Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell Host Microbe. Author manuscript; available in PMC 2022 July 28.

Published in final edited form as:

Cell Host Microbe. 2020 September 09; 28(3): 422–433.e7. doi:10.1016/j.chom.2020.07.020.

Multi-omic analysis of the interaction between *Clostridioides difficile* infection and pediatric inflammatory bowel disease

Frederic D. Bushman[4,10], **Maire Conrad**[1,8], **Yue Ren**[2], **Chunyu Zhao**[1], **Christopher Gu**[4], **Christopher Petucci**[9], **Min-Soo Kim**[9], **Arwa Abbas**[5,6], **Kevin J. Downes**[7,8], **Nina Devas**[1], **Lisa M. Mattei**[1], **Jessica Breton**[1,8], **Judith Kelsen**[1,8], **Sarah Marakos**[1], **Alissa Galgano**[1], **Kelly Kachelries**[1], **Jessi Erlichman**[1], **Jessica L. Hart**[6], **Michael Moraskie**[1], **Dorothy Kim**[1], **Huanjia Zhang**[1], **Casey E. Hofstaedter**[1], **Gary D. Wu**[3], **James D. Lewis**[3], **Joseph P. Zackular**[5,6], **Hongzhe Li**[2], **Kyle Bittinger**[1], **Robert Baldassano**[1,8]

[1]Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[2]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[3]Division of Gastroenterology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[4]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[5]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[6]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[7]Division of Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[8]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA

[9]Metabolomics Core, Cardiovascular Institute, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[10]Corresponding author (Lead Contact): Frederic Bushman bushman@pennmedicine.upenn.edu.

Author Contributions

Experimental design, F.D.B., M.C., U.R., C.Z., K.J.D., N.D., L.M.M., J.P.Z., H.L., K.B., R.B.; Experimental Execution, all authors; Specimen Collection, M.C., K.J.D., N.D., L.M.M., S.M., A.G., K.K., J.E., R.B.; Bioinformatic and Statistical Analysis, F.D.B., Y.R., C.Z., C.G., L.M.M., K.B.; Writing - Original Draft, F.D.B., M.C., Y.R., C.Z., K.J.D., L.M.M., J.P.Z., R.B.

Declaration of interests

The authors declare that they have no competing interests.

ADDITIONAL RESOURCES

Supplemental Information

Supplemental Information includes four figures and four tables and can be found online.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

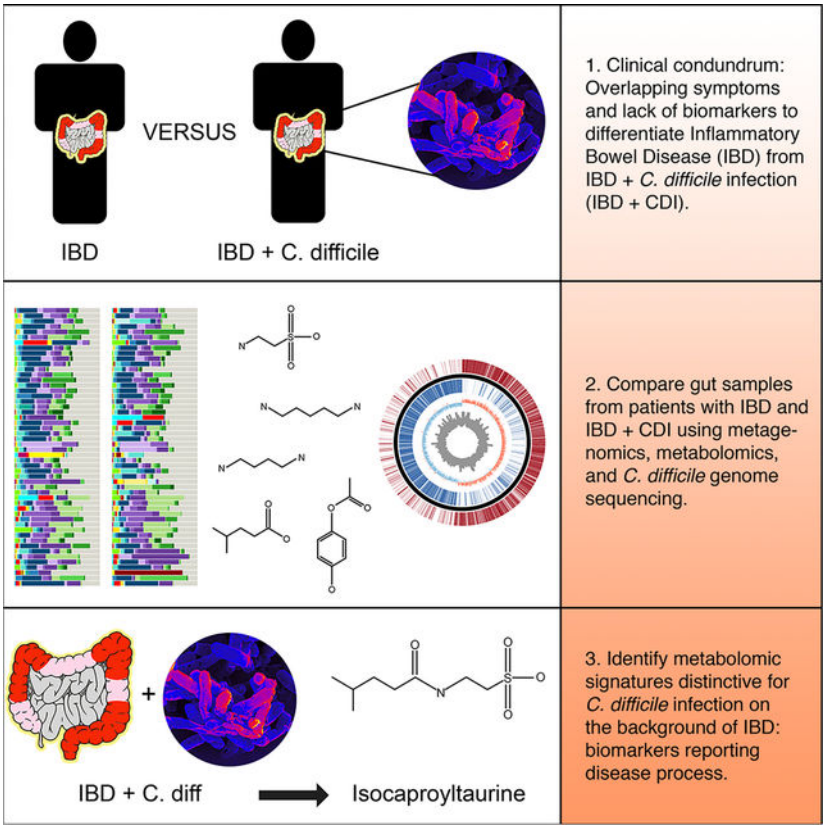
Summary

Children with inflammatory bowel diseases (IBD) are particularly vulnerable to infection with *Clostridioides difficile* (CDI). IBD and IBD+CDI have overlapping symptoms, but respond to distinctive treatments, highlighting the need for diagnostic biomarkers. Here we studied pediatric patients with IBD and IBD+CDI, comparing longitudinal data on the gut microbiome, metabolome and other measures. The microbiome is dysbiotic and heterogeneous in both disease states, but the metabolome reveals disease-specific patterns. The IBD group shows increased concentrations of markers of inflammation and tissue damage compared to healthy controls, and metabolic changes associate with susceptibility to CDI. In IBD+CDI, we detect both metabolites associated with inflammation/tissue damage and fermentation products produced by *C. difficile*. The most discriminating metabolite found is isocaproyltaurine, a covalent conjugate of a distinctive *C. difficile* fermentation product (isocaproate) and an amino acid associated with tissue damage (taurine), which may be useful as a joint marker of the two disease processes.

eTOC blurb

Children with inflammatory bowel diseases (IBD) are prone infection with *C. difficile* (CDI). Bushman et al. report the microbiome and metabolome of children with IBD or IBD+CDI, revealing distinctive metabolites associated with the two disease process that may be useful as biomarkers to guide targeted therapy.

Graphical Abstract



Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

Gram-positive; *Clostridioides difficile* ; microbiome; metabolome; inflammatory bowel disease; isocaproate; taurine; isocaproyltaurine

Introduction

Clostridioides difficile (previously *Clostridium difficile*) is a Gram-positive, spore-forming anaerobic bacteria of the phylum *Firmicutes*. *C. difficile* infection (CDI) is transmitted through persistent spores from toxigenic strains, leading in some cases to colitis, toxic megacolon and death (Abt et al., 2016; Burdon et al., 1978; George et al., 1978). Susceptibility to CDI is increased by antibiotic use, age, and comorbidities including inflammatory bowel disease (IBD), with increased vulnerability due in part to disruptions of normal gut microbial community structure. Incidence of CDI has been increasing, accompanied by greater disease severity and more frequent complications (Abt et al., 2016; Chitnis et al., 2013; Gerding and Lessa, 2015; Kim et al., 2008; Lessa et al., 2015; Ma and Lewis, 2018; Sandberg et al., 2015; Tschudin-Sutter et al., 2013).

The incidence of CDI is elevated in children with IBD, and the rate of CDI recurrence is higher as well (Hourigan et al., 2014; Kelsen et al., 2011; Lessa et al., 2015; Rodemann et al., 2007). This increase in CDI burden may in part be a result of therapies such as acid blocking medications, antibiotics, and immunosuppressive therapy, all of which add to the risk of developing CDI. An additional complicating factor is that it can be difficult to interpret a positive *C. difficile* assay in a patient with IBD, because symptoms overlap between IBD and CDI, and tests for CDI can be positive due to both pathogenic infection and benign colonization. Thus advanced patient phenotyping is needed to separate the influences of IBD and CDI.

Here we implemented a “multi-omics” approach to analyze microbial populations and disease activity in children with CDI and IBD. We and others have previously investigated the roles of the microbiome and metabolome in patients with IBD and in their responses to therapy (Braun et al., 2019; Chehoud et al., 2015; Franzosa et al., 2019; Gevers et al., 2014; Haberman et al., 2014; Kolho et al., 2017; Kostic et al., 2014; Lewis et al., 2015; Ni et al., 2017; Vich Vila et al., 2018). Here, we investigated the additional burden of CDI, with the goals of understanding disease mechanisms and developing potentially useful biomarkers. We analyzed a cohort of children with IBD and recent onset of CDI (IBD+CDI), and compared them to children with active IBD without CDI (IBD cohort), and to two non-IBD control groups, children with malignancy and CDI (M+CDI) and healthy children (HC). The data revealed molecular mechanisms active in each disease process, indicating that 1) IBD produces an environment conducive to subsequent CDI, 2) both IBD and IBD+CDI lead to production of compounds derived from inflammation and tissue destruction, and 3) *C. difficile* metabolic activity is correlated with production of distinctive fermentation products in patients with IBD+CDI. Unexpectedly, we detected a previously undescribed metabolite, isocaproyltaurine, that links taurine produced during inflammation/tissue breakdown with a distinctive *C. difficile* fermentation product (isocaproate). Follow up studies showed that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

C. difficile in pure culture is indeed able to produce isocaproyltaurine, but only in the presence of exogeneous taurine. These findings thus reveal distinctive metabolic processes and provide potential biomarkers of pathogenesis associated with high-level *C. difficile* metabolic activity in IBD.

Results

Clinical Characteristics

Ninety-seven children were enrolled in four study groups (Tables S1 to S3 and Figure S1): healthy children (HC; n=38), children with IBD alone (n=27), children with IBD+CDI (n=23), and children with malignancy plus CDI (M+CDI; n=9). Stool samples were collected at enrollment, which was at the time of positive *C. difficile* testing, and 4 and 8 weeks later. Inflammatory bowel disease activity was quantified at enrollment using the Pediatric Crohn's Disease Activity Index (PCDAI) and Pediatric Ulcerative Colitis Activity Index (PUCAI). The IBD patients were categorized as inactive (n=10), mild (n=17), moderate (n=17), and severe (n=6) based on validated ranges (Hyams et al., 1991; Turner et al., 2010; Turner et al., 2007). Seventy-four percent (17/23) of IBD+CDI patients had moderate or severe clinical disease activity at the baseline visit; 22% (6/27) of IBD group had moderate or severe activity. Disease activity was also measured at time of sample collection by fecal calprotectin (FCP), which was elevated at baseline in both the IBD (median: 1024) and IBD+CDI (median: 1964) groups (Figure 1A; p=NS for comparison between groups). Calprotectin levels decreased during the 8 week study period regardless of *C. difficile* infection (p=0.00026 for the IBD group and p=0.0026 for the IBD+CDI group; linear mixed effects models).

For the CDI groups, treatment of CDI included metronidazole and vancomycin (summarized in Figure S1). Clinical response to antibiotic therapy was seen in 88% (7/8) of the treated M-CDI patients but in only 50% (11/22) the treated IBD+CDI patients. The *C. difficile* burden was quantified using qPCR (Figure 1B) and plating assays (Table S3). *C. difficile* DNA copy numbers showed a trend toward reduced abundance associated with antibiotic treatment (Figure 1B).

To add to our understanding of disease phenotypes, we sequenced *C. difficile* genomes from patients for whom samples were available (n=27, Figure S2 and Table S4), including both IBD+CDI and M+CDI. High-quality genome sequences were generated using combined long read (Oxford Nanopore MinION) and short read (Illumina) sequence data (Table S4). Genomes were heterogeneous, with genome pairs separated by an average of 5879 single nucleotide variants, indicating that infection did not arise from a single nosocomial point source. Twenty-five of the 27 strains encoded toxins A and B.

Below we investigate aspects of the microbiome and metabolome data that differentiates the patient catagories (healthy, IBD, IBD+CDI, M+CDI), then focus particularly on metabolites that distinguish the IBD category from IBD+CDI.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Longitudinal metagenomic analysis

The gut microbiome is altered in IBD and is an established modulator of CDI (Abbas and Zackular, 2020), so we sought to characterize the extent to which gut microbiota structure correlated with health, IBD, IBD+CDI, and M+CDI. Seventeen IBD+CDI patients and nine M+CDI patients were compared to 27 patients with active IBD but without CDI, and to 38 healthy controls. Samples were collected and analyzed by shotgun metagenomic sequencing at enrollment, week 4 and week 8 (Figure 2 A–H). The disease groups were commonly treated with antibiotics, contributing to community disruption and vulnerability to *C. difficile* infection (Figure S1). Fifty-one percent of the 53 IBD, IBD+CDI, and M+CDI patients had recently been treated with antibiotics (assessing the latter two groups prior to CDI diagnosis).

