(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 June 2017 (15.06.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/100420 A1

(51) International Patent Classification:
G01N 33/569 (2006.01) *A61K 31/4164* (2006.01)

(21) International Application Number:
PCT/US2016/065566

(22) International Filing Date:
8 December 2016 (08.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/264,479 8 December 2015 (08.12.2015) US

(71) Applicant: **MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH** [US/US]; 200 First Street S.W., Rochester, Minnesota 55905 (US).

(72) Inventors: **KHANNA, Sahil**; 2032 Fox Valley Drive S.W., Rochester, Minnesota 55902-3447 (US). **KASHYAP, Purna C.**; 2126 Glenview Ln N.E., Rochester, Minnesota 55906-8398 (US). **PARDI, Darrell S.**; 4558 Mayo Woodlands Road S.W., Rochester, Minnesota 55902-3611 (US).

(74) Agents: **WILLIS, Margaret S. J.** et al.; Fish & Richardson P.C., PO Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: BIOMARKERS FOR PREDICTING CLOSTRIDIUM DIFFICILE INFECTION TREATMENT OUTCOME

(57) Abstract: Provided herein are biomarkers that can be altered in the gut of a mammal having CDI. One or more biomarkers can be used to predict Clostridium difficile infection (CDI) treatment outcome (e.g., response to primary CDI treatment and/or risk of recurrence after primary CDI treatment) in a mammal (e.g., a human) having CDI. Methods for using one or more biomarkers can be used to predict response to primary CDI treatment in a mammal having CDI. Methods for using one or more biomarkers can be used to predict risk of recurrence in a mammal having CDI. Also provided are methods of treating CDI.

WO 2017/100420 A1

BIOMARKERS FOR PREDICTING *CLOSTRIDIUM DIFFICILE* INFECTION TREATMENT OUTCOME

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Patent Application Serial No. 62/264,479, filed on December 8, 2015. The disclosure of the prior application is considered part of (and is incorporated by reference in) the disclosure of this application.

TECHNICAL FIELD

This document relates to methods and materials for predicting *Clostridium difficile* infection (CDI) treatment outcome. For example, this document provides methods for using one or more biomarkers to predict a response to primary CDI treatment and/or to predict risk of recurrence after primary CDI treatment in a mammal (*e.g.*, a human).

BACKGROUND

CDI is the most common hospital-acquired infection and a leading cause of diarrhea in the United States with an estimated 450,000 infections per year and over 29,000 deaths attributed to CDI (Kelly et al., 2008 *N Engl J Med* 359:1932-40; Magill et al., 2014 *N Engl J Med* 370:1198-208; and Lessa et al., 2015 *N Engl J Med* 372:825-34). Despite advancements in drug therapy and the emergence of standard treatment guidelines, CDI is associated with several adverse outcome including the development of severe-complicated CDI (up to 5%), treatment failure (up to 15%) and recurrent CDI (up to 60% or higher after 3 episodes) (Leffler et al., 2015 *N Engl J Med* 372:1539-48; Khanna et al., 2014 *Therap Adv Gastroenterol* 7:72-86). CDI also has a tremendous economic impact with attributable costs ranging from $8,911 to $30,049 for hospitalized patients with CDI (Nanwa et al., 2015 *Am J Gastroenterol* 110:511-9).

SUMMARY

There are several clinical models to predict the risk of recurrent CDI, although none are robust enough for routine clinical use and there are no models to predict treatment response after initial antibiotic therapy.

This document relates to biomarkers and methods for predicting CDI treatment outcome. There are no clinically used robust models to predict treatment failure or recurrent CDI after initial antibiotic treatment and as a result 5 – 35% patients have to undergo prolonged treatment trials prior to receiving effective treatment.

Provided herein are materials (*e.g.*, biomarkers) and methods to predict CDI treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment). As described herein, microbial communities in patients with CDI were characterized to identify biomarkers that can be used to predict a response to primary CDI treatment and/or to predict risk of recurrence after primary CDI treatment in a mammal (*e.g.*, a human). Biomarkers potentially allow identification of subsets patients that may be primarily treated with more effective therapies such as fecal microbiota transplantation or defined microbial consortia instead of a prolonged therapeutic trial with antibiotics.

This document provides methods for predicting response to primary CDI treatment in a mammal having CDI. Primary CDI treatment can include, for example, administration of a pharmaceutical (*e.g.*, metronidazole, vancomycin, fidaxomicin, and combinations thereof). In some cases, a method for predicting response to primary CDI treatment can include detecting a gut microbiota panel in a fecal sample obtained from the mammal and identifying the mammal as being likely to respond to a primary CDI treatment if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel (*e.g.* gut microbiota from one or more mammals that do not have CDI). Detecting a gut microbiota can include, for example, isolation of DNA from the fecal sample, identification of operational taxonomic units, and/or 16s sequencing. A gut microbiota panel for can include, for example, *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium* (*e.g.*, *Faecalibacterium prausnitzii*), and *Rothia* (*e.g.*, *Rothia mucilaginosa*).

In some cases, a method for predicting response to primary CDI can include detecting a clinical parameter panel for the mammal and identifying the mammal as being likely to have primary CDI treatment failure if one or more of the clinical parameters are detected. A clinical parameter panel can include, for example, severe CDI, prior antibiotic (*e.g.*, a β-lactam antibiotic such as a cephalosporin) use, and hospital-acquired *C. difficile*.

In some cases, a method for predicting response to primary CDI treatment can include detecting a mode of *C. difficile* acquisition and identifying the mammal as being likely to respond to a primary CDI treatment if the mode of *C. difficile* acquisition is community acquired *C. difficile*.

This document provides methods for predicting risk of recurrence in a mammal having CDI. In some cases, a method for predicting risk of recurrence in a mammal having CDI can include detecting a gut microbiota panel in a fecal sample obtained from the mammal and identifying the mammal as being likely to have CDI recurrence if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel (*e.g.*, gut microbiota from one or more mammals that do not have CDI). Detecting a gut microbiota can include, for example, isolation of DNA from the fecal sample, identification of operational taxonomic units, and/or 16s sequencing. A gut microbiota panel can include, for example, *Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides,* and *Lachnospiraceae*.

This document provides methods of treating CDI in a mammal. Methods of treating CDI can include primary CDI treatment or advanced CDI treatment. Primary CDI treatment can include, for example, administration of a pharmaceutical (*e.g.*, metronidazole, vancomycin, fidaxomicin, and combinations thereof). Advanced CDI treatment can include, for example, fecal transplantation, immunoglobulin therapy, defined microbial consortia, and combinations thereof.

In some cases, a method of treating CDI in a mammal can include detecting a gut microbiota panel in a fecal sample obtained from the mammal and administering a primary CDI treatment to the mammal if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel (*e.g.*, gut microbiota from one or more mammals that do not have CDI). Detecting a gut microbiota can include, for example, isolation of DNA from the fecal sample, identification of operational taxonomic units, and/or 16s sequencing. Primary CDI treatment can include, for example, administration of a pharmaceutical selected from the group consisting of metronidazole, vancomycin, fidaxomicin, and combinations thereof. A gut microbiota panel can include, for example, *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium*, and *Rothia*.

In some cases, a method of treating CDI can include detecting a clinical parameter (*e.g.*, a mode of *C. difficile* acquisition) for the mammal and administering a primary CDI treatment to the mammal if the mode of *C. difficile* acquisition is community acquired *C. difficile*.

In some cases a method for treating CDI in a mammal can include detecting a gut microbiota panel in a fecal sample obtained from the mammal and administering advanced CDI treatment to the mammal if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel (*e.g.*, gut microbiota from one or more mammals that do not have CDI). Detecting a gut microbiota can include, for example, isolation of DNA from the fecal sample, identification of operational taxonomic units, and/or 16s sequencing. A gut microbiota panel can include, for example, *Veillonella* (*e.g.*, *Veillonella dispar* or *Veillonella parvula*), *Enterobacteriaceae* (*e.g.*, *Enterobacteriaceae erwinia*), *Streptococcus, Parabacteroides* (*e.g.*, *Parabacteroides distasonis*), and *Lachnospiraceae*.

In some cases, a method of treating CDI in a mammal can include detecting a panel of clinical parameters for the mammal and administering advanced CDI treatment to the mammal if one or more of the clinical parameters are detected. A panel of clinical parameters can include, for example, severe CDI, prior antibiotic (*e.g.*, a β-lactam antibiotic such as a cephalosporin) use, and hospital-acquired *C. difficile*.

Methods provided herein can also include identifying the mammal (*e.g.*, a human) as having CDI.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Methods and materials are described herein for use in the present disclosure; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

Figure 1 shows a summary of the operational taxonomic units (OTUs) associated with response or non-response to treatment using Linear discriminant analysis Effect Size analysis. Representative bacteria with relative abundance of at least 1% with significant differences are represented.

Figure 2 shows a 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.

Figures 3 A-C show a risk index of non-response to treatment from this panel of OTUs that were significantly different between responders and non-responders. (A) CDI risk index based on the differentiated operational taxonomic units in non-responders, responders, and healthy controls calculated using the Mann–Whitney U test: ***: $p < 0.001$. The boxplots denote top quartile, median and bottom quartile and individual dots represent individual patient samples. (B) ROC curve analysis of the CDI risk index in pre-treatment fecal samples collected following 10-fold cross-validation (left graph), and ROC curve analysis of the CDI risk index in pre-treatment fecal samples collected following 10-fold cross-validation of all the held-out samples from the leave-one-out cross-validation analysis (right graph). Following the leave-one-out procedure, the ROC curves of each leave-one-out procedure are in blue and the mean ROC curve of all the procedure is in black. (C) Beeswarm plot of all the held-out samples from the leave-one-out (LOO) cross-validation ($p < 0.0001$).

Figures 4 A-B show Spearman correlations demonstrating taxon-taxon relative abundance including links with absolute value of correlation > 0.5 and False Discovery Rate-corrected p-value < 0.05. Network analyses displayed with Cytoscape using an edge-weighted spring embedded layout in responders (A) and non-responders (B). Positive correlations were in blue and the size of the edge is based on the strength of the correlation. The size of the node was based on the number of correlations associated with the corresponding taxon. The color of the node is based on the phylum level: Actinobacteria (blue), Bacteroidetes (green), Firmicutes (purple), Proteobacteria (yellow), Fusobacteria (pink) and Verrucomicrobia (grey). There is a three-fold decrease in the number of taxon

correlations in the non-responders and most of the decreased nodes were Firmicutes and Actinobacteria.

Figures 5 A-B show beta-diversity comparisons of the gut microbiomes of the fecal samples collected in 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) PCoA of weighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. This is no clustering between responder and non-responder samples (unweighted UniFrac distance metric: $R = -0.0888$, $p = 0.980$; weighted UniFrac distance metric: $R = -0.0035$, $p = 0.504$).

Figure 6 shows alpha-diversity comparisons of the gut microbiomes of the fecal samples collected from responders, non-responders, and healthy subjects 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. There is a trend for alpha diversity decrease in fecal samples collected in non-responders as compared with alpha diversity from samples collected in responders, observed with both phylogenetic (Faith's phylogenetic diversity) and nonphylogenetic (observed species, Shannon index, chao 1 index) richness metrics.

Figure 7 shows significantly increased carbohydrate-active enzymes GH70, GT38 (dextransucrase), and GH38 (α-mannosidase); and significantly decreased GH59 (β-galactosidase) and two carbohydrate-binding modules, CBM16 and CBM42 in non-responders ($n = 11$) compared to responders ($n = 77$) as shown by Mann–Whitney test. Boxplots denote top quartile, median and bottom quartile.

Figure 8 shows a summary of the OTUs associated with recurrence (n=22) versus non-recurrence (n=55) after successful treatment response using Linear discriminant analysis Effect Size analysis. Representative bacteria with relative abundance of at least 1% with significant differences are represented.

Figure 9 shows a 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.

Figures 10 A-C show a risk index of non-response to treatment from this panel of OTUs that were significantly different between patients with and without recurrence. (A) CDI risk index based on the differentiated OTUs in patients with success versus recurrence. Mann–Whitney U test: ***: $p < 0.001$. Boxplots denote top quartile, median and bottom quartile. (B) ROC curve analysis of the CDI risk index in fecal samples collected prior to treatment predicting recurrence. (C) ROC curve analysis of the CDI risk index in fecal samples collected prior to treatment. Following the leave-one-out procedure, the ROC curves of each leave-one-out procedure are in blue and the mean ROC curve of all the procedure is in black.

Figures 11 A-B show beta-diversity comparisons of the gut microbiomes of the fecal samples collected in responders and non-responders. Analyses performed on 16S rRNA V4 region data, with a rarefaction depth of 6,987 reads per sample. (A) PCoA of unweighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. (B) PCoA of weighted UniFrac distances. Proportion of variance explained by each principal coordinate axis is denoted in the corresponding axis label. This is no clustering between those with and without recurrence (unweighted UniFrac distance metric: $R = -0.0961$, $p = 0.993$; weighted UniFrac distance metric: $R = 0.0165$, $p = 0.331$).

Figure 12 shows alpha-diversity comparisons of the gut microbiomes of the fecal samples collected in responders and non-responders. 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 in fecal samples collected from patients with recurrence as compared with alpha diversity from samples collected patients without recurrence, observed with both phylogenetic (Faith's phylogenetic diversity) and nonphylogenetic (observed species, Shannon index, chao 1 index) richness metrics.

Figure 13 shows significantly increased carbohydrate-active enzymes GT30 (β-fucosidase) and CBM20 (carbohydrate-binding module) in patients with recurrence

compared to patients without recurrence as shown by Mann–Whitney test. Boxplots denote top quartile, median and bottom quartile.

DETAILED DESCRIPTION

This document provides methods and materials involved in predicting CDI treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment) in a mammal (*e.g.*, a human) having CDI. For example, this document provides one or more biomarkers that can be altered in the gut of a mammal having CDI. In some cases, this disclosure provides methods for using one or more biomarkers to predict response to primary CDI treatment in a mammal having CDI. In some cases, this disclosure provides methods to predict risk of recurrence in a mammal having CDI. In some cases, this disclosure provides methods of treating CDI.

Any type of mammal having CDI can be evaluated to predict CDI treatment outcome and/or treated as described herein. For example, humans and other primates such as monkeys having pancreatitis can be treated with one or more colipase inhibitors. In some cases, dogs, cats, horses, cows, pigs, sheep, rabbits, mice, and rats can be evaluated to predict CDI treatment outcome and/or treated as described herein.

Any appropriate method can be used to identify a mammal (*e.g.* a human) having CDI. For example, detection of bacterial polypeptides or toxins produced by *C. difficile* bacteria in a stool sample (*e.g*, by enzyme immunoassay, polymerase chain reaction, cell cytotoxicity assay, *etc.*) can be used to identify a human or other mammal as having CDI. In some cases, a CDI diagnosis may be confirmed using, for example, colon examinations and/or imaging tests.

