Running title: Microbiota predict *C. difficile* severity in humanized mice

Kaitlin J. Flynn[1], Nicholas Lesniak[1], Alyxandria M. Schubert[2], Hamide Sinani[?], and Patrick D. Schloss[1†]

† Corresponding author: pschloss@umich.edu

1. Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan 48109

2. Food and Drug Administration?

3. Department for Hamide?

Abstract

Clostridium difficile causes diarrheal disease when it successfully colonizes a dysbiotic gut microbial community. Current mouse models to study *C. difficile* infection (CDI) rely on pre-treatment with antibiotics to disrupt the mouse microbiome prior to inoculation. This model does not allow for analysis of human-associated microbial community members that modulate *C. difficile* colonization and expansion. To study human-associated microbes in the context of CDI, we inoculated germ-free C57BL/6 mice with one of 16 human fecal samples from diarrheal or healthy donors and challenged with C. difficile 14 days later. Five unique donor-mice combinations resulted in severe CDI while the remaining 11 only experienced mild disease. Both healthy and diarrheal donors were susceptible to colonization and severe symptoms of CDI. To determine if specific microbes were associated with disease severity outcomes, we built a classification Random Forest machine learning model based on relative abundance data of the communities prior to infection. The model identified a number of bacterial populations associated with the development of severe CDI, including *Bacilliales, Ruminococcaceae, Ruminococcus, Staphylococcus, Streptococcus and Bacteriodetes*. Additionally, a regression model accurately predicted colonization levels of *C. difficile* at one to ten days post-infection. This model explained 99% of the variance in the number of CFU isolated from mouse stool. Members of *Lachnospiraceae, Parabacteroides, Bacteroidales, Bacteroidetes, Porphyromonadaceae* and unclassified *Bacteria* families were predictive of future *C. difficile* colonization levels. Finally, challenging these mice with different strains of *C. difficile* revealed that susceptible human-associated microbial communities were prone to severe disease independent of strain type. Taken together these results suggest that human-associated microbial communities can be recapitulated in germ-free mice and used to characterize

dynamics of CDI. Because both healthy and diarrheal patients were susceptible to severe CDI, machine-learning models are useful to identify bacterial populations that allow colonization and contribute to the development of *C. difficile* associated disease in humans.

Introduction

Clostridium difficile is an opportunistic pathogen of the human lower gastrointestinal tract. Disruption of the native microbial community of the gut by antibiotics is the most common risk factor for development of C. difficile infection (CDI) (1). C. difficile is a spore-forming bacteria and can persist on abiotic surfaces and is not readily killed by ethanol-based hand-sanitizers, putting hospital patients particularly at risk. Indeed, ~12% of hospital acquired infections in the United States are due to C. difficile and result in up to 15,000 deaths annually (2). This is a test (@Britton12).

Murine models to study CDI typically rely on treating conventionally-raised mice with antibiotics either in drinking water or by injection to induce susceptibility (3, 4). This model provides a convenient way to study C. difficile pathogenesis and virulence factors. Numerous microbiome studies have been performed using this model to determine the antibiotic classes (5), starting microbial community (6) and metabolites (7) that impact development and severity of CDI. While informative, these studies are somewhat removed from human disease because they only examine mouse-associated microbial communities.

Gnotobiotic or germ-free mouse models have been used for a range of studies of CDI, including assessment of species-specific interactions between C. difficile and competing microbial community members (8), analysis of nutrient restriction (9), in vivo transcriptomics of C. difficile and examination of host immune response to CDI (10). Further, CDI therapeutics such as antibiotics and fecal microbiota transplants have been tested extensively in a gnotobiotic-piglet or piglet-to-gnotobiotic-mouse model of disease (11), (12). Pigs have a longer digestive tract with components more similar to humans than mice and are typically infected by strains typical in human infection (13). However, the murine

and porcine microbiomes typically do not resemble those of the human gut.

The power of the gnotobiotic models to study CDI has been further realized by first inoculating germ-free mice and piglets with human stool microbes. In one study, germ-free piglets were acutely colonized with human feces for one week and then treated with tigecycline. After challenge with C. difficile none of the antibiotic-treated piglets succumbed to infection, while some of the untreated human-colonized pigs did (11). Further, germ-free mice colonized with human feces were bred over several generations to create a cohort of mice with identical human-derived microbiomes (14). These mice were subsequently treated with a five-antibiotic cocktail to induce dysbiosis and then were successfully colonized by C. difficile (14). While informative, these studies were limited in their use of only one human donor as input inoculum. In order to best understand the impact of C. difficile pathogenesis on human disease, we must have a laboratory model that allows for study of a variety of human-derived microbiomes.

To test the impact of individual human microbiomes on CDI, we colonized germ-free mice with 16 different human stool donors. We then characterized human-associated microbiome response to C. difficile challenge. Additionally, the use of machine-learning models allowed us to build a predictive model that classified “at-risk” microbiomes prior to infection with C. difficile. These findings show that human-associated microbiomes can be at risk for CDI even in the absence of antibiotics and that study of mice colonized with human feces provides a range of clinical outcomes.

