Ubiquitous Detection of *Clostridioides difficile* in Outpatient Gut Microbiome Using Next Generation Shotgun Sequencing and Metagenomic Analysis: Non-Toxigenic C. *difficile* is Ubiquitous in Normal Population

Sabine Hazan[1]*, Sonya Dave[2], Andreas Papoutsis[1], Brad D. Barrows[1], Thomas J. Borody[3]

[1]*Progena Biome, Ventura, CA, USA;* [2]*Microbiome Research, Inc. Ventura, CA, USA;* [3]*Centre for Digestive Diseases, Five Dock, NSW, Australia*

ABSTRACT

Introduction: Toxigenic *Clostridioides Difficile* Infection (CDI) is the most common cause of nosocomial disease in the United States. However, the prevalence in the general population of toxigenic and non-toxigenic strains is poorly understood. In this cross-sectional study we sought to determine the presence of *Clostridioides difficile* colonizing a representative sample of 119 CDI-asymptomatic volunteers (health care providers, public with chronic conditions, and healthy public).

Materials and Methods: Next Generation Sequencing (NGS) was performed on fecal samples from participants (n=119) who clinically did not have CDI. Following stool collection, DNA was extracted, quantitated, and then normalized for downstream library fabrication utilizing shotgun methodology. Prepared and indexed libraries were subsequently pooled and sequenced on the Illumina Next Seq 550 System.

Results: 117 of the 119 subjects (98%) were found to possess C. *difficile* as identified by the Kraken bioinformatics meta genomic pipeline. The C. *difficile* normalized count was independent of health care setting exposure, age, probiotic use, or health history.

Conclusion: NGS provides a unique opportunity to increase the resolution and identification of the bacterium C. *difficile* compared to traditional categorizations. Using meta-genomics and a stringent read count criteria (>1000 counts), we deciphered species level resolution of bacteria present, finding C. *difficile* in 98% of subjects, which are likely non-toxigenic. This discovery suggests that non-pathogenic C. *difficile* may be an important component of the human commensal gut microbiome, possibly present since birth. This raises fundamental questions regarding assumptions of CDI transmission that need future exploration. Given that non-pathogenic C. *difficile* has been used for CDI treatment, determining levels of non-pathogenic C. *difficile* in stool may be predictive of resistance to development of CDI.

Keywords: *Clostridiodes difficile*, Prevalence, Non-Toxigenic, Infection, Next Generation Sequencing, Metagenomics

Abbreviations: *Clostridioides Difficile* Infection (CDI); *Clostridioides difficile* (C. *difficile*); Enzyme Immuno-Assays (EIA); Nucleic Acid Amplification Tests (NAAT), Next Generation Sequencing (NGS); National Center for Biotechnology Information (NCBI)

INTRODUCTION

Clostridioides difficile is a 10 million years old bacterium and the most common cause of healthcare associated infection in the US, with rates of infection (CDI) steadily increasing. The acute care direct cost associated with CDI is in the billions of dollars [1]. CDI was first discovered in 1935 as a stool commensal in a healthy neonate [2]. In 1978 the perception of *C. difficile* transitioned from innocent bystander bacterium to the pathogenic bacterium responsible for Pseudomembranous Colitis, a potentially fatal inflammatory condition of the colon causing severe diarrhea with dehydration and other complications including kidney failure and bowel perforation [3]. All of this makes CDI an urgent threat to public health

Many diagnostic tools have evolved from the conventional stool culture due to the high impact and prevalence of CDI. Toxin Enzyme Immuno-Assays (EIA) and Nucleic Acid Amplification Tests (NAAT; e.g. PCR tests) have become routine diagnostic methods. While NAAT has a somewhat high specificity and sensitivity [4-6], EIA has only approximately 60% sensitivity [7]. Similarly, the spectrum of current therapies continues to evolve including antibiotics, monoclonal antibodies, probiotics, and fecal micro biota transplantation [8].

