al. 2018; Chu et al. 2017; Rodríguez et al. 2015). These reasons, combined with the elevated asymptomatic carriage rate of *C. difficile* in infancy, indicate that *C. difficile* might be considered a hallmark of a desirable microbiome development in healthy infants.

Our study is, to the best of our knowledge, the largest metagenomic-based survey to date to investigate the microbial associations with *C. difficile* among healthy infants, and the first to suggest the commensal role of *C. difficile* in the gut microbiome of healthy infants.

3.2.5 Limited similarity between human and animal gut microbiomes harboring *C. difficile*

Investigations for *C. difficile* carriage in animal hosts predominantly focused on animals with clinical manifestations and/or on animals in direct contact with human populations. As no clear genetic distinction was found between *C. difficile* strains from humans and strains from animals (J. Scott Weese 2020), the study of *C. difficile* potential reservoirs is of paramount importance. *C. difficile* has been identified in a variety of animal species, ranging from domestic and farm animals to wildlife.

Rats living in urban areas were characterized by elevated carriage rate of human-associated ribotypes, suggesting that exposure to human waste can increase interspecies transmission of *C. difficile (“Carriage of Clostridium Difficile by Wild Urban Norway Rats (Rattus Norvegicus) and Black Rats (Rattus Rattus)” n.d.; Lim, Knight, and Riley 2020)*.

In healthy domestic animals, the majority of studies identified *C. difficile* carriage to be around or below 6% (J. Scott Weese 2020). However certain factors, such as hospitalization and direct contact with human elderly care facilities, are associated with increased carriage rates (up to 58%) (Sandra L. Lefebvre et al. 2009; S. L. Lefebvre et al. 2006).

While providing invaluable case-studies and statistics, these studies are usually performed on a limited geographical scale and are characterized by a reduced number of samples. To investigate *C. difficile* carriage in animal hosts on a global scale, 46 publicly available cohorts of shotgun metagenomic samples were collected, including 35 animal species spanning over 24 different countries (**Supplementary table 4** and **Figure 24A**).

The majority of samples (85.4%) were from healthy adult animals (**Figure 24B**), while only a small fraction (1.3%) were diseased. The remaining fraction (13.3%) did not have any available health status metadata. Taxonomic classification via mOTU2.0 identified *C. difficile* in 69 samples (1.7%). In line with previous studies, I found that dogs were characterized by the highest carriage rate (7%), followed by cattle and cats (3.6% and 3.2%, respectively) (**Figure 25A**). In healthy subjects, the average *C. difficile* carriage rate and relative abundance in humans was higher than in any other animal host, suggesting that future studies aiming at characterizing *C. difficile* role in healthy individuals should prioritize human subjects. Worth of note is that cows were the only animal species with *C. difficile* relative abundance comparable to humans (**Figure 25B**).

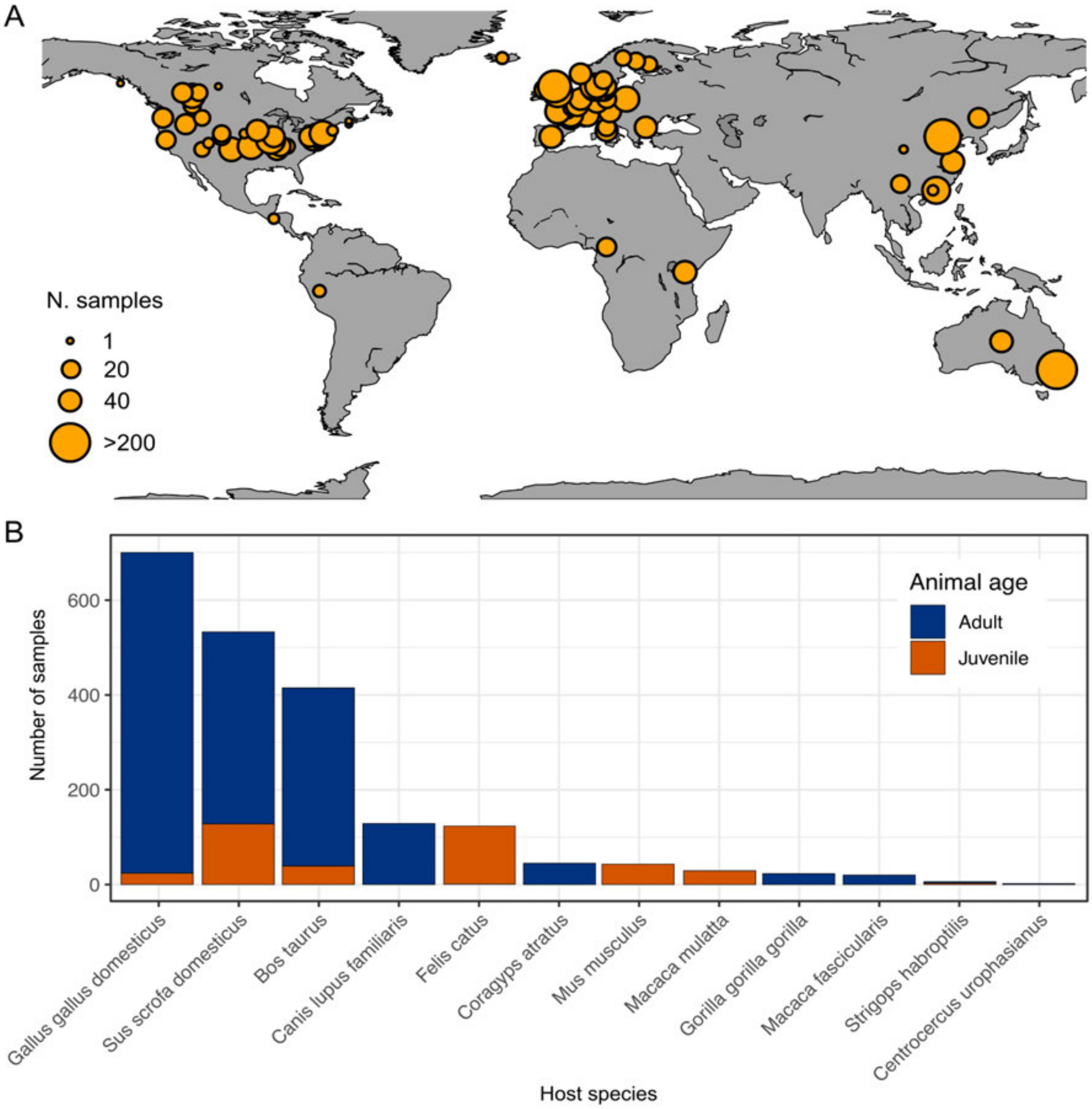


Figure 24. (A) Geographical and (B) age group distribution of the samples included in the global survey of *C. difficile* across animal species.

Overall, the species co-occurring with *C. difficile* in the gastrointestinal tract of humans did not significantly overlap with those of other hosts, with the sole exception of *Clostridium paraputrificum* and *Clostridium neonatale* (**Figure 25C-D**). Co-presence of *C. difficile* with *C. paraputrificum* has been associated with an increased biofilm production, compared to *C. difficile* alone (Normington et al. 2021).

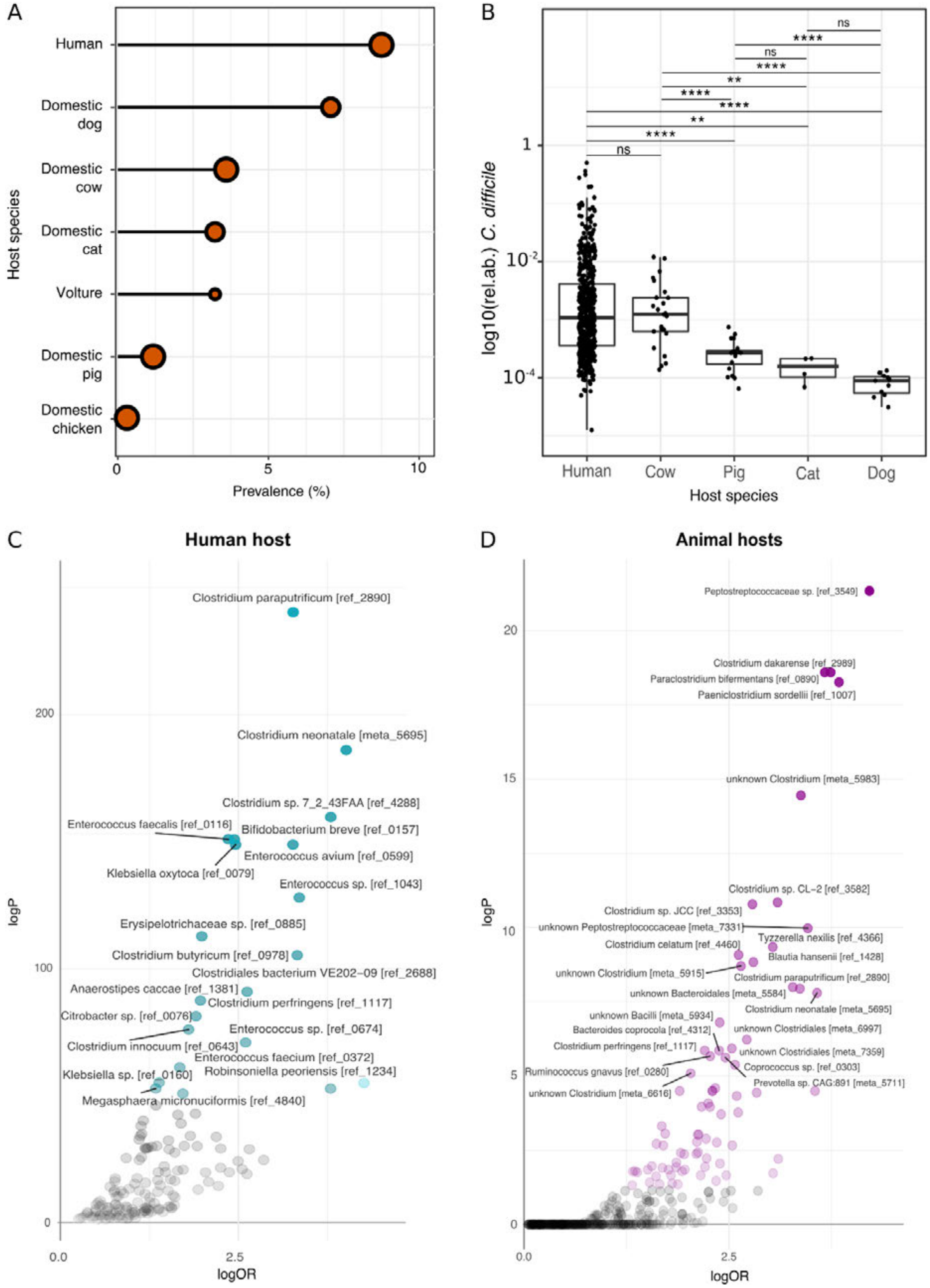


Figure 25. *C. difficile* (A) prevalence and (B) relative abundance in animal stool samples, compared to humans, as seen by mOTUs v2.0, using two marker genes. Only healthy animals and humans are included. Relative abundances are shown for only the *C. difficile* positive

samples. Mean comparison p-values calculated using t-test. Species co-occurring (logOR>0) with *C. difficile* across gut stool samples from (C) humans (n=14,095) and (D) animals (n=3,967).

The scarce overlap in terms of species composition between human and animal *C. difficile* positive samples can be attributed to the anatomical, physiological and biochemical differences in the digestive system structure between humans and other animals (Kararli 1995). Literature on *C. neonatale* and *C. paraputrificum* is extremely scarce (Kiu et al. 2017; Smith et al. 2011), highlighting the need to better characterize these species, especially in relation to *C. difficile* colonization, in both human and animal hosts.

3.3 Unexplored *C. difficile* within-species diversity and pathogenicity potential over lifetime

3.3.1 Metagenomic-based identification of a potentially novel clade of *C. difficile* exclusively found in infants

As described in detail in **Chapter 1**, *C. difficile* strains can be grouped into eight different monophyletic clades. To investigate within-species diversity, *C.difficile*-positive samples were analysed via metaSNV (Paul Igor Costea et al. 2017). Of the initial 2441 samples, 197 passed the requirements for robust SNV calling (8%) and 179 could be used to detect subspecies presence, see **Chapter 2** for detailed description). As only human stool samples passed this filtering stage, an adequate comparison in terms of subspecies between human, animal and environmental *C. difficile* positive samples could not be performed.

Two subspecies emerged (**Figure 26**): one composed of 121 samples where sub-structures were clearly present, called subspecies 1 and one with almost identical SNV profile, composed of 58 samples, hereafter called subspecies 2. The two subspecies could not be clearly separated by study population (not shown), geography, nor health status of the subjects. However, subspecies 1 was significantly enriched in *C. difficile* harbouring toxin genes (Fisher Test p-value: < 2.2e-16, odds ratio 62.0, **Figure 26**), and subspecies 2 was exclusively found in infant samples (Fisher Test p-value: 5e-05, odds ratio 14.7, **Figure 26-27**). Three samples, despite being within the thresholds for being grouped together with the other samples in subspecies 1, showed a divergent SNV profile (**Figure 26**). All three samples belonged to diseased subjects, one was an adult (affected by chronic kidney disease) and two were infants (one under antibiotic treatment, the other affected by cystic fibrosis).

In order to put our results in perspective of the current literature, we included reference genomes for all the eight currently known clades of *C. difficile (Knight et al. 2021)* (described in detail in **Chapter 1.2.3**). The genetic similarity between genomes of different clades was in line with previous studies (Knight et al. 2021). However, the *C. difficile* positive metagenomes obtained from our global survey showed a much wider genetic variability than the currently known clades of *C. difficile.* Many of the metagenomic samples in subspecies 1 did not cluster with any known clade, with few cases clustering with *C. difficile* genomes from Clade 1 and Clade 2.

