Gut microbiome predictors of treatment response and recurrence in primary Clostridium difficile infection

S. Khanna*†, E. Montassier‡†, B. Schmidt*, R. Patel§¶, D. Knights‡, D. S. Pardi* & P. C. Kashyap*

*Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA.
†EA 3826 Thérapeutiques Cliniques et Expérimentales des Infections, Faculté de Médecine, Université de Nantes, Nantes, France.
‡Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA.
§Division of Infectious Diseases, Mayo Clinic, Rochester, MN, USA.
¶Division of Clinical Microbiology, Department of Medicine, Mayo Clinic, Rochester, MN, USA.

Correspondence to:
Dr P. C. Kashyap and Dr D. S. Pardi,
Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA.
E-mails: kashyap.purna@mayo.edu and pardi.darrell@mayo.edu

1Denotes equal contribution

SUMMARY

Background
Clostridium difficile infection (CDI) may not respond to initial therapy and frequently recurs, but predictors of response and recurrence are inconsistent. The impact of specific alterations in the gut microbiota determining treatment response and recurrence in patients with CDI is unknown.

Aim
To assess microbial signatures as predictors of treatment response and recurrence in CDI.

Methods
Pre-treatment stool samples and clinical metadata including outcomes were collected prospectively from patients with their first CDI episode. Next generation 16s rRNA sequencing using MiSeq Illumina platform was performed and changes in microbial community structure were correlated with CDI outcomes.

Results
Eighty-eight patients (median age 52.7 years, 60.2% female) were included. Treatment failure occurred in 12.5% and recurrence after response in 28.5%. Patients who responded to treatment had an increase in Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium and Rothia compared to nonresponders. A risk-index built from this panel of microbes differentiated responders (mean 0.07 ± 0.24) from nonresponders (0.52 ± 0.42; P = 0.0002). Receiver operating characteristic (ROC) curve demonstrated that risk-index was a strong predictor of treatment response with an area under the curve (AUC) of 0.85. Among clinical parameters tested, only proton pump inhibitor use predicted recurrent CDI (OR 3.75, 95% CI 1.27–11.1, P = 0.01). Patients with recurrent CDI had statistically significant increases in Veillonella, Enterobacteriaceae, Streptococci, Parabacteroides and Lachnospiraceae compared to patients without recurrence and a risk index was able to predict recurrence (AUC = 0.78).

Conclusion
Gut microbiota signatures predict treatment response and recurrence potentially, allowing identification of patients with Clostridium difficile infection that may benefit from early institution of alternate therapies.

Aliment Pharmacol Ther
INTRODUCTION

*Clostridium difficile* infection (CDI) is a common infection in the USA with 450 000 infections and 29 000 deaths per year,1–3 with a tremendous economic impact (attributable costs ranging from $8911 to $30 049 per hospitalised patient).4 The pathophysiology of CDI is complex with alterations in gut microbiota playing an important role in susceptibility to CDI.5 A recent study reported an increase in Firmicutes, Proteobacteria and Actinobacteria and decreases in Bacteroidetes in CDI patients, and an altered ratio of Bacteroidetes to Firmicutes was found to be significant after controlling for confounding factors.6 Another study demonstrated that several bacterial species within the Ruminococcaceae, Lachnospiraceae, Bacteroides and Porphyromonadaceae were largely absent in CDI cases and highly associated with nondiarrhoeal controls.7 Similarly, alterations in *Ruminococcus gnavus*, certain Enterobacteriaceae, *Verrucomicrobia* and *Enterococcus* have been related to development of CDI.8–10 Risk factors for CDI include increasing age, antibiotic exposure, proton pump inhibitor use, hospitalisation, immunosuppression and comorbidities.10–15 Among these, gut microbiota changes associated with antibiotic use have been best studied in animal models and provide strong evidence for alterations in gut microbiota composition and function in susceptibility to primary and recurrent *C. difficile*.16

