



Meta-analyses of human gut microbes associated with obesity and IBD



William A. Walters^a, Zech Xu^b, Rob Knight^{b,c,d,*}

^a Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

^b BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA

^c Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309, USA

^d Howard Hughes Medical Institute, Boulder, CO 80309, USA

ARTICLE INFO

Article history:

Received 11 August 2014

Revised 25 September 2014

Accepted 29 September 2014

Available online 13 October 2014

Edited by Lloyd H. Kasper and Wilhelm Just

Keywords:

Gut microbiota

Metaanalysis

ABSTRACT

Recent studies have linked human gut microbes to obesity and inflammatory bowel disease, but consistent signals have been difficult to identify. Here we test for indicator taxa and general features of the microbiota that are generally consistent across studies of obesity and of IBD, focusing on studies involving high-throughput sequencing of the 16S rRNA gene (which we could process using a common computational pipeline). We find that IBD has a consistent signature across studies and allows high classification accuracy of IBD from non-IBD subjects, but that although subjects can be classified as lean or obese within each individual study with statistically significant accuracy, consistent with the ability of the microbiota to experimentally transfer this phenotype, signatures of obesity are not consistent between studies even when the data are analyzed with consistent methods. The results suggest that correlations between microbes and clinical conditions with different effect sizes (e.g. the large effect size of IBD versus the small effect size of obesity) may require different cohort selection and analysis strategies.

© 2014 The Authors. Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Recent advances in our ability to characterize the gut microbiota have led to tremendous interest in identifying organisms associated with different human diseases. Two general categories of disease that have attracted widespread interest are obesity, diabetes and metabolic syndrome, and inflammatory bowel disease including Crohn's disease and ulcerative colitis. In both cases, significant associations with disease have been reported by different studies. However, conflicting signals have been seen in the taxa associated with each disorder (as described in more detail below).

A major issue with comparing the results of different studies of the gut microbiota is that technical differences, including differences in DNA extraction protocols, PCR primers, etc. can often outweigh biological differences in the samples, especially when biological effect sizes are small [1]. Even computational differences, including which taxonomy database is used and which taxonomy assignment algorithm is used, can have surprisingly large effects [2]. These issues can be partially overcome for 16S rRNA

amplicon studies by downloading all the relevant data and analyzing them with consistent methods [1].

Accordingly, for studies of obesity and IBD, we utilized recent studies identifying microbial association with disease using amplicon sequencing, and analyzed the raw sequence data using a consistent set of methods to test whether these diseases, widely reported as having associations with the human gut microbiome, harbored consistent signatures associated with disease.

2. Methods

To test which trends at the broad phylogenetic level or at finer taxonomic resolution were consistent across 16S amplicon survey studies, we performed a meta-analysis of high-throughput amplicon sequencing studies. We processed demultiplexed data (i.e. sequences already assigned to their samples by barcode, downloaded from each study from the QIIME database, <http://www.microbio.me/qiime/>) using QIIME 1.8.0. Mapping files containing clinical information were also downloaded from the QIIME database. Samples lacking BMI or IBD data were filtered out, and BMI data were binned into categories of normal or obese based upon CDC BMI criteria (http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html). The data were clustered at 97% iden-

* Corresponding author at: BioFrontiers Institute, UCB 596, University of Colorado, Boulder, CO 80309, USA. Fax: +1 303 492 7744.

E-mail address: rob@colorado.edu (R. Knight).

tity against the Greengenes database (August 2013 release) [3] using UCLUST [4] v1.2.22q, discarding reads that failed to match the reference sequences, which is referred to as a “closed reference” approach to clustering. Taxonomies came directly from the reference database based upon the identity of the reference sequence clustered against. The 97% OTUs phylogenetic tree supplied with Greengenes was used for UniFrac [5] metrics. OTU tables were rarefied at 1000 sequences per sample for beta diversity (UniFrac), Kruskal–Wallis tests, and supervised learning. Unrarefied tables were repeatedly ($\times 10$) subsampled at 1000 sequences per sample for alpha diversity metrics (Observed Species, Shannon, and Phylogenetic Diversity). Data were not rarefied for taxonomic abundance comparisons (ratios of Firmicutes:Bacteroidetes), but low-abundance (<1000 sequences per sample) samples were filtered out to remove samples with questionable PCR or sequencing yields. The reason for taking the closed-reference approach is due to the various PCR primers used in different studies, which covered different regions of the 16S amplicon, so sequences from the same 16S gene would fall into different clusters because of the lack of overlap in the sequences if data were clustered de novo.