Once identified as having CDI, a mammal can be evaluated to predict CDI treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment). A mammal can be evaluated by, for example, detecting one or more biomarkers in a biological sample (*e.g.*, a fecal sample) obtained from the mammal.

As used herein, a "biomarker" is a characteristic that can be objectively measured and evaluated as a predictor of treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment). A biomarker that can be used to predict treatment outcome can include, for example, a bacterium (*e.g.*, a gut bacterium), a

clinical predictor (*e.g.*, severity of CDI, prior antibiotic use, and/or mode of *C. difficile* acquisition), a microbiota functional repertoire (*e.g.*, a polypeptide such as an enzyme) produced by a bacterium present in the gut, or combinations thereof. A biomarker panel can encompass any two or more (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more) of the biomarkers disclosed herein. For example, a biomarker panel can include one or more bacteria, one or more clinical parameters, one or more polypeptides present in the gut, or any combination thereof.

A gut bacterium biomarker for predicting treatment outcome (*e.g.*, response to primary CDI treatment) in a mammal (*e.g.*, a human) having CDI can be, for example, *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium,* or *Rothia*. In some cases, one or more gut bacterial biomarkers for predicting response to primary CDI treatment can be in a biomarker panel. For example, a biomarker panel for predicting treatment response to primary CDI treatment can include two or more (*e.g.*, two, three, four, five, or six) of *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium,* and *Rothia*. The *Faecalibacterium* can be *Faecalibacterium prausnitzii*. The *Rothia* can be *Rothia mucilaginosa*.

A gut bacterium biomarker for predicting treatment outcome (*e.g.*, risk of recurrence after primary CDI treatment) in a mammal (*e.g.*, a human) having CDI can be, for example, a *Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides,* or *Lachnospiraceae*. In some cases, one or more gut bacterial biomarkers for predicting risk of recurrence after primary CDI treatment can be in a biomarker panel. For example, a biomarker panel for predicting treatment response can include two or more (*e.g.*, two, three, four, or five) of *Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides,* and *Lachnospiraceae*. The *Enterobacteriaceae* can be *Enterobacteriaceae erwinia*. The *Veillonella* can be *Veillonella* dispar or *Veillonella parvula*. The *Parabacteroides* can be *Parabacteroides distasonis*.

A clinical parameter biomarker for predicting treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment) in a mammal (*e.g.*, a human) having CDI can be, for example, severity of CDI, prior antibiotic use (*e.g.*, β-lactam antibiotics such as cephalosporins, penicillins such as amoxicillin, and/or floroquinolones such as ciprofloxacin), the mode of *C. difficile* acquisition (*e.g.*, community acquired *C. difficile* or hospital acquired *C. difficile*), prior antineoplastic use (*e.g.*,

methotrexate and paclitaxel), prior proton pump inhibitor (PPI) use (*e.g.*, omeprazole, lansoprazole, dexlansoprazole, esomeprazole, pantoprazole, rabeprazole, and ilaprazole), and combinations thereof. In some cases, one or more clinical parameters for predicting response to primary CDI treatment can be in a biomarker panel. For example, a biomarker panel for predicting treatment response to primary CDI treatment can include two or more of severity of CDI, prior antibiotic use, and mode of *C. difficile* acquisition.

A polypeptide biomarker (*e.g.*, an enzyme produced by a bacterium present in the gut) for predicting treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment) in a mammal (*e.g.*, a human) having CDI can be, for example, a polypeptide can be a carbohydrate-active enzyme. Non-limiting examples of carbohydrate-active enzymes include polysialyltransferases (*e.g.*, GT38), dextransucrases (*e.g.*, GH70), α-mannosidases (*e.g.*, GH38), β-galactosidases (*e.g.*, GH59), and carbohydrate-binding modules (*e.g.*, CBM16, CBM42). In some cases, one or more (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more) polypeptides for predicting risk of CDI recurrence after primary CDI treatment can be in a biomarker panel. For example, a biomarker panel for predicting risk of recurrence after primary CDI treatment can include two or more of GT38, GH70, GH38, GH59, CBM16, and CBM42.

A polypeptide biomarker for predicting treatment outcome (*e.g.*, response to primary CDI treatment and/or risk of recurrence after primary CDI treatment) in a mammal (*e.g.*, a human) having CDI can be, for example, a carbohydrate-active enzyme. Non-limiting examples of carbohydrate-active enzymes include β-fucosidases (*e.g.*, GT30), and carbohydrate-binding modules (*e.g.*, CBM20). In some cases, one or more (*e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more) polypeptides for predicting risk of CDI recurrence after primary CDI treatment can be in a biomarker panel. For example, a biomarker panel for predicting risk of recurrence after primary CDI treatment can include at least one of GT30 and CBM20.

One or more gut microbiota biomarkers described herein and/or one or more polypeptides present in the gut described herein can be detected in a biological sample (*e.g.*, a fecal sample) using any suitable technology. Methods of detecting a bacterium can include, for example, identification of operational taxonomic units and 16s sequencing. In some cases, detecting a bacterium can also include isolation of DNA from the biological sample

(*e.g.*, a fecal sample). Methods of detecting a polypeptide can include, for example, immunoprecipitation, electrophoresis, western blot, spectrophotometry, and enzyme assay.

One or more clinical parameter biomarkers described herein can be detected by any suitable method. Methods of detecting a clinical parameter can include, for example, reviewing a medical record (*e.g.*, an electronic medical record or a physical medical record), obtaining information by word of mouth (*e.g.*, from a clinician or from the patient), and by clinical testing for the parameter.

Detection of one or more biomarkers described herein can be used to predict a response to primary CDI treatment in a mammal having CDI. For example, a mammal can be identified as being likely to respond to primary CDI treatment if one or more (*e.g.*, two or more, three or more, four or more, five or more, *etc.*) biomarkers described herein are altered (*e.g.*, increased or decreased) or identified relative to a control biomarker panel. As used herein, a response to primary CDI treatment means a mammal may show a reduction or resolution in CDI symptoms within 4-5 days of primary CDI treatment. In some cases, a mammal can be identified as being likely to response to primary CDI treatment if one or more (*e.g.*, two or more, three or more, four or more, or five or more) of *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium* (*e.g.*, *Faecalibacterium prausnitzii*), and *Rothia* (e.g., *Rothia mucilaginosa*) are increased. In some cases, a mammal can be identified as being likely to response to primary CDI treatment if the mammal is identified with community acquired *C. difficile*. In some cases, mammal can be identified as being likely to respond to primary CDI treatment if one or more (*e.g.*, two or more) of GT38, GH70, and GH38 are increased and one or more (*e.g.*, two or more) of GH59, CBM16, and CBM42 are decreased. For example, a mammal can be identified as being likely to have primary CDI treatment failure if one or more (*e.g.*, two or more, three or more, four or more, five or more, *etc.*) biomarkers described herein are altered (*e.g.*, increased or decreased) or identified relative to a control biomarker panel. As used herein, failure to respond to primary CDI treatment means a mammal may show no reduction in CDI symptoms within 4-5 days of primary CDI treatment. In some cases, a mammal can be identified as being likely to have primary CDI treatment failure if the mammal is identified with one or more (*e.g.*, two or more, or three or more) of severe CDI, prior antibiotic use (*e.g.*, β-lactam antibiotics such as

cephalosporins, penicillins such as amoxicillin, and/or floroquinolones such as ciprofloxacin), and hospital-acquired *C. difficile*.

Detection of one or more biomarkers described herein can be used to predict risk of recurrence after primary CDI treatment in a mammal having CDI. For example, a mammal can be identified as being at risk of recurrence after primary CDI treatment if one or more (*e.g.*, two or more, three or more, four or more, five or more, *etc.*) biomarkers described herein are altered (*e.g.*, increased or decreased) or identified relative to a control biomarker panel. As used herein, recurrence after primary CDI treatment means a mammal may show recurrence of CDI symptoms (*e.g.*, diarrhea) after reduction or resolution of symptoms at any time during the primary CDI treatment or within about 30 to 75 days (*e.g.*, within 35 to 70 days, within 40 to 65 days, within 45 to 60 days, within 50 to 55 days of the primary CDI treatment) of CDI onset. In some cases, a mammal can be identified as being at risk of recurrence if one or more (*e.g.*, two or more, three or more, or four or more) of *Veillonella* (*e.g.*, *Veillonella dispar* or *Veillonella parvula*), *Enterobacteriaceae* (*e.g.*, *Enterobacteriaceae erwinia*), *Streptococcus, Parabacteroides,* and *Lachnospiraceae* are increased. In some cases, a mammal can be identified as being at risk of recurrence if the mammal is identified with one or more (*e.g.*, two or more, three or more, or four or more) of increasing age, prior antibiotic (*e.g.*, metronidazole and/or vancomycin) use, decreased anti-toxin IgG levels, the presence of comorbidities, and use of gastric acid suppression medications. In some cases, a mammal can be identified as being at risk of recurrence if one or more of GT30 and CBM20 are increased.

Once identified as being likely to respond to primary CDI treatment, a mammal can be treated for CDI. For example, a mammal identified as being likely to respond to primary CDI treatment can be treated by, for example, administering one or more primary CDI therapies. Non-limiting examples of primary CDI therapies include oral delivery of antimicrobials (*e.g.*, metronidazole, vancomycin, fidaxomicin, and combinations thereof.

Once identified as being at risk of CDI recurrence after primary CDI treatment, a mammal can be treated for CDI. For example, a mammal identified as being at risk of CDI recurrence after primary CDI treatment can be treated by, for example, administering one or more advanced CDI therapies. Non-limiting examples of advanced CDI therapies include

pulsed antimicrobial regimens, fecal microbiota transplantation, defined microbial consortia, surgery (*e.g.*, to remove diseased portions of the colon), and combinations thereof.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1: Study Design

Patients with a first *Clostridium difficile* infection (CDI) episode were recruited. Consecutive patient stool samples that led to CDI diagnosis were collected.

Clinical data including demographics, hospitalization, concomitant medications, severity, laboratory parameters, prior and concomitant antibiotic use, initial CDI treatment, treatment response and recurrent CDI were obtained by a review of the electronic medical record.

CDI was defined as community-acquired if onset of symptoms occurred in the community or within 48 hours of admission to a hospital, provided symptom onset was >12 weeks after the last discharge from a hospital. CDI was defined as hospital acquired if onset of symptoms occurred after 48 hours of admission to a hospital or within 4 weeks of last hospital discharge.

CDI was considered indeterminate if symptom onset occurred between 4 and 12 weeks after a hospital dismissal (Khanna et al., 2012 *Am J Gastroenterol* 107:89-95; and Cohen et al., 2010 *Infect Control Hosp Epidemiol* 31:431-55). Severe CDI was defined as a peripheral white blood cell count ≥15,000 cells/mm3 or a serum creatinine rise of ≥50% from baseline within 7 days of diagnosis. CDI was classified as severe-complicated if the infection was complicated by sepsis, hypotension, ileus, toxic megacolon, perforation, need for intensive care unit admission, surgery for a CDI-related complication (*e.g.*, megacolon, perforation, refractory colitis), or death (Cohen et al., 2010 *Infect Control Hosp Epidemiol* 31:431-55).

Treatment failure was defined as non-response to treatment after 5 days of metronidazole treatment or 4 days of vancomycin therapy (Khanna et al., 2012 *Am J Gastroenterol* 107:89-95; Cohen et al., 2010 *Infect Control Hosp Epidemiol* 31:431-55; and Surawicz et al., 2013 *Am J Gastroenterol* 108:478-98). Recurrent CDI was defined as

recurrence of diarrhea after complete resolution of symptoms at any time during the treatment or within 56 days of onset with a positive CDI test (Cohen et al., 2010 *Infect Control Hosp Epidemiol* 31:431-55).

Stool samples were processed for next generation sequencing as described below and changes in microbial community structure and function were correlated to treatment response and risk of recurrence.

Example 2: Sequencing and Analytic Methods

16S sequencing data processing. Fecal samples were kept frozen at -80°C until they were processed. After fecal DNA isolation (MoBio, Carlsbad, CA fecal DNA kit), amplicons spanning the variable region 4 of bacterial 16S rRNA were generated and sequenced using illumina Mi Seq platform. 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 Quantitative Insights Into Microbial Ecology (QIIME) 1.8.0 software (Caporaso et al., 2010 *Nat Methods* 7:335-6). Closed reference OTU assignment was performed using 'uclust' with a 97% sequence similarity threshold against the Greengenes 13.8 database as a reference (DeSantis et al., 2006 *Appl Environ Microbiol* 72:5069-72).

Data analyses. Built-in functions of the QIIME pipeline were applied. Alpha diversity metrics were computed using non-phylogeny based metrics (Observed species, Chao1 index, Shannon index) and phylogeny-based metrics (Phylogenetic Distance whole tree). Differences between fecal samples from responders and non-responders or from patients with or without recurrence after initial response, were tested for differences with a Monte Carlo permuted t-test. Beta diversity metrics were computed using unweighted and weighted UniFrac distances and visualized with Principal Coordinate Analysis (PCoA) (Lozupone et al., 2011 *ISME J* 5:169-72). The differences between fecal samples from responders and non-responders or from patients with or without recurrence after initial response were treated with the Permutational Multivariate Analysis of Variance Using Distance Matrices (PERMANOVA) method (Anderson, 2001 *Austral Ecology* 26:32-46).

Comparisons of relative abundance of taxa between responders and non-responders, or between patients with and without recurrence after initial response, was performed using

Linear discriminant analysis Effect Size (LEfSe), a non-parametric 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 (Segata et al., 2011 *Genome Biol* 12:R60). A LDA score (log 10) > 2 was considered significant.

A risk index was built to differentiate responders from non-responders to initial treatment or patients with and without recurrence after initial response, based on taxonomy. A non-collapsed OTU table was used. 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 on the OTU table) were summed and the relative abundances of the taxa associated with non-responders to treatment (based on the LEfSe output) 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.

A leave-one-out cross-validation procedure was also conducted to evaluate the taxa that differentiated patients with or without response to treatment in the held-out patient, based on the LEfSe output of n-1 patients. The taxa that differentiated patients with or without response to treatment were identified using LEfSe in the n-1 dataset. Then, the relative abundances of the taxa associated with response to treatment (based on the LEfSe output of the n-1 dataset) were identified and the relative abundances of taxa associated with no response to treatment (based on the LEfSe output of the n-1 dataset) on the held-out patient summed. Then the difference between these two sums (relative abundance of taxa associated with no response to treatment minus relative abundance of taxa associated with response to treatment) was calculated and a risk index obtained for the held-out patient. The procedure was repeated n times to produce a risk index in each of the held-out patients based on the difference between the sum of the taxa associated with the no response to treatment minus the sum of the taxa associated with response to treatment found in each of the n-1 datasets. The significance of the difference in risk indexes between the patients with

response to treatment and patients without response to treatment was assessed using a permutation test, in which we permuted the labels of the outcome (that is response or non-response to treatment) 999 times, each time deriving new held-out predictions of the risk indexes for each subject for that outcome. The two procedures (creation of a risk index and leave-one-out cross-validation) were applied in the exact same way to compare patients with recurrence or no recurrence after initial response to treatment.