The management of CDI includes use of oral antibiotics but the rate of treatment failure (up to 35%) and recurrence (60% or higher after 3 episodes) is concerning.17, 18 Alternate treatment strategies such as faecal microbiota transplant are highly effective in patients with recurrent CDI.19, 20 Studies have assessed clinical features which may predict CDI response, but there is a lack of robust clinical or microbial biomarkers predictive of response to primary therapy.21–23 Severe CDI and hospital admission are clinical predictors of metronidazole failure, but these have not been validated. Increasing age, concomitant antibiotic use, decreased anti-toxin IgG levels, the presence of comorbidities and potentially the use of gastric acid suppression medications have been associated with recurrent CDI. Anti-toxin IgG levels are not clinically available.24, 25

Despite the identification of these risk factors, clinical models to predict the risk of recurrent CDI are not robust enough for routine clinical use and there are no models to predict response to initial antibiotic therapy.24, 26 Hence, there is a need for biomarkers and better models to predict these treatment outcomes. While the role gut microbiota alterations play in CDI susceptibility has been established,27 their role in determining outcomes of CDI treatment has not been investigated. We used next generation sequencing to characterise microbial communities in patients with primary CDI to identify differences in microbial community structure and key taxa that can predict treatment response and the risk of recurrence after successful treatment.

MATERIALS AND METHODS

Study design

We prospectively recruited 88 patients (median age 52.7 years, interquartile range 36.9–65.1; 60.2% female) with their first CDI episode (from 3/2012–9/2013) as identified from the Clinical Microbiology Laboratory at Mayo Clinic, Rochester, Minnesota and collected an aliquot from the stool samples that led to the diagnosis. Details on clinical data acquisition and analysis are outlined in Data S1.

SEQUENCING AND ANALYTIC METHODS

16S rRNA gene sequencing and data analyses

After faecal DNA isolation (fecal DNA kit; MoBio, Carlsbad, CA, USA), amplicons spanning the variable region 4 of bacterial 16S rRNA were generated and sequenced using MiSeq Illumina platform at the Mayo Clinic Medical Genome Facility, Rochester, MN, USA. We used 515F TATGGTAATTGTGTGCCAGCMGCCGCGGTAA and barcoded 806R primers AGTCAGTCAGCCGGAC-TACHVGGGTWTCTAAT. The amplicons were ~400 base pairs. The sequencing was bidirectional but only forward reads were used in the study due to low quality scores of reverse sequences. We applied built-in functions of the Quantitative Insights Into Microbial Ecology (QIIME) pipeline for data analysis.28 The 16S rRNA sequencing data from the Illumina runs were trimmed, demultiplexed, chimera filtered and assigned to operational taxonomic units (OTUs) using packages implemented in QIIME 1.8.0 software.28 Further details of 16S rRNA data analysis are outlined in Data S1.

In order to study differences in alpha diversity between our cohort of patients and healthy subjects, we included data from healthy subjects from a previously published study (global gut dataset).29 The differences in gut microbiota composition between healthy controls and patients with CDI have been previously described.3 We included adult US individuals and compared the following alpha diversity metrics with our cohort of CDI
patients: phylogenetic diversity whole tree, observed species and Chao 1 index. The methodology applied in both the studies was similar as samples were stored at −80 °C in both studies. DNA extraction was done using MO BIO power soil DNA isolation kit in both studies and the same protocol (Earth Microbiome Protocol) was followed including bead beating the faecal samples and the same region of 16S was amplified using the same primers and the same PCR protocol.30

In order to identify significant associations between microbial community abundances and clinical metadata, we applied a linear multivariate regression model specifically developed for microbiome data (MaAsLin, Multivariate microbial Association by Linear models).31 We used default parameters within MaAsLin as described by Morgan et al.31 Clinical data were selected by boosting, as microbial communities are known to be high dimensional, to identify those most associated with each microbial feature over potential covariates. Selected clinical data were then used in a general linear model including clinical data as predictor and taxonomic relative abundances as response.