3. Results

3.1. Microbes associated with obesity

Obesity, and its related comorbidities, is a globally prevalent issue, and is expected to affect over half a billion people by 2030 [6]. Although microbes are by no means the sole factor in the obesity epidemic, alterations in the gut microbiota have been observed in obese humans [7,8], microbial differences have been reported to classify people as lean or obese with $\sim 90\%$ accuracy [9,10], and in mouse models the obesity phenotype is partially transmissible via transmitting the microbes [11,12], establishing that microbes can contribute to obesity. However, the nature of the microbial changes associated with obesity is less clear, unlike the more obvious differences that are present in the case of IBD. Conflicting reports exist regarding broad, phyla-level shifts in obese human guts. Many studies show an increased ratio of Firmicutes:Bacteroidetes [8,13–19] although some other studies show no trend, or even the opposite trend [20–24]. However these studies vary in their methodology for sequencing and quantifying taxa (including amplicon sequencing, qPCR, rtPCR, PhyloChip, and HITChip). These

Table 1
Reported taxon enrichment and depletion in the obese gut as reported in prior studies.

Taxa	Abundance shifts in obese guts		Citation
	Increased	Decreased	
Actinobacteria			
<i>Bifidobacterium</i> (genus)		+	[23]
<i>Bifidobacterium animalis</i>		+	[46]
Euryarchaeota			
<i>Methanobrevibacter smithii</i>	+	+	[22,23]
Firmicutes			
<i>Oscillospira</i> (sp)		+	[24]
Clostridium cluster XIVa	+		[18]
<i>Roseburia intestinalis</i>	+		[18,24]
<i>Eubacterium rectale</i>	+		[15,24]
<i>Faecalibacterium prausnitzii</i>		+	[16,18]
<i>Lactobacillus</i> (genus)	+		[13,47]
<i>Lactobacillus casei/paracasei</i>		+	[46]
<i>Lactobacillus reuteri</i>	+		[46]
Bacteroidetes			
<i>Bacteroides</i> (genus)	+		[15,22]
<i>Bacteroides vulgates</i>		+	[13,18]
<i>Bacteroides uniformis</i>		+	[18]
<i>Alistipes</i> (genus)		+	[18]

Table 2
Studies used for BMI component of meta-analysis.

Study	Sequencing platform	Observations
Zupancic [25]	454	Lack of correlations to BMI, however, certain taxa were positively (e.g. <i>B. ovatus</i> , <i>R. torques</i>) or negatively (e.g. <i>F. prausnitzii</i> , <i>Clostridium glycolicum</i>) correlated to metabolic syndrome traits
Turnbaugh [8]	454	Reduced diversity in obese microbiome, shift to higher Firmicutes abundance in obese subjects
Human Microbiome Project [44]	454	Modest association of oral Pseudomonadaceae with BMI, not associations reported with gut microbiome
Wu [28]	454	<i>Oscillibacter</i> genus negatively correlated to BMI, Veillonellaceae positively correlated to BMI. Long-term diet shaped the <i>Prevotella</i> (high fiber) versus <i>Bacteroides</i> (high fat) abundances in the subjects
Yatsunenکو [45]	Illumina HiSeq	Reported results did not discuss BMI, although diet-specific metabolism differences were detected across populations

studies have shown significant (and occasionally contradictory) enrichment for more specific taxa (Table 1). The sequencing data utilized in this study are the subset of studies with high-throughput amplicon sequencing data, and these are summarized in Table 2.

The relative abundance of dominant gut taxa in each study is shown in Fig. 1. Variation on a per-study basis is considerable even at the phylum level. To address the question of whether the Firmicutes:Bacteroidetes ratio increase is significantly associated with obesity, we compared the means of the ratios for subjects with normal versus obese BMI (BMI category is determined using criteria from the Centers for Disease Control). These ratios are shown in Fig. 2.

In all studies except one [25], there is a trend showing an increase in the ratio of Firmicutes:Bacteroidetes in obese over lean subjects. However, no significant differences overall between obese and lean categories were found using Wilcoxon rank-sum tests in R 3.1.0 [26], nor was the difference in Firmicutes:Bacteroidetes ratio statistically significant using Fisher's Method for combining multiple independent tests of a hypothesis [27]. Because several studies [8,25,28] used the same PCR primers yet had opposite trends between lean and obese subjects in their respective

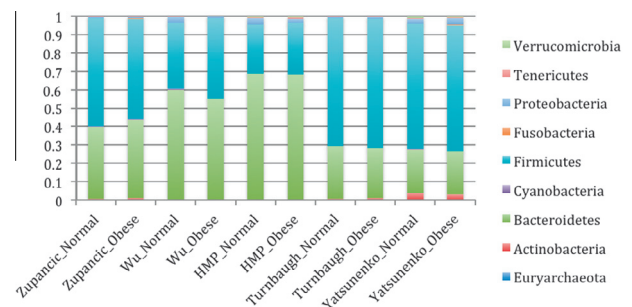


Fig. 1. Relative abundance of phylum-level gut microbial taxa. Studies listed below are Zupancic [25], Wu [28], Human microbiome project [44], Turnbaugh [8], and Yatsunenکو [45].