Clustering of samples using the Bray-Curtis dissimilarity metric (Figure 2A–C) emphasized that samples from all patient groups were heterogeneous and departed from the cluster of the healthy controls, with the IBD+CDI group typically most dysbiotic. Most IBD+CDI patients did not experience restoration of normal gut microbiota, despite antibiotic treatment, reduced disease activity (Figure 1A) and a trend toward reduced *C. difficile* burden (Figure 1B). The majority of the IBD+CDI and M+CDI patients were still on antibiotic therapy (11/18) at week 8—likely in part for this reason they clustered furthest from the healthy centroid (Figure 2A–C). Healthy control subjects were commonly high in *Bacteroides* (Figure 2 D–F) and showed higher species richness (Figure 2G), characteristic of reported healthy gut microbiota (Costea et al., 2018; Gill et al., 2006; Lewis et al., 2015; McDonald et al., 2018; Wu et al., 2011). Samples from patients with IBD and IBD+CDI were high in Proteobacteria such as *E. coli* and *K. pneumoniae*, as seen previously in many studies (Braun et al., 2019; Chehoud et al., 2015; Franzosa et al., 2019; Gevers et al., 2014; Haberman et al., 2014; Kolho et al., 2017; Kostic et al., 2014; Lewis et al., 2015; Ni et al., 2017; Vich Vila et al., 2018).

Comparison of the IBD+CDI group to the IBD group allowed assessment of the effects of CDI on microbial community structure. The abundance of *C. difficile* itself was higher in the IBD+CDI group ($p=0.001$; Wilcoxon rank sum test), as expected. However, no difference in species representation between IBD and IBD+CDI achieved significance after correction for multiple comparisons. This is in part a consequence of the heterogeneneity in the structure of the microbiota in individuals in both groups, resulting in insignificant differences in the comparisons of pooled subjects.

We also asked whether microbiome data at baseline could distinguish the pairs of groups using the Random Forest classifier (Liaw and Weiner, 2002). We found that healthy controls could be distinguished with some confidence from both the IBD (area under the curve (AUC)=0.79; out of bag error (OOB)=0.29) and IBD+CDI groups (AUC=0.95, OOB=0.15). However, the IBD group was not well distinguished from the IBD+CDI group (AUC=0.59, OOB=0.45).

Analysis of metagenomic data also allows investigation of gene types enriched or depleted within each subject group. Samples were annotated for gene content, then clustered based on gene representation (Figure S3). Most samples from the IBD and IBD+CDI groups

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

separated from the healthy controls. This separation was not diminished with antibiotic treatment for CDI, paralleling the lack of correction for taxonomic measures (Figure 2A–C).The Random Forest classifier was also used to assess clustering in the gene content data. Again healthy controls could be distinguished with some confidence from both the IBD (AUC=0.76; OOB=0.29) and IBD+CDI groups (AUC=0.90, OOB=0.09). However, the IBD group was not well distinguished from the IBD+CDI group (AUC=0.56, OOB=0.48).

In addition to analyzing bacterial content in the stool samples, we quantified the human DNA percentage, which is a potential biomarker for neutrophil activity, cell turnover and deposition of dead cells in the gut lumen (Figure 2H). Previously we reported that high amounts of human DNA in stool of pediatric IBD patients correlated with dysbiosis (Lewis et al., 2015); here, human DNA was also higher at baseline in the IBD group than in the healthy control ($p=4.0\times10^{-7}$; Wilcoxon rank sum test). At baseline, for the IBD+CDI group, human DNA in stool accounted for an average of 37% of all sequence reads. Human DNA was lower in the IBD alone group despite ongoing disease activity (p value= 0.008; Wilcoxon rank sum test). Human DNA was barely detectable in the samples from healthy controls. The proportion of human DNA in samples from the IBD+CDI group decreased significantly over the 8 weeks studied (p=0.002; linear mixed effect model), associated with the antibiotic treatment.

We thus asked whether data on the percentage of human DNA at baseline could distinguish the pairs of groups. We found that healthy controls could be distinguished with high confidence from both the IBD (AUC=0.87, logistic regression) and IBD+CDI groups (AUC=0.99, logistic regression). The IBD group was also well distinguished from the IBD+CDI group (AUC=0.79, logistic regression). Thus the presence of human DNA in stool provides significant discrimination between all subject groups and may provide a viable marker for disease severity in patients with IBD+CDI.

Longitudinal metabolomic analysis

We next interrogated metabolomic signatures of health, IBD, and IBD+CDI, motivated in part by previous reports of distinctive metabolomic signatures for IBD and CDI alone (Abt et al., 2016; Buffie et al., 2015; Nagao-Kitamoto et al., 2020; Thanissery et al., 2017). Untargeted metabolomic analysis of fecal material was carried out using four mass spectrometry methods to characterize samples from 86 subjects. For most of the subjects, samples were available for the three time points surveyed. A total of 1675 compounds were interrogated for differential representation in the IBD, IBD+CDI, M+CDI, and healthy control groups.

A principal component analysis based on the metabolomic data (Figure 3A–C) showed partitioning by group. Patients with IBD generally separated from the healthy controls, and patients with IBD+CDI mostly formed a broad cluster particularly distant from the healthy controls. The M+CDI patients overlapped with the IBD+CDI patients, emphasizing the potential role of *C. difficile* and antibiotic therapy in the observed metabolic abnormalities. The separation persisted at week 4 and week 8, despite treatment.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The mechanisms underlying differences in steady-state abundance can be complex, as changes in synthesis or degradation rates could be altered, or compounds could be redistributed into different physiological compartments. Though further study is warranted, many of the observed changes support specific mechanistic proposals.

Analysis of bile acid levels

Major primary and secondary bile acids were assessed individually (Figure 3B–E) because of their importance in CDI (Buffie et al., 2015; Sorg and Sonenshein, 2010; Theriot et al., 2014). The abundant primary bile acids, cholate and chenodeoxycholate, were elevated in samples from the CDI-positive groups compared to the CDI-negative groups, and this persisted over the study period ($p=4.0\times10^{-5}$ for cholate, $p=5.5\times10^{-5}$ for chenodeoxycholate; linear mixed effects model). Secondary bile acids showed a mixture of profiles, but the major *C. difficile*-inhibitory secondary bile acids, deoxycholic acid and lithocholic acid, were present at low abundance in the CDI groups ($p=6.2\times10^{-7}$ for deoxycholic acid, and $p=4.2\times10^{-7}$ for lithocholic acid; linear mixed effects model). Despite treatment, the primary bile acids remained at high levels, and the secondary bile acids at low levels at later time points, potentially consistent with ongoing antibiotic therapy in the CDI group inhibiting growth of bacteria that convert primary into secondary bile acids.

A comparison of samples from healthy controls and IBD patients at baseline showed that several primary bile acids were higher in abundance in IBD, including cholate and chenodeoxycholate (cholate: p=0.0001, q=0.0002; chenodeoxycholate: p=0.0013, q=0.0014, linear regression). Because primary bile acids can serve as germinants for *C. difficile* (Abt et al., 2016; Buffie et al., 2015), these findings suggest the IBD patients may have a greater vulnerability to CDI.

Metabolomic comparison of samples from healthy controls and IBD patients

Comparison of the metabolomic data from the healthy controls and IBD patients revealed numerous metabolites distinguishing the two groups (Figure 4A). Overall, a Random Forest model generated using metabolomic data discriminated the healthy subjects from IBD patients with good accuracy (AUC=0.90, OOB=0.21). The single most discriminatory molecule was ceramide, which was enriched in the IBD samples. Ceramide is a breakdown product of sphingomyelin, an abundant host cell lipid, which was reported to be associated with apoptosis and regulation of inflammation (Haimovitz-Friedman et al., 1994; Melum et al., 2019). Sphingomyelin and related molecules were also enriched (Franzosa et al., 2019), consistent with a model in which these compounds were released from lysing of host intestinal cells. A recent paper reported that ceramide and sphingolipids could be synthesized by Bacteriodetes, raising the possibility of a bacterial origin (Johnson et al., 2020)—however, high levels of Bacteriodetes are characteristic of the healthy gut and not IBD, and so it seems more likely that ceramide and sphingolipids were produced by destruction of intestinal cells. Another molecule enriched in the IBD samples was cadaverine, a polyamine produced by decarboxylation of lysine, which has been previously implicated in tissue degradation (Izquierdo et al., 2018). In addition, several amino acids were present at higher levels in the IBD samples compared to healthy controls. *C. difficile* can carry out an unusual metabolic process to generate energy by fermentation of amino

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

acids via the Stickland pathway, in which amino acid molecules are both oxidized and reduced to generate ATP. Thus the presence of increased levels of amino acids in IBD likely results in further vulnerability to *C. difficile* infection.

Metabolomic comparison of samples from healthy controls and IBD+CDI patients

Comparison of healthy controls to the IBD+CDI samples (Figure 4B) showed numerous compounds with strong differences between groups at the baseline visit. A Random Forest model comparing these two groups showed 99% accuracy in discriminating the two (AUC=0.99, OOB=0.01). The catabolic polyamine putrescine was strongly enriched in the IBD+CDI samples, as were multiple lipids associated with animal cells (e.g. sphingomyelins and lipids with polyunsaturated carbon chains), consistent with breakdown of human cells due to the disease process, paralleling the release of calprotectin and human DNA into stool. Heme was enriched in IBD+CDI samples and L-urobilin, a breakdown product of heme, was enriched in the healthy controls. This supports previous findings that CDI-mediated tissue damage increases heme levels in stool, and further suggests that CDI is associated with decreased capacity for heme degradation (Knippel et al., 2018). Taurine was also enriched in the IBD+CDI samples—taurine is an abundant sulfur-containing amino acid that is present at low concentrations in plasma but high concentrations in cells (Schuller-Levis and Park, 2003). Taurine is released from cells upon lysis, providing another marker of probable tissue damage.

Metabolomic comparison of samples from IBD patients and IBD+CDI patients

Metabolomic comparisons of the IBD group to the IBD+CDI group are of particular interest, because pathogenic processes due to CDI and IBD can be similar, but treatments differ, creating a need for distinguishing biomarkers (Figure 4C). The metabolomic data identified compounds that differed robustly between groups (Random Forest AUC=0.85; OOB=0.26). Taurine and several of its derivatives were enriched in the IBD+CDI group, likely reflecting inflammation and cell lysis associated with disease (Levy et al., 2015; Marcinkiewicz and Kontny, 2014). Multiple lipids characteristic of human cells were also detected, potentially also from host cell lysis. We also observed a modest enrichment in succinate, a metabolite previously associated with *C. difficile* colonization and virulence, in the IBD+CDI group (Ferreyra et al., 2014).

As mentioned above, *C. difficile* has the capacity to metabolize amino acids via Stickland fermentation during infection (Bouillaut et al., 2013; Neumann-Schaal et al., 2019; Robinson et al., 2019) and this is associated with colonization and virulence (Jenior et al., 2017; Martin-Verstraete et al., 2016). Thus, it has been postulated that amino acid availability in the gut is associated with susceptibility to CDI (Battaglioli et al., 2018; Bouillaut et al., 2013; Fletcher et al., 2018; Martin-Verstraete et al., 2016; Neumann-Schaal et al., 2019). Consistent with this, we observed two Stickland fermentation products, 4-hydroxyphenylacetate (fermentation product of tyrosine) and isocaproate (fermentation product of leucine), enriched in the IBD+CDI group (Neumann-Schaal et al., 2019). How much of these products is contributed by *C. difficile* versus other bacteria is unknown; however, we note that Stickland fermentation is not a common pathway in bacteria, consistent with *C. difficile* as a major source. Analysis of the genome sequences of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

27 *C. difficile* strains from our patients showed that all encoded the *ldhA/hadA* regulon that encodes the enzymes for Stickland fermentation of leucine (Kim et al., 2006).

To our surprise, we also detected a previously undescribed molecule, isocaproyltaurine, comprised of covalently joined isocaproate and taurine. Isocaproyltaurine was the single strongest discriminator between the IBD and IBD+CDI samples (Figure 5A). This molecule is of particular interest, since it potentially reports both inflammation and tissue degradation (taurine) and *C. difficile* metabolic activity (isocaproate). Longitudinal metabolomic analysis showed that isocaproate was seen at high levels in baseline samples selectively in the IBD+CDI and M+CDI groups, and that levels fell subsequently with antibiotic therapy for CDI (Figure 5B). For isocaproyltaurine (Figure 5C), levels were far higher in the IBD+CDI group than the M+CDI group, potentially reflecting a joint requirement for gut-mediated production of taurine and *C. difficile*-mediated production of isocaproate. Levels fell sharply with time after treatment of *C. difficile* infection. To investigate the origin of isocaproyltaurine further, we carried out targeted studies of *C. difficile* in pure culture.