Comparisons of relative abundance of taxa between responders and nonresponders, or between patients with and without recurrence after initial response, was performed using linear discriminant analysis effect size (LEfSe), a nonparametric Mann–Whitney U test applied to detect features with significant differential abundance with respect to the groups compared, followed by a linear discriminant analysis (LDA) to estimate the effect size of each differentially abundant feature.32 As proposed, an LDA score ($\log_{10}$ > 2) was considered significant.

A risk index was built to differentiate responders from nonresponders to initial treatment or patients with and without recurrence after initial response, based on taxonomy on a noncollapsed OTU table. All the taxa with a LDA score ($\log_{10}$ > 2) were included in the calculation of the risk index. In order to build the risk index to predict the patients who would respond or not respond to treatment, the relative abundances (arcsine square root transformed) of the taxa associated with the responders to treatment (based on the LEfSe output, all taxa with a LDA score ($\log_{10}$ > 2) were summed and the relative abundances of the taxa associated with nonresponders to treatment (based on the LEfSe output, all taxa with a LDA score ($\log_{10}$ > 2) were summed. Then, the difference between these two sums (relative abundance of the taxa associated with no response to treatment minus relative abundance of the taxa associated with response to treatment) was calculated, thereby obtaining a risk index. This procedure was repeated $n$ (overall sample size) times to obtain a risk index for each patient in the cohort. Therefore, a risk index was calculated for each patient. We had 11 risk indexes in the nonresponders and 77 risk indexes in the responders, 22 risk indexes in patients who had a recurrence and 55 risk indexes in patients who did not have a recurrence of CDI. The utility of a risk index to predict a clinical outcome was recently reported by our group.33

A leave-one-out cross-validation procedure was also conducted. This procedure calculates the risk index on the n-1 patients (taxa that differentiated patients based on the LEfSe output) and then tested the risk index in the held-out patient, that is, the risk index values are predicted for each patient using a panel of microbes retrained from the other patients. A detailed description of the leave-one-out cross-validation and receiver operating characteristic (ROC) curve analysis are outlined in Data S1.

Network analyses were carried out with Cytoscape using an edge-weighted spring-embedded layout.34 Importantly, we collapsed OTUs to the genus level and eliminated OTUs present in fewer than 25% of samples, to reduce the very large number of multiple hypotheses tested in correlation network analysis. We then performed Spearman correlation of taxon–taxon relative abundance and included only those links with absolute value of correlation $> 0.5$ and false discovery rate (FDR)-corrected $P < 0.05$.

In order to determine if competition for nutritional niches in the gut may play a role in determining response to treatment by providing C. difficile with a competitive disadvantage, we used Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) to predict Carbohydrate-Active Enzymes database–glycoside hydrolase (CAZY-GH) assignments. Details on PICRUSt analysis are outlined in Data S1.

RESULTS

Clinical characteristics fail to predict response to standard therapy

The rate of primary nonresponse following recommended treatment was 12.5%. Patients were treated with either metronidazole or vancomycin (Table 1). Among those who had an initial successful response, 28.5% had recurrent CDI. Prior antibiotic exposure was noted in 59.1% of patients, and 78.8% of those with antibiotic
exposure (or 36.1% of all 88 patients) were exposed to a single agent, 9.7% (or 5.7% of all 88 patients) to a combination of antibiotics and exact antibiotic name was unknown in the remaining six patients. Details on antibiotic use are summarised in Table S1. Chi-square univariate analyses demonstrated no significant differences in any clinical variables including initial treatment in primary nonresponders compared to responders (Table 1).

Gut microbiota signatures prior to treatment predict response to therapy in CDI

Of the 88 faecal samples collected, a total of 1 449 211 high-quality 16S rRNA gene-encoding sequences were identified, representing 7470 OTUs. The mean number of sequences obtained per sample was 16 468 /C64674.