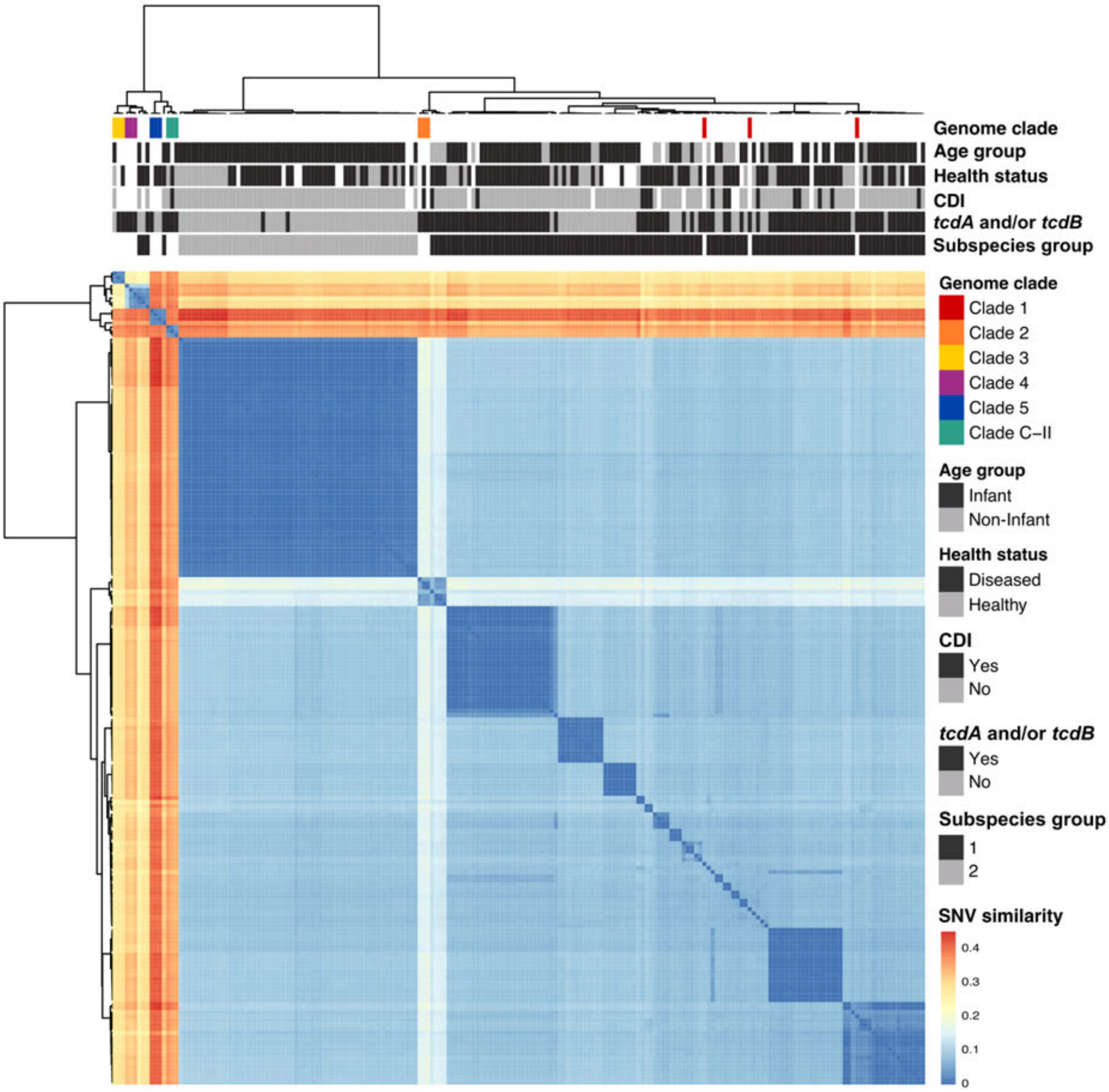


Figure 26. SNV similarity across *C. difficile* positive metagenomic samples from our global survey, and genomes from known *C. difficile* clades. Genomes from the cryptic clades C-I and C-III did not pass filtering steps. All metagenomes and genomes have ANI similarity >95%. In health status and age group, white is used for missing metadata. See Methods in **Chapter 2** for detailed analysis description.

The *C. difficile* metagenomic samples belonging to subspecies 2 were not represented by any of the known clades (**Figure 26**). The existence of a new clade of *C. difficile,* found exclusively in infants, can therefore be hypothesized. However, while subspecies 2 was observed exclusively in infant samples, infant samples were not exclusively identified in subspecies 2. In fact, several infant samples were clustered with Clade 1 genomes (**Figure 26**). The existence of this new *C. difficile* clade might have been previously overlooked as the investigative efforts on *C. difficile* pathogenicity mechanisms have been predominantly focused on the adult and elderly population, leaving the infant population (especially when asymptomatic) significantly under-studied and under-represented in the current *C. difficile* clade classification.

In three cases both subspecies of *C. difficile* co-existed within the same sample: in a 6 months old infant diagnosed with cystic fibrosis and in two premature infants at birth born via C-section (data not shown). In all three cases subspecies 1 was the most abundant one.

3.3.2 Extensive asymptomatic carriage of potentially toxigenic *C. difficile* over lifetime

While able to provide important information on the *C. difficile* population structure across age and health status, subspecies analysis does not necessarily reflect the finer-scale resolution differences in terms of toxigenic potential. As mentioned in **Chapter 1.2**, not all *C. difficile* strains carry the toxin genes encoding for Toxin A (TcdA) and Toxin B (TcdB). TcdA and TcdB are the major *C. difficile* toxins, however a small fraction, ranging from 17% to 23%, of *C. difficile* strains carries also an additional toxin called Binary Toxin (CDT, see **Chapter 1.2.3** for detailed explanation) (Eckert et al. 2015). Its presence has been associated with increased symptoms severity (Gerding et al. 2014), but A-/B-/CT+ (presence of CDT, lacking both TcdA and TcdB) are considered possible but rare (Eckert et al. 2015), and are usually found in animal hosts (Schneeberg et al. 2013; Knight, Squire, and Riley 2015). For these reasons, the results are from hereafter focused on assessing the presence of TcdA and TcdB toxin genes. Toxin genes detection was applied to *C. difficile* positive samples only, as in the smaller collection of CDI study populations discussed in **Chapter 3.1**, all samples with either one or both *C. difficile* toxin genes were *C. difficile* positive (**Figure 3.1**).

In our analysis, *C. difficile* toxin genes were found in both healthy and diseased subjects of all ages (**Figure 27A**). The prevalence of potentially toxigenic *C. difficile* in healthy infants (22.6%) was higher than previous estimations (Kubota et al. 2016; Rousseau et al. 2012), but comparable to the levels found in healthy adults (28.6%).

Diseased subjects were characterized by an increased carriage of toxin gene-carrying *C. difficile*, compared to healthy subjects (**Figure 27A**). In particular, *C. difficile* toxin genes were found in 50% of CDI patients (as previously mentioned in **Chapter 3.1**) 46% of cystic fibrosis patients, 45.5% of infants suffering from neonatal sepsis or NEC, 43.8% of Crohn's disease patients and in 42% of subjects taking antibiotics (**Supplementary figure 7**). None of the biopsy samples harboured *C. difficile* toxin genes (data not shown).

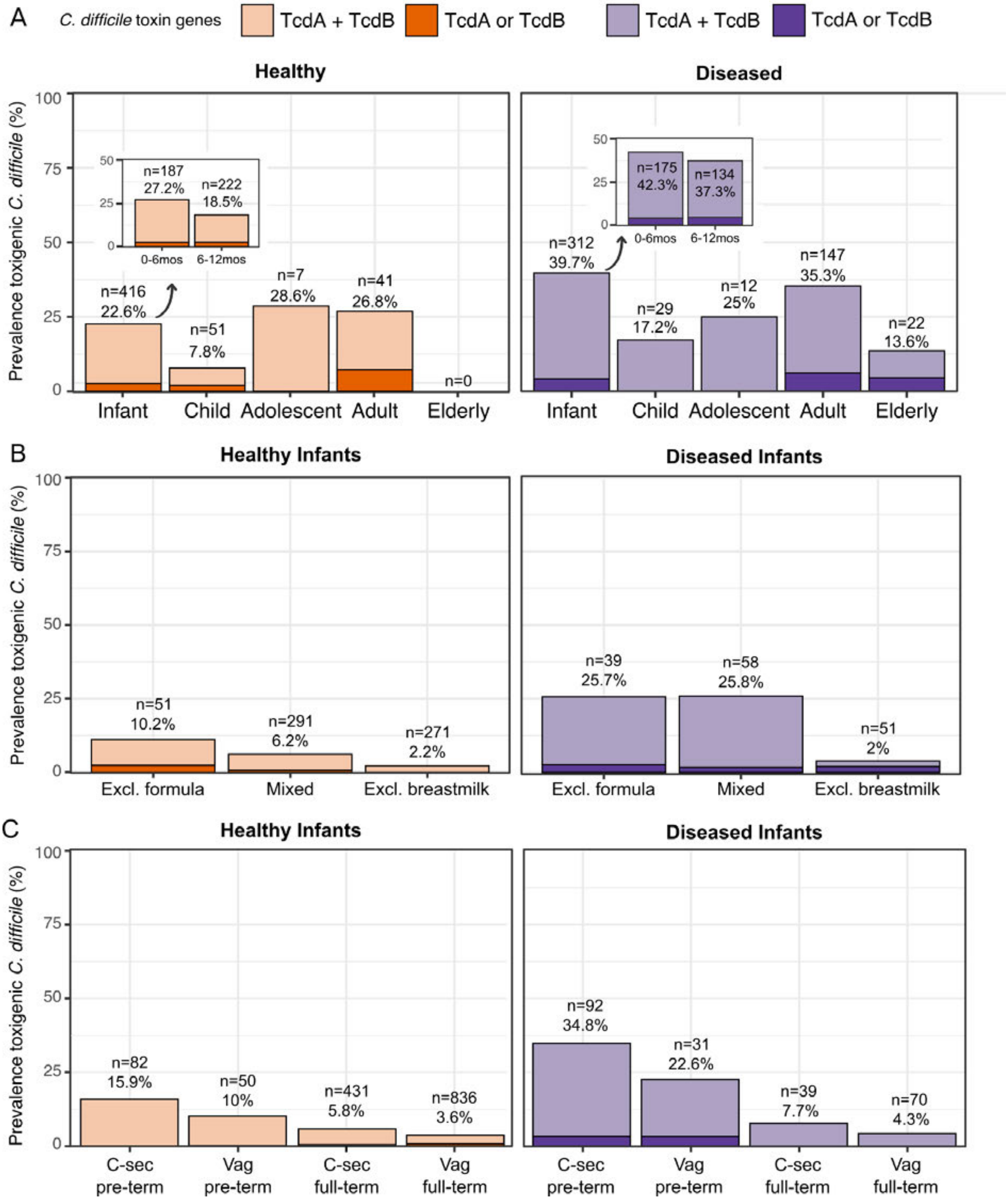


Figure 27. Prevalence of toxigenic *C. difficile* in the stools of (A) healthy and diseased across age groups, (B) diet during the first year of life, (C) delivery mode and gestational age. Toxigenic potential defined via detection of either one or both of *C. difficile* Toxin A (TcdA) or Toxin B (TcdB) genes. Prevalence is calculated on the number of *C. difficile* positive samples. *C. difficile* toxin genes were exclusively found among human stool samples.

The similarity in toxin burden between these two age groups is in contrast with the stark discrepancy observed in terms of *C. difficile* prevalence between infants and adults (**Figure 15**). Furthermore, while the average *C. difficile* prevalence was higher in the second semester of life compared to the first one, the *C. difficile* toxins carriage appears to be higher during the first semester, independently of the health status (+8.7% in healthy infants and +5% in diseased infants) (**Figure 27A**). The differential composition at the species level identified between healthy infants and subjects from different age groups and status, discussed in **Figure 27A**, was not influenced by *C. difficile* toxin genes presence (data not shown).

Healthy infants exclusively fed with formula milk were characterized by higher prevalence of potentially toxigenic *C. difficile* (10.2%), compared to infants fed a diet mixed of formula and breastmilk (6.2%) and infants exclusively fed with breast milk (2.2%) (**Figure 27B**). In diseased infants, while exclusive breast milk remained associated with the lowest *C. difficile* toxin burden, no significant difference was found when comparing exclusive formula with mixed feeding.

In terms of delivery mode and gestational age, the highest *C. difficile* toxin genes carriage rate was found in premature infants born via C-section, followed by preterms vaginally delivered, full-term babies born via C-section and full-terms born naturally (**Figure 27C**). This trend was observed in both healthy and diseased infants, suggesting that both gestational age and delivery mode have a higher impact on *C. difficile* toxin burden, than health status alone. This result is in line with what was previously described in **Figure 16C,** in terms of *C. difficile* prevalence across delivery mode and gestational age.

The parameters used for our toxin detection approach were conservative (see Methods in **Chapter 2**), to reduce at the minimum the possibility of false positives. However the presence of false negatives is possible, and therefore I suggest to consider these estimations on potentially toxigenic *C. difficile* carriage as lower limit estimations.

C. difficile was not the only species harbouring toxin genes found in healthy subjects: *S. flexneri* (16.2%% and 6.8% in infants and adults, respectively), *E. coli* (14.4% and 8.7%) and *C. perfringens* (13.2% and 1.6%) were found in both infants and adults at higher prevalence than *C. difficile* (**Figure 28**). Diseased subjects were characterized by a significant increase in carriage rate of potentially toxigenic species, compared to healthy subjects (**Figure 28**).

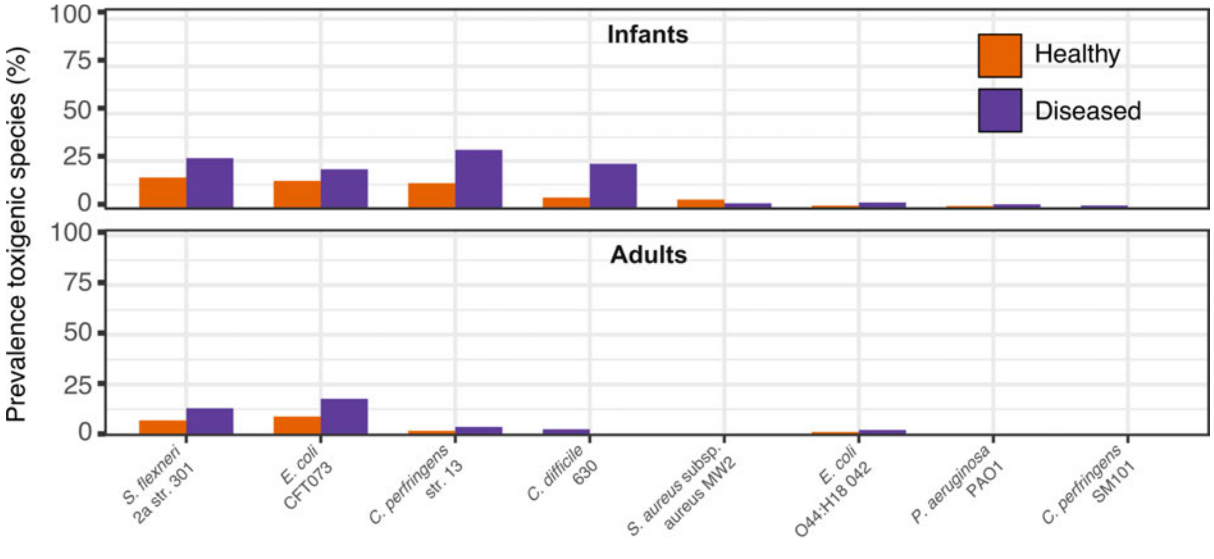


Figure 28. Prevalence of potentially toxigenic species other than *C. difficile*, in infants and adults, divided by health status. Toxigenic potential identified via detection of species-specific toxin genes published in VFDB (B. Liu et al. 2019) (see **Chapter 2** for detailed Methods description).

Known high prevalence of asymptomatic carriage of toxigenic *C. difficile* in infants is the main reason why *C. difficile* testing is generally discouraged by pediatric guidelines and why, even when in presence of diarrheal symptoms, the investigation of alternative etiologies is recommended (Schutze et al. 2013; McDonald et al. 2018).

While in our study was limited at the investigation of the toxins genes, *C. difficile* toxins are not uncommon in healthy infants (Viscidi, Willey, and Bartlett 1981; Kubota et al. 2016), and the apparent protection from CDI symptomatology has been so far ascribed to the lack of toxin receptors in the immature infant gut within the first two years of life. However, this hypothesis originated from a study from 1986 that showed no toxin uptake in the intestinal cells obtained from two fetuses aborted in the second trimester (Chang, Sullivan, and Wilkins 1986). Following studies on piglets and young rabbits came to contrasting results (Keel and Songer 2007; Eglow et al. 1992) and human-based validation of this claim is still missing. Therefore, while it is possible that the infant gut immaturity might play a role in the apparent protection of healthy infants from *C. difficile* infection, this should not prevent evaluation of alternative or complementary mechanisms.

Our analysis showed that no subspecies structure could be identified when comparing *C. difficile* found in healthy infants with *C. difficile* from diseased infants or older subjects. This, together with the presence of potentially toxigenic *C. difficile* in healthy infants indicates that presence of this species and its toxigenic potential are not enough to distinguish between asymptomatic and symptomatic subjects. Furthermore, multiple studies found that also *C. difficile* toxin concentration may not be a reliable proxy to infer disease presence and its severity (Jackson et al. 2016; Tang, Urrunaga, and von Rosenvinge 2016; Polage et al. 2015; Toltzis et al. 2012). In this context, the unique composition of the microbial community I found associated with *C. difficile* in healthy infants suggests that the gut microbiome might have an important and previously overlooked role in *C. difficile*-associated symptomatology.