Receiving operating characteristic (ROC) curves were plotted and the area under the curve (AUC) values were computed on a dataset containing 10 sets of predictions and corresponding labels obtained from 10-fold cross-validation using ROCR package in R (Sing et al., 2005 *Bioinformatics* 21:3940-1; R Development Core Team 2011 ISBN: 3-900051-07-0). A risk index threshold was also determined that best predicted response or no-response to treatment, or recurrence after initial response to treatment, with a leave-one-out cross-validation on the produced risk indexes from each patient. A ROC curve that included risk indexes from n-1 patients was built. The cutoff associated with a given specificity was determined. This cutoff was used to determine the outcome of the held-out patient. Thus, each held-out sample was treated as a new patient on whom the optimal risk index cutoff was tested to separate patients who respond to treatment and patient who did not respond to treatment. This procedure was repeated n times and we calculated the accuracy of the cutoff by summarizing the good prediction of the outcome of each held-out patient. Boxplots, beeswarms and two-dimensional PCoA plots were generated using R (R Development Core Team 2011 ISBN: 3-900051-07-0).

Network analyses were displayed with Cytoscape using an edge-weighted spring embedded layout (Smoot et al., 2011 *Bioinformatics* 27:431-2). OTUs were collapsed to the genus level to eliminate OTUs present in fewer than 25% of samples. Genera present in less than 25% of the samples were removed to reduce the very large number of multiple hypotheses tested in correlation network analysis. Spearman correlation of taxon-taxon relative abundance was performed and included only those links with absolute value of correlation > 0.5 and False Discovery Rate (FDR)-corrected p-value < 0.05.

Example 3: Clinical Characteristics Predict Response CDI Therapy and CDI Recurrence

Stool samples were collected from 88 patients (median age 52.7 years, interquartile range 36.9 – 65.1; 60.2% female) and clinical outcomes were recorded prospectively. The rate of primary non-response following recommended treatment was 12.5%, and was similar between metronidazole and vancomycin (Table 1).

In order to identify clinical characteristics that predict response to treatment in CDI, chi-square and univariate analyses were performed. There were no significant differences in age, sex, body-mass index, treatment regimen and concomitant systemic antibiotic exposure in primary non-responders compared to responders (Table 1). There were non-significant trends in prior antibiotic exposure, acquisition mode age, and severity of CDI in primary non-responders compared to responders (Table 1).

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/M^2)	27.5	27.4	28.1	0.75	27.7	26.6	0.44
Prior antibiotic exposure (%)	59	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	5.6	5.2	9.1	0.8	7.3	0	0.1

metronidazole							
Treatment failure (%)	12.5	—	—	—	—	—	—
Recurrence after successful treatment (%)	—	28.5	—	—	—	—	—

*: Denotes p-value for comparison of treatment responders versus treatment failures
**: Denotes p-value for comparison of recurrent infections versus non-recurrent infection
CDI = *Clostridium difficile* infection, BMI = Body mass index

These results demonstrate that clinical parameters of prior antibiotic exposure, community-acquired CDI, and/or severe CDI can predict patient response to primary CDI treatment.

Example 4: Gut Microbiota Signatures Can Predict CDI Response To Therapy

Of the 88 fecal samples collected, a total of 1,449,211 high-quality 16S rRNA gene-encoding sequences were identified, representing 7,470 OTUs. The mean number of sequences obtained per sample was 16,468 ± 4,674. Since samples contained between 6,987 and 35,494 sequences, diversity analyses were rarefied at 6,987 sequences per sample to avoid bias.

A panel of 36 OTUs that were significantly different between primary non-responders and responders was identified using LEfSe (significance corresponding to a LDA ($\log_{10}$)>2). Patients with successful response to treatment had a significant increase in OTUs within Ruminococcaceae, Rikenellaceae, *Bacteroides* species, *Faecalibacterium prausnitzii,* while primary non-responders had a significant increase in *Clostridiaceae*, Lachnospiraceae, *Blautia* species, *Coprococcus*, *Clostridium* species, *Clostridium perfringens, Streptococcus* species, *Bifidobacterium*, *Ruminococcus gnavus* and *Actinomyces* species (Figure 1).

The ability of this panel of microbes to discriminate between responders and non-responders was tested using jackknifed ROC curve analysis and found several individual OTUs to be strong predictors of response to treatment (Figure 2). Several individual OTUs were found to be strong predictors of response. *Faecalibacterium* species yielded a ROC-plot AUC value of 0.76, Ruminococcaceae yielded a ROC-plot AUC value of 0.86, *Bacteroides* species yielded a ROC-plot AUC value of 0.82, Rikenellaceae yielded a ROC-plot AUC value of 0.88, Lachnospiraceae yielded a ROC-plot AUC value of 0.86,

Streptococcus species yielded a ROC-plot AUC value of 0.86, *Clostridium* species yielded a ROC-plot AUC value of 0.79, *Clostridium perfringens* yielded a ROC-plot AUC value of 0.78 and *Ruminococcus gnavus* yielded a ROC-plot AUC value of 0.77. A risk index of non-response to treatment was built from this panel of OTUs that were significantly different between responders and non-responders. This risk index, calculated in each patient (responders and non-responders), corresponds to the difference between the sum of the relative abundance of all the OTUs associated with non-response to treatment and the sum the relative abundance of the OTUs associated with responses to treatment. This risk index was significantly different in the two groups of patients, with a mean score of 0.07 ± 0.24 in responders and 0.42 ± 0.52 in non-responders (Mann-Whitney U test, $p = 0.0002$) (Figure 3A). Moreover, the ROC curve analysis showed that this risk index was a strong predictor of treatment response, with an area under the curve of 0.85 (Figure 3B).

In order to determine the risk index threshold that best predicts response to treatment the leave-one-out cross-validation was used on our risk index. 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 non-responders. This LOO cross-validation procedure demonstrated that the sensitivity of a risk index threshold of 0.23 was 75% to predict the response to treatment in a new patient, at a specificity of 70%. Furthermore, another LOO procedure, where the risk index was built 88 times using n-1 samples each time and then tested on the held-out sample, showed the gut microbiota profile to be a strong predictor of response vs. non-response (permutation test; $p < 0.0001$; Figure 3C).

OTU networks are disrupted in primary non-responders.

Computed OTU networks at the genus level were computed to differentiate responders and non-responders. There was a three-fold (56 versus 16, ratio = 3.35) decrease in the number of strong taxon-taxon correlations (absolute value of correlation > 0.5 and FDR-corrected p-value < 0.05) in non-responders compared to responders, and most of the decreased nodes between responders and non-responders were members of the phylum Firmicutes (50 versus 10, ratio = 5) and Actinobacteria (7 versus 1, ratio = 7) (Figure 4).

Gut microbial diversity is not significantly different in primary non-responders compared to responders

Unweighted and weighted UniFrac based PCoA, based on 16S rRNA gene sequencing using MiSeq Illumina platform did not show significant differences between responders and non-responders (unweighted UniFrac distance metric: R = -0.0888, p = 0.98; weighted UniFrac distance metric: R = -0.0035, p = 0.50) (Figure 5). Similarly, although there was a trend towards decreased alpha diversity (phylogenetic diversity, observed species, shannon and chao1) in non-responders compared to responders, this difference was not significant (Figure 6).

Gut microbiota functional repertoire is significantly different in responders and non-responders.

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. We used PICRUSt with defaults parameters to normalize the OTU table by dividing each OTU by the known/predicted 16S copy number abundance (normalize_by_copy_number.py script). Then, we used the precalculated file containing the CAZY GH assignments, untitled cazy_13_5_precalculated.tab, and apply the predict_metagenomes.py script as recommended, specifying CAZY GH assignments in the type of prediction. The output file gave the CAZY GH assignment in each sample included in the OTU table. Using LEfSe, we found that GT38 GH70 (dextransucrase) and GH38 (α-mannosidase) were increased whereas GH59 (β-galactosidase) and two carbohydrate-binding modules, CBM16 and CBM42 were significantly decreased in non-responders compared to responders with LDA score ($\log_{10}$) >2 (Figure 7).

These results demonstrate that biomarkers, such as the gut microbiota signature and/or the gut microbiota functional repertoire (*e.g.*, polypeptides present in the gut) can predict response to primary CDI treatment, differentiating responders and non-responders.

Example 5: Gut Microbiota Signature Can Predict CDI Recurrence

Among the patients who initially responded to treatment, 28.5% had recurrent CDI (Table 1). There were no significant differences in age, sex, body-mass index, prior antibiotic exposure, acquisition mode, treatment regimen and concomitant systemic antibiotic exposure in patients with and without recurrent CDI (Table 1).

A panel of 11 OTUs that were significantly different between those with (n=22) and without (n=55) recurrence following successful treatment was identified using LEfSe. Patients with recurrent CDI had a significant decrease in OTUs related to *Veillonella* (*dispar, parvula*), Enterobacteriaceae (*Erwinia* species), *Streptococcus* species, *Parabacteroides distasonis* and Lachnospiraceae compared to patients without recurrence (Figure 8).

The ability of individual microbes to discriminate between patients who did or did not develop recurrent CDI was tested and several individual OTUs to be strong predictors of recurrence were found. *Veillonella dispar* yielded an AUC of 0.83, *Veillonella parvula* an AUC of 0.80, *Veillonella* an AUC of 0.82 and Lachnospiraceae an AUC of 0.81 (Figure 9).

A risk index of recurrence from this panel of microbes was built which differentiated between patients with and without recurrence. This risk index included all the OTUs that had a LDA score ($\log_{10}$) > 2. The risk index of recurrence is the difference between the sum of the relative abundance of the OTUs associated with recurrent CDI minus the sum of relative abundance of the OTUs associated with no recurrence. This risk index was significantly different in patients who did not have a recurrence (mean score 0.09 ±0.16) and those who did (0.19 ± 0.07) (Mann-Whitney U test, p-value = 0.0001) (Figure 10A). The ROC curve analysis showed that this risk index was a strong predictor of recurrence, with an AUC of 0.85 (Figure 10B).

A risk index threshold that best predicted recurrence with LOO cross-validation was determined. Each held-out sample was treated as a new patient on whom we tested and subsequently refined the optimal index cutoff to separate patients with recurrence vs. no recurrence. This procedure demonstrated that the sensitivity of a risk index cutoff of 0.036 to predict recurrence in a new patient was 82% at a specificity of 95%. Furthermore, using another LOO procedure, where the risk index was built 77 times using 76 samples and then tested on the held-out sample, the gut microbiota was a strong predictor of recurrence (permutation test performed on the difference between the mean of those without recurrence

and the mean of those with recurrence, of all the held-out samples, with 999 random permutations where the samples are all freely exchangeable; $p < 0.0001$; Figure 10C).

Pre-treatment gut microbial diversity is not significantly different in patients with and without recurrent CDI

Unweighted and weighted UniFrac based PCoA, based on 16S rRNA gene sequencing using MiSeq Illumina platform did not show significant differences between those with and without recurrence (unweighted UniFrac distance metric: R =-0.0961, p = 0.993; weighted UniFrac distance metric: R =0.0165, p =0.331) (Figure 11). Similarly, there was no significant difference in alpha diversity based on four different metrics between those with and without recurrence (Figure 12).

Gut microbiota functional repertoire is significantly different in those with and without recurrence

As described above, PICRUSt was used to predict CAZY GH assignments. Based on the LEfSe tool, it was found that GT30 (β-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 13).

These results demonstrate that biomarkers, such as the gut microbiota signature and/or the gut microbiota functional repertoire (*e.g.*, polypeptides present in the gut) can predict response to primary CDI treatment, differentiating patients with and without recurrence.

OTHER EMBODIMENTS

It is to be understood that while the disclosure has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the disclosure, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for predicting response to primary *Clostridium difficile* infection (CDI) treatment in a mammal having CDI, the method comprising:

detecting a gut microbiota panel in a fecal sample obtained from the mammal, wherein the gut microbiota panel comprises *Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium*, and *Rothia*; and

identifying the mammal as being likely to respond to a primary CDI treatment if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel.

2. A method of treating *Clostridium difficile* infection (CDI) in a mammal, the method comprising:

detecting a gut microbiota panel in a fecal sample obtained from the mammal, wherein the gut microbiota panel comprises *Ruminococcaceae*, *Rikenellaceae*, *Clostridiaceae*, *Bacteroides*, *Faecalibacterium*, and *Rothia*;

administering a primary CDI treatment to the mammal if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel.

3. A method for predicting response to primary *Clostridium difficile* infection (CDI) treatment in a mammal having CDI, the method comprising:

detecting a mode of *C. difficile* acquisition;

identifying the mammal as being likely to respond to a primary CDI treatment if the mode of *C. difficile* acquisition is community acquired *C. difficile*.

4. A method of treating *Clostridium difficile* infection (CDI) in a mammal, the method comprising:

detecting a clinical parameter for the mammal, wherein the clinical parameter is a mode of *C. difficile* acquisition;

administering a primary CDI treatment to the mammal if the mode of *C. difficile* acquisition is community acquired *C. difficile*.

5. A method for predicting response to primary *Clostridium difficile* infection (CDI) treatment in a mammal having CDI, the method comprising:

detecting a clinical parameter panel for the mammal, wherein the panel comprises severe CDI, prior antibiotic use, and hospital-acquired *C. difficile*;

identifying the mammal as being likely to have primary CDI treatment failure if one or more of the clinical parameters are detected.

6. The method of claim 5, wherein the antibiotic is a β-lactam antibiotic.

7. The method of claim 6, wherein the β-lactam antibiotic is a cephalosporin.

8. The method of claim 1 or claim 2, wherein *Faecalibacterium* is *Faecalibacterium prausnitzii*.

9. The method of claim 1 or claim 2, wherein *Rothia* is *Rothia mucilaginosa*.

10. The method of claims 1 - 5, wherein the primary CDI treatment comprises administration of a pharmaceutical selected from the group consisting of metronidazole, vancomycin, fidaxomicin, and combinations thereof.

11. A method for predicting risk of recurrence in a mammal having *Clostridium difficile* infection (CDI), the method comprising:

detecting a gut microbiota panel in a fecal sample obtained from the mammal, wherein the gut microbiota panel comprises *Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides,* and *Lachnospiraceae*; and

identifying the mammal as being likely to have CDI recurrence if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel.

12. A method for treating *Clostridium difficile* infection (CDI) in a mammal, the method comprising:

detecting a gut microbiota panel in a fecal sample obtained from the mammal, wherein the gut microbiota panel comprises *Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides,* and *Lachnospiraceae*;

administering advanced CDI treatment to the mammal if two or more gut microbiota in the panel are increased relative to a control gut microbiota panel.