Next Generation Sequencing (NGS) provides a unique opportunity to increase the sensitivity and specificity of detection and hone in to the level of species (or even subspecies) identification of *C. difficile* as compared to traditional categorizations, such as PCR ribo-types (e.g. RT027) [9]. This sensitivity is accomplished by the ability of NGS to differentiate species based on whole genome nucleotide sequences, as a part of sequencing the entire colon micro biota of an individual. Herein, we utilized the Kraken bioinformatics

Pipeline [10] to provide species level resolution of *C. difficile* by aligning with reference sequences from the NCBI (National Center for Biotechnology Information) database.

In recent years, such an application of NGS has started to become employed for molecular diagnositics of infectious diseases [11], including bacterial [12], as reviewed in Nature Reviews [13]. MNGS has been validated in clinical laboratories for diagnosing meningitis, encephalitis, sepsis and pneumonia. Specific to *C. difficile*, Zhou et al. [14] compare diagnosis via metagenomics shotgun NGS vs. PCR for symptomatic *C. difficile* and other patients and find *C. difficile* was detected, *via* mNGS, in 86.3% (17/20) of PCR-positive CDiff samples and four of five PCR-negative samples were , negative by mNGS. In short, mNGS diagnosis is established for various bacterial infection diagnoses and has been characterized for symptomatic/toxigenic *C. difficile*.

While *C. difficile* transmission is assumed to be the etiology of CDI, this has not been proven. Despite the efforts of many hospitals to reduce transmission of *C. difficile*, such as patient isolation and the use of personal protective equipment, rates of CDI continue to rise [15,16]. Additionally, research shows that only 24% [17] of CDI cases arose in hospital settings and 66% in all health care settings, raising the question of the source of

CDI for the remaining large portion of community associated cases.

The lack of effectiveness of decades of isolation and transmission reduction efforts, demonstrated through increasing rates of CDI [3,15,16], makes us question the paradigm that new cases arise from transmission. Perhaps *C. difficile* could be an innocent bystander triggered to become toxigenic (*i.e.*, to lead to CDI) in specific settings (opportunistic pathogen). Thus, CDI would only occur in individuals carrying the bacteria. To approach this question, we took a nontraditional approach of analyzing *C. difficile* prevalence and levels in a cohort of healthy subjects and those with GI and/or other conditions. We randomly recruited 121 volunteers and analyzed their gut bacterial micro biomes by shotgun sequencing with Kraken bioinformatics pipeline analysis.

MATERIALS AND METHODS

Stools were tested in 121 random volunteers during the year 2019 (with informed consent) from various clinical trials to understand the similarities in the gut flora of multiple diseases. A kit that included a questionnaire, consent and a tube filled with DNA/RNA shield material (Zymo tube) was given to subjects with specific instructions for sterile collection of stools. Stools were collected directly into the Zymo tube at the patient's home following standardized techniques and processes.

Volunteers were recruited from various parts of the world and had diverse health conditions and diet regimens, and there were no criteria on volunteer eligibility. A standardized process was used to create library preps of each sample. Once DNA was extracted from a 200 μl sample, NGS (shotgun methodology) was performed. Two participants were excluded from the 121 volunteers due to either symptomatic diarrhea from Crohn's Disease or chronic antibiotic use associated with the treatment of *Mycobacterium paratuberculosis* for 1 year. Both individuals had very little bacteria present in their samples for analysis. DNA of the remaining 119 volunteers was quantitated and normalized for downstream library fabrication. Prepared and indexed libraries were subsequently pooled and sequenced on the Illumina NextSeq 550 System. Sample FASTQ files were analyzed with the Kraken 2 taxonomic computational tool that profiles the microbial communities from metagenomic sequencing data with species level resolution. Finally, individual microbiome profiles were analyzed for the presence of *C. difficile*.