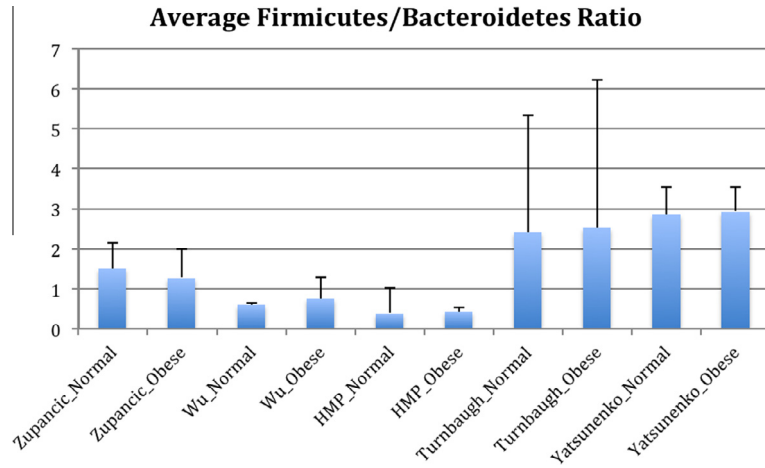


Fig. 2. Ratios of Firmicutes:Bacteroidetes in normal versus obese BMI subjects. Means of ratios for each study/BMI category are shown. Error bars are standard error of the mean. The Turnbaugh study includes a number of samples with extremely low Bacteroidetes, leading to large standard error values.

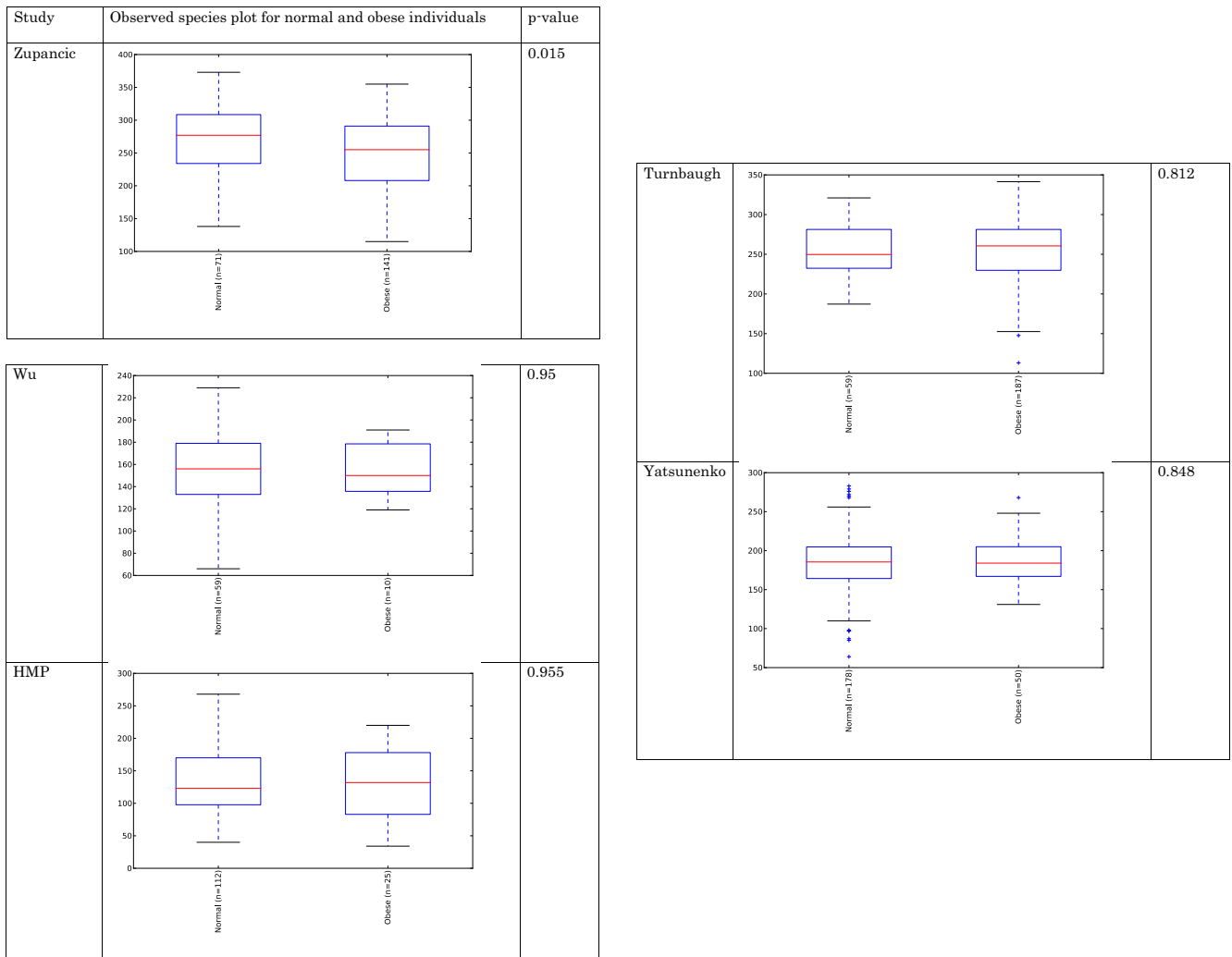


Fig. 3. Alpha diversity (observed species) across studies. Metric is observed species (counts of unique OTUs). Sequence depth is 1000 sequences per sample, and subsampling was performed 10 times. P-values were calculated by using a Monte Carlo simulation with 999 permutations.