13. A method of treating *Clostridium difficile* infection (CDI) in a mammal, the method comprising:

detecting a panel of clinical parameters for the mammal, wherein the panel comprises severe CDI, prior antibiotic use, and hospital-acquired *C. difficile*;

administering advanced CDI treatment to the mammal if one or more of the clinical parameters are detected.

14. The method of claim 13, wherein the antibiotic is a β-lactam antibiotic.

15. The method of claim 14, wherein the β-lactam antibiotic is a cephalosporin.

16. The method of claim 11 or claim 12, wherein the *Enterobacteriaceae* is *Enterobacteriaceae erwinia*.

17. The method of claim 11 or claim 12, wherein the *Veillonella* is *Veillonella dispar* or *Veillonella parvula*.

18. The method of claim 11 or claim 12, wherein the *Parabacteroides* is *Parabacteroides distasonis*.

19. The method of claim 12 or claim 13, wherein the advanced CDI treatment is selected from the group consisting of fecal transplantation, immunoglobulin therapy, and defined microbial consortia.

20. The method of claims 1, 2, 11 or 12, wherein the detecting comprises isolation of DNA from the fecal sample.

21. The method of claim 20, wherein the detecting comprises identification of operational taxonomic units.

22. The method of claim 20, wherein the detecting comprises 16s sequencing.

23. The method of claims 1, 2, 11 or 12, wherein the control gut microbiota panel comprises gut microbiota from one or more mammals that do not have CDI.

24. The method of claims 1 – 5 or claims 11 – 13, further comprising identifying the mammal as having CDI.

25. The method of claims 1 – 5 or claims 11 – 13, wherein the mammal is a human.

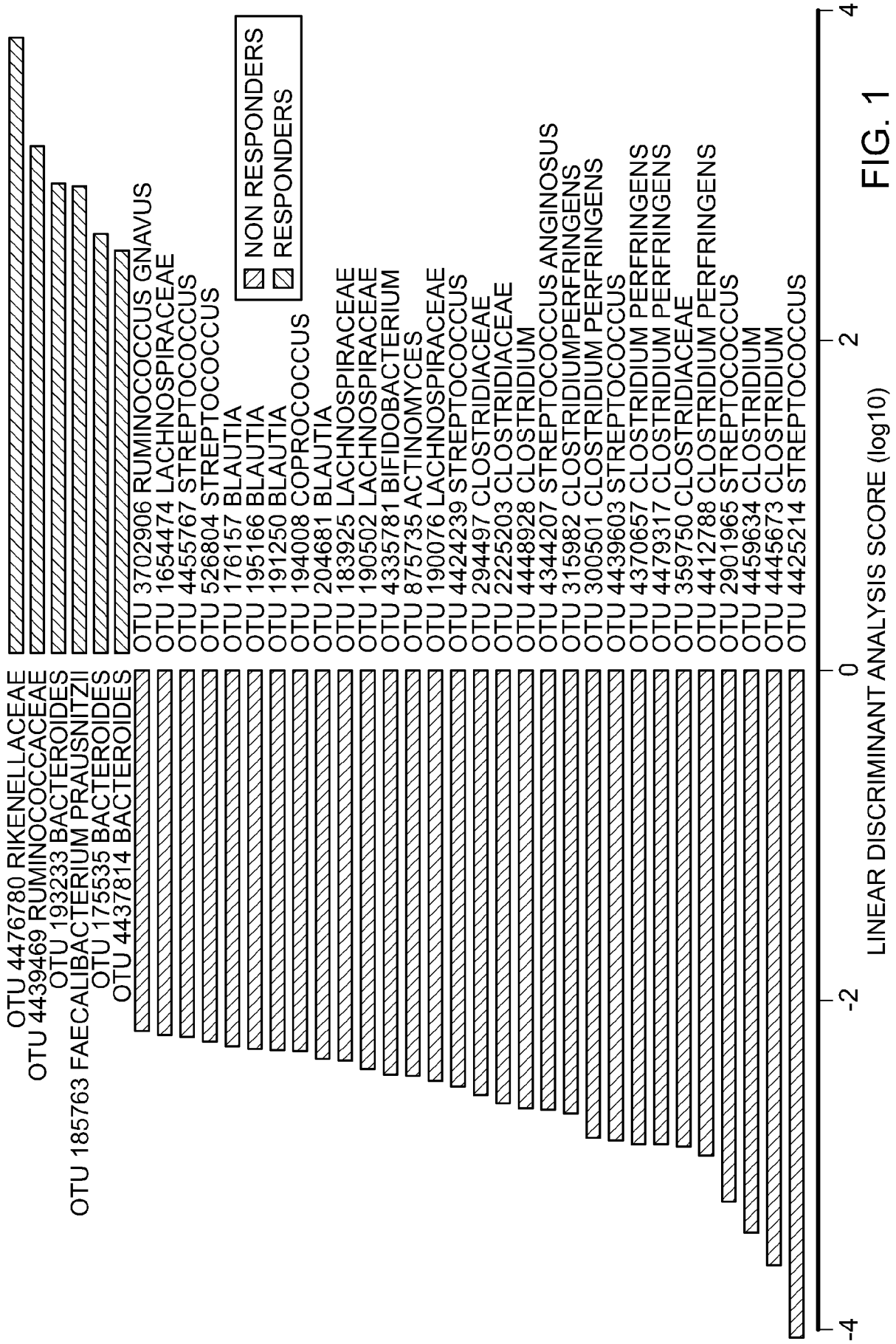

OTU 4476780 RIKENELLACEAE
OTU 4439469 RUMINOCOCCACEAE
OTU 193233 BACTEROIDES
OTU 185763 FAECALIBACTERIUM PRAUSNITZII
OTU 175535 BACTEROIDES
OTU 4437814 BACTEROIDES
OTU 3702906 RUMINOCOCCUS GNAVUS
OTU 1654474 LACHNOSPIRACEAE
OTU 4455767 STREPTOCOCCUS
OTU 526804 STREPTOCOCCUS
OTU 176157 BLAUTIA
OTU 195166 BLAUTIA
OTU 191250 BLAUTIA
OTU 194008 COPROCOCCUS
OTU 204681 BLAUTIA
OTU 183925 LACHNOSPIRACEAE
OTU 190502 LACHNOSPIRACEAE
OTU 4335781 BIFIDOBACTERIUM
OTU 875735 ACTINOMYCES
OTU 190076 LACHNOSPIRACEAE
OTU 4424239 STREPTOCOCCUS
OTU 294497 CLOSTRIDIACEAE
OTU 2225203 CLOSTRIDIACEAE
OTU 4448928 CLOSTRIDIUM
OTU 4344207 STREPTOCOCCUS ANGINOSUS
OTU 315982 CLOSTRIDIUMPERFRINGENS
OTU 300501 CLOSTRIDIUM PERFRINGENS
OTU 4439603 STREPTOCOCCUS
OTU 4370657 CLOSTRIDIUM PERFRINGENS
OTU 4479317 CLOSTRIDIUM PERFRINGENS
OTU 359750 CLOSTRIDIACEAE
OTU 4412788 CLOSTRIDIUM PERFRINGENS
OTU 2901965 STREPTOCOCCUS
OTU 4459634 CLOSTRIDIUM
OTU 4445673 CLOSTRIDIUM
OTU 4425214 STREPTOCOCCUS
NON RESPONDERS
RESPONDERS
-4
-2
0
2
4
LINEAR DISCRIMINANT ANALYSIS SCORE (log10)

FIG. 1

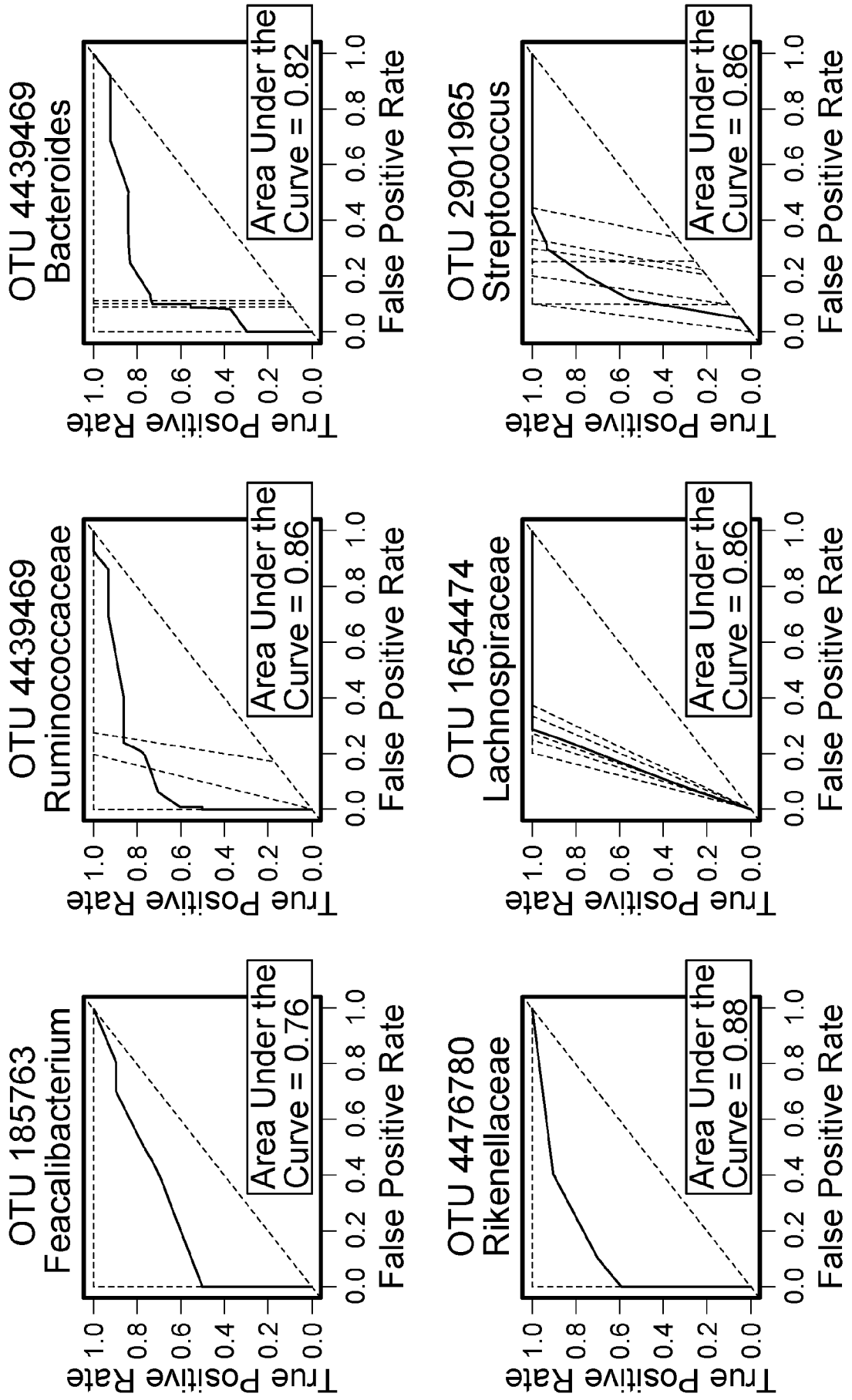
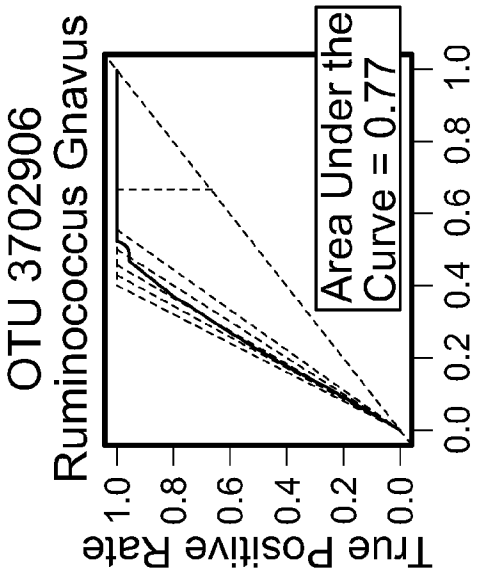
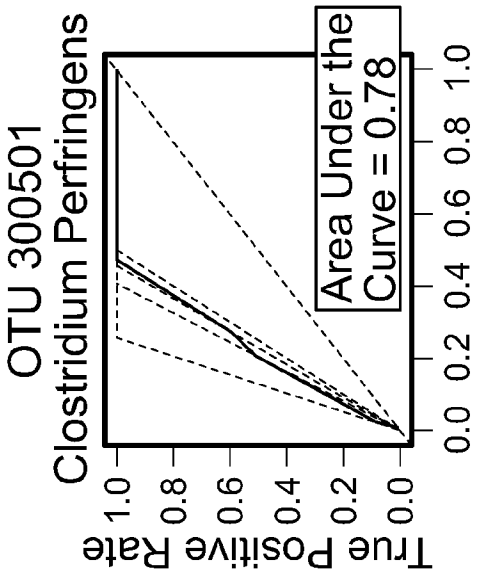

OTU 185763
Feacalibacterium
Area Under the Curve = 0.76
OTU 4439469
Ruminococcaceae
Area Under the Curve = 0.86
OTU 4439469
Bacteroides
Area Under the Curve = 0.82
OTU 4476780
Rikenellaceae
Area Under the Curve = 0.88
OTU 1654474
Lachnospiraceae
Area Under the Curve = 0.86
OTU 2901965
Streptococcus
Area Under the Curve = 0.86
True Positive Rate
False Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0

FIG. 2


OTU 3702906
Ruminococcus Gnavus
Area Under the Curve = 0.77
True Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0


OTU 300501
Clostridium Perfringens
Area Under the Curve = 0.78
True Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0

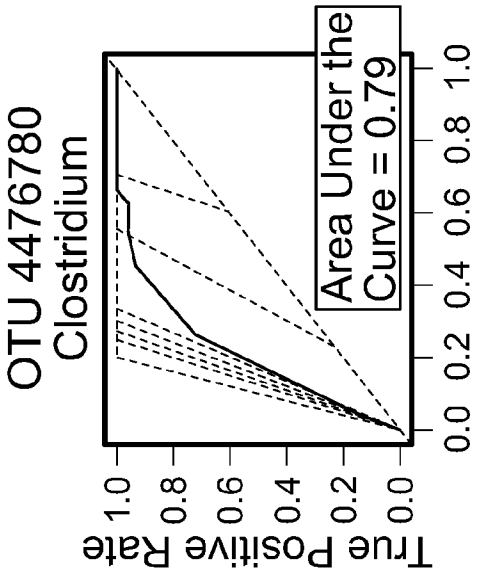

OTU 4476780
Clostridium
Area Under the Curve = 0.79
True Positive Rate
0.0 0.2 0.4 0.6 0.8 1.0

FIG. 2(Cont.)