The metagenomic approach, on which this study was based, offers unique advantages, such as the ability to provide a snapshot of the whole microbial community, with little to no a priori knowledge, allowing detection of a vast range of known or suspected pathogenic species that might co-occur with *C. difficile*. Another important advantage is that metagenomics has been extensively adoperated to investigate a wide range of age ranges, medical conditions, lifestyles, hosts and environments, which are publicly available and offer invaluable insight on the global characterization of *C. difficile*.

3.4 Study limitations and future perspectives

Our meta-analysis has several limitations associated with the use of metagenomics. First, the identification of toxin genes confirms the pathogenicity potential of a bacterial species, but cannot provide information as to whether the toxins are in fact produced, nor in which concentration (Polage et al. 2015). Second, our results refer to *C. difficile* in its vegetative form, and not in its endospore form, since the efficiency of many DNA extraction protocols from endospores is limited (Felczykowska et al. 2015). Third, metagenomic-based results can be affected by sequencing depth and therefore not detect species present at very low relative abundance ($<10^{-5}$), although we confirmed that sequencing depth was not associated with *C. difficile* detection in our dataset. Fourth, DNA-based detection of *C. difficile* could also derive from non viable cells, although the consistent detection of *C. difficile* over numerous timeseries suggests that *C. difficile* was continuously present over time and therefore likely indicative of viable cells. Finally, while our results indicated important associations between *C. difficile* and certain groups of other bacterial species, as well as multiple indicators of a desirable infant gut microbiome development, they do not prove causal relationships. Therefore, whether and how *C. difficile* might be directly involved in the maturation of the healthy infant gut microbiome remains the object of future studies.

While this study helped elucidate previously unknown aspects of *C. difficile* as a member of the gut microbiome, and its associations with other microbial species, several aspects require further investigation. In future studies, the combination of traditional laboratory-based tests with metagenomics will likely provide a better insight into the interpretation of *C. difficile* role in the gut microbiome across different life stages and health status. Nutrient availability, neighbouring bacterial species and stress level, which vary over lifetime, are able to profoundly influence *C. difficile* and its potential associated health outcome, and therefore need to be carefully considered. Further study is also required in the unbiased characterization of the potential roles of *C. difficile* toxins, beyond their capability to cause host damage. In fact, toxins have been suggested to be able to modulate niche establishment, motility and biofilm formation (Rudkin et al. 2017).

Chapter 4. Concluding remarks

Since its discovery almost a century ago, *C. difficile* has been extensively studied, with more than 18 thousand publications available on PubMed as of today. However, this imponent effort has been almost exclusively focused on *C. difficile* in diseased cohorts, in particular those affected by *C. difficile* infection, with only a small fraction focusing on *C. difficile* in healthy individuals. Low prevalence of *C. difficile* among healthy adults and the low prevalence of symptomatic carriage of toxigenic *C. difficile* in infants are among the main reasons for this biased investigative effort.

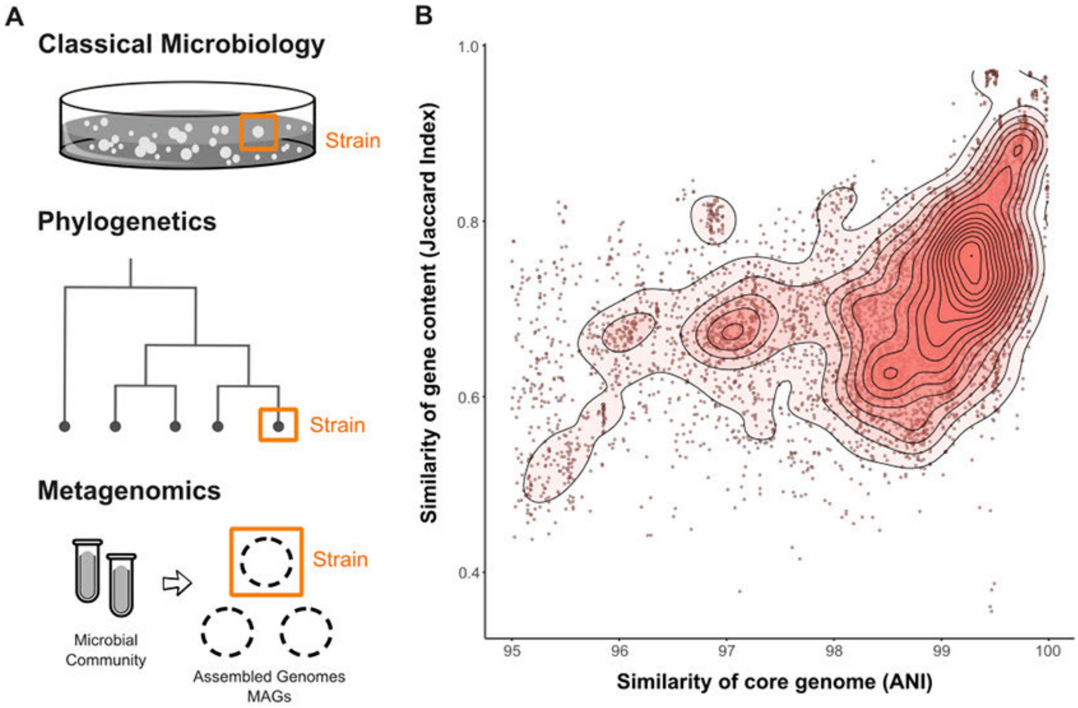
While more studies are needed to elucidate the precise role of *C. difficile* in human health, I here found that the definition of *C. difficile* as a pathogen falls short in multiple ways. I showed that *C. difficile* prevalence, abundance and biotic context varied over lifetime. While associated with a less desirable community composition in adults and elderly, *C. difficile* in healthy infants was associated with multiple beneficial indicators of a healthy gut microbiome development.

Such dichotomy is less typical of a pathogen, and more suited for an opportunistic pathogen or a pathobiont. While the distinction between these two terms remains blurry (Jochum and Stecher 2020), decoupling *C. difficile* from the purely pathogenic perspective is of paramount importance. Quoting the famous novelist Toni Morrison, "definitions belong to the definers, not the defined". How we label a species, in this case *C. difficile*, does not change the nature of the species itself, nor its interaction with the host, but it can affect our perception of the species and potentially how we decide to investigate it.

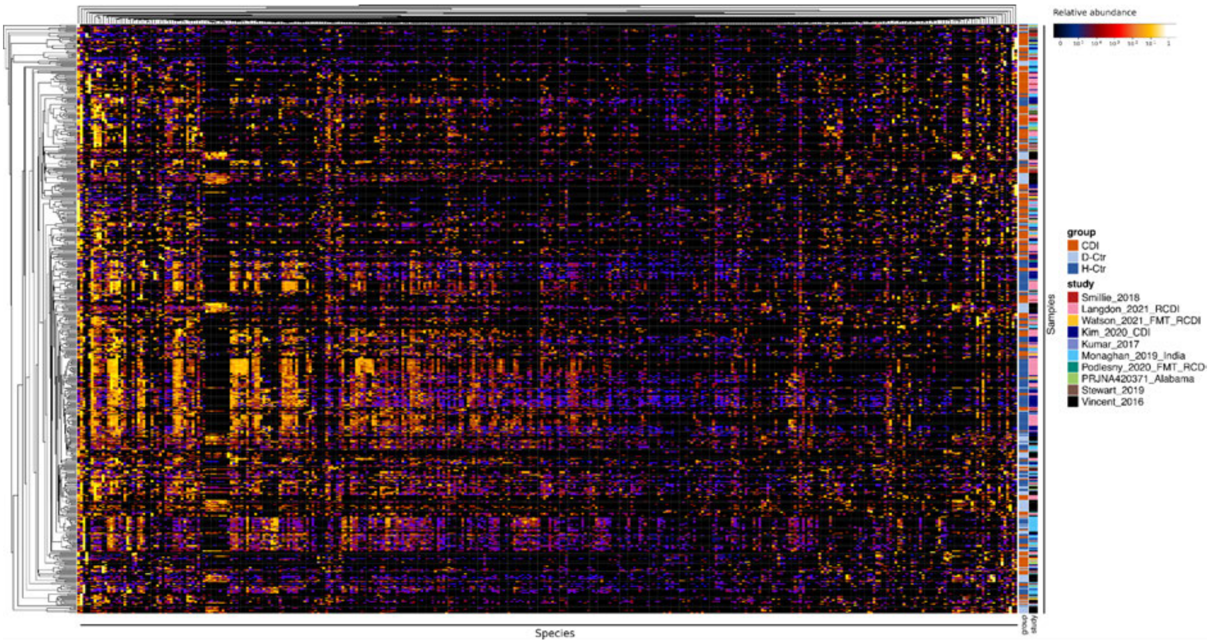
I therefore advocate for a shift from the so-far-adopted disease-centred investigation of *C. difficile,* to a more neutral species-centred investigation. Simultaneously, my results raise concerns for potential CDI overdiagnosis, and suggest that CDI-like symptomatology could be attributed to other, so far overlooked, pathogenic species. Thus, as it holds the potential to be associated with beneficial traits as well as with disease, *C. difficile* classification as pathogen needs to be reconsidered, and expanded beyond the disease-centred perspective.

Appendix

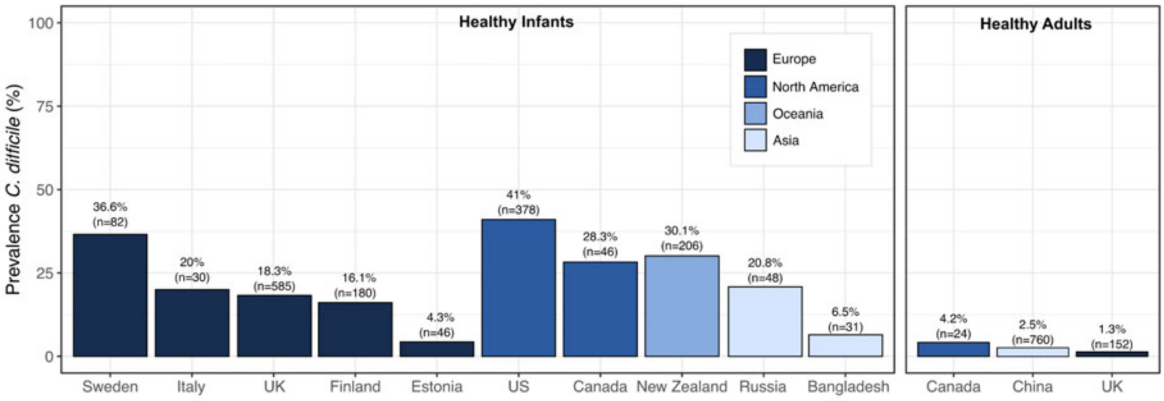
Supplementary figure 1. (A) Overview of the different operational definitions of “strain” in three fields of investigation: classical microbiology (as single cultured isolate), phylogenetics (a leaf of a phylogenetic tree) and metagenomics (a metagenomic assembled genome, MAG). (B) Correlation between similarity of the core genome (measure via ANI) and similarity of the gene content (measure via Jaccard Index) in conspecific isolate genomes from 155 bacterial species. Each dot is a pairwise comparison of one isolate genome against all other conspecific ones. While a correlation exists (Spearman correlation R=0.57, p < 2.2e-16), many genomes with core genome ANI >99% have <80% gene content in common, with isolated cases below 50%. Data from (Maistrenko et al. 2020), analysis performed by Dr.Oleksandr Maistrenko. Figure and caption adapted from (Van Rossum et al. 2020).



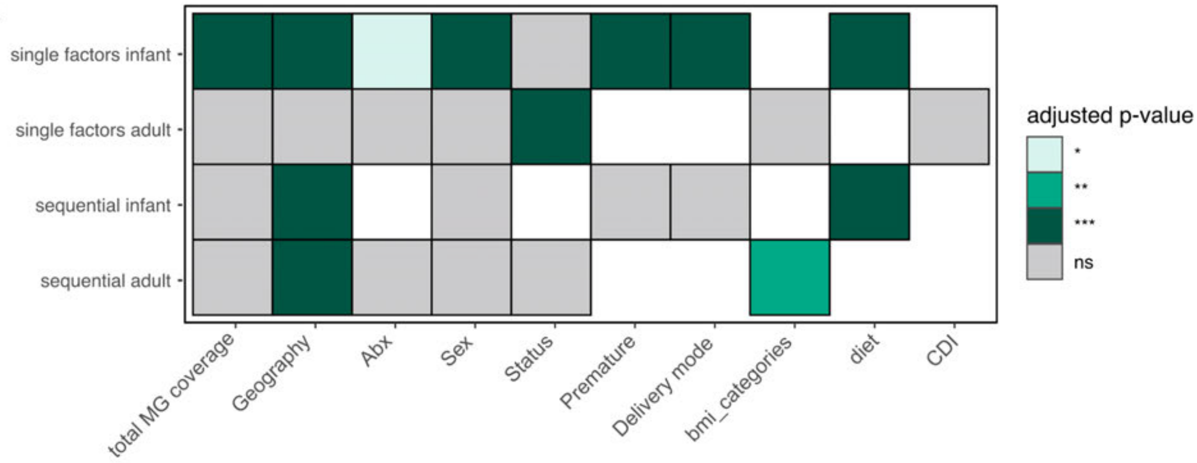
Supplementary figure 2. Taxonomic composition at the species level for CDI and control samples. Species with minimum relative abundance ≥ 0.01 and prevalence ≥ 0.1 are shown. Studies with at least 3 samples per study are shown.



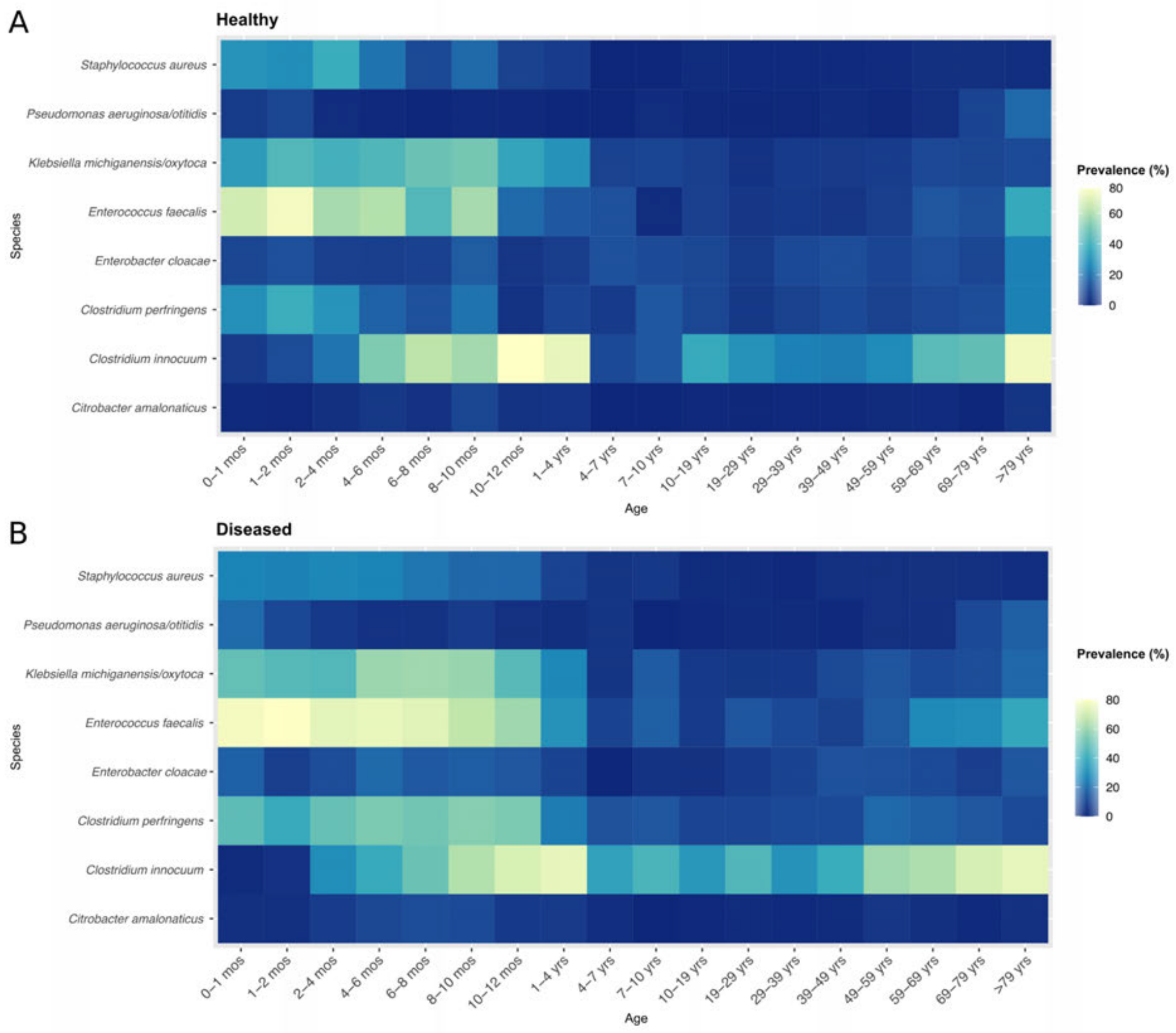
Supplementary figure 3. Geographical distribution of *C. difficile* prevalence in the stools of healthy infants (left) and adults (right) divided by country and continent. For healthy adults, only countries with *C. difficile* prevalence above 1% are shown.