Results

Germ-free mice inoculated with human feces as model for C. difficile infection. To generate mice with human-derived microbiomes, we inoculated one cage of gnotobiotic C57/BL6 mice with one of 16 different human fecal donors. Five donors were patients that had diarrhea that was not attributable to *C. difficile* infection while 11 donors were healthy at time of donation. Stool from a patient that was colonized with virulent *C. difficile* was used as a positive control. After inoculation with human stool, mice were allowed to equilibrate for 14 days. Prior to infection, stool samples were taken from each mouse to establish baseline. Then, the *C. difficile* strain isolated from the positive control patient's sample (strain 430) was used to infect each mouse with 100 spores. Mice were monitored for weight loss and clinical signs of disease. Fecal samples were taken to enumerate *C. difficile* CFU and for microbiome analysis every day for up to 10 days post-infection (Fig 1A). To ensure that the donors we selected represented a diverse array of human microbiomes, we sequenced the 16S rRNA genes from donor fecal inocula. Ordination of the distances between donor communities showed that the donors each had distinctly different communities, independent of whether the sample came from a sick or healthy person (Fig 1B). Likewise, the starting microbial communities of the mice on day 0 were characterizing by sequencing of fecal pellets DNA prior to infection. Ordination of all of the mouse communities on day 0 shows that mice were similar to each other within each cage and donor, but distinct from other donors (Fig 1C). This result confirmed that human-associated microbes were able to colonize gnotobiotic mice and provide distinct initial communities to test *C. difficile* dynamics.

C. difficile infection in mice with human-derived microbiota cause a range

of outcomes. *C. difficile* colonization was monitored by daily plating of stool pellets for *C. difficile* CFU. Nearly all of the mice were colonized to 10^5 – 10^7 CFU by one day post-infection and remained colonized at that level until the end of the experiment (Fig 2A). As one indicator of disease, mouse weights were taken each day post-infection and weight-loss was monitored alongside clinical signs of disease. When mice were judged to be too ill to continue they were humanely euthanized. Overall, disease phenotypes fell into two classes. Mice that became severely ill and lost 20% or more of their starting body weight within one to two days post-infection were classified as "severe" whereas mice that were colonized with *C. difficile* but did not show signs of disease or severe weight loss were considered to have "mild" disease (Fig 2A, 2B). Interestingly, *C. difficile* was able to cause severe disease in both mice that had been colonized with healthy stool and those colonized with diarrheal stool, suggesting susceptibility to CDI is dependent on the composition of the starting microbiome and not associated with donor clinical metadata.

Results to be written 1. Microbes present in the gut prior to infection are predictive of C. difficile CFU and severity a. Figure 3: Random forest to predict CFU, predictive OTUs b. Figure 4: Random Forest predicts CDI severity, predictive OTUs 2. Propensity for severe CDI is community-dependent and strain-independent a. Figure 5: Infection of mice with different C. difficile strains.

Discussion

• Restate results, • caveats about mouse weights • No donors were colonization resistant, discuss donor differences • Discuss prediction methods and outcomes • Discuss potential mechanisms for interesting OTUs • Discuss different strain results • Future work blah blah

Materials and Methods

• Mice ULAM number • Donor stool ERIN IRB shit • Bacteria/plating • Sequencing

• Data analysis, code availability • Machine learning models

Acknowledgments

Lab, sequencing core, Jhansi

Figure Legends

Figure 1. Germ-free mice inoculated with human feces as a model for *C. difficile infection.* A) Experimental design. Stool from 16 healthy, diarrheal and CDI patients were independently inoculated into 3-4 germ-free mice by oral gavage. 14 days later mice were orally infected with 100 spores of C. difficile strain 431. Weight and stool CFU were monitored for up to 10 days post infection. B) NMDS ordination of donor stool communities prior to inoculating mice. Each point represents one donor and are colored by clinical diagnosis. C) NDMS ordination of the stool communities on day 0. Each symbol represents one mouse and is colored by donor. Circles represent mice that experienced mild disease and triangles represent those that suffered severe disease.

Figure 2. *C. difficile* infection results in mild or severe disease. A) *C. difficile CFU* was enumerated by plating of mouse stool pellets daily. Each point represents a mouse and the lines represent the median CFU in each group. Error bars are interquartile ranges. Red lines and points correspond to mice that succumbed to severe disease whereas black lines and points correspond to mice that had mild or no disease. B) Mouse weights were recorded and daily percent weight loss calculated for each mouse. Data is presented as the median of each group and interquartile ranges. Mice that succumbed to severe infection typically lost a significant amount of weight by day 1 or 2 post infection. Red lines correspond to severely ill mice, black to mice with mild disease.

Figure 3. Random Forest prediction of *C. difficile* colonization level. A) OTUs above 1% relative abundance on day 0 were used to predict median $\log_{10}$ CFU of *C. difficile* after colonization. OTUs were chosen such that they were not predictive of cage or donor. Each point is a mouse colored by cage. B) Partial dependency plots of the top six predictive OTUs. Line displats the

partial dependence of $\log_{10}$ CFU on the relative abundance of each predictive OUT. Each median $\log_{10}$ CFU is plotted against its relative abundance for each predictive OTU.

Figure 4. Random Forest prediction of CDI severity. OTUs above 1% relative abundance on day 0 were used to predict disease severity. OTUs were chosen such that they were not predictive of cage or donor. Predictive classification tested via 10-fold (gray), leave-one-cage-out (purple dashed) or leave-one-mouse-out (blue dashed) models are displayed in (A). B) Partial dependency plots of most predictive OTUs. Line displays the partial dependence of $\log_{10}$ CFU on OTU relative abundance. Points are the OTU relative abundance of each mouse colored by outcome (red, severe, black, mild).

Figure 5. Infection of mice with different *C. difficile strains*. 3 strains of *C. difficile* were used to infect mice colonized with susceptible (DA00578) or resistant (DA00369, DA00430) human donor stool. A) *C. difficile* stool CFU was enumerated over 10 days. B) Percent weight loss was calculated each day for each mouse. In both plots, each mouse is a point and lines represent the mean of each cage.

Supplement Table S1: Mouse day 0 communities by donor genera (avg + stdev of cage)

References