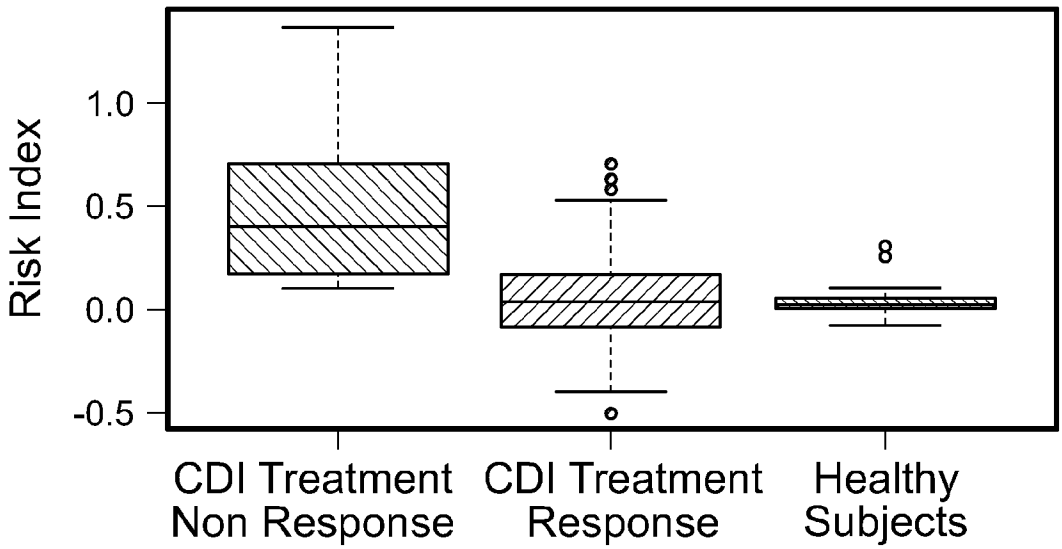

Risk Index
1.0
0.5
0.0
-0.5
CDI Treatment Non Response
CDI Treatment Response
Healthy Subjects

FIG. 3A

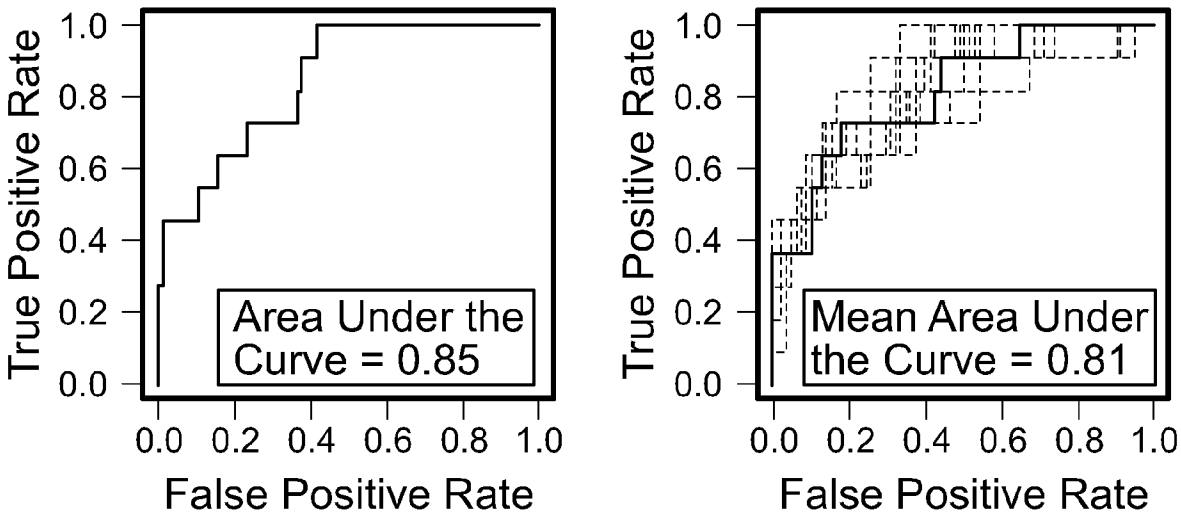

True Positive Rate
False Positive Rate
Area Under the Curve = 0.85
Mean Area Under the Curve = 0.81
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 3B

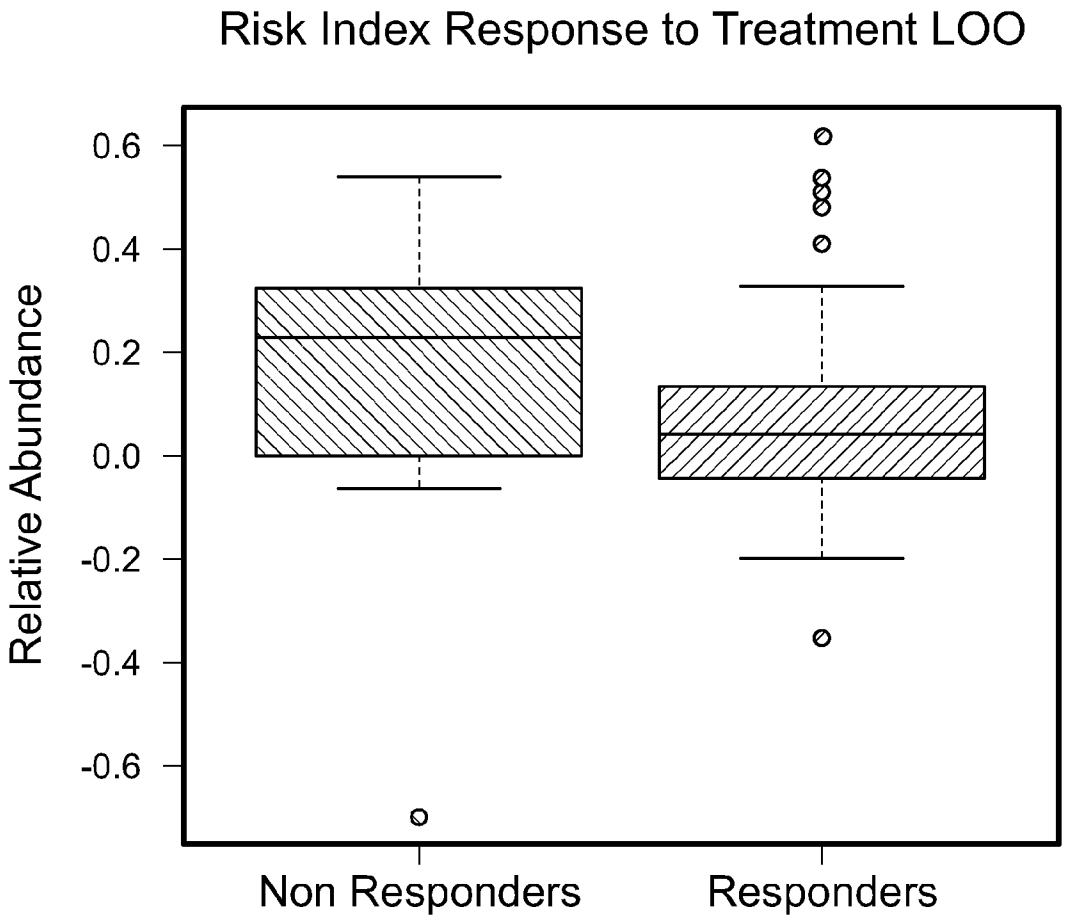

Risk Index Response to Treatment LOO
Relative Abundance
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
Non Responders
Responders

FIG. 3C

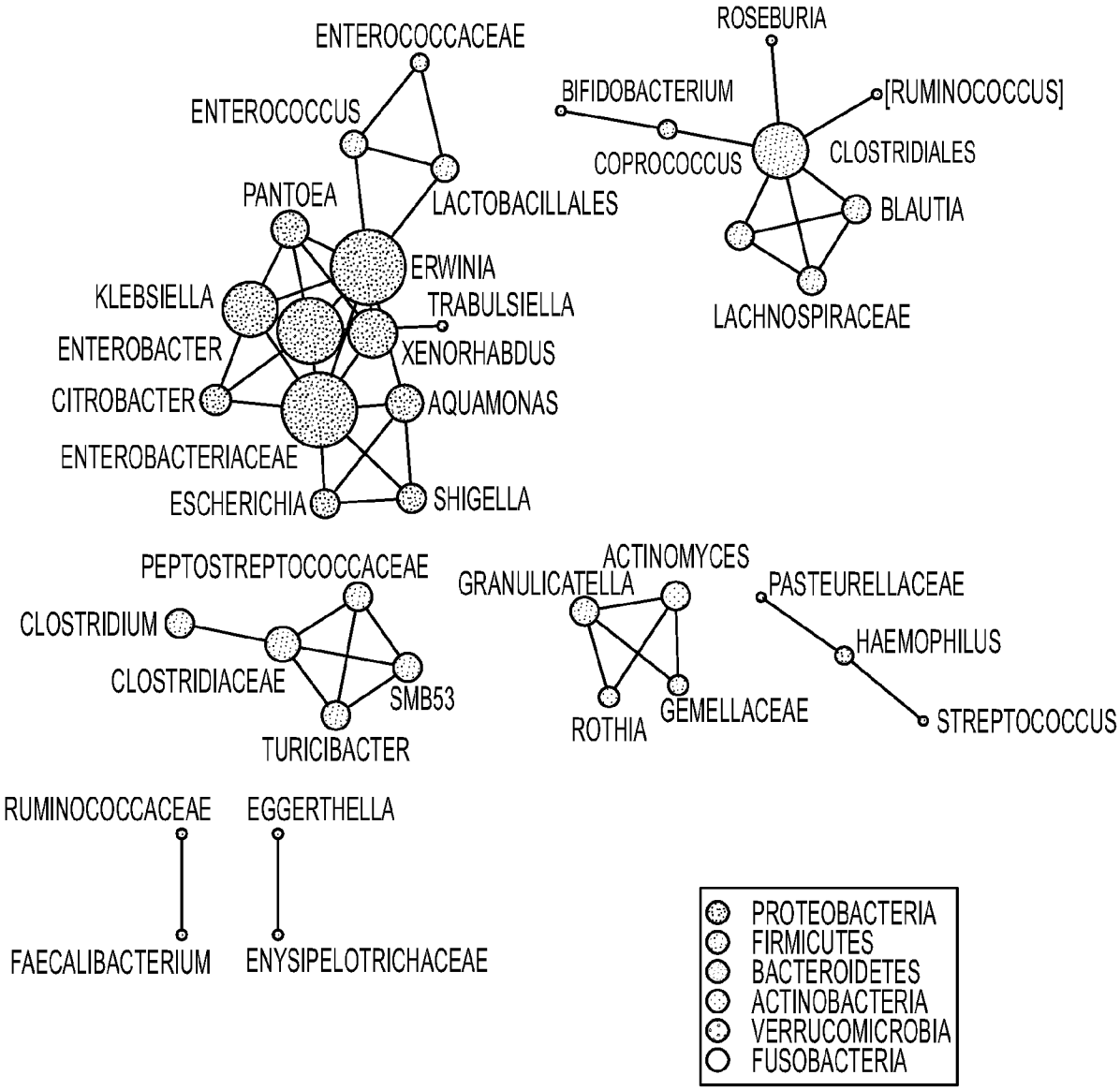

ENTEROCOCCACEAE
ENTEROCOCCUS
PANTOEA
LACTOBACILLALES
ERWINIA
KLEBSIELLA
TRABULSIELLA
ENTEROBACTER
XENORHABDUS
CITROBACTER
AQUAMONAS
ENTEROBACTERIACEAE
ESCHERICHIA
SHIGELLA
ROSEBURIA
BIFIDOBACTERIUM
[RUMINOCOCCUS]
COPROCOCCUS
CLOSTRIDIALES
BLAUTIA
LACHNOSPIRACEAE
PEPTOSTREPTOCOCCACEAE
CLOSTRIDIUM
CLOSTRIDIACEAE
SMB53
TURICIBACTER
ACTINOMYCES
GRANULICATELLA
ROTHIA
GEMELLACEAE
PASTEURELLACEAE
HAEMOPHILUS
STREPTOCOCCUS
RUMINOCOCCACEAE
EGGERTHELLA
FAECALIBACTERIUM
ENYSIPELOTRICHACEAE
PROTEOBACTERIA
FIRMICUTES
BACTEROIDETES
ACTINOBACTERIA
VERRUCOMICROBIA
FUSOBACTERIA

FIG. 4A

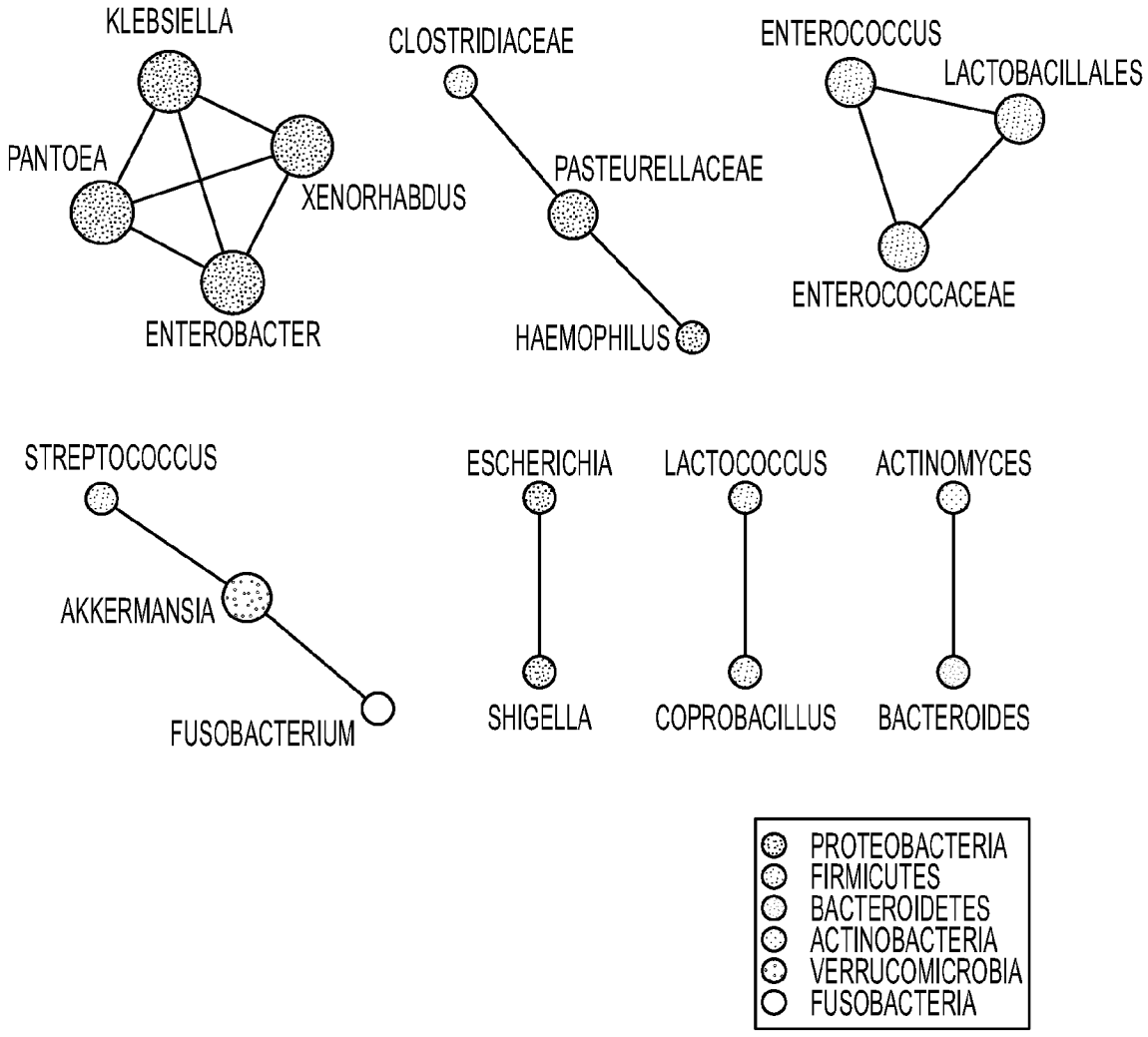

KLEBSIELLA
PANTOEA
XENORHABDUS
ENTEROBACTER
CLOSTRIDIACEAE
PASTEURELLACEAE
HAEMOPHILUS
ENTEROCOCCUS
LACTOBACILLALES
ENTEROCOCCACEAE
STREPTOCOCCUS
AKKERMANSIA
FUSOBACTERIUM
ESCHERICHIA
SHIGELLA
LACTOCOCCUS
COPROBACILLUS
ACTINOMYCES
BACTEROIDES
PROTEOBACTERIA
FIRMICUTES
BACTEROIDETES
ACTINOBACTERIA
VERRUCOMICROBIA
FUSOBACTERIA