populations, primer bias cannot drive the differences in observed Firmicutes:Bacteroidetes ratio or in the overall taxa summaries in Fig. 1. Differences in sample handling and extraction, or differences

among the populations studied, could account for the contradictory observations. In any case, the ratio changes between normal and obese individuals are not statistically significant overall and

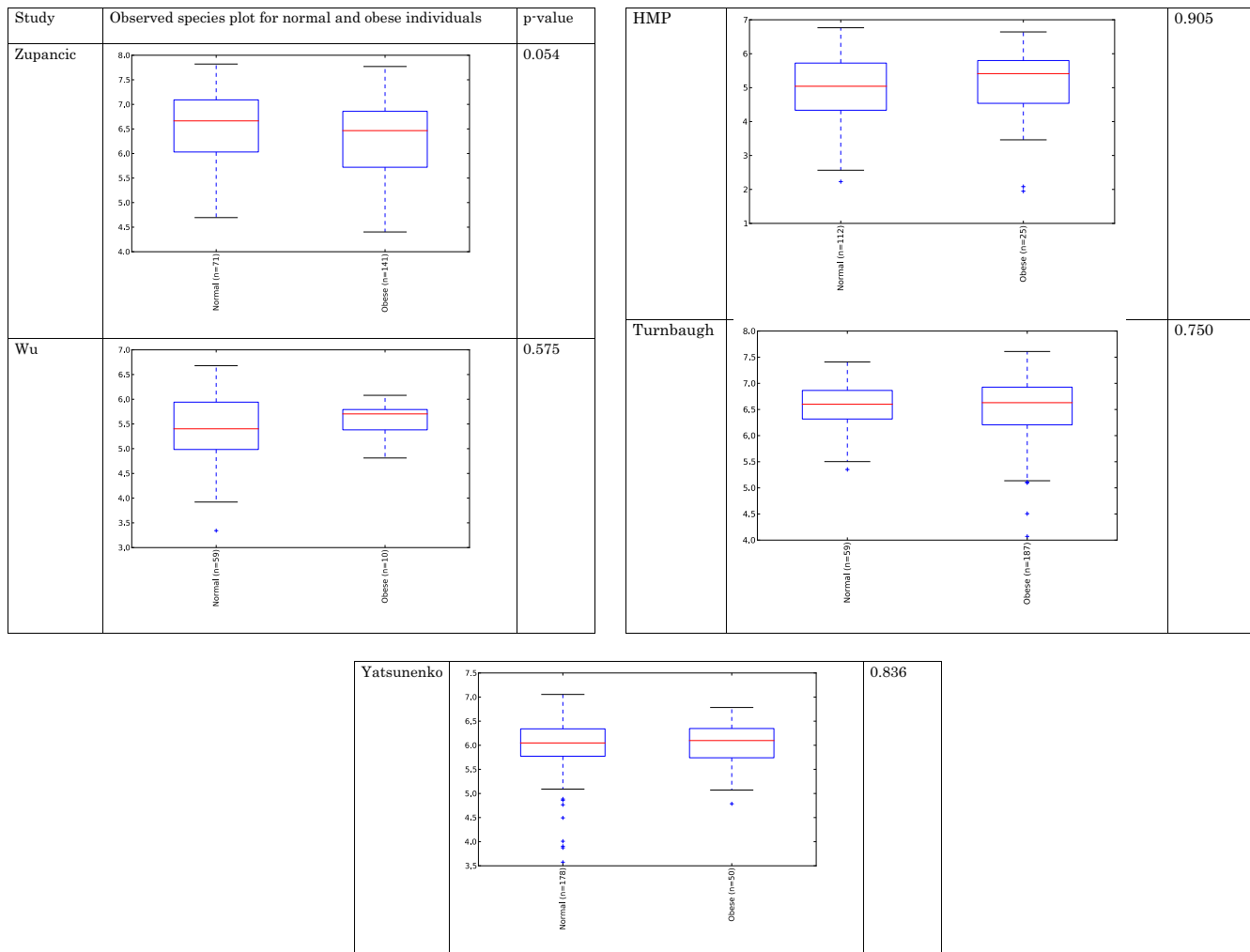


Fig. 4. Alpha diversity (Shannon) across studies. Metric is Shannon (abundance and evenness). Sequence depth is 1000 sequences per sample, and subsampling was performed 10 times. *P*-values were calculated by using a Monte Carlo simulation with 999 permutations.

therefore should not be considered a general feature distinguishing normal and obese human gut microbiota across populations.

Obese individuals have also been reported to have less diverse gut microbiota than normal weight individuals [8,10]. Differences are shown in Figs. 3 and 4 below for observed species (count of unique species) and the Shannon metric [29], a measure of species abundance and evenness, for each of the studies.