Production of isocaproyltaurine by *C. difficile* in pure culture in the presence of added taurine

To test whether *C. difficile* has the capacity to produce isocaproyltaurine, we grew *C. difficile* in pure culture in minimal media with or without added taurine and quantified accumulation of isocaproyltaurine (Figure 6). We hypothesized that the conjugation reaction to make isocaproyltaurine took place inside *C. difficile* cells following taurine import (Figure 6A), based in part on the finding that *Clostridium butyricum* (Xing et al., 2019) encodes a taurine importer. Analysis of the collection of *C. difficile* full genome sequences from our patients showed that genes encoding the putative taurine importer (tauABC) were present in all 27 *C. difficile* genome sequences (Table S4).

To assay isocaproyltaurine in culture, we first synthesized the molecule to generate a standard by activating isocaproate via conjugation of the carboxylic acid to EDC, then reacting the intermediate with taurine to form isocaproyltaurine (Figure S4). *C. difficile* minimal media contains leucine and other fermentable amino acids, but does not contain taurine. Comparison of the minimal media (Figure 6B) to the synthetic isocaproyltaurine standard (Figure 6C) showed that minimal media did not contain isocaproyltaurine, nor did isocaproyltaurine appear in media after growth of *C. difficile* to stationary phase (Figure 6D). However, supplementation of media containing *C. difficile* with 1 mM or 10 mM of taurine resulted in robust production of isocaproyltaurine (Figure 6E and F). More isocaproyltaurine was produced in the presence of the higher taurine concentration. Isocaproyltaurine was not present in the taurine preparation itself (Figure 6G). We conclude that *C. difficile* can take up exogenous taurine, direct its conjugation with isocaproate, and excrete the product molecule. Together these data suggest that isocaproyltaurine may provide a useful biomarker for active CDI in patients with IBD.

Discussion

Here we investigated *C. difficile* infection (CDI) in a particularly vulnerable population, children suffering from inflammatory bowel disease (IBD), and used multiple forms of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

high throughput analyses to assess biomarkers distinguishing disease processes in the IBD versus IBD+CDI groups. We further compared these groups to healthy control children and children with malignancy and CDI. The study combines longitudinal analysis of gut metagenomic and metabolomic profiles of over 97 subjects, *C. difficile* genome analysis, and follow up in vitro assays. Our results specify the data types most discriminatory among patient groups, identifying multiple disease processes and potential diagnostic biomarkers. Though we focused on IBD versus IBD+CDI, our results address numerous additional questions—we thus provide a detailed "Data Encyclopedia" containing many further comparisons in the ONLINE METHODS (Data S1: Data Encyclopedia), allowing extensive additional user-driven investigation.

We focused on signals that discriminated the IBD and IBD+CDI groups, because clinical symptoms can overlap between the two disease process, but therapies for IBD and CDI diverge. For an exacerbation of IBD, a clinician might increase immunosuppressive therapy, whereas in the presence of active CDI a clinician might make a different choice. We used the Random Forest classifier to assess the discrimination possible with data on gut microbial taxonomy, gut microbial gene content, human DNA content, and metabolomic analysis (Figure 5). Although all conclusions need confirmation in independent data sets, the most effective discriminator between IBD and IBD+CDI was the metabolomic data (AUC=0.85). Evidently effects on the microbiome due to IBD and IBD+CDI were sufficiently heterogeneous that a strong distinguishing signal was not evident; in addition, both groups had extensive antibiotic exposure, likely affecting the community composition in a similar fashion in each. In contrast, using metabolomics data, the IBD+CDI group was distinguished by increased production of both products of tissue breakdown (e.g. human-derived lipids, taurine) and products of *C. difficile* fermentation of amino acids by the Stickland pathway (e.g. isocaproate and 4-hydroxyphenylacetate). Consistent with this, the next most distinguishing signal after the metabolomic data was actually the percentage of human DNA in metagenomic sequence data (AUC=0.79), again reflecting increased breakdown of human tissue in the IBD+CDI group relative to the IBD group.

The single compound that best distinguished between the IBD+CDI and IBD groups was isocaproyltaurine, which has not been reported previously. Isocaproyltaurine was only present at high levels in the IBD+CDI group at baseline, and declined with successful therapy (Figure 5). We tested production of this molecule by *C. difficile* in pure culture and found that it was synthesized efficiently, but only in the presence of exogeneous taurine. A putative taurine importer is evident in our genome sequence data for *C. difficile,* suggesting that taurine is being imported into *C. difficile* cells, and synthesis of isocaproyltaurine is taking place within cells. Thus the concentration of isocaproyltaurine reports both metabolic activity of *C. difficile* (isocaproate) and damage to gut cells and inflammation (taurine). Analysis of 27 *C. difficile* full genome sequences from our patients showed that genes for Stickland fermentation of leucine to generate isocaproate were conserved in all isolates. We suggest that assays for isocaproyltaurine or similar products may be useful in distinguishing disease process involving IBD only (low concentrations) from disease involving both IBD and active CDI (high concentrations).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

We speculate that isocaproate covalently bound to acetyl-CoA may represent the intermediate in *C. difficile* that is attacked by taurine in the conjugation reaction. In vivo, Stickland fermentation of leucine is known to proceed through a covalent acetyl-CoA intermediate (Kim et al., 2006), potentially providing the substrate for attack by taurine. In possible support of this idea, we also observed two fatty acids conjugated to taurine that are enriched in IBD+CDI versus IBD alone, N-palmitoyltaurine and N-oleoyltaurine (Figure 4C). Fatty acids are also synthesized in pathways involving acetyl-CoA or acyl carrier protein-linked covalent intermediates, so synthesis of these two taurine-linked lipids may proceed by a similar reaction mechanism involving nucleophilic attack by taurine on the respective covalent intermediates.

This study has several limitations. The patients studied were heterogenous for various clinical features, and many had complex prior treatment histories. The antibiotic regimens used for treatment varied. For the malignancy + CDI (M+CDI) group, we had very few late time points, so no conclusions can be drawn for the longitudinal analysis of this group. Lastly, our sample set was from a single center, and no validation cohort was studied. Future work to address these limitations would be useful.

In summary, analysis of this large data set showed that metabolomic data was the best discriminator of patients with IBD versus IBD+CDI. The origin of differences were understandable as increased breakdown of human tissue, production of distinctive fermentation products by *C. difficile*, and covalent association of isocaproate and taurine to make isocaproyltaurine. These findings provide a framework for understanding and distinguishing disease process in IBD and CDI that may ultimately be useful for guiding therapeutic choices.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contacts, Frederic D. Bushman (bushman@pennmedicine.upenn.edu).

Materials Availability—This study did not generate new unique reagents.

Data and Code Availability—The 29 *C. difficile* genome sequences and their episomes were deposited at NCBI under PRJNA524299. Metagenomic DNA sequence data is deposited under PRJNA562600. Metabolomic data and computer code used in this study are available at https://github.com/zhaoc1/nanoflow, https://github.com/zhaoc1/coreSNPs, and https://github.com/reny1/cdiff.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population and classification of *C. difficile* severity—Subjects were recruited at the Children's Hospital of Philadelphia (CHOP) from September 2015 to April 2018 (IRB approval number 15-011817). Groups included 1) healthy children (HC), 2) children with IBD, (IBD) 3) children with IBD and concurrent CDI (IBD+CDI), and 4)

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

children with malignancy and CDI (M+CDI). Subject data is in Tables S1–S3. Healthy children and IBD only patients were age matched to those with CDI. For inclusion in the *C. difficile* infected groups, patients needed to 1) have an underlying diagnosis of inflammatory bowel disease or malignancy 2) have loose, watery stools at least 3 times per day for >24 hours or an increase in stool frequency from baseline and 3) be positive by glutamate dehydrogenase (GDH) testing assay or PCR for the presence of the toxin A/B gene (cycle of threshold <40). Clinical manifestations of *C. difficile* infection ranged from mild diarrhea to fulminant colitis and shock. Patients with CDI were identified by notification of positive testing for toxigenic *C. difficile* by the CHOP Microbiology laboratory. Exclusion criteria included history of hemophagocytic lymphohistiocytosis or Langerhan's cell histiocytosis or concurrent gastrointestinal infection proven by clinical testing. Prior history of *C. difficile* or recent antibiotic use were not exclusion criteria. Inclusion criteria for healthy controls included if they did not have a chronic diagnosis and could provide stool samples at all 3 time points. Healthy children were excluded if they had diarrhea, antibiotic use in 90 days prior to enrollment, or a family member with *C. difficile* enrolled into the study.

Patients with CDI were included with either primary infection or recurrent infections. There were 10/23 patients in the IBD+CDI cohort who had recurrent CDI at enrollment, three of whom had a prior positive clinical test for *C. difficile* within three months of enrollment. In the M+CDI cohort there were 3/9 patients with recurrent CDI, of whom two were within 3 months of enrollment.

Of the IBD patients, 89% of the IBD alone group had IBD involving the colon, while this was the case in 95% of the IBD+CDI group. Other disease characteristics are outlined in Table S2.

Antibiotic choice and duration of treatment was at the discretion of the treating physician (summarized in Figure S1). Clinical response to antibiotics was defined as resolution of diarrhea or return to baseline symptoms at the end of the antibiotic course and a sustained effect off antibiotics at the end of the study period. In the IBD+CDI cohort, 3 patients were treated with metronidazole, 16 were treated with enteral vancomycin, and 2 were treated with IV metronidazole initially for 3 days but completed treatment with enteral vancomycin. One patient did not receive treatment due to lack of clinical testing at time of enrollment but was found to be PCR positive on further non-clinical investigations and had increased symptoms at the time of study enrollment. In the M+CDI group, one patient was initially treated with metronidazole and was later changed to vancomycin. Seven patients were treated with vancomycin exclusively, and one patient with M+CDI did not receive treatment at the discretion of the treating physician. Treatment courses included 10 days, 14 days, 6 week taper, or prolonged (>4 weeks) dosing. One patient in the IBD group and one in the M+CDI group did not complete the treatment course. Of the three IBD+CDI patients who were the most similar to healthy controls by visit 3 (Figure 3), two had had a complete response to metronidazole and one had no treatment due to presumed colonization and not infection. The remainder of the IBD+CDI patients (n=12) with visit 3 samples available for analysis were treated with vancomycin. Seventeen percent of the IBD+CDI group sample longitudinally had antibiotic exposure within the preceeding 10 days, as did all of the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

M+CDI patients. Five of the IBD+CDI patients were treated with further antibiotics for additional indications including severe colitis (n=4) or pneumonia (n=1).

C. difficile testing was performed by the CHOP Microbiology Laboratory using a two-step process. First, samples were tested for glutamate dehydrogenase antigen and toxin A/B antigen using the C. DIFF QUIK CHEK COMPLETE® test (Alere North America, Inc.). Samples were considered positive if both GDH and toxin antigens were detected. Those that were inconclusive by this method (GDH antigen positive, toxin negative) were reflexed to a molecular detection method which evaluated the presence of toxin genes using loop-mediated isothermal DNA amplification (LAMP) via the illumigene® platform (Meridian Bioscience, Inc.). To confirm CDI classification, all assignments were then verified by PCR targeting the toxin B gene, resulting in revision of assignment for one patient. In addition *C. difficile* was cultured from stool samples to quantify colony forming units per gram (see main text and Figure S1). There was no significant difference in the enrollment *C. difficile* burden in the baseline samples between clinical responders or nonresponders to antibiotic therapy.There were only two patients on proton pump inhibitors and two patients on systemic steroids at enrollment, so it was not possible to assess the impact of these exposures on *C. difficile* burden. For the IBD+CDI group, enrollment and collection of the first sample was within 24 hours of the positive C. difficile test. The *C. difficile* burden was not correlated with FCP or clinical disease activity score (Spearman rho, *C. difficile* abundance versus FCP= −0.10, p = 0.75; rho *C. difficile* abundance versus clinical score=0.19, p = 0.47). Patients with CDI who had current or recent antibiotic use had a significantly higher *C. difficile* burden (p=0.007; Wilcoxon Rank Sum test).

C. difficile isolates were collected from stool of 27 pediatric patients with inflammatory bowel disease or malignancy who had positive testing for toxigenic *C. difficile* by the CHOP Microbiology Laboratory. For analysis of the microbiome and metabolome, longitudinal fecal samples were obtained from additional samples at time 0, then at 4 and at 8 weeks post-diagnosis (91 total subjects studied by metagenomics and 86 studied by metabolomics). *C. difficile* positive subjects with malignancy were included even if they only submitted baseline samples.