Importantly, since samples contained between 6987 and 35 494 sequences, diversity analyses were rarefied at 6987 sequences per sample to avoid bias.

We assessed the relationship between clinical metadata and microbial measurements by identifying associations using MaAsLin (see methods above). This model investigated the relationship between microbial taxa collapsed at genus level with metadata of interest while accounting for other covariates. In our cohort of patients, we did not find significant associations between gut microbiota and the clinical characteristics (for example, Age, Sex, BMI, Charlson comorbidity Index) (Table S2). A panel of 36 OTUs that were significantly different between primary nonresponders and responders using LEfSe [corresponding to an LDA (log_{10}) >2] were identified. Responders had a significant increase in relative abundance of OTUs within Ruminococcaceae, Rikenellaceae, Bacteroides and Faecalibacterium, while nonresponders had a significant increase in Clostridiaceae, Lachnospiraceae, Blautia, Coprococcus, Streptococcus, Bifidobacterium, Ruminococcus and Actinomyces (Figure 1).

Furthermore, the ability of this panel of microbes to discriminate between responders and nonresponders using ROC curve analysis demonstrated several individual OTUs to be strong predictors of response (Figure S1). A risk index of response to treatment was built from this panel of OTUs. The index, calculated in each patient (responders and nonresponders), corresponds to the difference between the sum of the relative abundance of all the OTUs associated with nonresponse to treatment [i.e. with an LDA (log_{10}) >2] and the sum of the relative abundance of all the OTUs associated with response to treatment [i.e. with an LDA (log_{10}) >2]. The mean risk index score was significantly different

| Table 1 | Clinical characteristics of all patients |
|---|---|---|---|---|---|
| | Overall (n = 88) | Treatment responder (n = 77) | Treatment failure (n = 11) | P value* | Treatment success with no recurrence (n = 55) | Treatment success with recurrence (n = 22) | P value† |
| Age (median) | 52.7 | 53.8 | 49.9 | 0.67 | 55.6 | 49.0 | 0.62 |
| Sex (% female) | 60.2 | 59.7 | 63.6 | 0.8 | 58.2 | 63.6 | 0.65 |
| BMI, mean (kg/m²) | 27.5 | 27.4 | 28.1 | 0.75 | 27.7 | 26.6 | 0.44 |
| Charlson comorbidity index | 1.32 | 1.32 | 1.27 | 0.53 | 1.34 | 1.27 | 0.49 |
| Prior antibiotic exposure (%) | 59.1 | 55.8 | 81.8 | 0.1 | 56.4 | 54.5 | 0.88 |
| Community-acquired (%) | 59.1 | 62.3 | 36.4 | 0.13 | 63.6 | 59.1 | 0.9 |
| Severe CDI (%) | 7 | 5.2 | 18.2 | 0.1 | 5.5 | 4.5 | 0.86 |
| Concomitant antibiotic exposure (%) | 28.4 | 27.2 | 36.4 | 0.5 | 27.3 | 27.3 | 1.0 |
| Concomitant PPI exposure (%) | 27.3 | 25.9 | 36.3 | 0.48 | 18.2 | 45.5 | 0.01 |
| Treatment with metronidazole (%) | 70.5 | 70.1 | 72.7 | 0.8 | 72.7 | 63.6 | 0.1 |
| Treatment with vancomycin (%) | 23.9 | 24.7 | 18.2 | 0.8 | 20.0 | 36.4 | 0.1 |
| Treatment with vancomycin and metronidazole | 5.6 | 5.2 | 9.1 | 0.8 | 7.3 | 0 | 0.1 |

CDI, Clostridium difficile infection; BMI, body mass index; PPI, proton pump inhibitor.