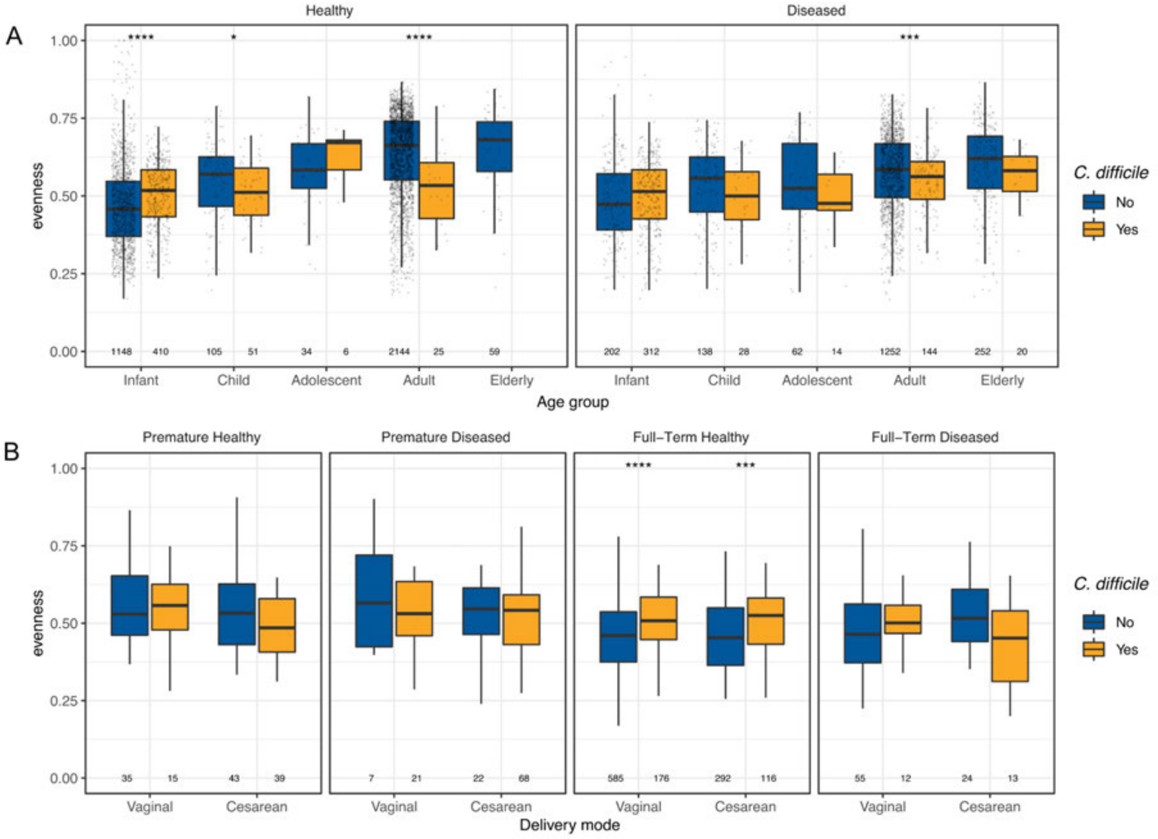
Supplementary figure 4. Logistic regression adjusted ANOVA p-values, divided by age group. Single models are indicated as “single factor”, while combined models are indicated as “sequential”. The order of factors used in the combined models is the same as the column order in the plot. “Total MG coverage” is used as a proxy for the sequencing depth.



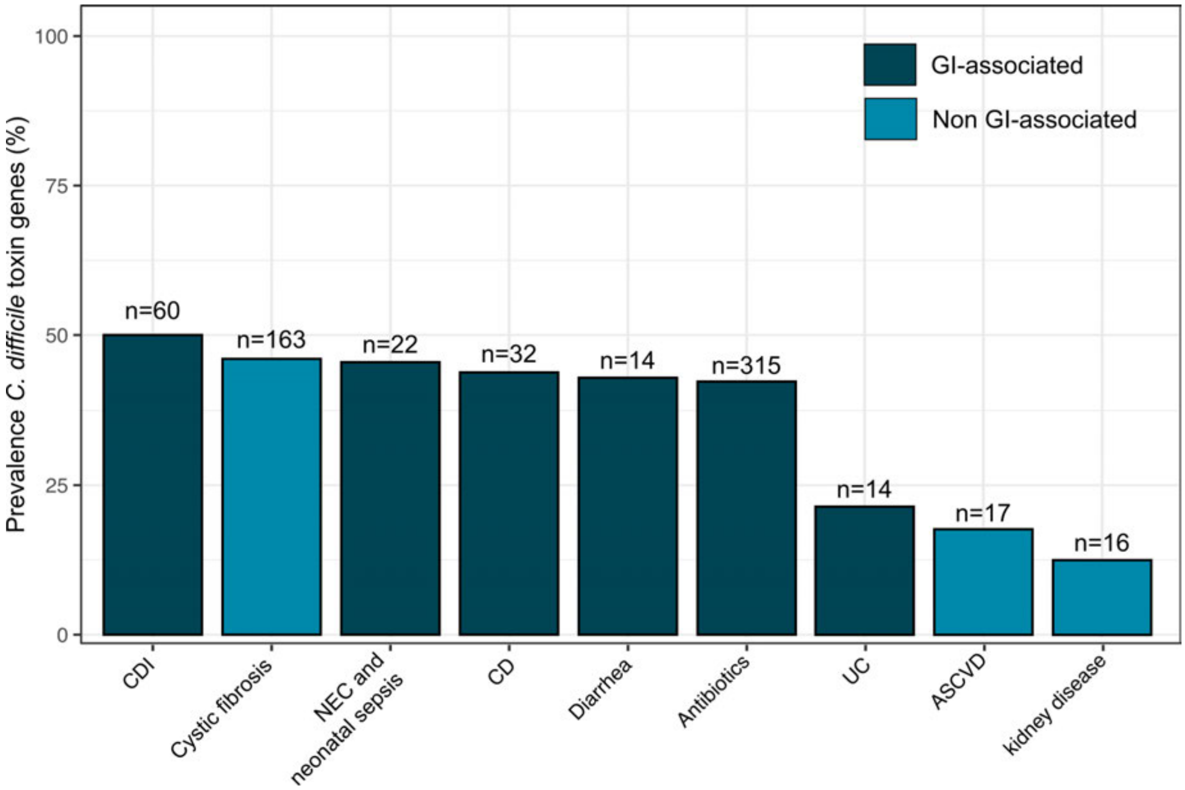
Supplementary figure 5. Prevalence of species known from the literature to be able to induce antibiotic-associated diarrhea (AAD), across (A) healthy and (B) diseased human gut stool samples over lifetime.



Supplementary figure 6. (A) Species evenness (as the inverse Simpson index divided by Richness) across age groups and health status. (B) Evenness in infant samples divided by delivery mode, prematurity and health status. The number of samples per group is shown under each boxplot. Mean comparison p-values calculated using t-test.



Supplementary figure 7. Prevalence of *C. difficile* toxin genes in the stools of diseased patients, divided by disease type or drugs. In dark blue are highlighted the diseases or drugs directly affecting the gastrointestinal (GI) tract, in light blue the ones affecting other body sites. Prevalence calculated on the per-disease total number of *C. difficile* positive samples. Only diseases with more than 10 *C. difficile* positive samples are shown. Abbreviations: NEC: “Necrotizing enterocolitis”; CD: “Crohn’s disease”; ASCVD: “atherosclerotic cardiovascular disease”; UC: “Ulcerative colitis”.



Supplementary table 1. List of publicly available CDI studies used for the *C. difficile* CDI meta-analysis. Abbreviations: “CDI”: *C. difficile* infection; “D-Ctr”: Diseased controls; “H-Ctr”: Healthy controls. See **Methods** for detailed description of sample groups.

Study accession on ENA	**Study name**	**N. samples**	**N. samples CDI**	**N. samples D-Ctr**	**N. samples H-Ctr**
PRJNA420371	PRJNA420371_Alabama	26	17	-	9
PRJNA339012	Kumar_2017	13	7	-	6
PRJNA564397	Monaghan_2020_India	93	32	22	39
PRJEB23489	Smillie_2018	26	22	-	4
PRJNA478949	Stewart_2019	24	10	-	14
SRP064400	Vincent_2016	96	4	92	-
PRJEB39023	Podlesny_2020_FMT	16	8	-	8
PRJEB33013; PRJEB35738	Kim_2020_CDI	87	26	-	61
PRJNA674880	Langdon_2021_RCDI	141	98	-	43
PRJNA701961	Watson_2021_FMT	12	10	-	2
	Total	**534**	**234**	**114**	**186**

Supplementary Table 2. Most prevalent toxin genes found in CDI samples (in order of prevalence).

Toxin gene	Toxin extended name	Species
senB	enterotoxin ShET2	*S. flexneri* 2a str. 301
sat	Secreted auto transporter toxin Sat	*E. coli* CFT073
toxA	toxin A TcdA	*C. difficile* 630
toxB	toxin B TcdB	*C. difficile* 630
nagH	hyaluronidase mu-toxin	*C. perfringens* str. 13
esaD	type VII secretion system secreted protein	*S. aureus* subsp. aureus MW2
nagK	hyaluronidase mu-toxin	*C. perfringens* str. 13
colA	collagenase kappa-toxin	*C. perfringens* str. 13
nagJ	hyaluronidase mu-toxin	*C. perfringens* str. 13
plc	phospholipase C alpha-toxin	*C. perfringens* str. 13
nagI	hyaluronidase mu-toxin	*C. perfringens* str. 13
lukD	leukotoxin leukocidin	*S. aureus* subsp. aureus MW2
nagL	hyaluronidase mu-toxin	*C. perfringens* str. 13
pfoA	perfringolysin O theta-toxin/PFO	*C. perfringens* str. 13
astA	exotoxin A precursor ExoA	*E. coli* O44:H18 042
toxA	heat-stable enterotoxin 1 EAST1	*P. aeruginosa* PAO1
tsst-1	toxic shock syndrome toxin-1 TSST-1	*S. aureus* subsp. aureus N315
cpe	enterotoxin Cpe CPE	*C. perfringens* SM101
sell	staphylococcal enterotoxin L precursor Enterotoxin-like L	*S. aureus* subsp. aureus MW2
sec	staphylococcal enterotoxin C precursor Enterotoxin C	*S. aureus* subsp. aureus MW2
esaG9	antitoxin protein EsaG homolog Type VII secretion system	*S. aureus* subsp. aureus MW2
cdtB	RTX toxin RtxA	*C. jejuni* subsp. jejuni NCTC 11168
cdtA	cytolethal distending toxin B	*C. jejuni* subsp. jejuni NCTC 11168

esaG1	cytolethal distending toxin A	*S. aureus* subsp. aureus MW2
esaG2	staphylococcal enterotoxin D precursor SE	*S. aureus* subsp. aureus MW2
sed	antitoxin protein EsaG Type VII secretion system	*S. aureus* RN4220
rtxA	antitoxin protein EsaG homolog Type VII secretion system	*V. cholerae* O1 biovar El Tor str. N16961

Supplementary table 3. List of the 253 publicly available studies used in the *C. difficile* global meta-analysis.

Supplementary table 4. List of animal species included in our global *C. difficile* survey, listing their scientific and common names, the number of samples per each species and the country of sampling.

Animal species scientific name	Animal species common name	N. samples	Country
Sus scrofa domesticus	Domestic Pig	1708	Netherlands (185), Denmark (220), China (295), Canada (67), Australia (902), Austria (939)
Gallus gallus domesticus	Domestic chicken	752	Netherlands (178), China (499), UK (74), Peru (1)
Bos taurus	Domestic cow	695	Canada (60), US (215), UK (281), Finland (13), Sweden (13), Iceland (6), Brazil (23), France (61), Italy (19), El Salvador (4)
Mus musculus	Mouse	213	Denmark (22), China (15), US (123), Sweden (20), Norway (33)
Canis lupus familiaris	Domestic dog	184	US (182), Peru (2)
Felis catus	Domestic cat	124	US (36), UK (88)
Apis mellifera carnica	Bee	54	Switzerland
Papio cynocephalus	Yellow baboon	48	Kenya
Coragyps atratus	Black vulture	45	na
Gymnogyps californianus	California condor	31	US
Macaca mulatta	Rhesus macaque	29	US
Elaphurus davidianus	Père David's deer	27	China
Gorilla gorilla gorilla	Gorilla	23	Cameroon
Macaca fascicularis	Crab-eating macaque	20	China
Ailuropoda melanoleuca	Giant panda	16	China
na	Oyster	15	US
Oncorhynchus mykiss	Rainbow trout	13	US
Ailurus fulgens	Red panda	6	China
Strigops habroptilus	Kakapo	4	New Zealand

Manis javanica	Sunda pangolin	2	China
Anser indicus	Bar-headed goose	2	China
Cavia porcellus	Guinea pig	2	Peru
Centrocercus urophasianus	Greater sage-grouse	2	US
na	Turtle	1	Peru
Indri indri	Indri lemur	1	Madagascar

Supplementary table 5. Metadata table for the 42,814 metagenomic samples used for the *C. difficile* global meta-analysis.

Supplementary table 6. mOTU2.0 taxonomic profiles (using 2 marker genes) for the 42,814 metagenomic samples used for the *C. difficile* global meta-analysis.

Supplementary tables 3, 5 and 6 are provided in the online version of the thesis.

References

Abt, Michael C., Peter T. McKenney, and Eric G. Pamer. 2016. "Clostridium Difficile Colitis: Pathogenesis and Host Defence." *Nature Reviews. Microbiology* 14 (10): 609–20.

Alasmari, Faisal, Sondra M. Seiler, Tiffany Hink, Carey-Ann D. Burnham, and Erik R. Dubberke. 2014. "Prevalence and Risk Factors for Asymptomatic Clostridium Difficile Carriage." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 59 (2): 216–22.

Allen, S. J., K. Wareham, D. Wang, C. Bradley, B. Sewell, H. Hutchings, W. Harris, et al. 2013. *Criteria for Severity of Clostridium Difficile Infection*. NIHR Journals Library.