Patients with IBD were categorized as inactive, mild, moderate, or severe inflammatory bowel disease based on baseline PCDAI or PUCAI scores according to previously published criteria and ranges (Hyams et al., 1991; Turner et al., 2007). Clinical manifestations of *C. difficile* infection (CDI), range from mild diarrhea to fulminant colitis and shock. Several scoring systems have been developed in adults based on the likelihood of developing complications, such as need for intensive care or colectomy or death, in order to guide treatment selection (i.e. metronidazole vs. vancomycin) (Surawicz et al., 2013; Zar et al., 2007). There are no well-established, equivalent scoring systems for children. Therefore, to categorize the severity of CDI in our study, we dichotomized patients into severe vs. non-severe according to the presence of two or more of the following criteria (adapted from adult scoring systems): a) abdominal pain, b) hypoalbuminemia for age, c) leukocytosis for age, d) hypotension for age, e) fever (>38.3 °C), and f) elevated serum creatinine for age. All criteria had to be present on the day that stool was collected for *C. difficile* testing. Only liquid stools were used for *C. difficile* testing, therefore quality/quantity of diarrhea was not

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

considered a criterion for our severity classification. We characterized CDI patients as severe (n=16) or non-severe (n=16) at presentation using a modified adult clinical scoring system that categorized patients based on their risk for complications.

Given the large scale of the data, we include a comprehensive “Data Encyclopedia” in the supplementary information, which presents analysis over many further parameters in the data, allowing investigation of a wealth of additional questions.

METHOD DETAILS

Calprotectin—Calprotectin levels were measured from fecal samples using the QUANTA Lite Calprotectin Extended Range ELISA assay kit (Inova Diagnostics, San Diego, CA), strictly in accordance with the manufacturer’s protocol. To extract samples, 100 mg feces was mixed 1:50 (w/v) with extraction buffer, vortexed for 30 seconds, and then homogenized for 25 minutes on a shaker. One milliliter of the homogenate was centrifuged for 20 minutes at 10,000 × g and the cleared supernatant was stored at −20 °C until the ELISAs were performed. The ELISAs were conducted on cleared supernatants diluted 1:400 in the provided dilution buffer. The supplied calibrators and controls were run alongside the samples on each plate. Absorbance measurements at 450 nm were taken on the EnSpire Multimodal Plate Reader (Perkin Elmer, Waltham, MA) and calprotectin levels were calculated based on a standard curve generated using the 4PL method.

Isolation of *C. difficile* strains—*C. difficile* was isolated from CDI-positive stool samples by inoculating CCMB-TAL media (Anaerobe Systems, Morgan Hill, CA, USA) with 0.5 mL stool that had been heat shocked for 10 minutes at 80 °C. After 48 hours of anaerobic incubation at 37 °C, the CCMB-TAL was subcultured to a CCFA-HT plate (Anaerobe Systems) and incubated anaerobically for an additional 48 hours. Colonies were then subcultured to a Brucella blood agar plate (Oxyrase, Mansfield, OH, USA) and incubated for 48 hours, after which bacterial lineages were identified by Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) Mass Spectometry using the Microflex instrument (Bruker Scientific, Billerica, MA, USA).

***C. difficile* genomic DNA sequence acquisition and assembly—**DNA from *C. difficile* isolates was extracted from 25 mL of overnight BHIS liquid cultures using a previously described protocol with modifications (Bouillaut et al., 2011). Briefly, cells were harvested by centrifugation, washed with TE and then incubated at 37 °C for 2 hours in 250 μL sucrose lysis buffer (0.8 M sucrose and 10 mg/ml lysozyme in TE). The mixture was incubated at 37 °C for an additional one hour after the addition of 100 μL 20% Sarkosyl and 15 μL RNAse A (10 mg/mL), and another 30 minutes after the addition of 15 μL proteinase K (10 mg/mL). The mixture was brought up to 600 μL with TE and then mixed with an equal volume of 25:24:1 (v/v/v) phenol/chloroform/isoamyl alcohol. After centrifugation for 10 minutes, the aqueous phase was mixed with an equal volume of 24:1 (v/v) chloroform/isoamyl alcohol to remove excess phenol. After centrifugation for 10 minutes, DNA was precipitated from the aqueous phase with 50 μL 3 M sodium acetate, pH 5.2 and 3 volumes of cold 95% ethanol. The precipitated DNA was washed once with 70% ethanol and the air-dried pellet was resuspended overnight in 50 μL EB (10 mM Tris, pH 8.5).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For long-read sequencing, libraries were prepared using versions SQK-RBK001 and SQK-RBK004 of the Rapid Barcoding Kit (Oxford Nanopore Technologies, Oxford, UK), according to the manufacturer's directions. Sequencing took place on the Oxford Nanopore MinION Mk1B using FLO-MIN106 flow cells and MinKNOW version 1.13.1.

For short-read sequencing, libraries were prepared with DNA that had been fragmented to an average of 550 bp on the Covaris E220evolution Focused-ultrasonicator (Covaris, Woburn, MA) using the TruSeq DNA PCR-Free Library Prep Kit (Illumina, San Diego, CA). Sequencing took place on the HiSeq 2500 (Illumina, San Diego, CA) to obtain 2×125 bp paired-end reads.

Twenty-seven genome sequences were determined. As controls for the frequency of error in the sequence determination, one genome was sequenced twice as a technical replicate (Cd4a and Cd4b), and one individual was sampled twice longitudinally (Cd12 and Cd13). These pairs serve as benchmarks to assess the numbers of SNVs accumulating due to error (Cd4a and Cd4b) or variation in the patient gut *C. difficile* population (Cd12 and Cd13).

Sequence assembly was carried out as follows. The Illumina short reads were quality controlled using the metagenomic analysis pipeline Sunbeam (Clarke et al., 2019). The long read sequencing data were processed using our in-house pipeline named nanoflow (https://github.com/zhaoc1/nanoflow) assisted by SAMtools (Li et al., 2009). The raw fast5 files were basecalled using ONT's Albacore command line tool (v.2.2.7), with barcode demultiplexing and fastq output. Adapter sequences were trimmed using Porechop (https://github.com/rrwick/Porechop)(Wick et al., 2017). Only reads for which Albacore and Porechop agreed on the barcode bin were retained. Only reads longer than 2 kbps was used. Each read set was subsampled down to 500 Mbps of high quality reads using Filtlong (https://github.com/rrwick/Filtlong). The hybrid assembly step used Unicycler (v0.4.6) with the option of '-- existing_long_read_assembly'. Briefly, Unicycler performs a short-reads only SPAdes assembly of the Illumina reads, and scaffolds the assembly graph using the long-read only assemblies. The long-read only assembly was carried out using Canu (Koren et al., 2017) and the consensus sequences were polished by Nanopolish (Wick et al., 2017) using signal-level data. As the last step, Unicycler (Wick et al., 2017) polishes its final assembly with Illumina reads using multiple rounds of Pilon (Walker et al., 2014) to reduce the rate of errors. For the hybrid assembled contigs, the largest linear/circular contigs were assigned as the chromosome contigs, and any smaller contigs that mapped back to the chromosome contigs were removed. The resulting 29 genome sequences and their episomes were deposited at NCBI under PRJNA524299.

Genome annotation, pangenome analysis and phylogenetic analysis—For comparison to our complete genome sequences, we assessed 1171 *C. difficile* genomes deposited in NCBI RefSeq (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/Clostridioides_difficile/). We used checkM (Parks et al., 2015) to estimate the genome quality of all 1171 genomes, and only kept draft genomes with *Completeness* > 0.97, *Contamination* < 0.02 and *Strain heterogeneity* < 0.25. To further removed any distantly related genomes, we removed any genomes with *GC* > 30 or *Genome_Size* > 4.6 Mbps. We kept 770 high quality *C. difficile* genomes for further analysis. To remove highly similar

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

genomes, we ordered the 770 genomes based on *Genome Score,* calculated by *dRep* (Olm et al., 2017), and then heuristically selected non-redundant genome sequences based on the pairwise nucleotide identity (*ANIm*) values. Any genomes with sequence similarity cutoff > 0.99 and minimum coverage of the larger genome > 0.95 to the representative genomes were removed. ANIm was used, which aligns whole genome fragments and calculates the nucleotide identity of aligned regions. After de-replication, we had 143 representative *C. difficile* genomes for downstream analysis. Multilocus sequencing typing (MLST) (Griffiths et al., 2010) category was determined for each of our genomes and the 143 database genomes using mlst_check.

Annotation of the assembled draft genome sequences was carried out using Prokka (Seemann, 2014) against a collection of annotated proteins from all the *C. difficile* strain from RefSeq as a trusted source for annotation. In addition, a set of genes of particular interest was verified and curated manually. To annotate the plasmids, we used *blastn* (Altschul et al., 1990) to compare the small assemblies to the RefSeq plasmids database (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/plasmid/).

The pangenomes of our 29 genomes and 143 non-redundant representative RefSeq genomes were built using Roary (Page et al., 2015). There are 2068 core genes, defined as being present in 99% among all the 172 *C. difficile* genomes, out of 16843 total genes. A phylogenetic tree based on the core genome alignments was build using Fasttree (FastTree -nt -gtr) (Price et al., 2009).

To quantify genome similarity between all our Clade 1 isolates, we quantified the single nucleotide variants SNVs for the core genomes among the 27 isolates, focusing on the 2498 core genes out of 8774 total genes. We identified SNVs using three steps: 1) multiple sequences alignment was carried out for each core gene using Muscle (Edgar, 2004); 2) for each MSA file, SNVs were identified using snp-sites (Page et al., 2016); 3) all the identified SNVs for each core gene for each sample were concatenated, and pairwise hamming distance were calculated. The code used is available on https://github.com/zhaoc1/coreSNPs.

Toxigenic strains of *C. difficile* are characterized by the presence of two toxins (tcdA/B) encoded in the 19.6 KB pathogenicity locus. Genomic analysis revealed that this region was present in all but two isolates (Cd-19 and Cd-27). Neither of these genomes were fully completed, and Cd-27 was positive by qPCR for tcdB. Thus, it is unclear whether these two were truly negative. An additional possibility is that multiple strains colonized one or both of these patients, and non-toxogenic strains were isolated by chance from a mixture.

The numbers of different episomes ranged from 0 to 4 per isolate. Of these, 47/49 were fully circular in the hybrid assemblies, suggestive of complete episomal genomes. Their apparent copy number relative to the main chromosome, estimated from coverage values, ranged from 0.89 to 13.59. Thirty-eight matched previously known episomes, while 11 were new here. Several annotated as containing multiple bacteriophage genes (including heads, tails, integrases, and partition functions), suggesting episomal phage replication. The presence of PAR gene homologs on low copy number phage-like episomes suggested these may maintain the prophage state as extrachromosomal circles, potentially paralleling mechanisms

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

known for *E. coli* phage P1 (Surtees and Funnell, 2003). Episome content was not strongly associated with disease severity.

To connect our findings to studies of CDI more widely, we also performed MLST by interrogating the genomic data. Much of the modern burden of CDI in hospital settings is associated with the rise of strain MLST1, ribotype 027 (RT027), which is a member of clade 2 (McDonald et al., 2005) and MLST11, ribotype 078, a member of clade 5 (Goorhuis et al., 2008). Among the 27 isolates reported here, Cd-1 belonged to clade 2 and Cd-17 belonged to clade 5, while the other 24 belong to clade 1. Cd-1 is a member of MLST188 and thus not a member of the hypervirulent MLST1/RT027 strains. Cd-17 is related to hypervirulent strains belonging to MLST11/RT078. This was consistent with disease severity as Cd-17 caused severe disease while Cd-1 did not. There was no association between clade and therapeutic response, though our power to detect differences was modest.

We compared our distribution of *C. difficile* genome sequences to a recently reported set of genome sequences recovered from older patients from Sloan-Kettering and Washington University hospitals, who acquired CDI during hematopoietic stem cell transplantation or other treatments (Lewis et al., 2017), and where infection was judged to be primarily community acquired. A statistical comparison to our patients showed no clear difference in clade distribution, though the numbers of patients available were modest. These data are consistent with Clade 1 being common community acquired CDI in both adults and children.