* Denotes P value for comparison of treatment responders vs. treatment failures
† Denotes P value for comparison of recurrent infections vs. nonrecurrent infection
[P < 0.001, one-way ANOVA (analysis of variance) with post-hoc Tukey HSD (honest significant difference)] between the responders (mean score 0.07 ± 0.2) and non-responders (0.5 ± 0.4) as well as healthy subjects (0.03 ± 0.1, adult US subjects from global gut study with BMI under 25, n = 60) and non-responders (Figure 2). The risk index was not significantly different between healthy subjects and responders (P = 0.57, one-way ANOVA with post hoc Tukey HSD) suggesting that the difference in risk index in responders and non-responders is not due to variability in gut microbiome among individuals.

Furthermore, ROC curve analysis showed that this risk index was a strong predictor of treatment response, with an area under the curve (AUC) of 0.85. A cut-off of 0.21 was associated with a sensitivity of 77% and a specificity of 73% (Figure 3a). Importantly, we did not find a correlation between the risk index and previous antibiotic treatment in terms of response to treatment (Pearson’s product-moment correlation = 0.09, P = 0.37). Thus, nonresponse was not associated with prior antibiotic administration.

In order to assess our risk index, a leave-one-out cross-validation was performed, where the risk index was built 88 times using n-1 samples each time and then tested on the held-out sample. Thus, each held-out patient was treated as a new patient, independently from the initial cohort, on whom we tested and subsequently refined the optimal index cut-off to separate responders and nonresponders. We showed that the risk index was a strong predictor of response vs. nonresponse.
(permutation test performed on the difference between the mean of those without response and the mean of those with response to treatment with 999 random permutations, $P < 0.0001$; Figure 3b). We also determined with this leave-one-out procedure that a CDI risk index threshold of 0.11 best predicts response to treatment, yielding a sensitivity of 70% at a specificity of 73% (mean AUC = 0.81). Thus, we found that our risk index, based on a panel of 36 OTUs that were significantly different between primary nonresponders and responders using LEfSe, accurately identified patients with CDI, likely to not respond to conventional treatment.

OTU networks are disrupted in primary nonresponders
Correlations between OTU networks at the genus level were computed to differentiate responders and nonresponders and demonstrated a threefold (56 vs. 16, ratio = 3.35) decrease in the number of strong taxon–taxon correlations (absolute value of Spearman correlation $>0.5$ and false discovery rate-corrected $P < 0.05$) in nonresponders compared to responders, and most of the decreased nodes between responders and nonresponders were members of the phylum Firmicutes (50 vs. 10, ratio = 5) and Actinobacteria (7 vs. 1, ratio = 7) (Figure 4).

Gut microbial diversity is not significantly different in primary nonresponders compared to responders
Unweighted and weighted UniFrac-based principal coordinate analysis (PCoA) did not show significant differences in beta diversity between respondents and nonresponders (unweighted UniFrac distance metric: $R = -0.09$, $P = 0.9$; weighted UniFrac distance metric: $R = -0.004$, $P = 0.5$) (Figure S2). Patients with CDI (responders and nonresponders) had significantly decreased alpha diversity ($P < 0.001$ for the 3 measures phylogenetic diversity whole tree, observed species and Chao 1 index; one-way ANOVA with post hoc Tukey HSD) compared to healthy subjects (adult US subjects from global gut study$^{29}$). Among patients with CDI, although there was a trend towards decreased alpha diversity in nonresponders compared to responders, this difference was not significant for the 3 measures (phylogenetic diversity whole tree, observed species and Chao 1 index; one-way ANOVA with post hoc Tukey HSD) (Figure S3).

Gut microbiota functional repertoire is significantly different in responders and nonresponders
We imputed functional aspects of the microbiota from 16S rRNA data using PICRUSt to predict CAZY-GH assignments. Based on LEfSe, we found that GH70 (dextransucrase) and GH38 ($\alpha$-mannosidase) were increased, whereas GH59 ($\beta$-galactosidase) and two carbohydrate-binding modules, CBM16 and CBM42, were significantly decreased in nonresponders compared to responders with LDA score ($\log_{10}$) $>2$ (Figure S4).