FIG. 4B

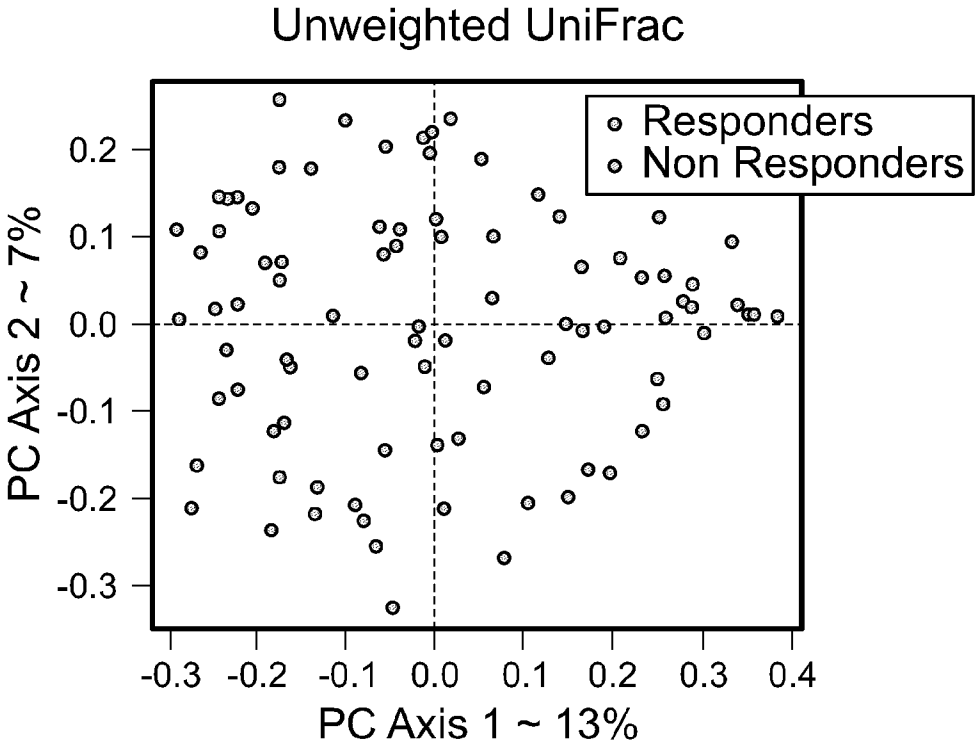

Unweighted UniFrac
Responders
Non Responders
PC Axis 2 ~ 7%
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.3 -0.2 -0.1 0.0 0.1 0.2 0.3 0.4
PC Axis 1 ~ 13%

FIG. 5A

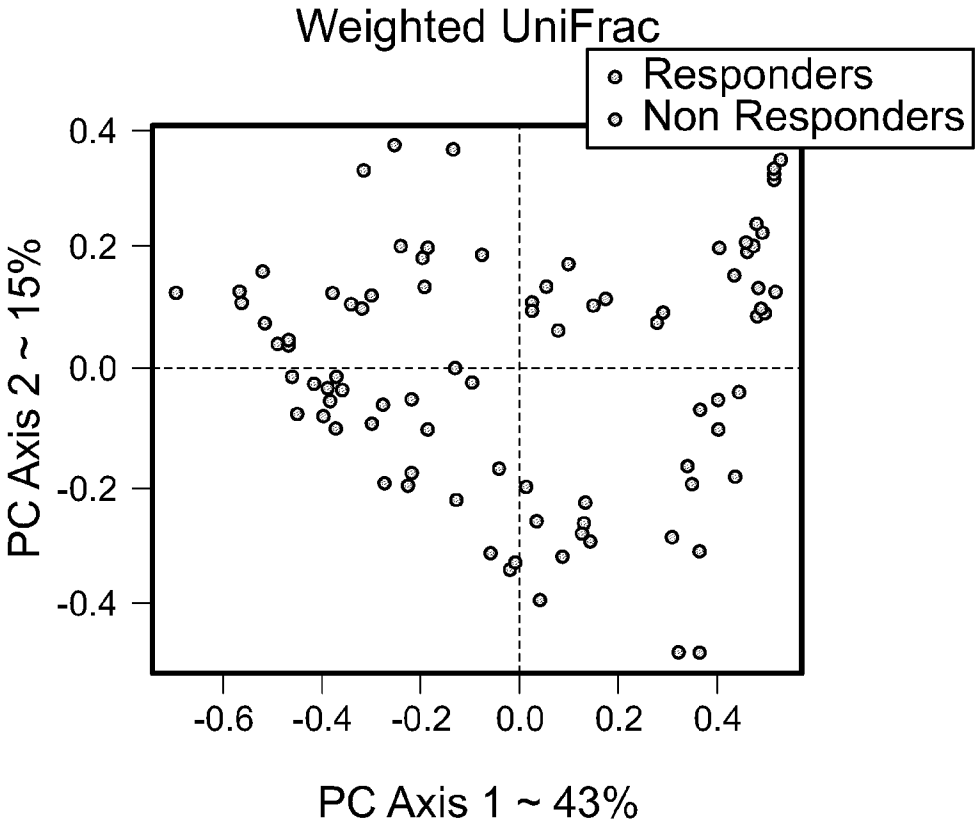

Weighted UniFrac
Responders
Non Responders
PC Axis 2 ~ 15%
0.4
0.2
0.0
-0.2
-0.4
-0.6 -0.4 -0.2 0.0 0.2 0.4
PC Axis 1 ~ 43%

FIG. 5B

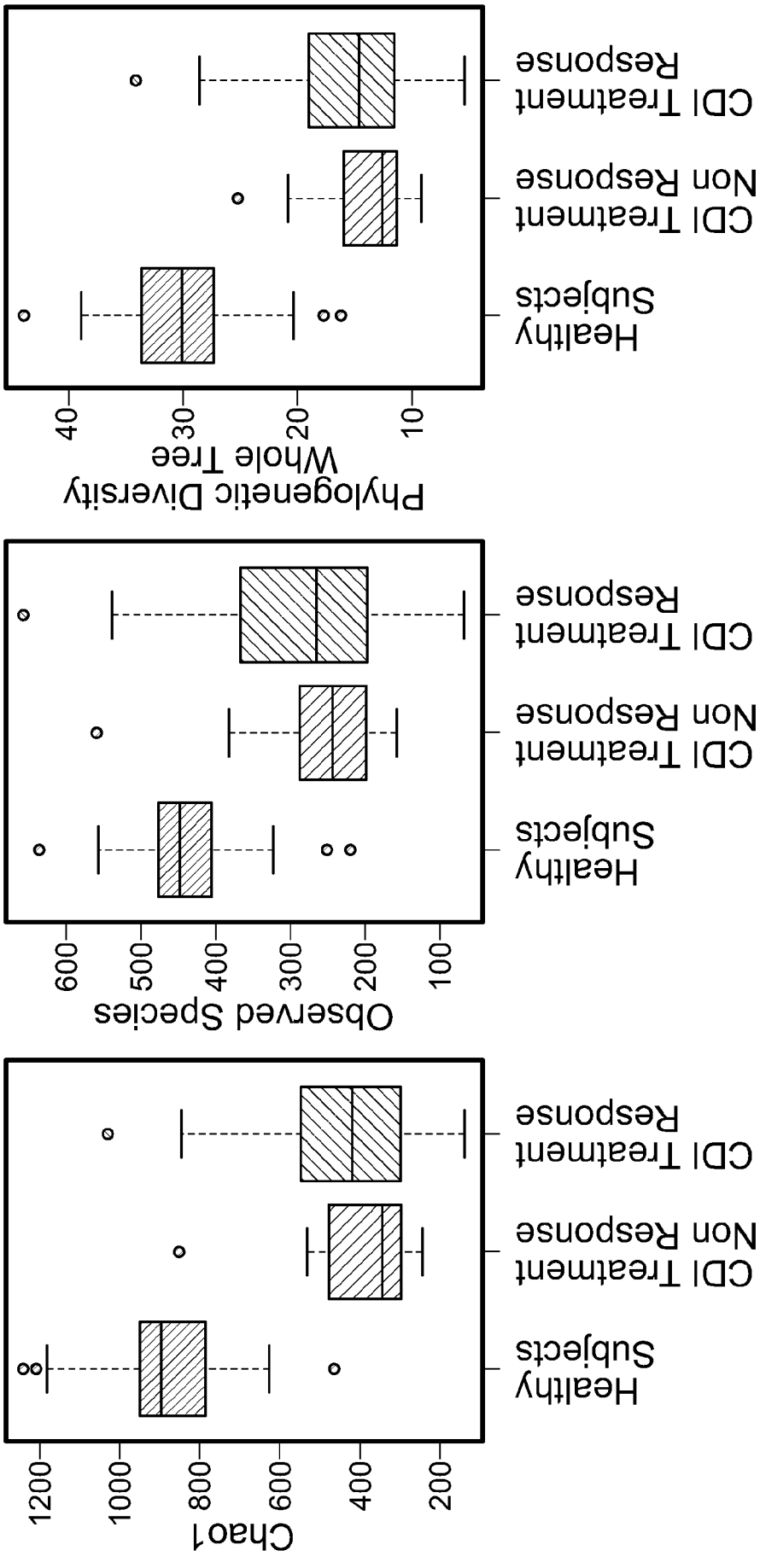

Chao1
1200
1000
800
600
400
200
Observed Species
600
500
400
300
200
100
Phylogenetic Diversity
Whole Tree
40
30
20
10
Healthy Subjects
CDI Treatment Non Response
CDI Treatment Response

FIG. 6

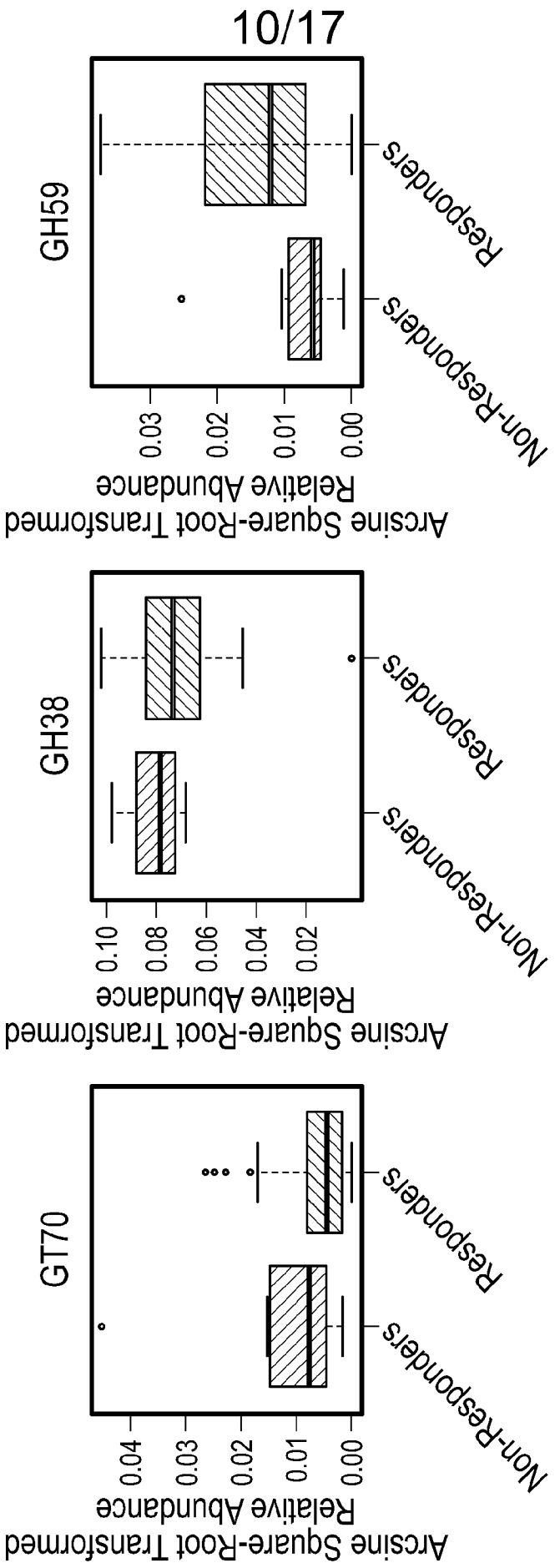

GH59
Responders
Non-Responders
0.03
0.02
0.01
0.00
Arcsine Square-Root Transformed
Relative Abundance
GH38
Responders
Non-Responders
0.10
0.08
0.06
0.04
0.02
Arcsine Square-Root Transformed
Relative Abundance
GT70
Responders
Non-Responders
0.04
0.03
0.02
0.01
0.00
Arcsine Square-Root Transformed
Relative Abundance