We created thresholds for calling identity based on two pairs of control sequences. Cd12 and Cd13 were isolated from the same patient, and Cd4a and Cd4b were technical replicates, that is, genome sequences determined twice from the same preparation of genomic DNA. For SNVs, our replicate pairs showed sequence divergences for Cd12 and Cd13 (biological replicates) of 7, and for Cd4a and Cd4b (technical replicates), the number of SNVs was 16. A model based on rates of sequence error suggested that 17 SNVs represented a 95% threshold for the number attributable to error. For comparisons based on gene content, the Hamming distance for accessory genes for Cd4a and Cd4b was 105 genes; and for Cd12 and Cd13, the Hamming distance was 78 genes (calculated based on presence or absence of accessory genes). For sharing of episomes, episomes were first clustered to create clusters of similar sequences, then episome presence or absence was called based on cluster membership. Neither of the above pairs contained any episomes.

Most strains showed unique signatures. Eighteen different patterns were seen for the distributions of episome groups by strain. Our most closely related pairs of strains were Cd11 versus Cd15, and Cd2 versus Cd8. For Cd11 versus Cd15, the number of pairwise SNVs was 17, pangenome divergence was 191 genes, and all episomes were shared. For Cd2 versus Cd8, the number of pairwise SNVs was 9, pangenome divergence was 69 genes, and all episomes were shared.

The patient harboring Cd2 was treated at a satellite outpatient location, and the patient harboring Cd8 was hospitalized at the main hospital 11 months later. Any transmission thus appears to have potentially taken place in the community and not in the hospital system.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Strains Cd11 and Cd15 were both from hospitalized patients, but the two were inpatients on different floors within the hospital 12 weeks apart. The two patients were both in the Emergency Department 11 weeks apart, which is within the length of time of *C. difficile* spore survival. The patients harboring Cd11 and Cd15 were both scored as severe CDI. The SNV and episome measures called the two strains as indistinguishable; the pangenome divergence was slightly above our threshold for calling identity. Thus, it is possible that this represents a transmission within the hospital system, though transmission in the community is also possible. We do not have any further data that would allow more investigation of possible transmission mechanisms.

***tcdB* qPCR—**A quantitative PCR (qPCR) assay was used to detect the presence of toxin B (*tcdB*) genes from *C. difficile* in fecal samples (Luna et al., 2011). DNA was extracted from fecal samples using the DNeasy PowerSoil kit (Qiagen, Germantown, MD) according to the manufacturer's protocol. Reactions were performed in triplicate in a total volume of 20 μL containing the following: 5 μL DNA (diluted 1:10), 1 X TaqMan™ Fast Universal PCR Master Mix, no AmpErase™ UNG (Thermo Fisher Scientific, Waltham, MA), 0.6 μM *tcdB* forward primer, 0.6 μM *tcdB* reverse primer, and 0.1 μM *tcdB* probe. Quantitative PCRs were conducted on a QuantStudio6 instrument (Applied Biosystems, Waltham, MA) using the following conditions: 20 seconds at 95 °C followed by 40 cycles of 3 seconds at 95 °C and 30 seconds at 60 °C. The primers and probes were from Integrated DNA Technologies (Skokie, Illinois). The sequences are as follows:

Primer Name	Primer Sequence (5' to 3')
tcdB-For	GAAAGTCCAAGTTTACGCTCAAT
tcdB-Rev	GCTGCACCTAAACTTACACCA
tcdB-Probe	/5HEX/ GCTGCA /ZEN/ CCTAAACTTACACCA /3IABkFQ/

/5HEX/	5' Hexachlorofluorescein
/ZEN/	Internal ZEN quencher
/3IABkFQ/	3' Iowa Black® FQ

C. difficile* burden—C. difficile* burdens (vegetative cells + spores) were quantified from frozen stool. Stool samples were weighed and thoroughly homogenized in sterile PBS. Samples were then serially diluted and plated on taurocholate cycloserine cefoxitin fructose agar (TCCFA) at 37 °C in anaerobic conditions (85% nitrogen, 10% hydrogen, 5% carbon dioxide). Burdens were enumerated as colony forming units (CFUs) per gram of stool (Table S3).

Metagenomic sequence acquisition—DNA was extracted from stool samples using the DNeasy PowerSoil Kit (Qiagen, Germantown, MD, USA). Sequencing libraries were generated using the NexteraXT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced on a HiSeq 2500 using 2×125 bp chemistry. To assess environmental

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and reagent contamination, extraction blanks and DNA free water were processed in parallel with the experimental samples. As a positive control, libraries were made from a laboratory-generated mock community consisting of DNA from *Vibrio campbellii* and Lambda phage.

Metabolomic data acquisition—Extraction of metabolites, ultra-high performance liquid chromatography/tandem high resolution and accurate mass spectrometry using reverse phase and HILIC chromatography was performed by Metabolon, Inc. (North Carolina, USA). Four LC/MS methods were used to analyze stool sample extracts. All methods used a Waters ACQUITY ultra-performance liquid chromatograph (UPLC) and a Thermo Scientific Q-Exactive high resolution accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible with each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was eluted by gradient conditions from a C18 column (Waters Aquity BEH C18,2.1×100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions; however, it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was eluted from the C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were eluted from the column using methanol and water, and with 6.5 mM ammonium bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters Acquity BEH Amide 2.1×150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10 mM ammonium formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic exclusion. The scan range varied slightly between methods but covered *m/z* 70–1000. A total of 1675 compounds were quantified.

Synthesis of Isocaproyltaurine.—A 100 μL aliquot of 0.2 mM taurine in water was added to a 100 μL aliquot of 0.1 mM isocaproate in water. Then, a 10 μL aliquot of 2 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) was added to the solution, and the reaction mixture was allowed to react for 10 minutes at room temperature, followed by LC/MS analysis.

LC/MS of Isocaproyltaurine in Synthetic Mixture, Control Media, and *C. difficile* Media.—*C. difficile* strain R20291 (ribotype 027) was grown anaerobically in 5 mL liquid *C. difficile* minimal media (CDMM) (Cartman and Minton, 2010) supplemented with 0, 1 or 10 mM taurine (Sigma Aldrich, T0625) at 37°C until stationary phase (30 hours). Cultures were centrifuged at 4000×g for 10 minutes at 4°C. Supernatant was then passed through 0.22 micron filters (Sigma Millipore) and immediately stored at −80°C

Aliquots (100 μL) of control media and *C. difficile* media samples were extracted with a mixture of 100 μL of water and 600 μL of ethyl acetate. This mixture was vigorously

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

vortexed for approximately 10 seconds and centrifuged at 14000 rpm for 10 min at 4 °C. A 100 μL aliquot of ethylacetate from each sample was transferred to a 96 well plate and dried under nitrogen at 30°C. Dried residues were reconstituted in 100 μL of water and analyzed by LC/MS.

Isocaproyltaurine was separated from components in each mixture by injecting 5 μL of each sample onto an Agilent 1290 Infinity UHPLC using an 8 minute linear gradient (A: 0.1% formic acid, B: ACN/0.1% formic acid) on a Waters Acquity BEH C18 2 × 100 mm, 1.7 μm column. The HPLC flow was directed into the electrospray ionization source of an Agilent 6495B triple quadrupole mass spectrometer run in negative ion mode at 2500 V. Product ion spectra were generated by collision induced dissociation of parent molecular ions of isocaproyltaurine (*m/z* 222) with nitrogen gas at 29 V collision energies to match product ions of the standard to those in the extracted media. In addition, single reaction monitoring spectra of each sample were obtained by monitoring the parent to product ion transition of *m/z* 222 to 80 to demonstrate agreement of retention times and mass transitions of isocaproyltaurine in the standard and extracted media. In addition, samples were analyzed using the chromatographic method above on a Thermo Vantage UHPLC/Orbitrap ID-X mass spectrometer in negative electrospray ionization mode (2500 V) at 120,000 resolution for accurate mass and molecular formula determinations.

Mass accuracies obtained for the negative molecular ions of the standard (*m/z* 222.07968, −3.9 ppm) and media extract (*m/z* 222.08051, −0.18 ppm) resulted in predicted molecular formulas in agreement with the theoretical negative ion of isocaproyltaurine (*m/z* 222.08055) with a molecular formula of $C_8H_{16}O_4NS$ (Supplementary Figure 4).

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic analysis of microbial communities—Shotgun metagenomics sequencing data were processed using the Sunbeam pipeline (Clarke et al., 2019) based on Snakemake (Koster and Rahmann, 2012). Paired-end reads were quality-filtered and Illumina adapter sequences were removed using Trimmomatic v 0.3322. Low complexity reads that fell below the default threshold were marked and removed using Komplexity (Clarke et al., 2019). Reads that aligned to the human genome (hg38) or to the genome of phage phiX174 (which is used in sequencing library prep) using BWA were removed. Taxonomic classification was carried out using Kraken (Wood and Salzberg, 2014) and used to analyze 256 fecal samples, resulting in detection of 4371 putative species. For each sample, the read counts were normalized by total bacterial reads for that sample to calculate relative abundance. The PCoA plot was made based on Bray-Curtis Distance. Random Forest (R (Ihaka and Gentleman, 1996) package Random Forest) was used to perform classification between pairs of groups and the variable importance scores were used to rank the variables important for classification. In each comparison, species with maximum abundance <=0.01% were removed. The under-sampling strategy was used to balance the sample size. At each iteration, the Random Forest model was applied to samples at the Baseline Visit for 500 times to reduce the randomness of the Random Forest. The average out-of-bag (OOB) errors and the top important features were reported.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Metabolomic analysis—For the metabolomic analysis, we used the normalized data set from Metabolon, where each biochemical in the raw data set was rescaled to set the median to 1 and then the missing values were imputed with minimum value across samples. The 20 metabolites with a variance of 0 were removed from the data set. There were 1274 known metabolites and 381 unknowns. A total of 213 fecal samples were analyzed. The data was log10-transformed before downstream analysis. Classification between the healthy group, IBD and IBD+CDI was carried out using the Random Forest model from the R package Random Forest. In each pair of the groups, the under-sampling strategy was used to balance the sample size. At each iteration, the Random Forest was run 500 times using the known metabolites and samples at the baseline visit. The average OOB error and the top important features were reported.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Laurie Zimmerman for digital art and help with the manuscript. We thank Dr. Erin Graf and the CHOP Clinical Microbiology Laboratory for isolation of *C. difficile* strains. This Tobacco Formula grant is under the Commonwealth Universal Research Enhancement (C.U.R.E) program with the grant number SAP # 4100068710.

References

Abbas A, and Zackular JP (2020). Microbe-microbe interactions during Clostridioides difficile infection. Curr Opin Microbiol 53, 19–25. [PubMed: 32088581]

Abt MC, McKenney PT, and Pamer EG (2016). Clostridium difficile colitis: pathogenesis and host defence. Nat Rev Microbiol 14, 609–620. [PubMed: 27573580]

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. J Mol Biol 215, 403–410. [PubMed: 2231712]

Battaglioli EJ, Hale VL, Chen J, Jeraldo P, Ruiz-Mojica C, Schmidt BA, Rekdal VM, Till LM, Huq L, Smits SA, et al. (2018). Clostridioides difficile uses amino acids associated with gut microbial dysbiosis in a subset of patients with diarrhea. Sci Transl Med 10.

Bouillaut L, McBride SM, and Sorg JA (2011). Genetic manipulation of Clostridium difficile. Curr Protoc Microbiol Chapter 9, Unit 9A 2.

Bouillaut L, Self WT, and Sonenshein AL (2013). Proline-dependent regulation of Clostridium difficile Stickland metabolism. J Bacteriol 195, 844–854. [PubMed: 23222730]

Braun T, Di Segni A, BenShoshan M, Neuman S, Levhar N, Bubis M, Picard O, Sosnovski K, Efroni G, Farage Barhom S, et al. (2019). Individualized Dynamics in the Gut Microbiota Precede Crohn's Disease Flares. Am J Gastroenterol.

Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature 517, 205–208. [PubMed: 25337874]

Burdon DW, Brown JD, George RH, Arabi Y, Alexander-Williams J, and Keighley MR (1978). Pseudomembranous colitis caused by Clostridia. The New England journal of medicine 299, 48–49.