As Figs. 3 and 4 show, there is no consistent alpha diversity trend across 16S amplicon surveys of human BMI, and only one study showed significant differences with the observed species metric. In addition to the OTU-level alpha diversity calculations, significance was tested with taxonomic levels from phylum to species across each study with the Shannon and observed species metric. Only the phylum level Wu et al. data have a significant difference (*P*-value 0.039) with the Shannon metric. Turnbaugh et al. [8] detected increased phylogenetic diversity (QIIME's PD_whole_tree metric) in lean samples versus obese samples. The closed-reference data generated during this analysis did not yield significantly different alpha diversity with the phylogenetic diversity metric, although the Turnbaugh data processed in a de novo fashion (pick_otus script with default settings, using UCLUST v1.2.22q, followed by alignment of representative sequences with PyNast 1.2, filtering, and phylogenetic tree construction with Fast-Tree 2.1.3, all using default QIIME parameters) shows a significant difference in diversity, replicating the observation of Turnbaugh

et al. (2009), as shown in Fig. 5. The observed species and Shannon metric did not yield significant differences with the de novo processed data (not shown).

These results indicate that the overall diversity of the community, at least by 16S amplicon surveys, does not generally distinguish communities when closed-reference OTU picking is employed (requiring sequences to match what is already in the sequence database), although the difference seen in the de novo approach could indicate that novel taxa, not represented in the reference database, may differentiate these communities. Additionally, overall genomic content, as observed by Le Chatelier et al. [10], but not by Turnbaugh et al. [8], could still be more diverse in lean communities due to genomic variation that is not detected using the 16S amplicon survey approach.

Clustering of the data showed no trend based upon BMI (the data clustered strongly by study, but no trends were observed within each study according to BMI). The PCoA plot (using unweighted UniFrac distances) in Fig. 6 shows clustering colored by study and by normal/obese BMI categories (BMI categories not significantly different via PERMANOVA test). These results indicate that per-study effects are much larger than the biological effects separating lean from obese individuals, and point to a need to control for variation among studies.

Next, differences between taxa (examining all taxonomic levels from phylum down to species) within the studies were tested

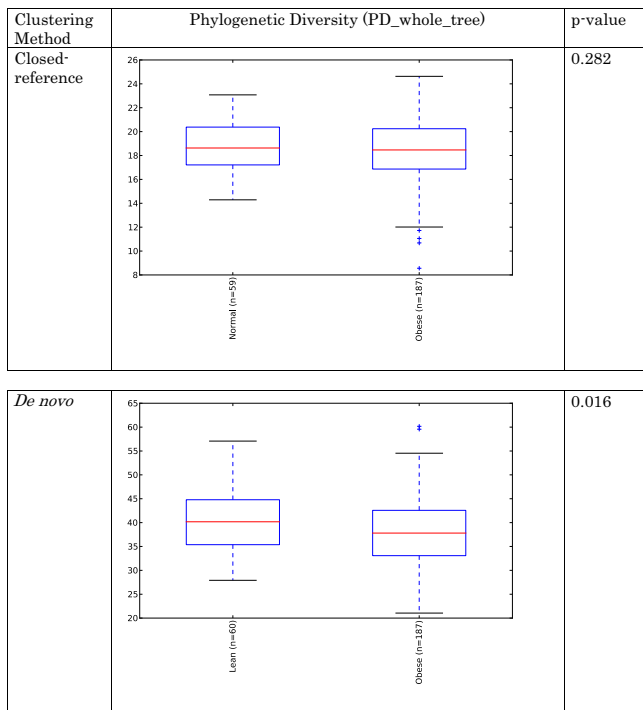


Fig. 5. Alpha diversity (PD) for Turnbaugh et al. data [8] across clustering methods. Metric is phylogenetic diversity (a measure of branch length of the phylogenetic tree occupied by the sequences present in the samples). Sequence depth is 1000 sequences per sample, and subsampling was performed 10 times. *P*-values were calculated by using a Monte Carlo simulation with 999 permutations.

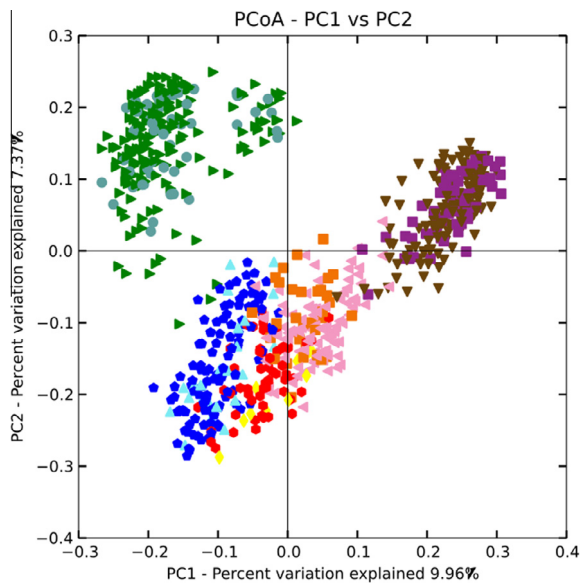


Fig. 6. Clustering of BMI samples with unweighted UniFrac. Shape/color: study/BMI category; purple square: Zupancic normal; brown triangle: Zupancic obese; orange square: Turnbaugh normal; pink triangle: Turnbaugh obese; red circle: Wu normal; yellow diamond: Wu obese; dark blue circle: HMP normal; light blue triangle: HMP obese; green triangle: Yatsunenکو normal; grey circle: Yatsunenکو obese.