Andreani, Nadia Andrea, Elze Hesse, and Michiel Vos. 2017. "Prokaryote Genome Fluidity Is Dependent on Effective Population Size." *The ISME Journal* 11 (7): 1719–21.

Antharam, Vijay C., Eric C. Li, Arif Ishmael, Anuj Sharma, Volker Mai, Kenneth H. Rand, and Gary P. Wang. 2013. "Intestinal Dysbiosis and Depletion of Butyrogenic Bacteria in Clostridium Difficile Infection and Nosocomial Diarrhea." *Journal of Clinical Microbiology* 51 (9): 2884–92.

Argamany, Jacqueline R., Samuel L. Aitken, Grace C. Lee, Natalie K. Boyd, and Kelly R. Reveles. 2015. "Regional and Seasonal Variation in Clostridium Difficile Infections among Hospitalized Patients in the United States, 2001-2010." *American Journal of Infection Control* 43 (5): 435–40.

Aviello, G., and U. G. Knaus. 2017. "ROS in Gastrointestinal Inflammation: Rescue Or Sabotage?" *British Journal of Pharmacology* 174 (12): 1704–18.

Bäckhed, Fredrik, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, et al. 2015. "Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life." *Cell Host & Microbe* 17 (6): 852.

Bagdasarian, Natasha, Krishna Rao, and Preeti N. Malani. 2015. "Diagnosis and Treatment of Clostridium Difficile in Adults: A Systematic Review." *JAMA: The Journal of the American Medical Association* 313 (4): 398–408.

Bapteste, Eric, Maureen A. O'Malley, Robert G. Beiko, Marc Ereshefsky, J. Peter Gogarten, Laura Franklin-Hall, François-Joseph Lapointe, et al. 2009. "Prokaryotic Evolution and the Tree of Life Are Two Different Things." *Biology Direct*. https://doi.org/10.1186/1745-6150-4-34.

Bartlett, J. G., N. Moon, T. W. Chang, N. Taylor, and A. B. Onderdonk. 1978. "Role of Clostridium Difficile in Antibiotic-Associated Pseudomembranous Colitis." *Gastroenterology* 75 (5): 778–82.

Bartlett, John G., and Dale N. Gerding. 2008. "Clinical Recognition and Diagnosis of Clostridium Difficile Infection." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 46 Suppl 1 (January): S12–18.

Bauer, M. P., A. Farid, M. Bakker, R. A. S. Hoek, E. J. Kuijper, and J. T. van Dissel. 2014. "Patients with Cystic Fibrosis Have a High Carriage Rate of Non-Toxigenic Clostridium Difficile." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 20 (7): O446–49.

Berkell, Matilda, Mohamed Mysara, Basil Britto Xavier, Cornelis H. van Werkhoven, Pieter Monsieurs, Christine Lammens, Annie Ducher, et al. 2021. "Microbiota-Based Markers Predictive of Development of Clostridioides Difficile Infection." *Nature Communications* 12 (1): 2241.

Blaser, M. J., G. I. Perez-Perez, H. Kleanthous, T. L. Cover, R. M. Peek, P. H. Chyou, G. N. Stemmermann, and A. Nomura. 1995. "Infection with Helicobacter Pylori Strains Possessing cagA Is Associated with an Increased Risk of Developing Adenocarcinoma of the Stomach." *Cancer Research* 55 (10): 2111–15.

Bobay, Louis-Marie, and Howard Ochman. 2017. "Biological Species Are Universal across Life's Domains." *Genome Biology and Evolution*, February. https://doi.org/10.1093/gbe/evx026.

———. 2018. "Factors Driving Effective Population Size and Pan-Genome Evolution in Bacteria." *BMC Evolutionary Biology* 18 (1): 153.

Bridgman, Sarah L., Meghan B. Azad, Catherine J. Field, Andrea M. Haqq, Allan B. Becker, Piushkumar J. Mandhane, Padmaja Subbarao, et al. 2017. "Fecal Short-Chain Fatty Acid

Variations by Breastfeeding Status in Infants at 4 Months: Differences in Relative versus Absolute Concentrations.” *Frontiers in Nutrition* 4 (April): 11.

Brouwer, Michael S. M., Peter Mullany, Elaine Allan, and Adam P. Roberts. 2016. “Investigating Transfer of Large Chromosomal Regions Containing the Pathogenicity Locus Between Clostridium Difficile Strains.” *Methods in Molecular Biology* 1476: 215–22.

Brouwer, Michael S. M., Adam P. Roberts, Haitham Hussain, Rachel J. Williams, Elaine Allan, and Peter Mullany. 2013. “Horizontal Gene Transfer Converts Non-Toxigenic Clostridium Difficile Strains into Toxin Producers.” *Nature Communications* 4: 2601.

Brown, Kevin Antoine, Laura K. MacDougall, Kim Valenta, Andrew Simor, Jennie Johnstone, Samira Mubareka, George Broukhanski, Gary Garber, Allison McGeer, and Nick Daneman. 2018. “Increased Environmental Sample Area and Recovery of Clostridium Difficile Spores from Hospital Surfaces by Quantitative PCR and Enrichment Culture.” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 39 (8): 917–23.

Burke, D. G., M. J. Harrison, C. Fleming, M. McCarthy, C. Shortt, I. Sulaiman, D. M. Murphy, et al. 2017. “Clostridium Difficile Carriage in Adult Cystic Fibrosis (CF); Implications for Patients with CF and the Potential for Transmission of Nosocomial Infection.” *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* 16 (2): 291–98.

Burke, Kristin E., and J. Thomas Lamont. 2014. “Clostridium Difficile Infection: A Worldwide Disease.” *Gut and Liver* 8 (1): 1–6.

Caro-Quintero, Alejandro, and Konstantinos T. Konstantinidis. 2012. “Bacterial Species May Exist, Metagenomics Reveal.” *Environmental Microbiology* 14 (2): 347–55.

“Carriage of Clostridium Difficile by Wild Urban Norway Rats (Rattus Norvegicus) and Black Rats (Rattus Rattus).” n.d. Accessed July 17, 2021. https://journals.asm.org/doi/abs/10.1128/AEM.03609-13.

Chang, T. W., N. M. Sullivan, and T. D. Wilkins. 1986. “Insusceptibility of Fetal Intestinal Mucosa and Fetal Cells to Clostridium Difficile Toxins.” *Zhongguo Yao Li Xue Bao = Acta Pharmacologica Sinica* 7 (5): 448–53.

Charlesworth, Brian. 2009. “Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation.” *Nature Reviews. Genetics* 10 (3): 195–205.

Chaun, H. 2001. “Colonic Disorders in Adult Cystic Fibrosis.” *Canadian Journal of Gastroenterology = Journal Canadien de Gastroenterologie* 15 (9): 586–90.

Chen, Shuyi, Chunli Sun, Haiying Wang, and Jufang Wang. 2015. “The Role of Rho GTPases in Toxicity of Clostridium Difficile Toxins.” *Toxins* 7 (12): 5254–67.

Chernikova, Diana A., Juliette C. Madan, Molly L. Housman, Muhammad Zain-Ul-Abideen, Sara N. Lundgren, Hilary G. Morrison, Mitchell L. Sogin, et al. 2018. “The Premature Infant Gut Microbiome during the First 6 Weeks of Life Differs Based on Gestational Maturity at Birth.” *Pediatric Research* 84 (1): 71–79.

Chia, J-H, Y. Feng, L-H Su, T-L Wu, C-L Chen, Y-H Liang, and C-H Chiu. 2017. “Clostridium Innocuum Is a Significant Vancomycin-Resistant Pathogen for Extraintestinal Clostridial Infection.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 23 (8): 560–66.

Chia, J-H, T-S Wu, T-L Wu, C-L Chen, C-H Chuang, L-H Su, H-J Chang, et al. 2018. “Clostridium Innocuum Is a Vancomycin-Resistant Pathogen That May Cause Antibiotic-Associated Diarrhoea.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 24 (11): 1195–99.

Chu, Derrick M., Jun Ma, Amanda L. Prince, Kathleen M. Antony, Maxim D. Seferovic, and Kjersti M. Aagaard. 2017. “Maturation of the Infant Microbiome Community Structure and Function across Multiple Body Sites and in Relation to Mode of Delivery.” *Nature Medicine* 23 (3): 314–26.

Claro, Tânia, Stephen Daniels, and Hilary Humphreys. 2014. “Detecting Clostridium Difficile Spores from Inanimate Surfaces of the Hospital Environment: Which Method Is Best?” *Journal of Clinical Microbiology* 52 (9): 3426–28.

Coelho, Luis Pedro, Renato Alves, Paulo Monteiro, Jaime Huerta-Cepas, Ana Teresa Freitas, and

Peer Bork. 2019. “NG-Meta-Profiler: Fast Processing of Metagenomes Using NGLess, a Domain-Specific Language.” *Microbiome* 7 (1): 84.

Cornick, Steve, Adelaide Tawiah, and Kris Chadee. 2015. “Roles and Regulation of the Mucus Barrier in the Gut.” *Tissue Barriers* 3 (1-2): e982426.

Costea, Paul I., Luis Pedro Coelho, Shinichi Sunagawa, Robin Munch, Jaime Huerta-Cepas, Kristoffer Forslund, Falk Hildebrand, Almagul Kushugulova, Georg Zeller, and Peer Bork. 2017. “Subspecies in the Global Human Gut Microbiome.” *Molecular Systems Biology* 13 (12): 960.

Costea, Paul Igor, Robin Munch, Luis Pedro Coelho, Lucas Paoli, Shinichi Sunagawa, and Peer Bork. 2017. “metaSNV: A Tool for Metagenomic Strain Level Analysis.” *PloS One* 12 (7): e0182392.

Crobach, M. J. T., T. Planche, C. Eckert, F. Barbut, E. M. Terveer, O. M. Dekkers, M. H. Wilcox, and E. J. Kuijper. 2016. “European Society of Clinical Microbiology and Infectious Diseases: Update of the Diagnostic Guidance Document for Clostridium Difficile Infection.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 22 Suppl 4 (August): S63–81.

Crobach, Monique J. T., Amoe Baktash, Nikolas Duszenko, and Ed J. Kuijper. 2018. “Diagnostic Guidance for C. Difficile Infections.” *Advances in Experimental Medicine and Biology* 1050: 27–44.

Cuevas-Ramos, Gabriel, Claude R. Petit, Ingrid Marcq, Michèle Boury, Eric Oswald, and Jean-Philippe Nougayrède. 2010. “Escherichia Coli Induces DNA Damage in Vivo and Triggers Genomic Instability in Mammalian Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (25): 11537–42.

Dai, Wenkui, Heping Wang, Qian Zhou, Dongfang Li, Xin Feng, Zhenyu Yang, Wenjian Wang, et al. 2019. “An Integrated Respiratory Microbial Gene Catalogue to Better Understand the Microbial Aetiology of Mycoplasma Pneumoniae Pneumonia.” *GigaScience* 8 (8). https://doi.org/10.1093/gigascience/giz093.

Daniel, Noëmie, Emelyne Lécuyer, and Benoit Chassaing. 2021. “Host/microbiota Interactions in Health and Diseases-Time for Mucosal Microbiology!” *Mucosal Immunology*, March. https://doi.org/10.1038/s41385-021-00383-w.

Daquigan, Ninalynn, Anna Maria Seekatz, K. Leigh Greathouse, Vincent B. Young, and James Robert White. 2017. “High-Resolution Profiling of the Gut Microbiome Reveals the Extent of Clostridium Difficile Burden.” *NPJ Biofilms and Microbiomes* 3 (December): 35.

Deakin, Laura J., Simon Clare, Robert P. Fagan, Lisa F. Dawson, Derek J. Pickard, Michael R. West, Brendan W. Wren, Neil F. Fairweather, Gordon Dougan, and Trevor D. Lawley. 2012. “The Clostridium Difficile spo0A Gene Is a Persistence and Transmission Factor.” *Infection and Immunity* 80 (8): 2704–11.

Deane, Jennifer, Fiona Fouhy, Nicola J. Ronan, Mary Daly, Claire Fleming, Joseph A. Eustace, Fergus Shanahan, et al. 2021. “A Multicentre Analysis of Clostridium Difficile in Persons with Cystic Fibrosis Demonstrates That Carriage May Be Transient and Highly Variable with Respect to Strain and Level.” *The Journal of Infection* 82 (3): 363–70.

Devaux, Christian Albert, Matthieu Million, and Didier Raoult. 2020. “The Butyrogenic and Lactic Bacteria of the Gut Microbiota Determine the Outcome of Allogenic Hematopoietic Cell Transplant.” *Frontiers in Microbiology* 11 (July): 1642.

Dharmasena, Muthu, and Xiuping Jiang. 2018. “Isolation of Toxigenic Clostridium Difficile from Animal Manure and Composts Being Used as Biological Soil Amendments.” *Applied and Environmental Microbiology* 84 (16). https://doi.org/10.1128/AEM.00738-18.

Diaz, Cristina Rodriguez, Christian Seyboldt, and Maja Rupnik. 2018. “Non-Human C. Difficile Reservoirs and Sources: Animals, Food, Environment.” *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/978-3-319-72799-8_13.

Donaldson, Gregory P., S. Melanie Lee, and Sarkis K. Mazmanian. 2016. “Gut Biogeography of the Bacterial Microbiota.” *Nature Reviews. Microbiology* 14 (1): 20–32.

Doolittle, W. Ford. 2012. “Population Genomics: How Bacterial Species Form and Why They Don’t Exist.” *Current Biology: CB* 22 (11): R451–53.

Drall, Kelsea M., Hein M. Tun, Nadia P. Morales-Lizcano, Theodore B. Konya, David S. Guttman, Catherine J. Field, Rupasri Mandal, et al. 2019. “Clostridioides Difficile Colonization Is

Differentially Associated With Gut Microbiome Profiles by Infant Feeding Modality at 3-4 Months of Age." *Frontiers in Immunology* 10 (December): 2866.

Eckert, C., A. Emirian, A. Le Monnier, L. Cathala, H. De Montclos, J. Goret, P. Berger, et al. 2015. "Prevalence and Pathogenicity of Binary Toxin-Positive Clostridium Difficile Strains That Do Not Produce Toxins A and B." *New Microbes and New Infections* 3 (January): 12–17.

Eglow, R., C. Pothoulakis, S. Itzkowitz, E. J. Israel, C. J. O'Keane, D. Gong, N. Gao, Y. L. Xu, W. A. Walker, and J. T. LaMont. 1992. "Diminished Clostridium Difficile Toxin A Sensitivity in Newborn Rabbit Ileum Is Associated with Decreased Toxin A Receptor." *The Journal of Clinical Investigation* 90 (3): 822–29.

Eisenstein, Michael. 2020. "The Hunt for a Healthy Microbiome." *Nature* 577 (7792): S6–8.

Engevik, Melinda A., Heather A. Danhof, Jennifer Auchtung, Bradley T. Endres, Wenly Ruan, Eugénie Bassères, Amy C. Engevik, et al. 2021. "Fusobacteriumnucleatum Adheres to Clostridioides Difficile via the RadD Adhesin to Enhance Biofilm Formation in Intestinal Mucus." *Gastroenterology* 160 (4): 1301–14.e8.

Engevik, Melinda Anne, Alexandra Chang-Graham, Joseph M. Hyser, and James A. Versalovic. 2020. "The Enteric Pathogen Clostridium Difficile Chemotaxes towards Mucin Glycans during the Early Colonization Phase of Infection." *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 34 (S1): 1–1.

Eyre, David W., David Griffiths, Alison Vaughan, Tanya Golubchik, Milind Acharya, Lily O'Connor, Derrick W. Crook, A. Sarah Walker, and Tim E. A. Peto. 2013. "Asymptomatic Clostridium Difficile Colonisation and Onward Transmission." *PloS One* 8 (11): e78445.