FIG. 7

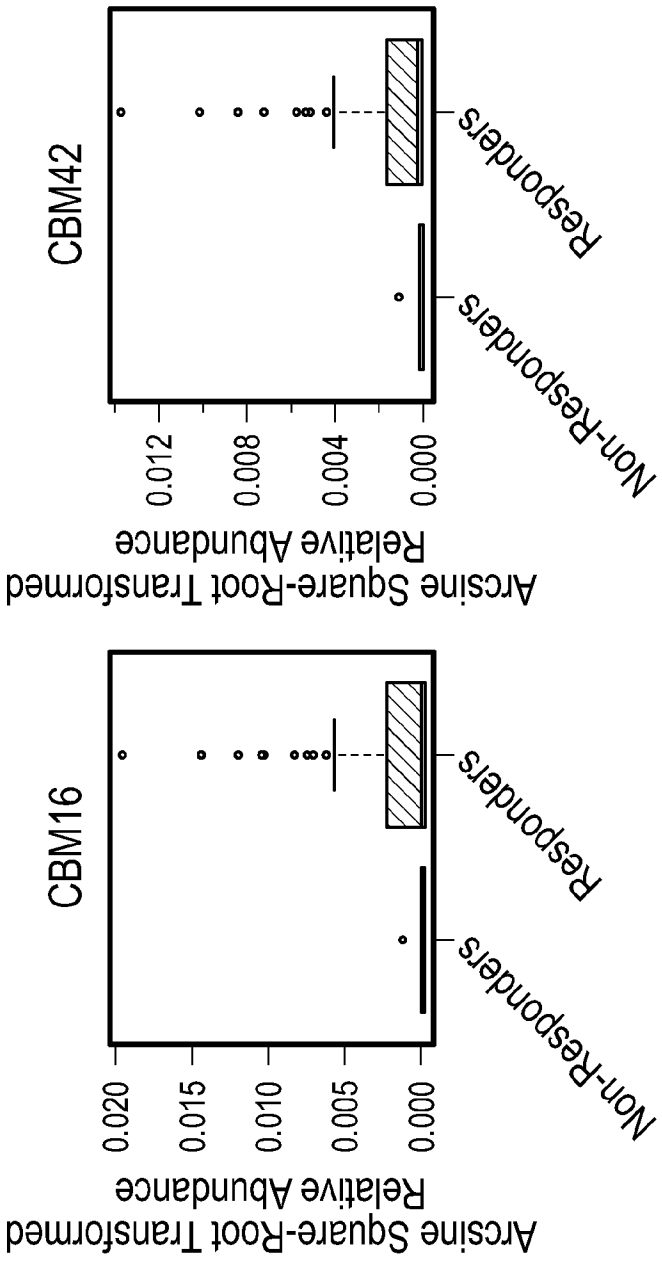

CBM42
Arcsine Square-Root Transformed Relative Abundance
0.012
0.008
0.004
0.000
Responders
Non-Responders
CBM16
Arcsine Square-Root Transformed Relative Abundance
0.020
0.015
0.010
0.005
0.000
Responders
Non-Responders

FIG. 7(Cont.)

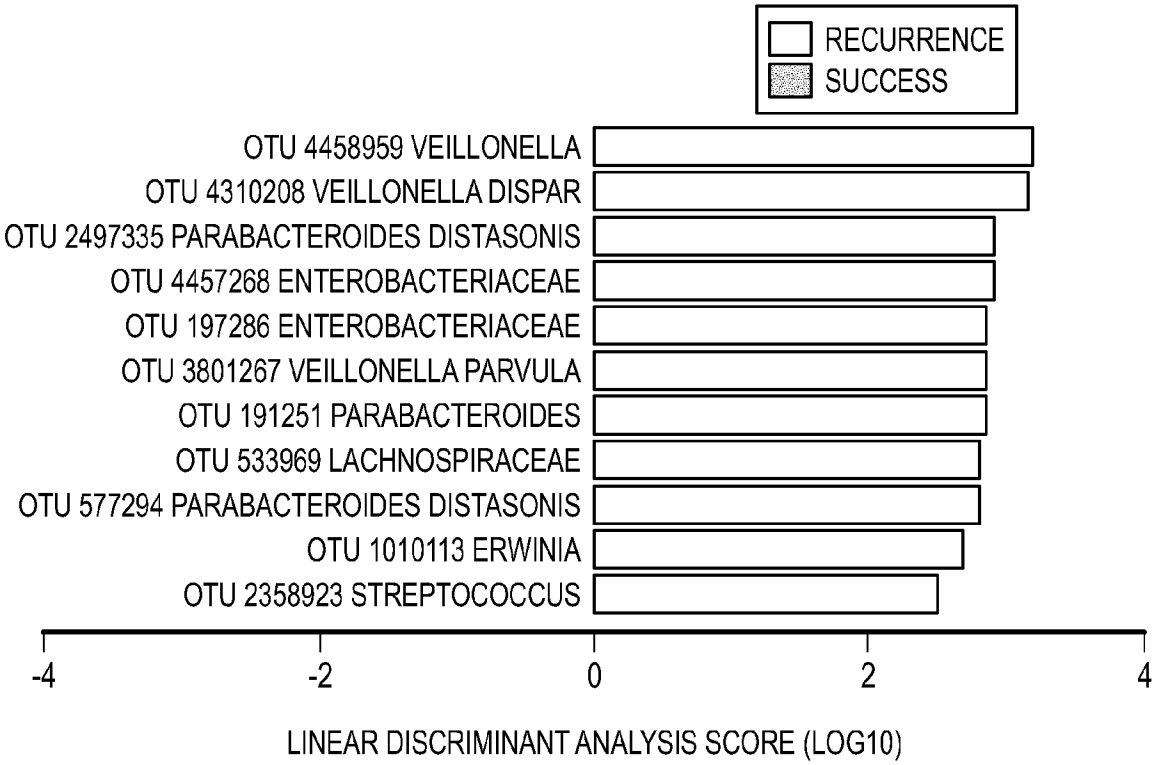

RECURRENCE
SUCCESS
OTU 4458959 VEILLONELLA
OTU 4310208 VEILLONELLA DISPAR
OTU 2497335 PARABACTEROIDES DISTASONIS
OTU 4457268 ENTEROBACTERIACEAE
OTU 197286 ENTEROBACTERIACEAE
OTU 3801267 VEILLONELLA PARVULA
OTU 191251 PARABACTEROIDES
OTU 533969 LACHNOSPIRACEAE
OTU 577294 PARABACTEROIDES DISTASONIS
OTU 1010113 ERWINIA
OTU 2358923 STREPTOCOCCUS
-4
-2
0
2
4
LINEAR DISCRIMINANT ANALYSIS SCORE (LOG10)

FIG. 8

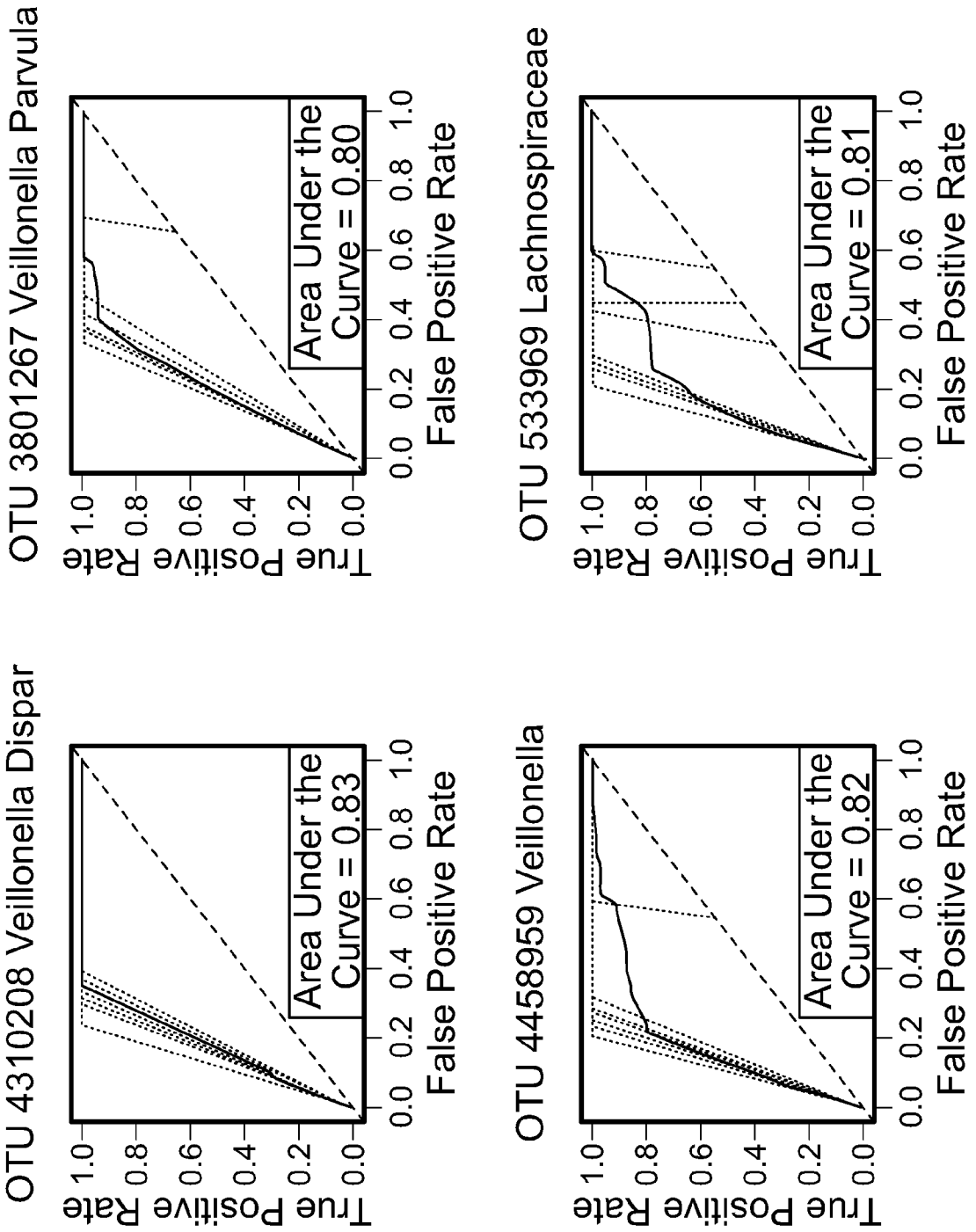

OTU 3801267 Veillonella Parvula
Area Under the Curve = 0.80
OTU 533969 Lachnospiraceae
Area Under the Curve = 0.81
OTU 4310208 Veillonella Dispar
Area Under the Curve = 0.83
OTU 4458959 Veillonella
Area Under the Curve = 0.82
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0