Gut microbiota signatures predict CDI recurrence
Among the patients who initially responded to treatment, 28.5% had recurrent CDI. The median time to recurrence was 23 days (range 15–56 days). There were
no significant differences in patients with and without recurrent CDI (Table 1) among the clinical variables analysed except PPI use which predicted recurrent CDI on univariate analysis (odds ratio 3.75, 95% confidence interval 1.27–11.1, \( P = 0.01 \)) and multivariable analysis after controlling for age (\( P = 0.0007 \)) and comorbidities (\( P = 0.0009 \)) in separate multivariable models. However, PPI use was not associated with alterations in the gut microbiota (Table S2).

Relative abundance of 11 OTUs was significantly different between those with and without recurrence. Patients with recurrence had a significant increase in Veillonella, Enterobacteriaceae (Erwinia), Streptococcus, Parabacteroides and Lachnospiraceae using LEfSe (Figure 5). Several individual microbes were strong predictors of recurrence (Figure S5).

A risk index of recurrence built from this panel of microbes differentiated between patients with and without recurrence. This index included all OTUs that had a LDA score (\( \log_{10} \)) >2, as previously described. This index was significantly different in patients who did not have a recurrence (mean score 0.09 ± 0.08) and those who did (0.19 ± 0.12) (Mann–Whitney \( U \) test, \( P = 0.0001 \)) (Figure 6a). The ROC curve analysis showed that this risk index was a strong predictor of recurrence, with an AUC of 0.78 (Figure 6b). Moreover, a cut-off of 0.13 had a sensitivity of 78% and a specificity of 68%. As described previously, a leave-one-out cross-validation was performed, where the risk index was built 77 times using n-1 samples each time and then tested on the held-out sample. This procedure showed that the gut microbiota was a strong predictor of recurrence (permutation test performed on the difference between the mean of those with recurrence and the mean of those without recurrence with 999 random permutations, \( P < 0.0001 \); Figure 6c). This procedure also determined that a CDI risk index classification threshold of 0.11 best predicts recurrence with a sensitivity of 75% at a specificity of 69% (mean AUC = 0.75). Thus, we found that our risk index can identify patients with CDI who are likely to have recurrence after initial response. Importantly, a supervised learning method using a Random Forest model was also applied but was unable to accurately assign samples to their source population based on taxonomic profiles at the OTU level and was outperformed by the above risk index approach.
Pre-treatment Gut microbial diversity is not significantly different in patients with and without recurrent CDI
Unweighted and weighted UniFrac-based PCoA did not show significant differences in beta diversity between patients with and without recurrence (unweighted UniFrac distance metric: $R = -0.1$, $P = 0.9$; weighted UniFrac distance metric: $R = 0.02$, $P = 0.3$) (Figure S6). Moreover, there was no significant difference in alpha diversity based on three different metrics between CDI patients with and without recurrence (Figure S7).

Gut microbiota functional repertoire is significantly different in those with and without recurrence
As described above, we again imputed functional aspects of the microbiota from 16S rRNA data using PICRUSt to predict CAZY-GH assignments. Based on the LEfSe tool, GT30 ($\beta$-fucosidase) and a carbohydrate-binding module (CBM20) were increased in patients who had a recurrence compared to patients who did not with LDA score ($\log_{10} > 2$) (Figure S8).