Cartman ST, and Minton NP (2010). A mariner-based transposon system for in vivo random mutagenesis of Clostridium difficile. Appl Environ Microbiol 76, 1103–1109. [PubMed: 20023081]

Chehoud C, Albenberg LG, Judge C, Hoffmann C, Grunberg S, Bittinger K, Baldassano RN, Lewis JD, Bushman FD, and Wu GD (2015). Fungal Signature in the Gut Microbiota of Pediatric Patients With Inflammatory Bowel Disease. Inflamm Bowel Dis 21, 1948–1956. [PubMed: 26083617]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, Farley MM, Dumyati GK, Wilson LE, Beldavs ZG, et al. (2013). Epidemiology of community-associated Clostridium difficile infection, 2009 through 2011. JAMA Intern Med 173, 1359–1367. [PubMed: 23780507]

Clarke EL, Taylor LJ, Zhao C, Connell A, Lee JJ, Fett B, Bushman FD, and Bittinger K (2019). Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. Microbiome 7, 46. [PubMed: 30902113]

Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, de Vos WM, Ehrlich SD, Fraser CM, Hattori M, et al. (2018). Enterotypes in the landscape of gut microbial community composition. Nat Microbiol 3, 8–16. [PubMed: 29255284]

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113. [PubMed: 15318951]

Ferreyra JA, Wu KJ, Hryckowian AJ, Bouley DM, Weimer BC, and Sonnenburg JL (2014). Gut microbiota-produced succinate promotes C. difficile infection after antibiotic treatment or motility disturbance. Cell Host Microbe 16, 770–777. [PubMed: 25498344]

Fletcher JR, Erwin S, Lanzas C, and Theriot CM (2018). Shifts in the Gut Metabolome and Clostridium difficile Transcriptome throughout Colonization and Infection in a Mouse Model. mSphere 3.

Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol 4, 293–305. [PubMed: 30531976]

George RH, Symonds JM, Dimock F, Brown JD, Arabi Y, Shinagawa N, Keighley MR, Alexander-Williams J, and Burdon DW (1978). Identification of Clostridium difficile as a cause of pseudomembranous colitis. Br Med J 1, 695. [PubMed: 630301]

Gerding DN, and Lessa FC (2015). The epidemiology of Clostridium difficile infection inside and outside health care institutions. Infect Dis Clin North Am 29, 37–50. [PubMed: 25582647]

Gevers D, Kugathasan S, Denson LA, Vazquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe 15, 382–392. [PubMed: 24629344]

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, and Nelson KE (2006). Metagenomic analysis of the human distal gut microbiome. Science 312, 1355–1359. [PubMed: 16741115]

Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, Bergwerff AA, Dekker FW, and Kuijper EJ (2008). Emergence of Clostridium difficile infection due to a new hypervirulent strain, polymerase chain reaction ribotype 078. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47, 1162–1170. [PubMed: 18808358]

Griffiths D, Fawley W, Kachrimanidou M, Bowden R, Crook DW, Fung R, Golubchik T, Harding RM, Jeffery KJ, Jolley KA, et al. (2010). Multilocus sequence typing of Clostridium difficile. J Clin Microbiol 48, 770–778. [PubMed: 20042623]

Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldassano RN, Noe JD, Rosh J, Markowitz J, et al. (2014). Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Invest 124, 3617–3633. [PubMed: 25003194]

Haimovitz-Friedman A, Kan CC, Ehleiter D, Persaud RS, McLoughlin M, Fuks Z, and Kolesnick RN (1994). Ionizing radiation acts on cellular membranes to generate ceramide and initiate apoptosis. J Exp Med 180, 525–535. [PubMed: 8046331]

Hourigan SK, Oliva-Hemker M, and Hutfless S (2014). The prevalence of Clostridium difficile infection in pediatric and adult patients with inflammatory bowel disease. Dig Dis Sci 59, 2222–2227. [PubMed: 24788321]

Hyams JS, Ferry GD, Mandel FS, Gryboski JD, Kibort PM, Kirschner BS, Griffiths AM, Katz AJ, Grand RJ, Boyle JT, et al. (1991). Development and validation of a pediatric Crohn's disease activity index. J Pediatr Gastroenterol Nutr 12, 439–447. [PubMed: 1678008]

Ihaka R, and Gentleman R (1996). R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5, 299–314.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Izquierdo C, Gomez-Tamayo JC, Nebel JC, Pardo L, and Gonzalez A (2018). Identifying human diamine sensors for death related putrescine and cadaverine molecules. PLoS Comput Biol 14, e1005945. [PubMed: 29324768]

Jenior ML, Leslie JL, Young VB, and Schloss PD (2017). Clostridium difficile Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. mSystems 2.

Johnson EL, Heaver SL, Waters JL, Kim BI, Bretin A, Goodman AL, Gewirtz AT, Worgall TS, and Ley RE (2020). Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. Nat Commun 11, 2471. [PubMed: 32424203]

Kelsen JR, Kim J, Latta D, Smathers S, McGowan KL, Zaoutis T, Mamula P, and Baldassano RN (2011). Recurrence rate of clostridium difficile infection in hospitalized pediatric patients with inflammatory bowel disease. Inflamm Bowel Dis 17, 50–55. [PubMed: 20722068]

Kim J, Darley D, Selmer T, and Buckel W (2006). Characterization of (R)-2-hydroxyisocaproate dehydrogenase and a family III coenzyme A transferase involved in reduction of L-leucine to isocaproate by Clostridium difficile. Appl Environ Microbiol 72, 6062–6069. [PubMed: 16957230]

Kim J, Smathers SA, Prasad P, Leckerman KH, Coffin S, and Zaoutis T (2008). Epidemiological features of Clostridium difficile-associated disease among inpatients at children's hospitals in the United States, 2001–2006. Pediatrics 122, 1266–1270. [PubMed: 19047244]

Knippel RJ, Zackular JP, Moore JL, Celis AI, Weiss A, Washington MK, DuBois JL, Caprioli RM, and Skaar EP (2018). Heme sensing and detoxification by HatRT contributes to pathogenesis during Clostridium difficile infection. PLoS pathogens 14, e1007486. [PubMed: 30576368]

Kolho KL, Pessia A, Jaakkola T, de Vos WM, and Velagapudi V (2017). Faecal and Serum Metabolomics in Paediatric Inflammatory Bowel Disease. J Crohns Colitis 11, 321–334. [PubMed: 27609529]

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, and Phillippy AM (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27, 722–736. [PubMed: 28298431]

Koster J, and Rahmann S (2012). Snakemake--a scalable bioinformatics workflow engine. Bioinformatics 28, 2520–2522. [PubMed: 22908215]

Kostic AD, Xavier RJ, and Gevers D (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology 146, 1489–1499. [PubMed: 24560869]

Lessa FC, Winston LG, McDonald LC, and Emerging Infections Program C.d.S.T. (2015). Burden of Clostridium difficile infection in the United States. The New England journal of medicine 372, 2369–2370. [PubMed: 26061850]

Levy M, Thaiss CA, Zeevi D, Dohnalova L, Zilberman-Schapira G, Mahdi JA, David E, Savidor A, Korem T, Herzig Y, et al. (2015). Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling. Cell 163, 1428–1443. [PubMed: 26638072]

Lewis BB, Carter RA, Ling L, Leiner I, Taur Y, Kamboj M, Dubberke ER, Xavier J, and Pamer EG (2017). Pathogenicity Locus, Core Genome, and Accessory Gene Contributions to Clostridium difficile Virulence. MBio 8.

Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, et al. (2015). Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe 18, 489–500. [PubMed: 26468751]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Liaw A, and Weiner M (2002). Classification and Regression by randomForest. R News 3, 18–22.

Luna RA, Boyanton BL Jr., Mehta S, Courtney EM, Webb CR, Revell PA, and Versalovic J (2011). Rapid stool-based diagnosis of Clostridium difficile infection by real-time PCR in a children's hospital. J Clin Microbiol 49, 851–857. [PubMed: 21209161]

Ma GK, and Lewis JD (2018). Increasing Incidence of Multiply Recurrent Clostridium difficile Infection. Ann Intern Med 168, 308.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Marcinkiewicz J, and Kontny E (2014). Taurine and inflammatory diseases. Amino Acids 46, 7–20. [PubMed: 22810731]

Martin-Verstraete I, Peltier J, and Dupuy B (2016). The Regulatory Networks That Control Clostridium difficile Toxin Synthesis. Toxins (Basel) 8.

McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, et al. (2018). American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems 3.

McDonald LC, Killgore GE, Thompson A, Owens RC Jr., Kazakova SV, Sambol SP, Johnson S, and Gerding DN (2005). An epidemic, toxin gene-variant strain of Clostridium difficile. The New England journal of medicine 353, 2433–2441. [PubMed: 16322603]

Melum E, Jiang X, Baker KD, Macedo MF, Fritsch J, Dowds CM, Wang J, Pharo A, Kaser A, Tan C, et al. (2019). Control of CD1d-restricted antigen presentation and inflammation by sphingomyelin. Nat Immunol 20, 1644–1655. [PubMed: 31636468]

Nagao-Kitamoto H, Leslie JL, Kitamoto S, Jin C, Thomsson KA, Gillilland MG 3rd, Kuffa P, Goto Y, Jenq RR, Ishii C, et al. (2020). Interleukin-22-mediated host glycosylation prevents Clostridioides difficile infection by modulating the metabolic activity of the gut microbiota. Nat Med.

Neumann-Schaal M, Jahn D, and Schmidt-Hohagen K (2019). Metabolism the Difficile Way: The Key to the Success of the Pathogen Clostridioides difficile. Front Microbiol 10, 219. [PubMed: 30828322]

Ni J, Shen TD, Chen EZ, Bittinger K, Bailey A, Roggiani M, Sirota-Madi A, Friedman ES, Chau L, Lin A, et al. (2017). A role for bacterial urease in gut dysbiosis and Crohn's disease. Sci Transl Med 9.

Olm MR, Brown CT, Brooks B, and Banfield JF (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME journal 11, 2864–2868. [PubMed: 28742071]

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, and Parkhill J (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. [PubMed: 26198102]

Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, and Harris SR (2016). SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2, e000056. [PubMed: 28348851]

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, and Tyson GW (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25, 1043–1055. [PubMed: 25977477]

Price MN, Dehal PS, and Arkin AP (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26, 1641–1650. [PubMed: 19377059]

Robinson JI, Weir WH, Crowley JR, Hink T, Reske KA, Kwon JH, Burnham CD, Dubberke ER, Mucha PJ, and Henderson JP (2019). Metabolomic networks connect host-microbiome processes to human Clostridioides difficile infections. J Clin Invest 130.

Rodemann JF, Dubberke ER, Reske KA, Seo DH, and Stone CD (2007). Incidence of Clostridium difficile infection in inflammatory bowel disease. Clin Gastroenterol Hepatol 5, 339–344. [PubMed: 17368233]

Sandberg KC, Davis MM, Gebremariam A, and Adler J (2015). Disproportionate rise in Clostridium difficile-associated hospitalizations among US youth with inflammatory bowel disease, 1997–2011. J Pediatr Gastroenterol Nutr 60, 486–492. [PubMed: 25419679]

Schuller-Levis GB, and Park E (2003). Taurine: new implications for an old amino acid. FEMS Microbiol Lett 226, 195–202. [PubMed: 14553911]

Seemann T (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. [PubMed: 24642063]

Sorg JA, and Sonenshein AL (2010). Inhibiting the initiation of Clostridium difficile spore germination using analogs of chenodeoxycholic acid, a bile acid. J Bacteriol 192, 4983–4990. [PubMed: 20675492]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Surawicz CM, Brandt LJ, Binion DG, Ananthakrishnan AN, Curry SR, Gilligan PH, McFarland LV, Mellow M, and Zuckerbraun BS (2013). Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections. Am J Gastroenterol 108, 478–498; quiz 499. [PubMed: 23439232]

Surtees JA, and Funnell BE (2003). Plasmid and chromosome traffic control: how ParA and ParB drive partition. Curr Top Dev Biol 56, 145–180. [PubMed: 14584729]

Thanissery R, Winston JA, and Theriot CM (2017). Inhibition of spore germination, growth, and toxin activity of clinically relevant C. difficile strains by gut microbiota derived secondary bile acids. Anaerobe 45, 86–100. [PubMed: 28279860]

Theriot CM, Koenigsknecht MJ, Carlson PE Jr., Hatton GE, Nelson AM, Li B, Huffnagle GB, J, Z.L., and Young VB (2014). Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun 5, 3114. [PubMed: 24445449]

Tschudin-Sutter S, Tamma PD, Naegeli AN, Speck KA, Milstone AM, and Perl TM (2013). Distinguishing community-associated from hospital-associated Clostridium difficile infections in children: implications for public health surveillance. Clin Infect Dis 57, 1665–1672. [PubMed: 24046303]

Turner D, Griffiths AM, Walters TD, Seah T, Markowitz J, Pfefferkorn M, Keljo D, Otley A, Leleiko NS, Mack D, et al. (2010). Appraisal of the pediatric Crohn's disease activity index on four prospectively collected datasets: recommended cutoff values and clinimetric properties. Am J Gastroenterol 105, 2085–2092. [PubMed: 20372111]