Subjects were grouped by disease type for examination of the impact due to their chronic conditions. Treated CDI refers to former infection with CDI. Non-CDI GI refers to any other GI condition or infection. Neuropsych/drugs of abuse refer to any neurological or psychiatric condition or use of recreational drugs. Metabolic-syndrome associated refers to conditions associated with metabolic syndrome, including diabetes or pre-diabetes, hypertension, fatty liver disease, or increased cholesterol or triglycerides. Thyroid refers to hyper- or hypo-thyrodism or thyroid surgery or cancer. Cancer refers to any type of cancer. Subjects frequently were classified in multiple such groups.

A stringent cutoff of 1000 reads was established to qualify as positive for presence of *C. difficile* bacteria in this analysis. A normalization factor was calculated by dividing the *C. difficile* read count of an individual by the individuals' total bacterial read count.

This value was then multiplied by the lowest total bacterial read count among all subjects. Groups were compared using one-way ANOVA and post-hoc as specified.

RESULTS

Figure 1 shows *C. difficile* count (non-normalized figures) above the 1000 count threshold in 117 of 119 (98%) subjects, demonstrating its ubiquitous presence. Figure. 1B compares health care providers (most health care setting exposure), non-providers with chronic illness, and healthy non-providers (least health care setting exposure).

There is no significant difference in *C. difficile* normalized read count among the three groups (p=0.90), demonstrating *C. difficile* normalized read count is independent of health care exposure (similar results seen with *C. difficile* non-normalized count, data not shown).
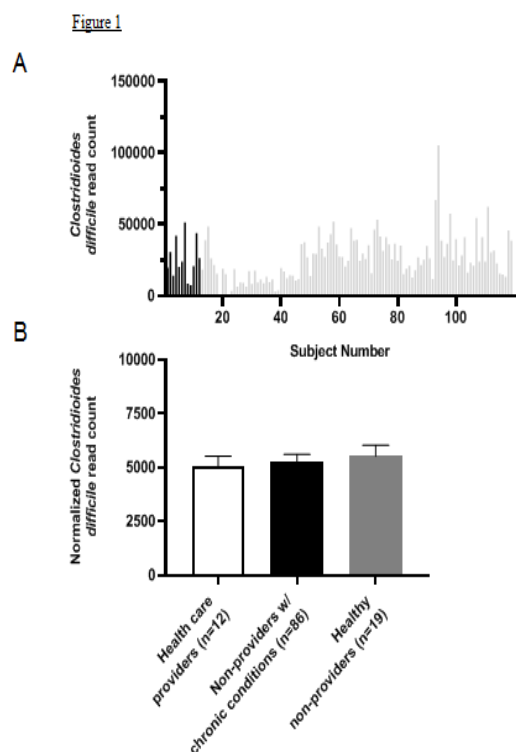

Figure 2: Normalized *C. difficile* read counts were independent of A. Subject age (r2=0.071) or B. Probiotic usage (p=0.39).

Figure 2A shows that *C. difficile* normalized read count is independent of subject age (r2=0.071 for linear regression). Probiotic use also Figure 2B did not affect normalized read count (p=0.39).