DISCUSSION
In this study, we report specific gut microbiota signatures associated with the initial response to treatment and recurrence after successful treatment in patients with primary CDI. We have developed a risk index based on compositional differences among patients with CDI, which can help predict response to treatment and recurrence in patients with CDI. This is a step towards identifying CDI patients who would be candidates for early alternative therapies such as faecal microbiota...
Several studies indicate that CDI is accompanied by a depletion of butyrogenic bacteria, suggesting either an inhibitory effect of butyrate or the loss of a nutrient niche, which can be occupied by *C. difficile*. Thus, the presence of *F. prausnitzii* and Ruminococcaceae may improve the response to treatment by providing niche competition or direct inhibition. An increase in Bacteroidetes represented by greater abundance of the families Bacteroidaceae, Rikenellaceae and Porphyromonadaceae has been reported following successful FMT for treatment of CDI. Interestingly a higher relative abundance of *Bacteroides* and Rikenellaceae was seen in primary responders in our study, suggesting they may synergise with conventional treatment for exclusion of *C. difficile*. While the mechanism of action remains unclear, recent studies have shown that certain *Bacteroides* spp. may play a role in preventing infection with *C. difficile*. *B. fragilis* influences the development of the immune response and *B. theta-taomicron* stimulates Paneth cells to produce antibacterial peptides, which may prevent pathogens from colonising. While indirect, these data suggest a synergistic role for members of the gut microbial community in augmenting treatment response.

Interestingly, even though overall gut microbial diversity is not significantly different in responders compared to nonresponders, there is a threefold increase in the number of strong taxon–taxon correlations in responders as compared to nonresponders, and most of the decreased nodes in nonresponders were members of the phylum Firmicutes. This suggests that in addition to differences in individual taxa, the overall microbial community structure plays an important role in augmenting response to primary therapy by more effectively excluding *C. difficile*.

In patients with recurrent CDI we found increases in taxa that were different than in primary nonresponders. Several of these have been previously associated with recurrent CDI or potentially promote the growth of *C. difficile*. There was an increase in Enterobacteraeae, Veillonella and Parabacteroides among others in patients with recurrent CDI compared to patients without recurrence. Several studies both on mice and on humans demonstrate an increase of Enterobacteraeae in CDI. A persistent expansion of Enterobacteraeae has also been shown after treatment with clindamycin suggesting it may play a role in susceptibility to colonisation with *C. difficile*. *Parabacteroides diastonis*, a succinate producer, has been previously associated with CDI perhaps since increased succinate availability facilitates expansion of *C. difficile*. The low concentration of succinate present in the microbiota of conventional mice is transiently elevated.
upon antibiotic treatment or chemically induced intestinal motility disturbance, and *C. difficile* exploits this succinate spike to expand in the perturbed intestine.\(^3^5\) Further studies are needed to identify mechanisms by which these bacteria may contribute to failure of treatment or recurrence after successful treatment.

The primary strength of our study is the ability to predict the clinically important outcomes of treatment failure and recurrence in patients with primary CDI by characterising the gut microbiota of pre-treatment stool samples. It is possible that strain level differences in *C. difficile* contribute to the outcome but these were not assessed in this study. This study helps define a new diagnostic paradigm, but as with any initial finding, there are limitations. The risk index developed in our study likely needs to be prospectively validated in a larger cohort of CDI patients given the relatively small sample size. We did not have information on the patients’ dietary histories at the time of diagnosis so we were unable to assess how diet may influence microbial composition and CDI outcomes. Additionally, detailed analyses comparing microbiome changes due to prior antibiotic exposure would need to be considered in future studies. We also used healthy controls from a previously published study and there could be a study effect given the samples were run at different times as has been pointed out in previous studies. This is usually attributed to differences in DNA extraction protocol, primers and data analysis.\(^4^9\) However, as these were not different between our study and the healthy controls, we do not expect a study effect. Prior studies, which have investigated study effects have looked at healthy controls from different studies, where study-specific effects can be seen. However, as we are investigating differences in healthy controls and patients with CDI, we expect the true biological effect to overcome inter-individual study differences. While this was not a primary aim of our study we do note that our findings are similar to those previously reported.\(^7\) Both these studies used Illumina platform for sequencing, we used Miseq and the healthy controls were sequenced on Hiseq. A cross-validation has been previously published showing that the data from these two platforms do not introduce a bias in the results.\(^3^0\) The risk index determined in our study and applied to the healthy controls was similar in responders and healthy controls (from different studies) and different from nonresponders, further supporting the validity of our finding.