using a Kruskal–Wallis test. These are shown in Table 3. The depletion of *Faecalibacterium prausnitzii* in the gut microbiota of obese individuals is consistent with prior studies. The species of *Megasphaera* that is increased in obese subjects is unknown, but

it may fill the same niche in the obese gut as the short-chain fatty acid producers *Eubacterium* and *Roseburia* (fermenting excess carbohydrates into fatty acids absorbed by the host) [30] that were previously observed to be increased in obese human fecal microbiota (Table 1). There is a slight, but significant depletion of an unclassified OTU (Clostridiaceae family) for obese subjects in the Yatsunenکو study. The nearest NCBI blast matches of cultured taxa include members of Clostridium cluster XI, which do not have unified metabolic properties. The failure to replicate the prior results may be due to actual population differences or alternative detection/quantification methods (e.g. amplicon sequencing versus qPCR) yielding different results.

Application of supervised learning (a form of machine learning where each sample is placed into a category that is used to create a training data set) with the random forest algorithm [31], i.e. decision trees based upon features, such as abundance of particular taxa, were used to build a model for placement of samples into categories for each study (across multiple levels of taxa, from phylum to the OTU level) for the normal and obese BMI categories. These had limited predictive power over random guessing (data not shown) using the QIIME supervised_learning.py script with default settings. Previous work [9] showed that one was able to classify the samples from the Turnbaugh dataset [8] into lean or obese subjects with the highest accuracies when clustering was done at lower thresholds than the standard 97% OTU threshold (the optimal threshold was ~82%). To replicate this approach, we clustered the demultiplexed sequence data using the pick_otus.py script utilizing UCLUST [4] v1.2.22q as the algorithm, and specified an identity of clustering of 0.600–0.995 in 0.05% increments, which is a finer gradation but includes the original percent identities utilized by Knights [9]. These analyses revealed that the current implementation of supervised learning in QIIME has an important limitation for datasets which have categories whose composition are highly similar, such as lean and obese humans: the classifier is quite sensitive to disparities in the sample sizes, as shown below. The Turnbaugh data were evenly sampled at 1000 sequences per sample, and processed with QIIME's supervised_learning.py (which uses random forests as its method) using 10× cross-fold validation for error estimation. Predicted error ratios are shown below in Fig. 7, which show far better results for an even subset of the data matching the samples utilized originally by Knights (61 lean samples and 61 obese samples, identified in Appendix A), and marginally better results for a random subset (30 lean versus 30 obese samples randomly chosen 10× for each percent identity, with the average error ratio plotted) versus the entirety of the data (61 lean samples versus 196 obese samples).

To handle uneven sample abundances with data that are highly similar, such as in the obese versus lean human fecal 16S sequence, and still generate models that can accurately classify samples into categories, we used the receiver operator characteristic (ROC) curve approach. This approach uses an area under the curve (AUC) calculation to optimize feature (in this case OTU) selection, which maximizes sensitivity and specificity for categorization of samples into lean and obese groups. The commands are listed in Appendix B (the R scripts are provided as Supplementary files). In contrast to the results in Fig. 7, when ROC-based optimization was used, the entire Turnbaugh dataset can be classified with over 80% accuracy at a range of OTU thresholds, as shown in Fig. 8. A ROC AUC value of 0.5 is no better than random guess, and 1.0 corresponds to perfect sensitivity and specificity. The OTU threshold identities that resulted in the best classification ranged from 77.5% to 98.0%.

Application of the ROC-based supervised learning approach to the closed-reference 97% OTUs across each BMI study resulted in classification that was better than random, although only

Table 3
Significant taxa associated with BMI via Kruskal–Wallis test. FDR multiple comparison correction used. Only taxa with at least 0.5% relative abundance in at least one of the BMI categories were included.

Study	Taxa	Obese mean	Normal mean	P-value
Zupancic	<i>Faecalibacterium prausnitzii</i>	0.031914894	0.0447	0.023
Turnbaugh	<i>Megasphaera</i> (sp)	0.006372549	3.13×10^{-5}	0.004
Yatsunenko	<i>Clostridiaceae</i>	0.00214	0.005011236	0.022

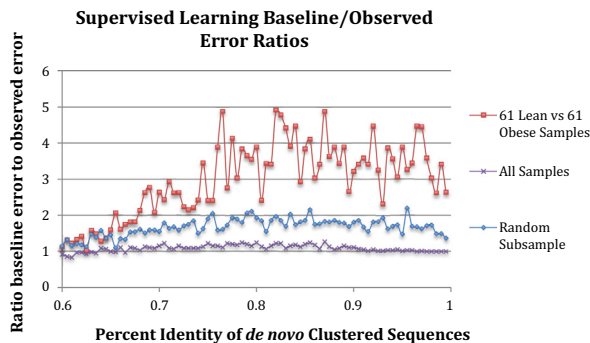


Fig. 7. Comparison of supervised learning error ratios to clustering identity of data. Samples matching those used in Knights et al. [9] replicated the improved classifications relative to random guessing (value of 1) for lean and obese subjects in the Turnbaugh [8] study, and are shown as the red line. The average error ratio for a random subsample of 30 obese and lean samples (10× sample at each percent identity, average ratio is shown) is depicted in blue. The purple line shows the classification error ratio when all samples (61 lean versus 196 obese samples), which is essentially no better than random guess for any clustering identity. The sequences were clustered using a de novo approach for each percent identity listed.

marginally in some cases. The ROC values are listed in Table 4, and the confusion matrices for each study are shown in Table S1.