Faber, Franziska, Lisa Tran, Mariana X. Byndloss, Christopher A. Lopez, Eric M. Velazquez, Tobias Kerrinnes, Sean-Paul Nuccio, et al. 2016. "Host-Mediated Sugar Oxidation Promotes Post-Antibiotic Pathogen Expansion." *Nature* 534 (7609): 697–99.

Fachi, José Luís, Jaqueline de Souza Felipe, Laís Passariello Pral, Bruna Karadi da Silva, Renan Oliveira Corrêa, Mirella Cristiny Pereira de Andrade, Denise Morais da Fonseca, et al. 2019. "Butyrate Protects Mice from Clostridium Difficile-Induced Colitis through an HIF-1-Dependent Mechanism." *Cell Reports* 27 (3): 750–61.e7.

Felczykowska, Agnieszka, Anna Krajewska, Sylwia Zielińska, and Joanna M. Łoś. 2015. "Sampling, Metadata and DNA Extraction - Important Steps in Metagenomic Studies." *Acta Biochimica Polonica* 62 (1): 151–60.

Ferraris, Laurent, Jeanne Couturier, Catherine Eckert, Johanne Delannoy, Frédéric Barbut, Marie-José Butel, and Julio Aires. 2019. "Carriage and Colonization of C. Difficile in Preterm Neonates: A Longitudinal Prospective Study." *PloS One* 14 (2): e0212568.

Ferretti, Pamela, Edoardo Pasolli, Adrian Tett, Francesco Asnicar, Valentina Gorfer, Sabina Fedi, Federica Armanini, et al. 2018. "Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome." *Cell Host & Microbe* 24 (1): 133–45.e5.

Figueroa, Iris, Stuart Johnson, Susan P. Sambol, Ellie J. C. Goldstein, Diane M. Citron, and Dale N. Gerding. 2012. "Relapse versus Reinfection: Recurrent Clostridium Difficile Infection Following Treatment with Fidaxomicin or Vancomycin." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 55 Suppl 2 (August): S104–9.

Fletcher, Joshua R., Colleen M. Pike, Ruth J. Parsons, Alissa J. Rivera, Matthew H. Foley, Michael R. McLaren, Stephanie A. Montgomery, and Casey M. Theriot. 2021. "Clostridioides Difficile Exploits Toxin-Mediated Inflammation to Alter the Host Nutritional Landscape and Exclude Competitors from the Gut Microbiota." *Nature Communications* 12 (1): 462.

Fortier, Louis-Charles, and Ognjen Sekulovic. 2013. "Importance of Prophages to Evolution and Virulence of Bacterial Pathogens." *Virulence* 4 (5): 354–65.

Fraser, Christophe, William P. Hanage, and Brian G. Spratt. 2007. "Recombination and the Nature of Bacterial Speciation." *Science* 315 (5811): 476–80.

Freeman, J., and M. H. Wilcox. 2003. "The Effects of Storage Conditions on Viability of Clostridium Difficile Vegetative Cells and Spores and Toxin Activity in Human Faeces." *Journal of Clinical Pathology* 56 (2): 126–28.

Gabriel, L., and A. Beriot-Mathiot. 2014. "Hospitalization Stay and Costs Attributable to Clostridium Difficile Infection: A Critical Review." *The Journal of Hospital Infection* 88 (1): 12–21.

Garrido, Daniel, Daniela Barile, and David A. Mills. 2012. “A Molecular Basis for Bifidobacterial Enrichment in the Infant Gastrointestinal Tract.” *Advances in Nutrition* 3 (3): 415S – 21S.

Gateau, C., J. Couturier, J. Coia, and F. Barbut. 2018. “How to: Diagnose Infection Caused by Clostridium Difficile.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 24 (5): 463–68.

Gaudier, E., A. Jarry, H. M. Blottière, P. de Coppet, M. P. Buisine, J. P. Aubert, C. Laboisse, C. Cherbut, and C. Hoebler. 2004. “Butyrate Specifically Modulates MUC Gene Expression in Intestinal Epithelial Goblet Cells Deprived of Glucose.” *American Journal of Physiology. Gastrointestinal and Liver Physiology* 287 (6): G1168–74.

George, W. L., V. L. Sutter, D. Citron, and S. M. Finegold. 1979. “Selective and Differential Medium for Isolation of Clostridium Difficile.” *Journal of Clinical Microbiology* 9 (2): 214–19.

Gerding, Dale N., Stuart Johnson, Maja Rupnik, and Klaus Aktories. 2014. “Clostridium Difficile Binary Toxin CDT: Mechanism, Epidemiology, and Potential Clinical Importance.” *Gut Microbes* 5 (1): 15–27.

Gerhard, Ralf. 2017. “Receptors and Binding Structures for Clostridium Difficile Toxins A and B.” *Current Topics in Microbiology and Immunology* 406: 79–96.

Ghimire, Sudeep, Chayan Roy, Supapit Wongkuna, Linto Antony, Abhijit Maji, Mitchel Chan Keena, Andrew Foley, and Joy Scaria. 2020. “Identification of Clostridioides Difficile-Inhibiting Gut Commensals Using Culturomics, Phenotyping, and Combinatorial Community Assembly.” *mSystems* 5 (1). https://doi.org/10.1128/mSystems.00620-19.

Giloteaux, Ludovic, Julia K. Goodrich, William A. Walters, Susan M. Levine, Ruth E. Ley, and Maureen R. Hanson. 2016. “Reduced Diversity and Altered Composition of the Gut Microbiome in Individuals with Myalgic Encephalomyelitis/chronic Fatigue Syndrome.” *Microbiome* 4 (1): 30.

Goldenberg, Simon D., and Gary L. French. 2011. “Lack of Association of tcdC Type and Binary Toxin Status with Disease Severity and Outcome in Toxigenic Clostridium Difficile.” *The Journal of Infection* 62 (5): 355–62.

Golić, Nataša, Katarina Veljović, Nikola Popović, Jelena Djokić, Ivana Strahinić, Igor Mrvaljević, and Amarela Terzić-Vidojević. 2017. “In Vitro and in Vivo Antagonistic Activity of New Probiotic Culture against Clostridium Difficile and Clostridium Perfringens.” *BMC Microbiology* 17 (1): 108.

Goris, Johan, Konstantinos T. Konstantinidis, Joel A. Klappenbach, Tom Coenye, Peter Vandamme, and James M. Tiedje. 2007. “DNA-DNA Hybridization Values and Their Relationship to Whole-Genome Sequence Similarities.” *International Journal of Systematic and Evolutionary Microbiology* 57 (Pt 1): 81–91.

Green, R. H. 1974. “The Association of Viral Activation with Penicillin Toxicity in Guinea Pigs and Hamsters.” *The Yale Journal of Biology and Medicine* 47 (3): 166–81.

Guittar, John, Ashley Shade, and Elena Litchman. 2019. “Trait-Based Community Assembly and Succession of the Infant Gut Microbiome.” *Nature Communications* 10 (1): 512.

Gupta, Shaan, Emma Allen-Vercoe, and Elaine O. Petrof. 2016. “Fecal Microbiota Transplantation: In Perspective.” *Therapeutic Advances in Gastroenterology* 9 (2): 229–39.

Haiser, Henry J., David B. Gootenberg, Kelly Chatman, Gopal Sirasani, Emily P. Balskus, and Peter J. Turnbaugh. 2013. “Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium Eggerthella Lenta.” *Science* 341 (6143): 295–98.

Hall, Ivan C., and Elizabeth O’toole. 1935. “INTESTINAL FLORA IN NEW-BORN INFANTS: WITH A DESCRIPTION OF A NEW PATHOGENIC ANAEROBE, BACILLUS DIFFICILIS.” *American Journal of Diseases of Children* 49 (2): 390–402.

Hamre, Dorothy M., Geoffrey Rake, Clara M. Mckee, Harold B. Macphillamy, and Others. 1943. “The Toxicity of Penicillin as Prepared for Clinical Use.” *American Journal of Medical Sciences* 206 (5): 642–52.

Hensgens, M. P. M., O. M. Dekkers, A. Demeulemeester, A. G. M. Buiting, P. Bloembergen, B. H. B. van Benthem, S. Le Cessie, and E. J. Kuijper. 2014. “Diarrhoea in General Practice: When Should a Clostridium Difficile Infection Be Considered? Results of a Nested Case-Control Study.” *Clinical Microbiology and Infection: The Official Publication of the European Society of*

Clinical Microbiology and Infectious Diseases 20 (12): O1067–74.

Hey, J. 2001. “The Mind of the Species Problem.” *Trends in Ecology & Evolution* 16 (7): 326–29.

Hoppe, Brigitte. 1983. “Die Biologie Der Mikroorganismen von F. J. Cohn (1828-1898): Entwicklung Aus Forschungen über Mikroskopische Pflanzen Und Tiere.” *Sudhoffs Archiv* 67 (2): 158–89.

Houghteling, Pearl D., and W. Allan Walker. 2015. “Why Is Initial Bacterial Colonization of the Intestine Important to Infants’ and Children's Health?” *Journal of Pediatric Gastroenterology and Nutrition* 60 (3): 294–307.

Hugenholtz, Philip, Maria Chuvochina, Aharon Oren, Donovan H. Parks, and Rochelle M. Soo. 2021. “Prokaryotic Taxonomy and Nomenclature in the Age of Big Sequence Data.” *The ISME Journal* 15 (7): 1879–92.

Jackson, Melissa, Sidney Olefson, Jason T. Machan, and Colleen R. Kelly. 2016. “A High Rate of Alternative Diagnoses in Patients Referred for Presumed Clostridium Difficile Infection.” *Journal of Clinical Gastroenterology* 50 (9): 742–46.

Jain, Chirag, Luis M. Rodriguez-R, Adam M. Phillippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2018. “High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries.” *Nature Communications* 9 (1): 5114.

Jangi, Sushrut, and J. Thomas Lamont. 2010. “Asymptomatic Colonization by Clostridium Difficile in Infants: Implications for Disease in Later Life.” *Journal of Pediatric Gastroenterology and Nutrition* 51 (1): 2–7.

Jasemi, Seyedesomaye, Mohammad Emaneini, Mohammad Sadegh Fazeli, Zahra Ahmadinejad, Bizhan Nomanpour, Fatemah Sadeghpour Heravi, Leonardo A. Sechi, and Mohammad Mehdi Feizabadi. 2020. “Toxigenic and Non-Toxigenic Patterns I, II and III and Biofilm-Forming Ability in Bacteroides Fragilis Strains Isolated from Patients Diagnosed with Colorectal Cancer.” *Gut Pathogens* 12 (June): 28.

Jenior, Matthew L., Jhansi L. Leslie, Vincent B. Young, and Patrick D. Schloss. 2017. “Clostridium Difficile Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes.” *mSystems* 2 (4). https://doi.org/10.1128/mSystems.00063-17.

Jochum, Lara, and Bärbel Stecher. 2020. “Label or Concept - What Is a Pathobiont?” *Trends in Microbiology* 28 (10): 789–92.

Juge, Nathalie, Louise Tailford, and C. David Owen. 2016. “Sialidases from Gut Bacteria: A Mini-Review.” *Biochemical Society Transactions* 44 (1): 166–75.

Karanika, Styliani, Suresh Paudel, Fainareti N. Zervou, Christos Grigoras, Ioannis M. Zacharioudakis, and Eleftherios Mylonakis. 2016. “Prevalence and Clinical Outcomes of Clostridium Difficile Infection in the Intensive Care Unit: A Systematic Review and Meta-Analysis.” *Open Forum Infectious Diseases* 3 (1): ofv186.

Kararli, T. T. 1995. “Comparison of the Gastrointestinal Anatomy, Physiology, and Biochemistry of Humans and Commonly Used Laboratory Animals.” *Biopharmaceutics & Drug Disposition* 16 (5): 351–80.

Keel, M. K., and J. G. Songer. 2007. “The Distribution and Density of Clostridium Difficile Toxin Receptors on the Intestinal Mucosa of Neonatal Pigs.” *Veterinary Pathology* 44 (6): 814–22.

Keighley, M. R., D. W. Burdon, J. Alexander-Williams, N. Shinagawa, Y. Arabi, H. Thompson, D. Youngs, S. Bentley, and R. H. George. 1978. “Diarrhoea and Pseudomembranous Colitis after Gastrointestinal Operations. A Prospective Study.” *The Lancet* 2 (8101): 1165–67.

Kim, Pamela, Akash Gadani, Heitham Abdul-Baki, Ricardo Mitre, and Marcia Mitre. 2019. “Fecal Microbiota Transplantation in Recurrent Clostridium Difficile Infection: A Retrospective Single-Center Chart Review.” *JGH Open : An Open Access Journal of Gastroenterology and Hepatology* 3 (1): 4–9.

Kiu, Raymond, Shabhonam Caim, Cristina Alcon-Giner, Gusztav Belteki, Paul Clarke, Derek Pickard, Gordon Dougan, and Lindsay J. Hall. 2017. “Preterm Infant-Associated Clostridium Tertium, Clostridium Cadaveris, and Clostridium Paraputrificum Strains: Genomic and Evolutionary Insights.” *Genome Biology and Evolution* 9 (10): 2707–14.

Kiu, Raymond, and Lindsay J. Hall. 2018. “An Update on the Human and Animal Enteric Pathogen Clostridium Perfringens.” *Emerging Microbes & Infections* 7 (1): 141.

Klare, Ingo, Carola Konstabel, Guido Werner, Geert Huys, Vanessa Vankerckhoven, Gunnar

Kahlmeter, Bianca Hildebrandt, Sibylle Müller-Bertling, Wolfgang Witte, and Herman Goossens. 2007. "Antimicrobial Susceptibilities of Lactobacillus, Pediococcus and Lactococcus Human Isolates and Cultures Intended for Probiotic or Nutritional Use." *The Journal of Antimicrobial Chemotherapy* 59 (5): 900–912.

Knight, Daniel R., Korakrit Imwattana, Brian Kullin, Enzo Guerrero-Araya, Daniel Paredes-Sabja, Xavier Didelot, Kate E. Dingle, David W. Eyre, César Rodríguez, and Thomas V. Riley. 2021. "Major Genetic Discontinuity and Novel Toxigenic Species in Clostridioides Difficile Taxonomy." *eLife* 10 (June). https://doi.org/10.7554/eLife.64325.

Knight, Daniel R., Michele M. Squire, and Thomas V. Riley. 2015. "Nationwide Surveillance Study of Clostridium Difficile in Australian Neonatal Pigs Shows High Prevalence and Heterogeneity of PCR Ribotypes." *Applied and Environmental Microbiology* 81 (1): 119–23.

Knippel, Reece J., Aaron G. Wexler, Jeanette M. Miller, William N. Beavers, Andy Weiss, Valérie de Crécy-Lagard, Katherine A. Edmonds, David P. Giedroc, and Eric P. Skaar. 2020. "Clostridioides Difficile Senses and Hijacks Host Heme for Incorporation into an Oxidative Stress Defense System." *Cell Host & Microbe* 28 (3): 411–21.e6.

Konstantinidis, Konstantinos T., and Edward F. DeLong. 2008. "Genomic Patterns of Recombination, Clonal Divergence and Environment in Marine Microbial Populations." *The ISME Journal* 2 (10): 1052–65.

Konstantinidis, Konstantinos T., and James M. Tiedje. 2005. "Genomic Insights That Advance the Species Definition for Prokaryotes." *Proceedings of the National Academy of Sciences of the United States of America* 102 (7): 2567–72.

Kubota, Hiroyuki, Hiroshi Makino, Agata Gawad, Akira Kushiro, Eiji Ishikawa, Takafumi Sakai, Takuya Akiyama, et al. 2016. "Longitudinal Investigation of Carriage Rates, Counts, and Genotypes of Toxigenic Clostridium Difficile in Early Infancy." *Applied and Environmental Microbiology* 82 (19): 5806–14.