Turner D, Otley AR, Mack D, Hyams J, de Bruijne J, Uusoue K, Walters TD, Zachos M, Mamula P, Beaton DE, et al. (2007). Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. Gastroenterology 133, 423–432. [PubMed: 17681163]

Vich Vila A, Imhann F, Collij V, Jankipersadsing SA, Gurry T, Mujagic Z, Kurilshikov A, Bonder MJ, Jiang X, Tigchelaar EF, et al. (2018). Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Sci Transl Med 10.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one 9, e112963. [PubMed: 25409509]

Wick RR, Judd LM, Gorrie CL, and Holt KE (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13, e1005595. [PubMed: 28594827]

Wood DE, and Salzberg SL (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15, R46. [PubMed: 24580807]

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science 334, 105–108. [PubMed: 21885731]

Xing M, Wei Y, Hua G, Li M, Nanjaraj Urs AN, Wang F, Hu Y, Zhai W, Liu Y, Ang EL, et al. (2019). A gene cluster for taurine sulfur assimilation in an anaerobic human gut bacterium. Biochem J 476, 2271–2279. [PubMed: 31350331]

Zar FA, Bakkanagari SR, Moorthi KM, and Davis MB (2007). A comparison of vancomycin and metronidazole for the treatment of Clostridium difficile-associated diarrhea, stratified by disease severity. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 45, 302–307. [PubMed: 17599306]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- Multi-omics reveals markers of *C. difficile* infection (CDI) in pediatric IBD patients
- Identification of metabolites reveals distinctive features for IBD and CDI
- Isocaproyltaurine is made by *C. difficile* and associates with active IBD
- Identifies biomarkers potentially useful for distinguishing disease processes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

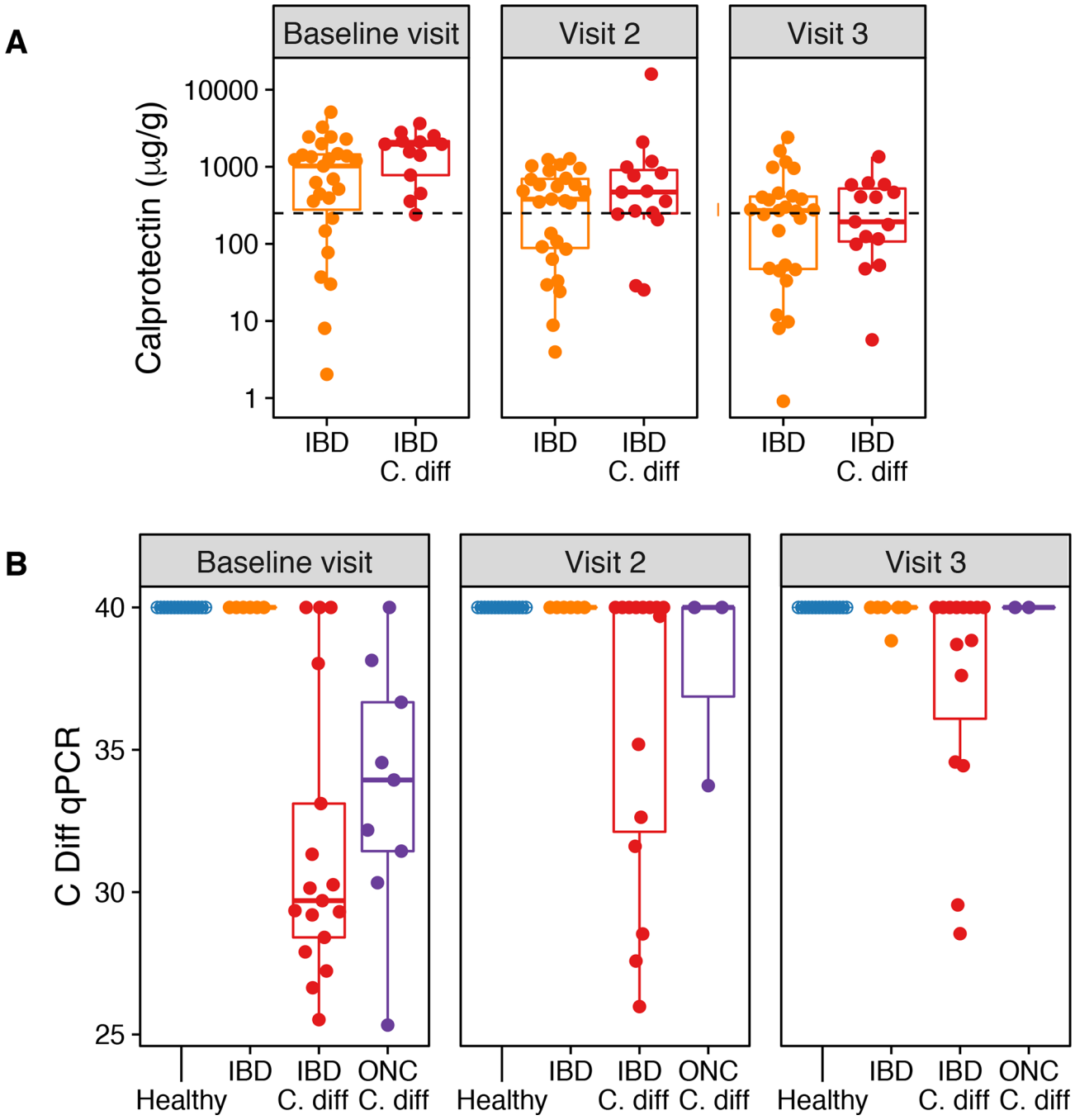


Figure 1.

Patient response to therapy. A) Calprotectin values for the groups studied. Values are shown for the IBD group (yellow) and the IBD+CDI group (red). Times analyzed are shown above each plot. The dashed line indicates the healthy cut-off value. B) Levels of *C. difficile* DNA quantified using qPCR. The x-axis specifies the groups studied and the y-axis shows the qPCR cycle of threshold, so that lower values respresent higher amounts of *C. difficile* DNA. Here and below, "ONC" indicates the group with CDI and malignancy.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

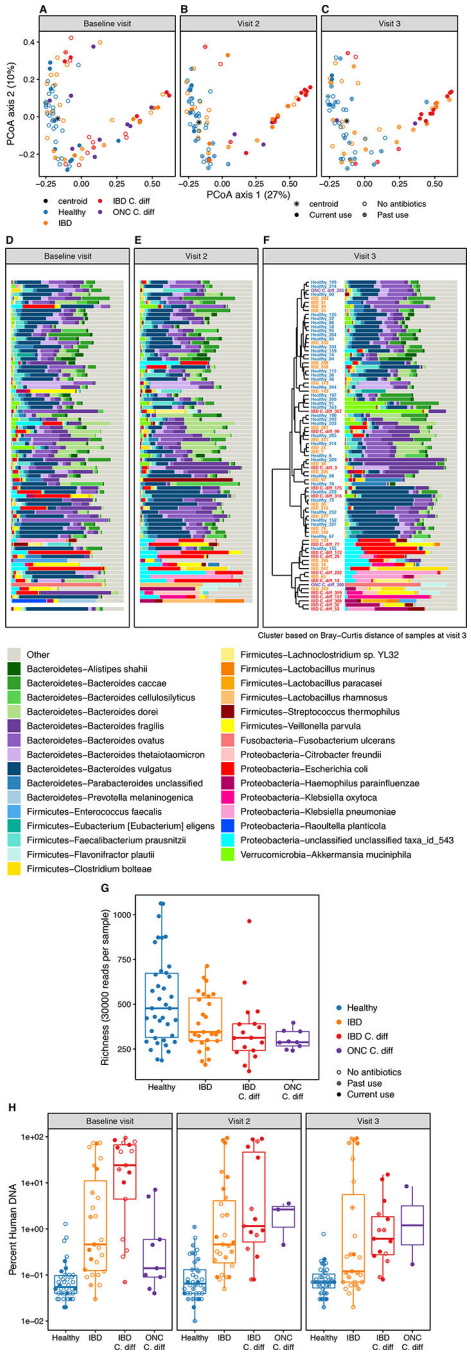


Figure 2.

Longitudinal analysis of the fecal microbiome in patients with CDI and controls. A-C: PCoA plots based on the Bray-Curtis dissimilarity assessed using the microbiome data. All samples were used to make a common plot, then samples collected at study entry (baseline visit), 4 weeks (visit 2), and 8 weeks (visit 3) are each displayed separately. Samples are color coded to indicate the four study groups. The centroid of the healthy controls is marked by a black asterisk. The shape of the symbol indicates antibiotic use. D-F: Major taxa present in each sample at each time point. Bacterial lineages are color coded (bottom) and

presented as stacked bar graphs. Clustering is based on the visit 3 time point. The names for each study group are color coded on the dendrogram. G: Microbial richness at baseline in each of the study groups, compared using normalized numbers of microbial reads. H: Longitudinal abundance of the proportion of human DNA in the samples studied. Points are coded by whether patients were on antibiotic therapy previously or at the time of sampling.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

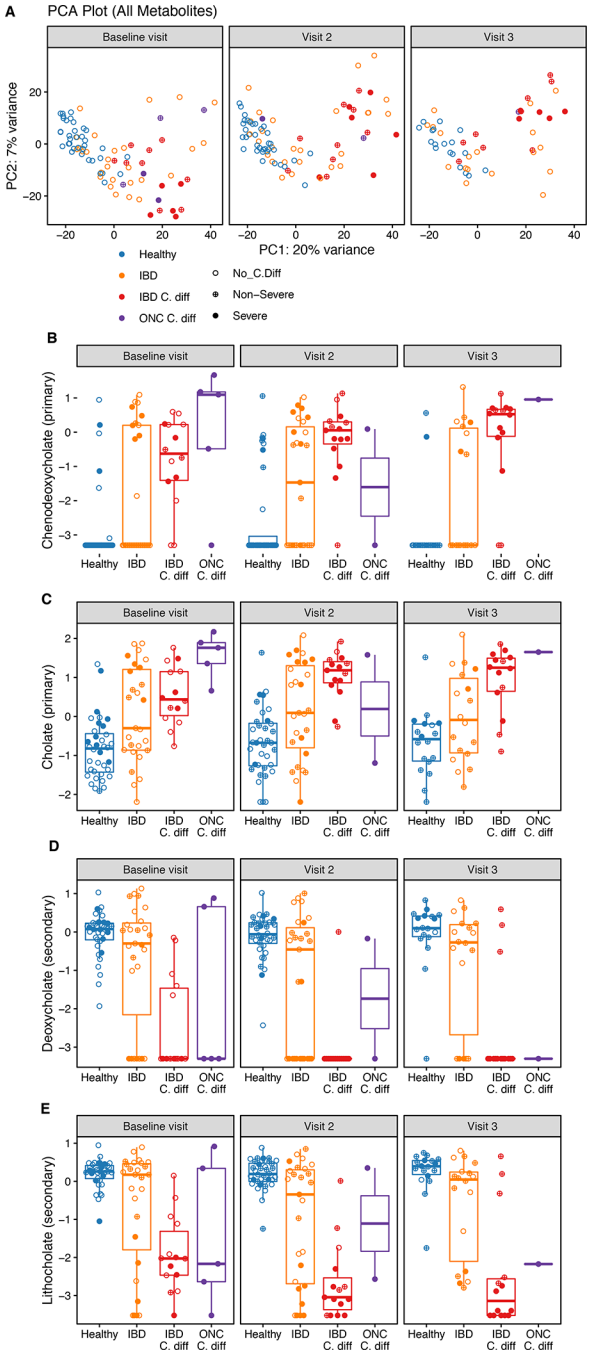


Figure 3.

Longitudinal comparisons of fecal metabolite profiles for the four study groups. A: Principal component analysis comparing metabolic content in the four study groups at time 0 (baseline visit), week 4 (visit 2), and week 8 (visit 3). The study groups are color coded as indicated to the right of the plots. CDI status is indicated by the shapes of the points. B-E: Longitudinal comparisons of the relative abundances of bile acids in fecal samples from the four study groups. Included are the primary bile acids B) chenodeoxycholate and C) cholate,

and the secondary bile acids D) deoxycholate and E) lithocholate. Groups are color coded as above, and results are shown for each of the three sampling times.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 4.