Taken together, examination of the broad taxonomic shifts (e.g. the Firmicutes:Bacteroidetes ratios), alpha diversity of communities, clustering of obese versus lean individuals, and shifts of taxa within communities, suggests that there are only weak and, for the most part, non-significant associations of particular taxa or overall diversity with the obese human gut that hold true across different studies. However, using supervised learning with receiver operator curves to maximize sensitivity and specificity, one can categorize subjects according to lean and obese states with in some cases considerable accuracy, indicating that there are features present that discriminate between the lean and obese human fecal

Table 4

Receiver operator characteristic curve values for 97% closed-reference OTUs by study. Values listed are average values of 5× repeated ROC analyses, using random forest method, with 10-fold cross validation.

Study	ROC AUC values	Standard deviation
Turnbaugh	0.7250379	0.1488683
Amish	0.6077041	0.1103585
HMP	0.6656818	0.1295004
Wu	0.8623333	0.1638126
Yatsunenko	0.6259477	0.1033857

microbiota within each study but that are not consistent across different studies. As these features are not detected by other significance tests, this indicates that numerous small differences are present in the communities, rather than large differences in a few microbial taxa.

The methodologies for sample handling, extraction, PCR amplification, and sequencing technologies could all contribute to the differences observed between studies (Figs. 1 and 2). However, any prior observed differences due to algorithms or reference databases have been eliminated in this analysis. It is clear that the obesity phenotype is transmissible via microbes [12,32] and hence there exist real differences between these microbial communities that confer physiologically relevant effects. The actual differences, if detectable with 16S surveys, are likely to be found with indicator taxa, such as *F. prausnitzii*, or alternatively, by measurements of minute changes over the community.

3.2. Microbes associated with IBD

Inflammatory bowel disease (IBD), which includes Crohn's Disease and Ulcerative Colitis [33], has clearer reported taxonomic shifts than obesity, including a depletion of Firmicutes and Bacteroides and enrichment in Proteobacteria and Actinobacteria

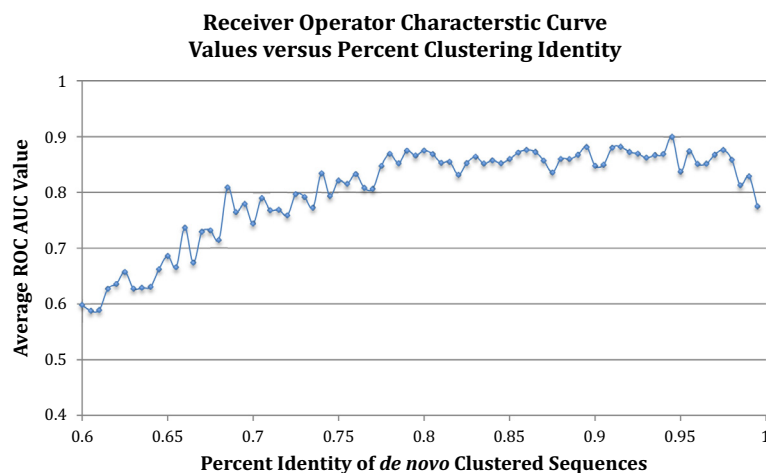


Fig. 8. Receiver operator characteristic curve values for all Turnbaugh lean and obese subjects. The average ROC area under the curve value (using random forest method) for each clustering identity was calculated by the averaging the 5× repeated (with 10-fold cross-validation) optimized ROC values. A 0.5 value indicates no better than random guess, while 1.0 indicates perfect sensitivity and specificity. The sequences were clustered using a de novo approach for each percent identity listed.

Table 5
Studies used for IBD component of meta-analysis.