Kumar, Ranjit, Nengjun Yi, Degui Zhi, Peter Eipers, Kelly T. Goldsmith, Paula Dixon, David K. Crossman, et al. 2017. "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 (June): 12.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*. https://doi.org/10.18637/jss.v082.i13.

Lane, Nick. 2015. "The Unseen World: Reflections on Leeuwenhoek (1677) 'Concerning Little Animals.'" *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370 (1666): 20140344.

Larcombe, Sarah, Melanie L. Hutton, Thomas V. Riley, Helen E. Abud, and Dena Lyras. 2018. "Diverse Bacterial Species Contribute to Antibiotic-Associated Diarrhoea and Gastrointestinal Damage." *The Journal of Infection* 77 (5): 417–26.

Lawler, Amelia J., Peter A. Lambert, and Tony Worthington. 2020. "A Revised Understanding of Clostridioides Difficile Spore Germination." *Trends in Microbiology* 28 (9): 744–52.

Lawson, Paul A., and Fred A. Rainey. 2016. "Proposal to Restrict the Genus Clostridium Prazmowski to Clostridium Butyricum and Related Species." *International Journal of Systematic and Evolutionary Microbiology* 66 (2): 1009–16.

Lee, Helen S., Kamryn Plechot, Shruti Gohil, and Jennifer Le. 2021. "Clostridium Difficile: Diagnosis and the Consequence of Over Diagnosis." *Infectious Diseases and Therapy* 10 (2): 687–97.

Lees, E. A., F. Miyajima, M. Pirmohamed, and E. D. Carrol. 2016. "The Role of Clostridium Difficile in the Paediatric and Neonatal Gut - a Narrative Review." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 35 (7): 1047–57.

Lefebvre, Sandra L., Richard J. Reid-Smith, David Waltner-Toews, and J. Scott Weese. 2009. "Incidence of Acquisition of Methicillin-Resistant Staphylococcus Aureus, Clostridium Difficile, and Other Health-Care-Associated Pathogens by Dogs That Participate in Animal-Assisted Interventions." *Journal of the American Veterinary Medical Association* 234 (11): 1404–17.

Lefebvre, S. L., D. Waltner-Toews, A. S. Peregrine, R. Reid-Smith, L. Hodge, L. G. Arroyo, and J. S. Weese. 2006. “Prevalence of Zoonotic Agents in Dogs Visiting Hospitalized People in Ontario: Implications for Infection Control.” *The Journal of Hospital Infection* 62 (4): 458–66.

Ley, Ruth E. 2014. “Harnessing Microbiota to Kill a Pathogen: The Sweet Tooth of Clostridium Difficile.” *Nature Medicine* 20 (3): 248–49.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.

Lim, S. C., D. R. Knight, and T. V. Riley. 2020. “Clostridium Difficile and One Health.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 26 (7): 857–63.

Litvin, Marina, Kimberly A. Reske, Jennie Mayfield, Kathleen M. McMullen, Peter Georgantopoulos, Susan Copper, Joan E. Hoppe-Bauer, Victoria J. Fraser, David K. Warren, and Erik R. Dubberke. 2009. “Identification of a Pseudo-Outbreak of Clostridium Difficile Infection (CDI) and the Effect of Repeated Testing, Sensitivity, and Specificity on Perceived Prevalence of CDI.” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 30 (12): 1166–71.

Liu, Bo, Dandan Zheng, Qi Jin, Lihong Chen, and Jian Yang. 2019. “VFDB 2019: A Comparative Pathogenomic Platform with an Interactive Web Interface.” *Nucleic Acids Research* 47 (D1): D687–92.

Liu, C., D. N. Frank, M. Horch, S. Chau, D. Ir, E. A. Horch, K. Tretina, K. van Besien, C. A. Lozupone, and V. H. Nguyen. 2017. “Associations between Acute Gastrointestinal GvHD and the Baseline Gut Microbiota of Allogeneic Hematopoietic Stem Cell Transplant Recipients and Donors.” *Bone Marrow Transplantation* 52 (12): 1643–50.

Loman, Nicholas J., Chrystala Constantinidou, Martin Christner, Holger Rohde, Jacqueline Z-M Chan, Joshua Quick, Jacqueline C. Weir, et al. 2013. “A Culture-Independent Sequence-Based Metagenomics Approach to the Investigation of an Outbreak of Shiga-Toxigenic Escherichia Coli O104:H4.” *JAMA: The Journal of the American Medical Association* 309 (14): 1502–10.

Lopez, Christopher A., Tess P. McNeely, Kamila Nurmakova, William N. Beavers, and Eric P. Skaar. 2020. “Clostridioides Difficile Proline Fermentation in Response to Commensal Clostridia.” *Anaerobe* 63 (June): 102210.

Lozupone, Catherine A., Jesse Stombaugh, Antonio Gonzalez, Gail Ackermann, Doug Wendel, Yoshiki Vázquez-Baeza, Janet K. Jansson, Jeffrey I. Gordon, and Rob Knight. 2013. “Meta-Analyses of Studies of the Human Microbiota.” *Genome Research* 23 (10): 1704–14.

Lozupone, Catherine A., Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. 2012. “Diversity, Stability and Resilience of the Human Gut Microbiota.” *Nature* 489 (7415): 220–30.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, et al. 1998. “Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (6): 3140–45.

Maistrenko, Oleksandr M., Daniel R. Mende, Mechthild Luetge, Falk Hildebrand, Thomas S. B. Schmidt, Simone S. Li, João F. Matias Rodrigues, et al. 2020. “Disentangling the Impact of Environmental and Phylogenetic Constraints on Prokaryotic within-Species Diversity.” *The ISME Journal*. https://doi.org/10.1038/s41396-020-0600-z.

Martínez-Meléndez, Adrián, Adrián Camacho-Ortiz, Rayo Morfin-Otero, Héctor Jesús Maldonado-Garza, Licet Villarreal-Treviño, and Elvira Garza-González. 2017. “Current Knowledge on the Laboratory Diagnosis of Clostridium Difficile Infection.” *World Journal of Gastroenterology: WJG* 23 (9): 1552–67.

Mayden, R. L. 1997. “A Hierarchy of Species Concepts: The Denouement in the Saga of the Species Problem.” In *Species: The Units of Diversity,* edited by M. F. Claridge, H. A. Dawah, and M. R. Wilson, 381–423. Chapman & Hall.

McDonald, L. Clifford, Dale N. Gerding, Stuart Johnson, Johan S. Bakken, Karen C. Carroll, Susan E. Coffin, Erik R. Dubberke, et al. 2018. “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: An Official Publication of the Infectious Diseases Society of America* 66 (7): e1–48.

Mende, Daniel R., Ivica Letunic, Jaime Huerta-Cepas, Simone S. Li, Kristoffer Forslund, Shinichi Sunagawa, and Peer Bork. 2017. “proGenomes: A Resource for Consistent Functional and Taxonomic Annotations of Prokaryotic Genomes.” *Nucleic Acids Research* 45 (D1): D529–34.

Mende, Daniel R., Shinichi Sunagawa, Georg Zeller, and Peer Bork. 2013. “Accurate and Universal Delineation of Prokaryotic Species.” *Nature Methods* 10 (9): 881–84.

Milanese, Alessio, Daniel R. Mende, Lucas Paoli, Guillem Salazar, Hans-Joachim Ruscheweyh, Miguelangel Cuenca, Pascal Hingamp, et al. 2019. “Microbial Abundance, Activity and Population Genomic Profiling with mOTUs2.” *Nature Communications*. https://doi.org/10.1038/s41467-019-08844-4.

Moldovan, Mikhail A., and Mikhail S. Gelfand. 2018. “Pangenomic Definition of Prokaryotic Species and the Phylogenetic Structure of Prochlorococcus Spp.” *Frontiers in Microbiology* 9 (March): 428.

Moore, Rebecca E., and Steven D. Townsend. 2019. “Temporal Development of the Infant Gut Microbiome.” *Open Biology* 9 (9): 190128.

Natarajan, Mukil, Seth T. Walk, Vincent B. Young, and David M. Aronoff. 2013. “A Clinical and Epidemiological Review of Non-Toxigenic Clostridium Difficile.” *Anaerobe* 22 (August): 1–5.

Navaneethan, Udayakumar, and Ralph A. Giannella. 2009. “Thinking beyond the Colon-Small Bowel Involvement in Clostridium Difficile Infection.” *Gut Pathogens* 1 (1): 7.

Na, Xi, and Ciaran Kelly. 2011. “Probiotics in Clostridium Difficile Infection.” *Journal of Clinical Gastroenterology* 45 Suppl (November): S154–58.

Ng, Katharine M., Jessica A. Ferreyra, Steven K. Higginbottom, Jonathan B. Lynch, Purna C. Kashyap, Smita Gopinath, Natasha Naidu, et al. 2013. “Microbiota-Liberated Host Sugars Facilitate Post-Antibiotic Expansion of Enteric Pathogens.” *Nature* 502 (7469): 96–99.

Normington, Charmaine, Ines B. Moura, Jessica A. Bryant, Duncan J. Ewin, Emma V. Clark, Morgan J. Kettle, Hannah C. Harris, et al. 2021. “Biofilms Harbour Clostridioides Difficile, Serving as a Reservoir for Recurrent Infection.” *NPJ Biofilms and Microbiomes* 7 (1): 16.

Olm, Matthew R., Alexander Crits-Christoph, Spencer Diamond, Adi Lavy, Paula B. Matheus Carnevali, and Jillian F. Banfield. 2020. “Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries.” *mSystems* 5 (1). https://doi.org/10.1128/mSystems.00731-19.

Park, Seon-Young, and Geom Seog Seo. 2021. “Fecal Microbiota Transplantation: Is It Safe?” *Clinical Endoscopy* 54 (2): 157–60.

Pereira, Fátima C., Kenneth Wasmund, Iva Cobankovic, Nico Jehmlich, Craig W. Herbold, Kang Soo Lee, Barbara Sziranyi, et al. 2020. “Rational Design of a Microbial Consortium of Mucosal Sugar Utilizers Reduces Clostridiodes Difficile Colonization.” *Nature Communications* 11 (1): 5104.

Planche, Timothy, and Mark Wilcox. 2011. “Reference Assays for Clostridium Difficile Infection: One or Two Gold Standards?” *Journal of Clinical Pathology* 64 (1): 1–5.

Polage, Christopher R., Clare E. Gyorke, Michael A. Kennedy, Jhansi L. Leslie, David L. Chin, Susan Wang, Hien H. Nguyen, et al. 2015. “Overdiagnosis of Clostridium Difficile Infection in the Molecular Test Era.” *JAMA Internal Medicine* 175 (11): 1792–1801.

Polage, Christopher R., Jay V. Solnick, and Stuart H. Cohen. 2012. “Nosocomial Diarrhea: Evaluation and Treatment of Causes Other than Clostridium Difficile.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 55 (7): 982–89.

Pruitt, Rory N., and D. Borden Lacy. 2012. “Toward a Structural Understanding of Clostridium Difficile Toxins A and B.” *Frontiers in Cellular and Infection Microbiology* 2 (March): 28.

Pruss, K. M., A. Marcobal, A. M. Southwick, D. Dahan, S. A. Smits, J. A. Ferreyra, S. K. Higginbottom, et al. 2021. “Mucin-Derived O-Glycans Supplemented to Diet Mitigate Diverse Microbiota Perturbations.” *The ISME Journal* 15 (2): 577–91.

Qin, Nan, Fengling Yang, Ang Li, Edi Prifti, Yanfei Chen, Li Shao, Jing Guo, et al. 2014. “Alterations of the Human Gut Microbiome in Liver Cirrhosis.” *Nature* 513 (7516): 59–64.

Richter, Michael, and Ramon Rosselló-Móra. 2009. “Shifting the Genomic Gold Standard for the

Prokaryotic Species Definition." *Proceedings of the National Academy of Sciences of the United States of America* 106 (45): 19126–31.

Rigottier-Gois, Lionel. 2013. "Dysbiosis in Inflammatory Bowel Diseases: The Oxygen Hypothesis." *The ISME Journal* 7 (7): 1256–61.

Rivera-Chávez, Fabian, Lillian F. Zhang, Franziska Faber, Christopher A. Lopez, Mariana X. Byndloss, Erin E. Olsan, Gege Xu, et al. 2016. "Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella." *Cell Host & Microbe* 19 (4): 443–54.

Rivière, Audrey, Marija Selak, David Lantin, Frédéric Leroy, and Luc De Vuyst. 2016. "Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut." *Frontiers in Microbiology* 7 (June): 979.

Rodríguez, Juan Miguel, Kiera Murphy, Catherine Stanton, R. Paul Ross, Olivia I. Kober, Nathalie Juge, Ekaterina Avershina, et al. 2015. "The Composition of the Gut Microbiota throughout Life, with an Emphasis on Early Life." *Microbial Ecology in Health and Disease* 26 (February): 26050.

Rodriguez-R, Luis M., Chirag Jain, Roth E. Conrad, Srinivas Aluru, and Konstantinos T. Konstantinidis. 2021. "Reply to: 'Re-Evaluating the Evidence for a Universal Genetic Boundary among Microbial Species.'" *Nature Communications*.

Roger, Laure C., Adele Costabile, Diane T. Holland, Lesley Hoyles, and Anne L. McCartney. 2010. "Examination of Faecal Bifidobacterium Populations in Breast- and Formula-Fed Infants during the First 18 Months of Life." *Microbiology* 156 (Pt 11): 3329–41.

Roger, Laure C., and Anne L. McCartney. 2010. "Longitudinal Investigation of the Faecal Microbiota of Healthy Full-Term Infants Using Fluorescence in Situ Hybridization and Denaturing Gradient Gel Electrophoresis." *Microbiology* 156 (Pt 11): 3317–28.

Roswall, Josefine, Lisa M. Olsson, Petia Kovatcheva-Datchary, Staffan Nilsson, Valentina Tremaroli, Marie-Christine Simon, Pia Kiilerich, et al. 2021. "Developmental Trajectory of the Healthy Human Gut Microbiota during the First 5 Years of Life." *Cell Host & Microbe* 29 (5): 765–76.e3.

Rousseau, Clotilde, Florence Levenez, Charlène Fouqueray, Joël Doré, Anne Collignon, and Patricia Lepage. 2011. "Clostridium Difficile Colonization in Early Infancy Is Accompanied by Changes in Intestinal Microbiota Composition." *Journal of Clinical Microbiology* 49 (3): 858–65.

Rousseau, Clotilde, Isabelle Poilane, Loic De Pontual, Anne-Claire Maherault, Alban Le Monnier, and Anne Collignon. 2012. "Clostridium Difficile Carriage in Healthy Infants in the Community: A Potential Reservoir for Pathogenic Strains." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 55 (9): 1209–15.

Rudkin, Justine K., Rachel M. McLoughlin, Andrew Preston, and Ruth C. Massey. 2017. "Bacterial Toxins: Offensive, Defensive, or Something Else Altogether?" *PLoS Pathogens* 13 (9): e1006452.

Rupnik, Maja. 2008. "Heterogeneity of Large Clostridial Toxins: Importance of Clostridium Difficile Toxinotypes." *FEMS Microbiology Reviews* 32 (3): 541–55.

Ruppitsch, Werner. 2016. "Molecular Typing of Bacteria for Epidemiological Surveillance and Outbreak investigation/Molekulare Typisierung von Bakterien Für Die Epidemiologische Überwachung Und Ausbruchsabklärung." *Food and Environmental Virology* 67: 199–224.

Rutayisire, Erigene, Kun Huang, Yehao Liu, and Fangbiao Tao. 2016. "The Mode of Delivery Affects the Diversity and Colonization Pattern of the Gut Microbiota during the First Year of Infants' Life: A Systematic Review." *BMC Gastroenterology* 16 (1): 86.