Comparison of metabolites distinguishing the study groups. Metabolites shown are the 30 most distinguishing features from each Random Forest model. For each panel, columns indicate samples and metabolites are in the rows. Metadata is shown along the top using the color code at the right. The relative abundance of metabolites is summarized by the color scale beside the figure panel. A) Analysis of the metabolites distinguishing the healthy and IBD groups. B) Analysis of metabolites distinguishing the healthy and IBD+CDI groups. C) Analysis of the metabolites distinguishing the IBD and IBD+CDI groups.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

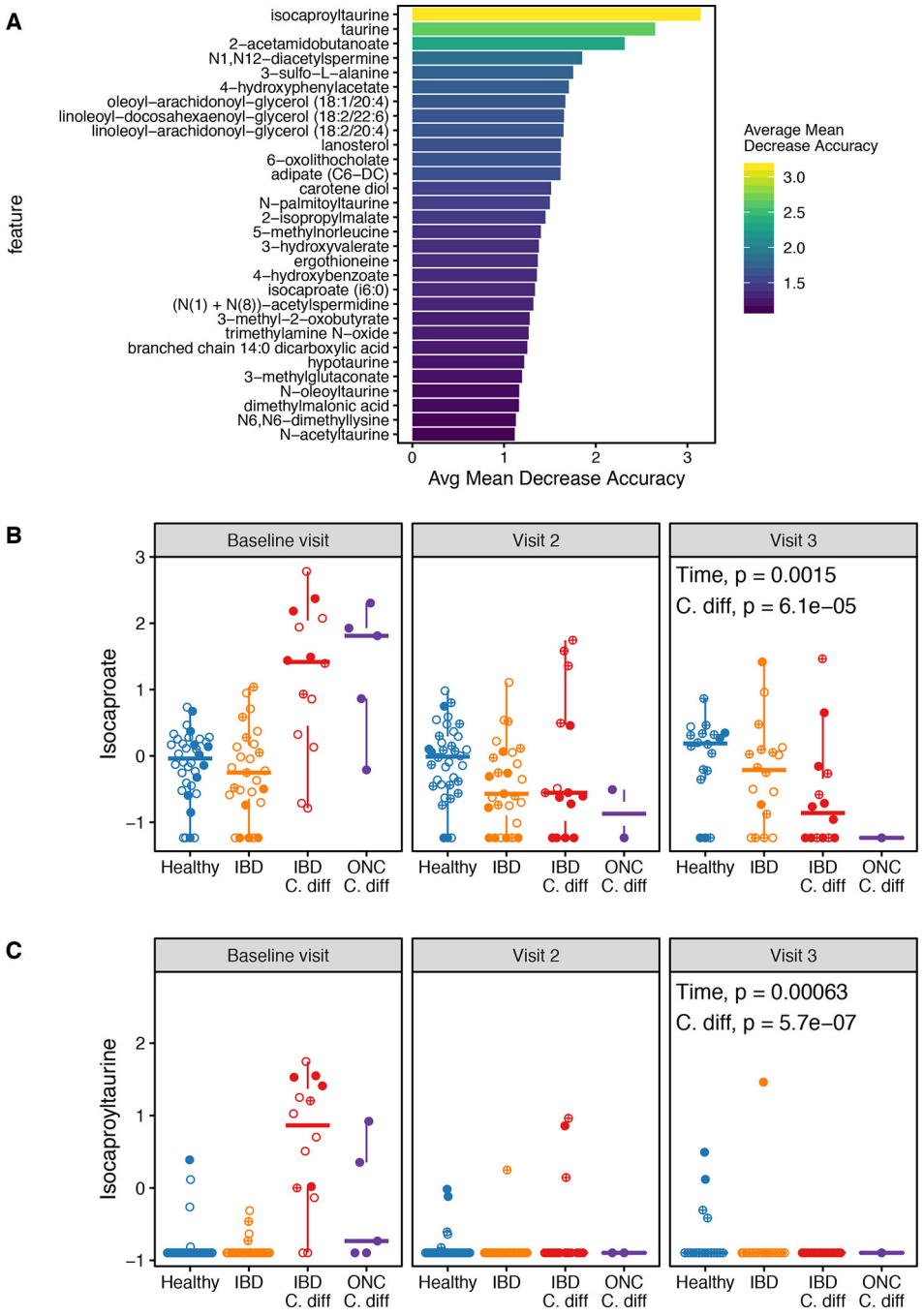


Figure 5.

Assessing the factors that most strongly distinguish the IBD from the IBD+CDI groups. A) Graph of the factors most contributing to discrimination in the Random Forest model. B) Longitudinal abundance of isocaproate in stool samples from each of the study groups. C) Longitudinal abundance of isocaproyltaurine in each of the study groups.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

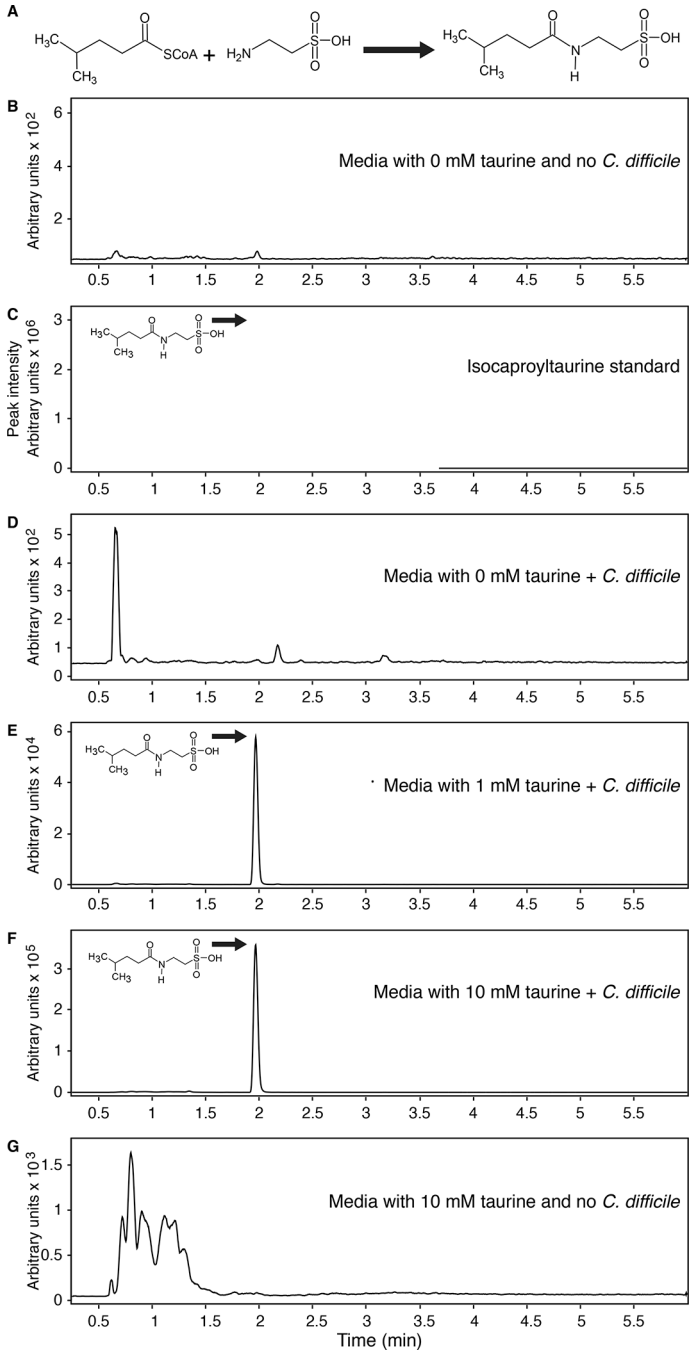


Figure 6.

Production of isocaproyltaurine by *C. difficile* selectively in the presence of added taurine. A) Proposed pathway for production of isocaproyltaurine by *C. difficile* by reaction of isocaproyl-CoA (Stickland fermentation intermediate) in the presence of taurine. B-G) Single ion monitoring mass spectra (*m/z* 222$^-$ to 80$^-$) assaying the presence of isocaproyltaurine in media. B) Trace showing control blank media. C) Synthetic isocaproyltaurine standard. D) Culture of *C. difficile* in media without added taurine. E) Culture of *C. difficile* in media with 1 mM taurine added. F) Culture of *C. difficile* in media

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

with 10 mM taurine added. G) Control showing lack of isocaproyltaurine in media with added taurine but no *C. difficile*. Thus E) and F) show accumulation of a compound with the single ion mass transition and retention time (1.96 min) of the isocaproyltaurine standard selectively in culture media containing taurine and *C. difficile*.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Vibrio campbellii	Bushman lab strain collection	N/A
Lambda phage	Bushman lab strain collection	N/A
Biological Samples		
Stool samples	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
Cycloserine Cefoxitin Mannitol Broth with Taurocholate Lysozyme Cysteine – CCMB-TAL	Anaerobe Systems	AS-8216
Cycloserine-Cefoxitin Fructose Agar with Horse Blood and Taurocholate – CCFA-HT	Anaerobe Systems	AS-2136
Brucella blood agar Plate	Oxyrase, Inc.	P-BRU-BA
Critical Commercial Assays		
C. DIFF QUIK CHEK COMPLETE®	Abbott Laboratories	30550C
Alethia™ C. difficile (previously illumigene®)	Meridian Bioscience	480050
QUANTA Lite® Calprotectin Extended Range	Inova Diagnostics	704860
Rapid Barcoding Kit	Oxford Nanopore Technologies	SQK-RBK001, SQK-RBK004
MinION flow cell	Oxford Nanopore Technologies	FLO-MIN106
DNeasy PowerSoil Kit	QIAGEN	12888-100
TruSeq DNA PCR-Free Library Prep Kit	Illumina	20015963
TruSeq DNA CD (combinatorial dual) Indexes	Illumina	20015949
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
Nextera XT Index Kit v2, Set A/B/C/D	Illumina	FC-131-2001/2/3/4
Quant-iT™ PicoGreen™ dsDNA Assay Kit	ThermoFisher Scientific	P7589
High Sensitivity NGS Fragment Analysis Kit (1–6000 bp), 500 Samples	Agilent Technologies, Inc.	DNF-474-0500
HiSeq PE Cluster Kit v4 cBot	Illumina	PE-401-4001
HiSeq SBS Kit V4 250 Cycle Kit	Illumina	FC-401-4003
TaqMan™ Fast Universal PCR Master Mix (2X), no AmpErase™ UNG	ThermoFisher Scientific	4352042
Oligonucleotides		
tcdB qPCR forward primer GAAAGTCCAAGTTTACGCTCAAT	Luna et al, 2011	IDT
tcdB qPCR reverse primer GCTGCACCTAAACTTACACCA	Luna et al, 2011	IDT
tcdB qPCR probe /5HEX/ GCTGCA /ZEN/ CCTAAACTTACACCA /3IABkFQ/	Luna et al, 2011	IDT

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
C difficile genome accession numbers	NCBI	PRJNA524299
Metagenomic sequence accession numbers	NCBI	PRJNA562600
Metabolomics data	This study	https://github.com/reny1/cdiff
Software and Algorithms		
Albacore	Oxford Nanopore	https://community.nanoporetech.com
Filtlong	N/A	https://github.com/rrwick/Filtlong
Porechop	Wick et al, 2017	https://github.com/rrwick/Porechop
SAMtools	Li et al., 2009	http://samtools.sourceforge.net/
R	Ihaka and Gentleman, 1996	https://www.r-project.org/
Sunbeam	Clarke et al., 2019	https://github.com/sunbeam-labs/sunbeam
Kraken	Wood et al., 2014	https://github.com/DerrickWood/kraken
BLASTn	Altschul et al., 1990	https://www.ncbi.nlm.nih.gov/books/NBK279690/
Nanoflow	This study	https://github.com/zhaoc1/nanoflow,
Snakemake	Koster and Rahmann, 2012	https://snakemake.readthedocs.io/en/stable/
Canu	Koren et al., 2017	https://github.com/marbl/canu
Nanopolish	Wick et al, 2017	https://github.com/jts/nanopolish
Unicycler	Wick et al., 2017	https://github.com/rrwick/Unicycler
Pilon	Walker et al., 2014	https://github.com/broadinstitute/pilon/
CoreSNPs	This study	https://github.com/zhaoc1/coreSNPs
MUSCLE	Edgar, 2004	https://www.drive5.com/muscle/
checkM	Parks et al., 2015	https://github.com/Ecogenomics/CheckM/
Prokka	Seemann et al., 2014	https://github.com/tseemann/prokka
Roary	Page et al., 2015	https://github.com/sanger-pathogens/Roary
snp-sites	Page et al., 2016	https://github.com/sanger-pathogens/snp-sites
Fasttree	Price et al., 2010	www.microbesonline.org/fasttree