FIG. 9

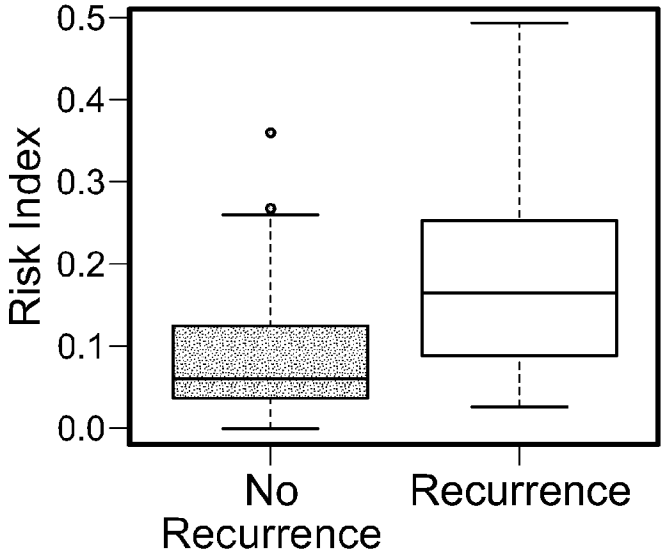


FIG. 10A

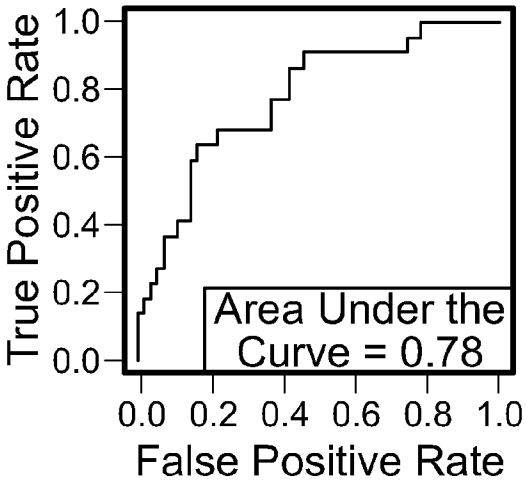


FIG. 10B

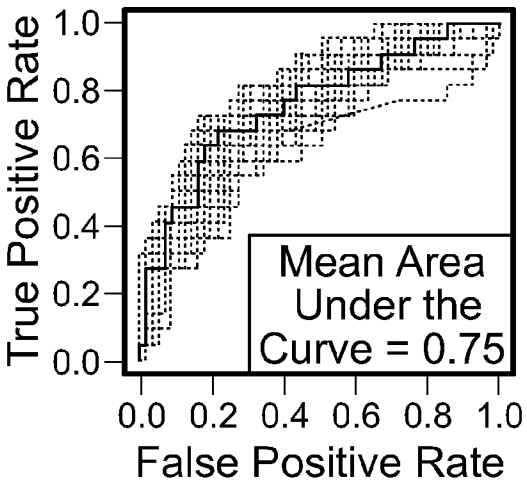


FIG. 10C

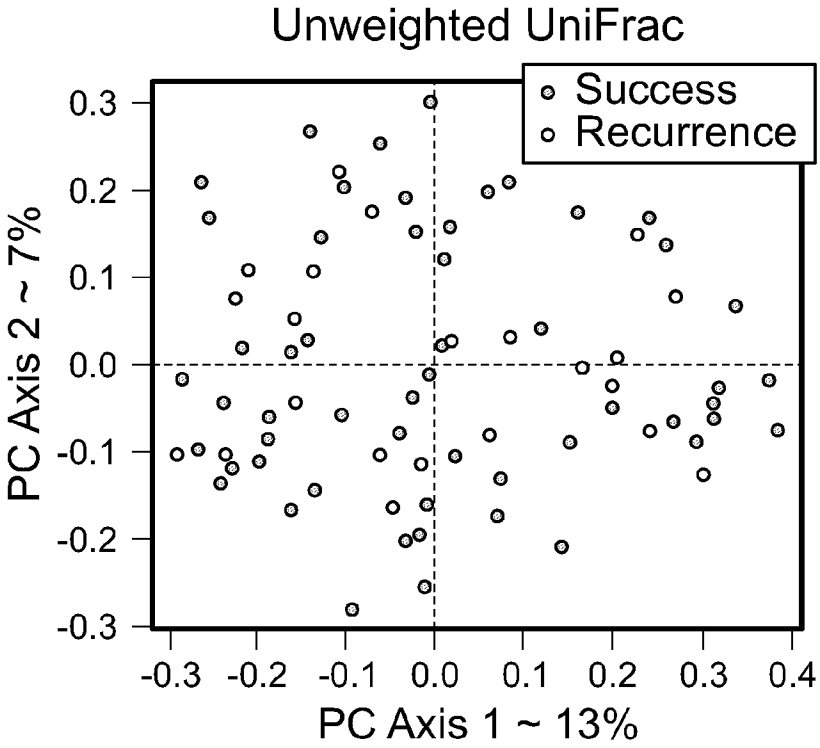


FIG. 11A

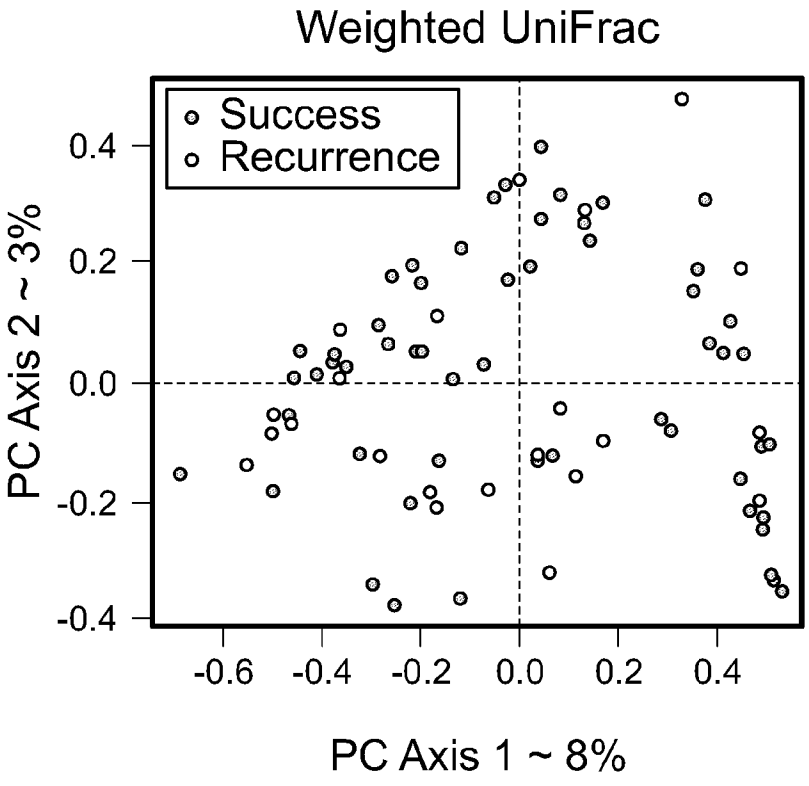


FIG. 11B

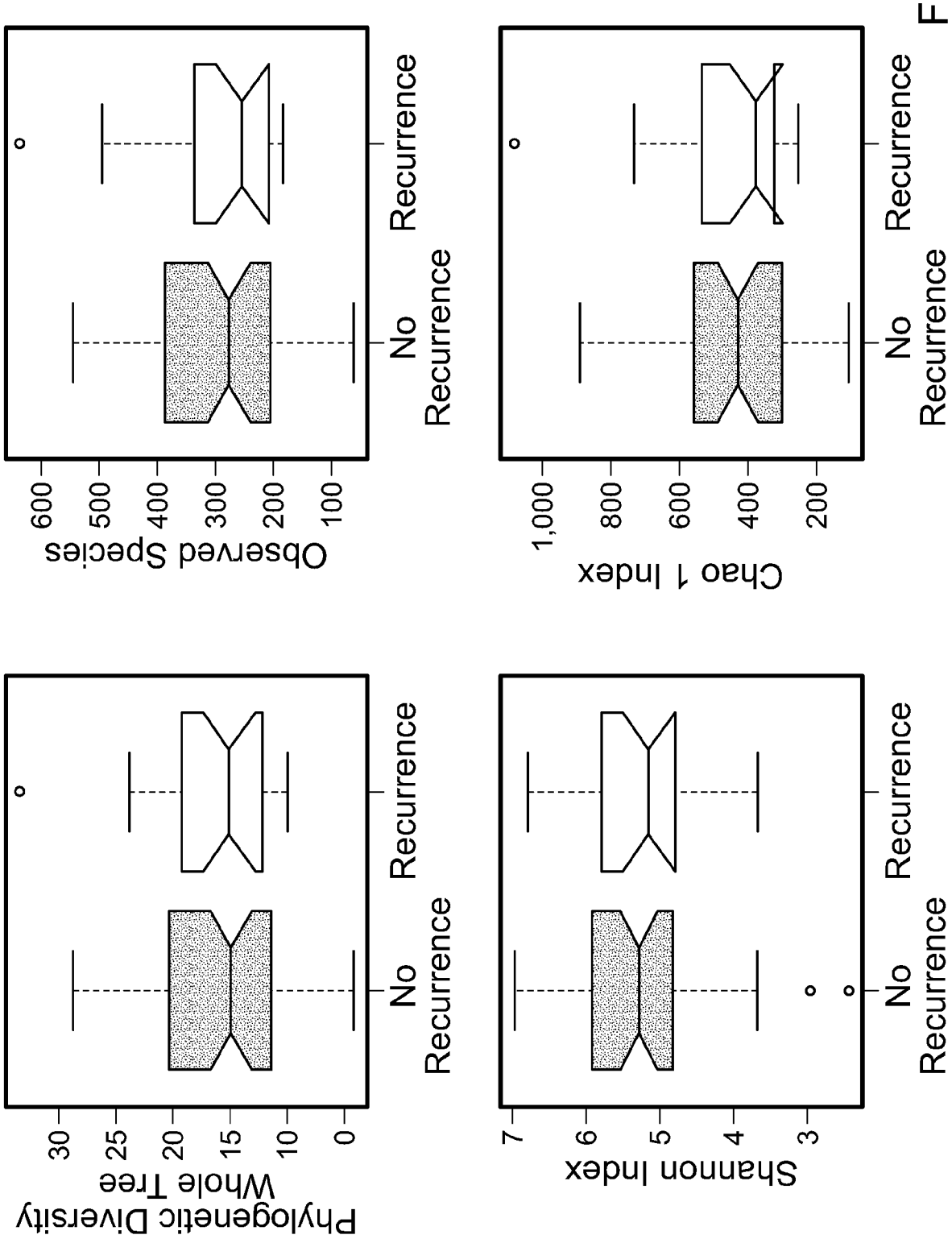

Observed Species
600
500
400
300
200
100
No Recurrence
Recurrence
Chao 1 Index
1,000
800
600
400
200
No Recurrence
Recurrence
Phylogenetic Diversity Whole Tree
30
25
20
15
10
0
No Recurrence
Recurrence
Shannon Index
7
6
5
4
3
No Recurrence
Recurrence

FIG. 12

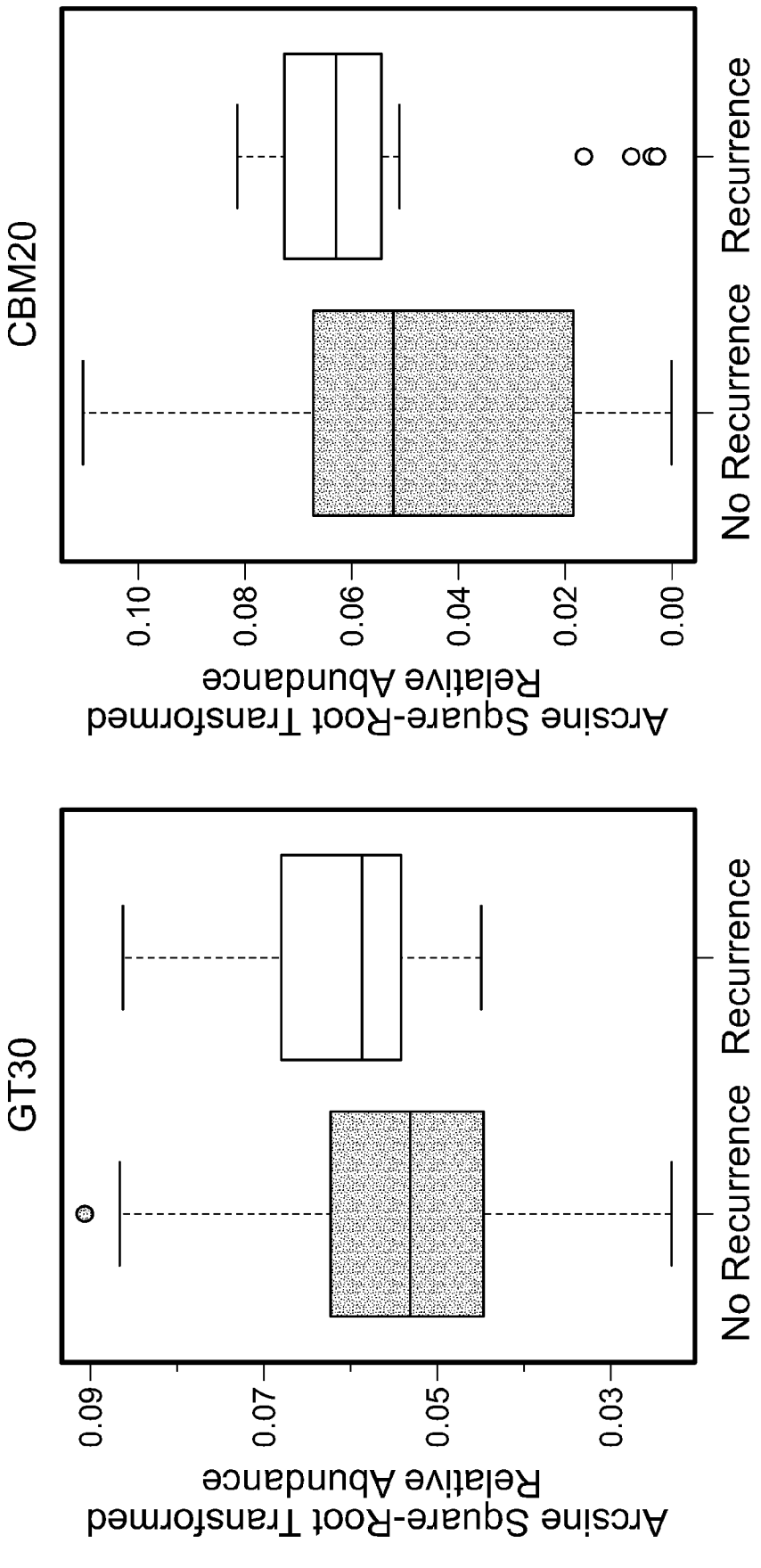

CBM20
Arcsine Square-Root Transformed Relative Abundance
0.10
0.08
0.06
0.04
0.02
0.00
No Recurrence
Recurrence
GT30
Arcsine Square-Root Transformed Relative Abundance
0.09
0.07
0.05
0.03
No Recurrence
Recurrence

FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/65566

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G01N 33/569, A61K 31/4164 (2017.01)
CPC - G01N 33/56911, A61K 38/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2015/0344940 A1 (BAYLOR COLLEGE OF MEDICINE) 03 December 2015 (03.12.2015) para [0008]; [0014]; [0026]; [0028]; [0039]; [0040]; [0075].	1, 2, 8, 9, 11, 12, 16-18, 19/12, 20-23
Y	US 2014/0199281 A1 (SERES HEALTH INC) 17 July 2014 (17.07.2014) para [0025]; [0029]; [0220]; Tables 1, 8, 9.	1, 2, 8, 9, (20-23)/(1,2)
Y	WO 2014/153194 A2 (SERES HEALTH INC) 25 September 2014 (25.09.2014) para [020]; [049] -[051]; [071]-[077]; [089]; [235].	11, 12, 16-18, 19/12, (20-23)/(11,12)
Y	US 2013/0224155 A1 (KAPLAN et al.) 29 August 2013 (29.08.2013) para [0187]; [0188].	16
A	US 2014/0219966 A1 (TECHLAB, INC.) 07 August 2014 (07.08.2014) para [0011]-[0012]; [0019]-[0020]; claim 1.	1

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 March 2017	17 APR 2017
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/65566

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 10, 24, 25
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1, 2, 8, 9, 11, 12, 16-18, 19(in part), and 20-23, directed to a method for detecting a gut microbiota panel in a fecal sample obtained from mammals infected with Clostridium difficile.

Group II, claims 3-7, 13-15, and 19(in part), directed to a method for detecting a mode of C. difficile acquisition in mammals infected with Clostridium difficile.

The inventions listed as Groups I and II do not relate to a single special technical feature under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

--continued on first extra sheet attached hereto--

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 2, 8, 9, 11, 12, 16-18, 19(in part), and 20-23

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/65566

--continuation of Box No III: Observations where unity of invention is lacking--

Special technical features
Group I has the special technical feature of detecting a gut microbiota panel in a fecal sample, wherein the gut microbiota panel comprises Veillonella, Enterobacteriaceae, Streptococcus, Parabacteroides, and Lachnospiraceae, or Ruminococcaceae, Rikenellaceae, Clostridiaceae, Bacteroides, Faecalibacterium, and/or Rothia, that is not required by Group II.

Group II has the special technical feature of detecting a mode of C. difficile acquisition comprising detecting a clinical parameter panel comprising severe CDI, prior antibiotic use, and hospital-acquired C. difficile, that is not required by Group I.

Common technical features:
Groups I-II share the common technical feature of a method comprising detecting a CDI marker panel in a mammal subject (sample), and using the results (1) for predicting response to primary CDI treatment in a mammal having CDI; (2) for treating CDI by administering a primary CDI treatment to the mammal, or (3) for predicting risk of recurrence of CDI. However, this shared technical feature does not represent a contribution over prior art, because this shared technical feature is anticipated by US 2014/0219966 A1 to Techlab, Inc. (hereinafter Techlab).

Techlab teaches a method comprising detecting a CDI marker panel in a mammal subject sample (para [0012] "test methods for aiding in stratifying patients based on severity of C. difficile disease. Stratifying patients with disease based on severity using a panel of biomarkers"; claim 11 "measuring the fecal sample for a level of one or more of lactoferrin or calprotectin"), and using the results (1) for predicting response to primary CDI treatment in a mammal having CDI (para [0019] "Knowing whether toxigenic C. difficile is present in combination with a lactoferrin concentration will help to determine disease severity to optimize treatment"; [0020] "the presence of one or more biomarkers (e.g. GDH, toxins A and/or B) provides an indicator of the bacteria, which may be used to determine a patient's response to treatment"); (2) for treating CDI by administering a primary CDI treatment to the mammal (claim 1 "A method of treating a patient with C. Difficile disease, the method comprising: administering a therapeutically effective amount of treatment shown to be effective...to a patient who has tested positive for an elevated level"; abstract "Mild and moderate cases may be treated with metronidazole while moderate-to-severe and relapsing cases are often treated with vancomycin or fidaxomicin"), or (3) for predicting risk of recurrence of CDI (para [0011] "[0011] FIG. 4C depicts a summary of biomarker results for patients with clinical recurrence or no cure (return of symptoms and C. difficile during and/or after initial treatment) according to embodiments of the invention").

As the technical features were known in the art at the time of the invention, they cannot be considered special technical features that would otherwise unify the groups.

Therefore, Group I-II inventions lack unity under PCT Rule 13 because they do not share the same or corresponding special technical feature.