Figure 1: *C. difficile* bacterium is present in 98% of subjects analyzed (117/119). (A) Individual *C. difficile* non-normalized read counts for 119 subjects. Black bars indicate health care providers; Light grey bars indicate general population cohort. (B) There was no difference in *C. difficile* normalized read count for health care providers (n=12), non-providers with chronic conditions (n=86), and healthy non-providers (n=19) (*via* one-way ANOVA, p=0.90).
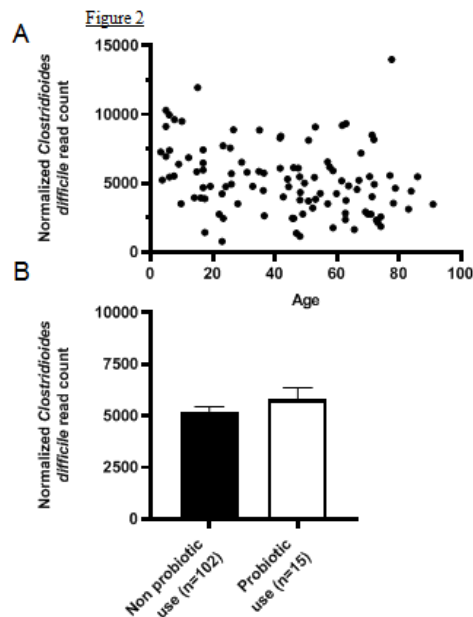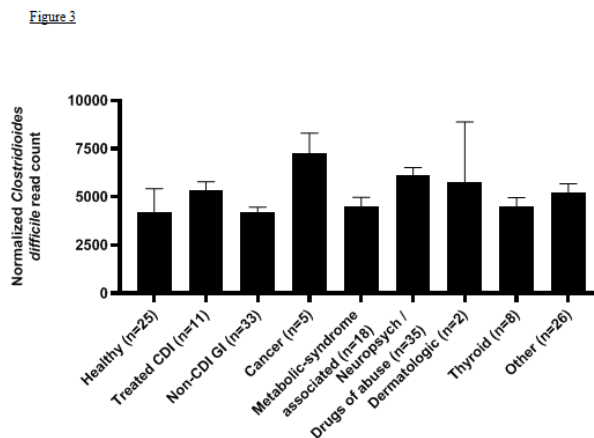
Figure 3: Normalized *Clostridioides difficile* read count was not significantly different based on subjects underlying health conditions (see methods for definitions). Analyzed via one-way ANOVA with Dunnett's post-hoc, no disease group showed significance (p>0.6 for all disease groups, except cancer p=0.10, neuropsych/drugs of abuse p=0.11), although the overall ANOVA was significant at p=0.02.

Figure 3 compares normalized read count for *C. difficile* in subjects with various health histories. When compared individually with control (one-way ANOVA with Dunnett's post-hoc), no disease group showed significance (p>0.6 for all disease groups, except cancer p=0.10 and neuropsych/drugs of abuse

p=0.11), although the overall ANOVA was significant at p=0.02. Thus, levels of *C. difficile* appear to be independent of an individual's health history.

DISCUSSION

Using NGS shotgun sequencing with Kraken meta-genomic analysis, 98% (117/119) of our representative sample was found to be *C. difficile* positive, *i.e.* harbor DNA to achieve at least 1000 reads that align with the *C. difficile* genome Figure 1A. The presence of *C. difficile* was independent of healthcare-setting exposure Figure 1B, age Figure 2A, probiotic use Figure 2B or chronic health conditions Figure 3. Even the two subjects that were *C. difficile* negative still had high read numbers (880 and 785) but below our threshold of 1000 reads. This discovery suggests that non-pathogenic *C. difficile* may make up an important component of the human commensal gut micro biome.

A proper application of NGS is essential for accurate specificity and sensitivity. We are applying kraken meta-genomics type NGS, using a cutoff of 1000 reads, which is more stringent than the typical 500 read cutoff. As reviewed by Chiu and Miller, background signal, whose impact is determined by read cutoff, is a critical determinant of specificity and senility of NGS. Thus, we are using a read cutoff that leads to low sensitivity and high specificity, and still discover a high prevalence of non-toxigenic *C. difficile.*

We hypothesize that *C. difficile* may be present in subjects shortly after birth. Micro biota may be transferred from the mother during delivery, and have lasting effect on future gut flora development [18,19]. Perhaps *C. difficile* is maternally transferred or otherwise acquired in the process of normal gut flora development in nearly every human after birth and remains through life. In this scenario, CDI may not arise from transmission, but rather through activation of toxin production in an opportunistic pathogen as a response to a disturbance in the normal gut flora (e.g. antibiotic use). We realize future studies are needed to explore such a hypothesis.