In conclusion, gut microbiota signatures can be used to predict response to and recurrence after initial treatment in patients with CDI. This finding will need to be validated in a larger cohort and future work will focus on understanding interaction of individual taxa with *C. difficile* to understand mechanisms by which they may determine outcome of treatment. Nevertheless, these biomarkers potentially allow identification of subsets patients that may be initially treated with more effective therapies such as newer antibiotics, faecal microbiota transplantation or defined microbiota consortia instead of a prolonged therapeutic trial.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Receiver operating characteristic (ROC) curve analysis of the most distinctive operational taxonomic units in fecal samples collected prior to treatment following 10-fold jack-knifing. The 10 ROC curves are in blue and the mean ROC curve is depicted in black.

**Figure S2.** Difference in beta-diversity among responders and non-responders. Analyses performed on 16S rRNA V4 region data, with a rarefaction depth of 6,987 reads per sample. (A) Principal Coordinate Analysis (PCoA) of unweighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. (B) Principal Coordinate Analysis (PCoA) of weighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. There are no significant differences in beta diversity among responders and non-responders samples (un-weighted UniFrac distance metric: \(R = -0.09, p = 0.9\); weighted UniFrac distance metric: \(R = -0.004, p = 0.5\)).

**Figure S3.** Difference in alpha-diversity among responders and non-responders using both phylogenetic (Faith’s phylogenetic diversity) and non-phylogenetic (observed species, Shannon index, chao 1 index) richness metrics. Analyses were performed on 16S rRNA V4 region data, with a rarefaction depth of 6,987 reads per sample. Whiskers in the boxplot represent the range of minimum and maximum alpha diversity values within a population, excluding outliers.

**Figure S4.** Differences in glycoside hydrolases: Significantly different Carbohydrate-Active Enzymes database - glycoside hydrolases (CAZY GH) in samples collected prior to treatment between responders \((n = 77)\) and non-responders \((n = 11)\) using Linear discriminant analysis Effect Size analysis (LEfSe). Boxplots denote top quartile, median and bottom quartile.

**Figure S5.** Receiver operating characteristic analysis of the most distinctive operational taxonomic units in fecal samples collected prior to treatment following 10-fold
jack-knifing. The 10 ROC curves are in blue and the mean ROC curve is in black.

**Figure S6.** Difference in beta-diversity among patients with and without recurrence after successful treatment. Analyses performed on 16S rRNA V4 region data, with a rarefaction depth of 6,987 reads per sample. (A) Principal Coordinate Analysis (PCoA) of unweighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. (B) Principal Coordinate Analysis (PCoA) of weighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. There are no significant differences between those with and without recurrence (unweighted UniFrac distance metric: $R = -0.1, p = 0.9$; weighted UniFrac distance metric: $R = 0.02, p = 0.3$).

**Figure S7.** Difference in alpha-diversity among patients with and without recurrence after successful treatment. Analyses were performed on 16S rRNA V4 region data, with a rarefaction depth of 6,987 reads per sample. Whiskers in the boxplot represent the range of minimum and maximum alpha diversity values within a population, excluding outliers. There was no significant difference in alpha diversity based on four different metrics between those with and without recurrence.

**Figure S8.** Differences in glycoside hydrolases: Significantly different GT30 (β-fucosidase) and a carbohydrate-binding module (CBM20) in patients who had a recurrence compared to patients who did not with LDA score ($\log_{10} > 2$) using Linear discriminant analysis Effect Size analysis (LEfSe). Boxplots denote top quartile, median and bottom quartile.

**Table S1.** Prior antibiotic exposure in patients with primary *Clostridium difficile* infection.

**Table S2.** Correlations between clinical characteristics and bacterial taxa

**Data S1.** Data acquisition and analyses.

**Data S2.** R Code to generate risk index.

**REFERENCES**


45. Mazmanian SK, Kasper DL. The love-hate relationship between bacterial polysaccharides and the host immune