Saif, N. al, and J. S. Brazier. 1996. "The Distribution of Clostridium Difficile in the Environment of South Wales." *Journal of Medical Microbiology* 45 (2): 133–37.

Sartelli, Massimo, Stefano Di Bella, Lynne V. McFarland, Sahil Khanna, Luis Furuya-Kanamori, Nadir Abuzeid, Fikri M. Abu-Zidan, et al. 2019. "2019 Update of the WSES Guidelines for Management of Clostridioides (Clostridium) Difficile Infection in Surgical Patients." *World Journal of Emergency Surgery: WJES* 14 (February): 8.

Schmidt, Thomas Sb, Matthew R. Hayward, Luis P. Coelho, Simone S. Li, Paul I. Costea, Anita Y. Voigt, Jakob Wirbel, et al. 2019. "Extensive Transmission of Microbes along the Gastrointestinal

Tract." *eLife* 8 (February). https://doi.org/10.7554/eLife.42693.

Schneeberg, Alexander, Heinrich Neubauer, Gernot Schmoock, Ernst Grossmann, and Christian Seyboldt. 2013. "Presence of Clostridium Difficile PCR Ribotype Clusters Related to 033, 078 and 045 in Diarrhoeic Calves in Germany." *Journal of Medical Microbiology* 62 (Pt 8): 1190–98.

Schubert, Alyxandria M., Mary A. M. Rogers, Cathrin Ring, Jill Mogle, Joseph P. Petrosino, Vincent B. Young, David M. Aronoff, and Patrick D. Schloss. 2014. "Microbiome Data Distinguish Patients with Clostridium Difficile Infection and Non-C. Difficile-Associated Diarrhea from Healthy Controls." *mBio* 5 (3): e01021–14.

Schubl, Sebastian D., Loveita Raymond, R. Jonathan Robitsek, and Farshad Bagheri. 2016. "Isolated Clostridium Difficile Small Bowel Enteritis in the Absence of Predisposing Factors." *Surgical Infections Case Reports* 1 (1): 38–40.

Schutze, Gordon E., Rodney E. Willoughby, Committee on Infectious Diseases, and American Academy of Pediatrics. 2013. "Clostridium Difficile Infection in Infants and Children." *Pediatrics* 131 (1): 196–200.

Seekatz, Anna Maria, Krishna Rao, Kavitha Santhosh, and Vincent Bensan Young. 2016. "Dynamics of the Fecal Microbiome in Patients with Recurrent and Nonrecurrent Clostridium Difficile Infection." *Genome Medicine* 8 (1): 47.

Shetty, N., M. W. D. Wren, and P. G. Coen. 2011. "The Role of Glutamate Dehydrogenase for the Detection of Clostridium Difficile in Faecal Samples: A Meta-Analysis." *The Journal of Hospital Infection* 77 (1): 1–6.

Smelov, Vitaly, Alison Vrbanac, Eleanne F. van Ess, Marlies P. Noz, Raymond Wan, Carina Eklund, Tyler Morgan, et al. 2017. "Chlamydia Trachomatis Strain Types Have Diversified Regionally and Globally with Evidence for Recombination across Geographic Divides." *Frontiers in Microbiology* 8 (November): 2195.

Smith, Birgitte, Susan Bodé, Bodil L. Petersen, Tim K. Jensen, Christian Pipper, Julie Kloppenborg, Mette Boyé, Karen A. Krogfelt, and Lars Mølbak. 2011. "Community Analysis of Bacteria Colonizing Intestinal Tissue of Neonates with Necrotizing Enterocolitis." *BMC Microbiology* 11 (April): 73.

Sorg, Joseph A., and Abraham L. Sonenshein. 2008. "Bile Salts and Glycine as Cogerminants for Clostridium Difficile Spores." *Journal of Bacteriology* 190 (7): 2505–12.

Stabler, Richard A., Lisa F. Dawson, Esmeralda Valiente, Michelle D. Cairns, Melissa J. Martin, Elizabeth H. Donahue, Thomas V. Riley, et al. 2012. "Macro and Micro Diversity of Clostridium Difficile Isolates from Diverse Sources and Geographical Locations." *PloS One* 7 (3): e31559.

Stare, Barbara Geric, Michel Delmée, and Maja Rupnik. 2007. "Variant Forms of the Binary Toxin CDT Locus and tcdC Gene in Clostridium Difficile Strains." *Journal of Medical Microbiology* 56 (Pt 3): 329–35.

Stewart, Christopher J., Nadim J. Ajami, Jacqueline L. O'Brien, Diane S. Hutchinson, Daniel P. Smith, Matthew C. Wong, Matthew C. Ross, et al. 2018. "Temporal Development of the Gut Microbiome in Early Childhood from the TEDDY Study." *Nature* 562 (7728): 583–88.

Stoesser, Nicole, David W. Eyre, T. Phuong Quan, Heather Godwin, Gemma Pill, Emily Mbuvi, Alison Vaughan, et al. 2017. "Epidemiology of Clostridium Difficile in Infants in Oxfordshire, UK: Risk Factors for Colonization and Carriage, and Genetic Overlap with Regional C. Difficile Infection Strains." *PloS One* 12 (8): e0182307.

Sun, Xingmin, Tor Savidge, and Hanping Feng. 2010. "The Enterotoxicity of Clostridium Difficile Toxins." *Toxins* 2 (7): 1848–80.

Swidsinski, Alexander, Vera Loening-Baucke, Franz Theissig, Holger Engelhardt, Stig Bengmark, Stefan Koch, Herbert Lochs, and Yvonne Dörffel. 2007. "Comparative Study of the Intestinal Mucus Barrier in Normal and Inflamed Colon." *Gut* 56 (3): 343–50.

Taft, Diana H., Jinxin Liu, Maria X. Maldonado-Gomez, Samir Akre, M. Nazmul Huda, S. M. Ahmad, Charles B. Stephensen, and David A. Mills. 2018. "Bifidobacterial Dominance of the Gut in Early Life and Acquisition of Antimicrobial Resistance." *mSphere* 3 (5). https://doi.org/10.1128/mSphere.00441-18.

Tailford, Louise E., Emmanuelle H. Crost, Devon Kavanaugh, and Nathalie Juge. 2015. "Mucin

Glycan Foraging in the Human Gut Microbiome." *Frontiers in Genetics* 6 (March): 81.

Tamma, Pranita D., and Thomas J. Sandora. 2012. "Clostridium Difficile Infection in Children: Current State and Unanswered Questions." *Journal of the Pediatric Infectious Diseases Society* 1 (3): 230–43.

Tang, Derek M., Nathalie H. Urrunaga, and Erik C. von Rosenvinge. 2016. "Pseudomembranous Colitis: Not Always Clostridium Difficile." *Cleveland Clinic Journal of Medicine* 83 (5): 361–66.

Tenaillon, Olivier, David Skurnik, Bertrand Picard, and Erick Denamur. 2010. "The Population Genetics of Commensal Escherichia Coli." *Nature Reviews. Microbiology* 8 (3): 207–17.

Tenover, Fred C., Ellen Jo Baron, Lance R. Peterson, and David H. Persing. 2011. "Laboratory Diagnosis of Clostridium Difficile Infection Can Molecular Amplification Methods Move Us out of Uncertainty?" *The Journal of Molecular Diagnostics: JMD* 13 (6): 573–82.

Theriot, Casey M., Alison A. Bowman, and Vincent B. Young. 2016. "Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium Difficile Spore Germination and Outgrowth in the Large Intestine." *mSphere* 1 (1). https://doi.org/10.1128/mSphere.00045-15.

Theriot, Casey M., and Vincent B. Young. 2015. "Interactions Between the Gastrointestinal Microbiome and Clostridium Difficile." *Annual Review of Microbiology* 69: 445–61.

Thevenon, Olivier, Willem Adema, and Chris Clarke. 2016. "Background Brief on Fathers' Leave and Its Use." OECD. https://doi.org/10.13140/RG.2.2.27717.24808.

Tibshirani, Robert, and Guenther Walther. 2005. "Cluster Validation by Prediction Strength." *Journal of Computational and Graphical Statistics*. https://doi.org/10.1198/106186005x59243.

Toltzis, Philip, Michelle M. Nerandzic, Elie Saade, Mary Ann O'Riordan, Sarah Smathers, Theoklis Zaoutis, Jason Kim, and Curtis J. Donskey. 2012. "High Proportion of False-Positive Clostridium Difficile Enzyme Immunoassays for Toxin A and B in Pediatric Patients." *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 33 (2): 175–79.

Truper, H. G., and J. P. Euzeby. 2009. "International Code of Nomenclature of Prokaryotes. Appendix 9: Orthography." *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*. https://doi.org/10.1099/ijs.0.016741-0.

Tschudin-Sutter, Sarah, Olivier Braissant, Stefan Erb, Anne Stranden, Gernot Bonkat, Reno Frei, and Andreas F. Widmer. 2016. "Growth Patterns of Clostridium Difficile - Correlations with Strains, Binary Toxin and Disease Severity: A Prospective Cohort Study." *PloS One* 11 (9): e0161711.

Tynkkynen, S., K. V. Singh, and P. Varmanen. 1998. "Vancomycin Resistance Factor of Lactobacillus Rhamnosus GG in Relation to Enterococcal Vancomycin Resistance (van) Genes." *International Journal of Food Microbiology* 41 (3): 195–204.

(u.s.), Centers For Disease Control And Prevention, and Centers for Disease Control and Prevention (U.S.). 2019. "Antibiotic Resistance Threats in the United States, 2019." https://doi.org/10.15620/cdc:82532.

Van Rossum, Thea, Pamela Ferretti, Oleksandr M. Maistrenko, and Peer Bork. 2020. "Diversity within Species: Interpreting Strains in Microbiomes." *Nature Reviews. Microbiology* 18 (9): 491–506.

Vincent, Caroline, Mark A. Miller, Thaddeus J. Edens, Sudeep Mehrotra, Ken Dewar, and Amee R. Manges. 2016. "Bloom and Bust: Intestinal Microbiota Dynamics in Response to Hospital Exposures and Clostridium Difficile Colonization or Infection." *Microbiome* 4 (March): 12.

Viscidi, R., S. Willey, and J. G. Bartlett. 1981. "Isolation Rates and Toxigenic Potential of Clostridium Difficile Isolates from Various Patient Populations." *Gastroenterology* 81 (1): 5–9.

Voth, Daniel E., and Jimmy D. Ballard. 2005. "Clostridium Difficile Toxins: Mechanism of Action and Role in Disease." *Clinical Microbiology Reviews* 18 (2): 247–63.

Weese, J. S. 2010. "Clostridium Difficile in Food--Innocent Bystander or Serious Threat?" *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 16 (1): 3–10.

Weese, J. Scott. 2020. "Clostridium (Clostridioides) Difficile in Animals." *Journal of Veterinary Diagnostic Investigation: Official Publication of the American Association of Veterinary*

Laboratory Diagnosticians, Inc 32 (2): 213–21.

Welkon, C. J., S. S. Long, C. M. Thompson Jr, and P. H. Gilligan. 1985. “Clostridium Difficile in Patients with Cystic Fibrosis.” *American Journal of Diseases of Children* 139 (8): 805–8.

Whitman, William B., Fred Rainey, Peter Kämpfer, Martha Trujillo, Jonsik Chun, Paul DeVos, Brian Hedlund, Svetlana Dedysh, and O. I. Nedashkovskaya. 2016. “Bergey’s Manual of Systematics of Archaea and Bacteria.” https://elibrary.ru/item.asp?id=27223331.

Wilkins, John S. 2003. “How to Be a Chaste Species Pluralist-Realist: The Origins of Species Modes and the Synapomorphic Species Concept.” *Biology and Philosophy* 18 (5): 621–38.

Willemsen, L. E. M., M. A. Koetsier, S. J. H. van Deventer, and E. A. F. van Tol. 2003. “Short Chain Fatty Acids Stimulate Epithelial Mucin 2 Expression through Differential Effects on Prostaglandin E(1) and E(2) Production by Intestinal Myofibroblasts.” *Gut* 52 (10): 1442–47.

Wirbel, Jakob, Konrad Zych, Morgan Essex, Nicolai Karcher, Ece Kartal, Guillem Salazar, Peer Bork, Shinichi Sunagawa, and Georg Zeller. 2021. “Microbiome Meta-Analysis and Cross-Disease Comparison Enabled by the SIAMCAT Machine Learning Toolbox.” *Genome Biology* 22 (1): 93.

Wu, Kuan-Sheng, Ling-Shan Syue, Aristine Cheng, Ting-Yu Yen, Hsien-Meng Chen, Yu-Hsin Chiu, Yu-Lung Hsu, et al. 2020. “Recommendations and Guidelines for the Treatment of Clostridioides Difficile Infection in Taiwan.” *Journal of Microbiology, Immunology, and Infection = Wei Mian Yu Gan Ran Za Zhi* 53 (2): 191–208.

Yang, Xiaoping, Kwame Oteng Darko, Yanjun Huang, Caimei He, Huansheng Yang, Shanping He, Jianzhong Li, Jian Li, Berthold Hocher, and Yulong Yin. 2017. “Resistant Starch Regulates Gut Microbiota: Structure, Biochemistry and Cell Signalling.” *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* 42 (1): 306–18.

Yassour, Moran, Tommi Vatanen, Heli Siljander, Anu-Maaria Hämäläinen, Taina Härkönen, Samppa J. Ryhänen, Eric A. Franzosa, et al. 2016. “Natural History of the Infant Gut Microbiome and Impact of Antibiotic Treatment on Bacterial Strain Diversity and Stability.” *Science Translational Medicine* 8 (343): 343ra81.

Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, et al. 2012. “Human Gut Microbiome Viewed across Age and Geography.” *Nature* 486 (7402): 222–27.

Young, Vincent Bensan. 2021. “Unexpected Results From a Phase 2 Trial of a Microbiome Therapeutic for Clostridioides Difficile Infection: Lessons for the Future.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*.

Young, Vincent B., and Thomas M. Schmidt. 2008. “Overview of the Gastrointestinal Microbiota.” *Advances in Experimental Medicine and Biology* 635: 29–40.

Yutin, Natalya, and Michael Y. Galperin. 2013. “A Genomic Update on Clostridial Phylogeny: Gram-Negative Spore Formers and Other Misplaced Clostridia.” *Environmental Microbiology* 15 (10): 2631–41.

Zhang, Jiahui, Ting Wang, Rongrong Li, Wei Ji, Yongdong Yan, Zhichao Sun, Jiahong Tan, Jinfeng Wu, Li Huang, and Zhengrong Chen. 2021. “Prediction of Risk Factors of Bronchial Mucus Plugs in Children with Mycoplasma Pneumoniae Pneumonia.” *BMC Infectious Diseases* 21 (1): 67.

Zhao, Hailong, David C. Nickle, Zhen Zeng, Pierra Y. T. Law, Mark H. Wilcox, Lan Chen, Ye Peng, et al. 2021. “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): 853–70.

Zheng, Leon, Caleb J. Kelly, and Sean P. Colgan. 2015. “Physiologic Hypoxia and Oxygen Homeostasis in the Healthy Intestine. A Review in the Theme: Cellular Responses to Hypoxia.” *American Journal of Physiology. Cell Physiology* 309 (6): C350–60.

Zhou, Yanjiao, Kristine M. Wylie, Rana E. El Feghaly, Kathie A. Mihindukulasuriya, Alexis Elward, David B. Haslam, Gregory A. Storch, and George M. Weinstock. 2016. “Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens.” *Journal of Clinical Microbiology* 54 (2): 368–75.

Zollner-Schwetz, Ines, Christoph Högenauer, Martina Joainig, Paul Weberhofer, Gregor Gorkiewicz, Thomas Valentin, Thomas A. Hinterleitner, and Robert Krause. 2008. “Role of Klebsiella Oxytoca in Antibiotic-Associated Diarrhea.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 47 (9): e74–78.