One cannot ignore that there are numerous studies correlating use of sanitary procedures to reduced CDI. If such stringent procedures are in place, why does the rate of CDI continue to rise? Studies correlating such measures with reduction in infection must be interpreted with caution [20]. There may be a source of increasing CDI, which is outpacing any potential decrease due to sanitary measures. Perhaps this source is induction of toxigenicity within an opportunistic pathogen, *C. difficile*, which we demonstrate as ubiquitously present in human gut flora. Given that presence of *C. difficile* is so high Figure 1A in this outpatient/community study, one could speculate that induction of *C. difficile* toxigenicity could be particularly relevant to the noted increase in rate of community-associated toxigenic CDI [10].

Since *C. difficile* appears to be present in most individuals, one should ask who develops CDI and why? The use of antibiotics (that are not used to treat CDI) is a well-known risk factor for CDI. Could a high endogenous level of *C. difficile* protect the subject, especially in light of its use in treatment [21]? A highly

effective treatment for CDI, Fecal Microbial Transplant (FMT), replaces the gut micro biome with that of a healthy donor and supplies often deficient Bacteroidetes and Formicates comprising the Clostridia class. Increased diversity of the micro biome of the donor in FMT frequently leads to higher success rates of FMT [22]. Thus, one may hypothesize that microbial diversity comprising high levels of Clostridiales mitigates the conversion to toxin production by *C. difficile* that leads to CDI.

Future studies are essential and may help demonstrate that the lack of gut flora microbial diversity is indeed the inciting factor in activating toxin production by this opportunistic pathogen. For instance, one can assess correlation between increased micro biome diversity and toxigenicity or disease severity.

In short, we present a pivotal study demonstrating pervasive presence of *C. difficile* in 98% of subjects analyzed from a community setting, despite health care exposure, age, or pre-existing conditions. The implications of this finding and levels of Clostridiales could be profound, and future studies are needed to ascertain presence and levels of *C. difficile* to better understand modes of transmission or origin of CDI.

Conclusion

NGS provides a unique opportunity to increase the resolution and identification of the bacterium *C difficile* compared to traditional categorizations. Using meta-genomics and a stringent read count criteria (>1000 counts), we deciphered species level resolution of bacteria present, finding *C difficile* in 98% of subjects, which are likely non-toxigenic. This discovery suggests that non-pathogenic *C difficile* may be an important component of the human commensal gut microbiome, possibly present since birth. This raises fundamental questions regarding assumptions of CDI transmission that need future exploration. Given that non-pathogenic *C difficile* has been used for CDI treatment, determining levels of non-pathogenic *C difficile* in stool may be predictive of resistance to development of CDI.

ACKNOWLEDGEMENT

CONFLICTS OF INTEREST

SH, BDB, AP and TJB have corporate affiliation to Progena BiomeTM. SH has additional corporate affiliation to Ventura Clinical Trials. TJB has additional corporate affiliation to Finch Therapeutics Inc, Topelia Therapeutics Inc, Atopic Inc, RedHill BioPharma Ltd. SD has corporate affiliation to McKesson Specialty Health and North End Advisory, LLC.

ETHICAL STATEMENTS

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The trial was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by New

England IRB (now known as WCG IRB; NO: IRB00000533) as protocol PRG-002, and informed consent was taken from all individual participants.

REFERENCES

1. Ghantoji SS, Sail K, Lairson DR, DuPont HL, Garey KW. Economic healthcare costs of Clostridium difficile infection: A systematic review. J Hosp Infect. 2010;74(4): 309-18.

2. Bartlett JG. Historical perspectives on studies of Clostridium difficile and C. difficile infection. Clin Infect Dis. 2008;46(1): 4-11.

3. Schaffler H, Breitruck A. Clostridium difficile from colonization to infection. Front Microbiol. 2018;9: 646.

4. Doing KM, Hintz MS. Prospective evaluation of the Meridian Illumigene loop-mediated amplification assay and the Gen Probe Pro-Gastro Cd polymerase chain reaction assay for the direct detection of toxigenic Clostridium difficile from fecal samples. Diagn Microbiol Infect Dis. 2012;72(1): 8-13.

5. Knetsch CW, Bakker D, Boer RF, Sanders I, Hofs S, Kooistra AM, et al. Comparison of real-time PCR techniques to cytotoxigenic culture methods for diagnosing Clostridium difficile infection. J Clin Microbiol. 2011;49(1): 227-31.

6. Lalande V, Barrault L, Wadel S, Eckert C, Petit JC, Barbut F, et al. Evaluation of a loop-mediated isothermal amplification assay for diagnosis of Clostridium difficile infections. J Clin Microbiol. 2011;49(7): 2714-6.

7. Humphries RM, Uslan DZ, Rubin Z. Performance of Clostridium difficile toxin enzyme immunoassay and nucleic acid amplification tests stratified by patient disease severity. J Clin Microbiol. 2013;51(3): 869-73.

8. Kampouri E, Croxatto A, Prod'hom G, Guery B. Clostridioides difficile infection, still a long way to go. J Clin Med. 2021;10(3).

9. Goyal M, Hauben L, Pouseele H, Jaillard M, De Bruyne K, van Belkum A, et al. Retrospective definition of Clostridioides difficile PCR ribotypes on the basis of whole genome polymorphisms: A proof of principle study. Diagnostics. 2020;10(12).

10. Wood DE, Salzberg SL. Kraken: Ultrafast meta-genomic sequence classification using exact alignments. Genome Biol. 2014;15(3): 46.

11. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol. 2019;14: 319-38.

12. Kujiraoka M, Kuroda M, Asai K, Sekizuka T, Kato K, Watanabe M, et al. Comprehensive diagnosis of bacterial infection associated with acute cholecystitis using meta-genomic approach. Front Microbiol. 2017;8: 685.

13. Chiu CY, Miller SA. Clinical meta-genomics. Nat Rev Genet. 2019;20(6): 341-55.

14. Zhou Y, Wylie KM, Feghaly RE, Mihindukulasuriya KA, Elward A, Haslam DB, et al. Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens. J Clin Microbiol. 2016;54(2): 368-75.

15. De Roo AC, Regenbogen SE. Clostridium difficile infection: An epidemiology update. Clin Colon Rectal Surg. 2020;33(2): 49-57.

16. Redelings MD, Sorvillo F, Mascola L. Increase in Clostridium difficile-related mortality rates, United States, 1999-2004. Emerg Infect Dis. 2007;13(9): 1417-9.

17. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, et al. Burden of Clostridium difficile infection in the United States. N Engl J Med. 2015;372(9): 825-34.

18. Abrahamsson TR, Wu RY, Jenmalm MC. Gut microbiota and allergy: The importance of the pregnancy period. Pediatr Res. 2015;77(1-2): 214-9.

19. Jain N. The early life education of the immune system: Moms, microbes and (missed) opportunities. Gut Microbes. 2020;12(1): 1824564.

20. Barker AK, Ngam C, Musuuza JS, Vaughn VM, Safdar N. Reducing Clostridium difficile in the inpatient setting: A systematic review of the adherence to and effectiveness of C. difficile prevention bundles. Infect Control Hosp Epidemiol. 2017;38(6): 639-50.

21. Seal D, Borriello SP, Barclay F, Welch A, Piper M, Bonnycastle M, et al. Treatment of relapsing Clostridium difficile diarrhoea by administration of a non-toxigenic strain. Eur J Clin Microbiol. 1987;6(1): 51-3.

22. Zanella MC, Simonet ML, Bichard P, Frossard JL. Recurrent Clostridium difficile infections: the importance of the intestinal microbiota. World J Gastroenterol. 2014;20(23): 7416-23.