Study	Sequencing platform	Observations
Papa [37]	454	The authors observed decreased diversity in IBD patients, depletion in Firmicutes, increased Proteobacteria and Actinobacteria in IBD subjects
Momozawa [38]	454	The purpose of study was to examine the differences between the microbiome of intestinal biopsies and stool and how extraction techniques altered the microbiome. A subset of subjects, with IBD, were present and used for this meta-analysis, but no conclusions regarding IBD itself were reported in this study
Morgan [39]	454	The authors observed decreased Firmicutes abundance and increase Proteobacteria abundance in IBD subjects
Willing [40]	454	The authors found differences among categories of IBD (e.g. ulcerative colitis and ileal Crohn's disease) versus healthy controls. The smallest differences were observed in ulcerative colitis, and the largest in small intestinal forms of IBD
Lamendella (unpublished)	Illumina MiSeq	This study is a Swedish cohort, with longitudinal stool samples along with flare-up and remission data

[34]. Epithelial-associated microbes of the small and large intestine are different than fecal microbiota [35,36], and are more likely to be key players in the etiology of IBD because of their more direct interaction with the affected tissues and the mucosal immune system. Despite the disparity in the fecal and epithelial-associated microbiota, detectable changes in fecal microbiota are still present in individuals with IBD [37]. This raises two clinically important questions: can particular forms of IBD be diagnosed from fecal samples alone (and avoid invasive medical procedures), and are such potential diagnostics consistent across populations studied (i.e., what dataset, if any, can be used to test an incoming sample from a random individual and have reasonably high accuracy in predicting if the subject has a particular form of IBD)? To address these questions, we examined studies of IBD based on human fecal 16S high-throughput sequencing surveys [37–40] plus a study of a Swedish cohort led by Gina Lamendella that is currently unpublished (summarized in Table 5). These data were processed with QIIME 1.8.0, using the same software packages/settings with the closed-reference approach described in Methods above, except for the even sampling depth used for beta diversity, Kruskal–Wallis tests, and supervised learning, which was 1004 sequences per sample.

These resulting taxonomic distributions for the three variants of IBD (UC = ulcerative colitis, CCD = colonic Crohn's disease, ICD = ileal Crohn's disease) and healthy controls (HC) are shown in Fig. 9, and match the expected enrichment (Actinobacteria, Proteobacteria) and depletion (Bacteroidetes, Firmicutes) previously observed.

Clustering of healthy controls versus IBD samples are significantly different with a P -value < 0.050 with PERMANOVA tests of unweighted UniFrac distances. These PCoA plots are shown in Fig. 10. A study effect is also apparent, which is seen in other PCoA

axes (not shown). The small intestinal Crohn's disease samples clustered the most distinctly from the healthy controls, and there was substantially more overlap between healthy subjects and ulcerative colitis samples, which again reflects previous observations including a recent large cross-cohort analysis [41].

Next, the significant taxa associated with IBD states were tested using the Kruskal–Wallis test. These taxa are listed in Table 6. These taxa appear generally consistent with prior observations of enrichment, although there is an interesting result of increased *Bifidobacterium adolescentis* and *Lactobacillus* in ICD and CCD samples, as these taxa have been used to treat chemically induced colitis [42]. Most of the prior studies recorded decreases in the overall abundance of the Bacteroidetes phylum, which was observed in this analysis as well (Fig. 9), although these were not detected as significantly different at the phylum level with Kruskal–Wallis tests.

Mycobacterium avium subspecies *paratuberculosis* has been implicated in IBD [43]. Two subjects in this analysis had detectable *Mycobacterium* sequences, but these were exceedingly rare (19 out of 496038 sequences in a subject with ileal Crohn's disease, and 1 out of 252119 sequences in a subject with ulcerative colitis). The PCR primers used are a perfect match for *M. avium*, and hence should have amplified this taxon. If *Mycobacterium* species are present in these IBD patients, they are not detectable in the feces of the subjects. Case studies of fecal microbiota transplant treatment of recalcitrant IBD show a sizable success rate [46], indicating a causal role for gut microbiota in causing or exacerbating the symptoms of IBD. The microbes may be shifted to a state of dysbiosis by various factors (e.g. antibiotic usage or host immune system) in the IBD subjects. Depletion of certain commensal taxa (*Akkermansia muciniphila*, *F. prausnitzii*, *Bacteroides uniformis*) in IBD (trend is different according to disease type) could be respon-

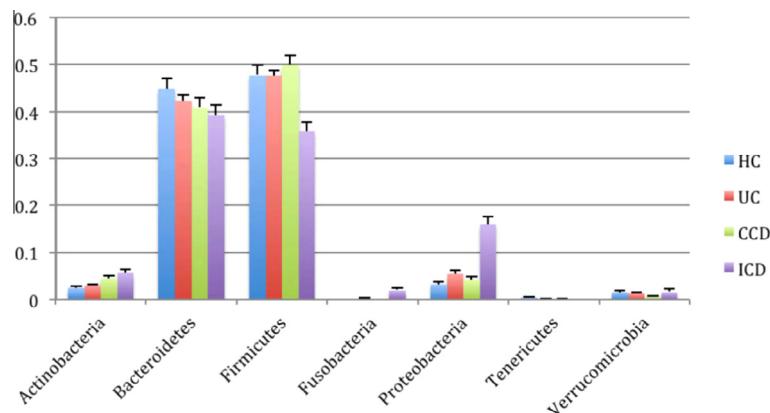


Fig. 9. Phylum-level taxa plots for IBD subjects versus healthy controls. Seven most abundant phyla shown. HC = healthy controls, UC = ulcerative colitis, CCD = colonic Crohn's disease, ICD = ileal Crohn's disease. Error bars indicate standard error of the mean.

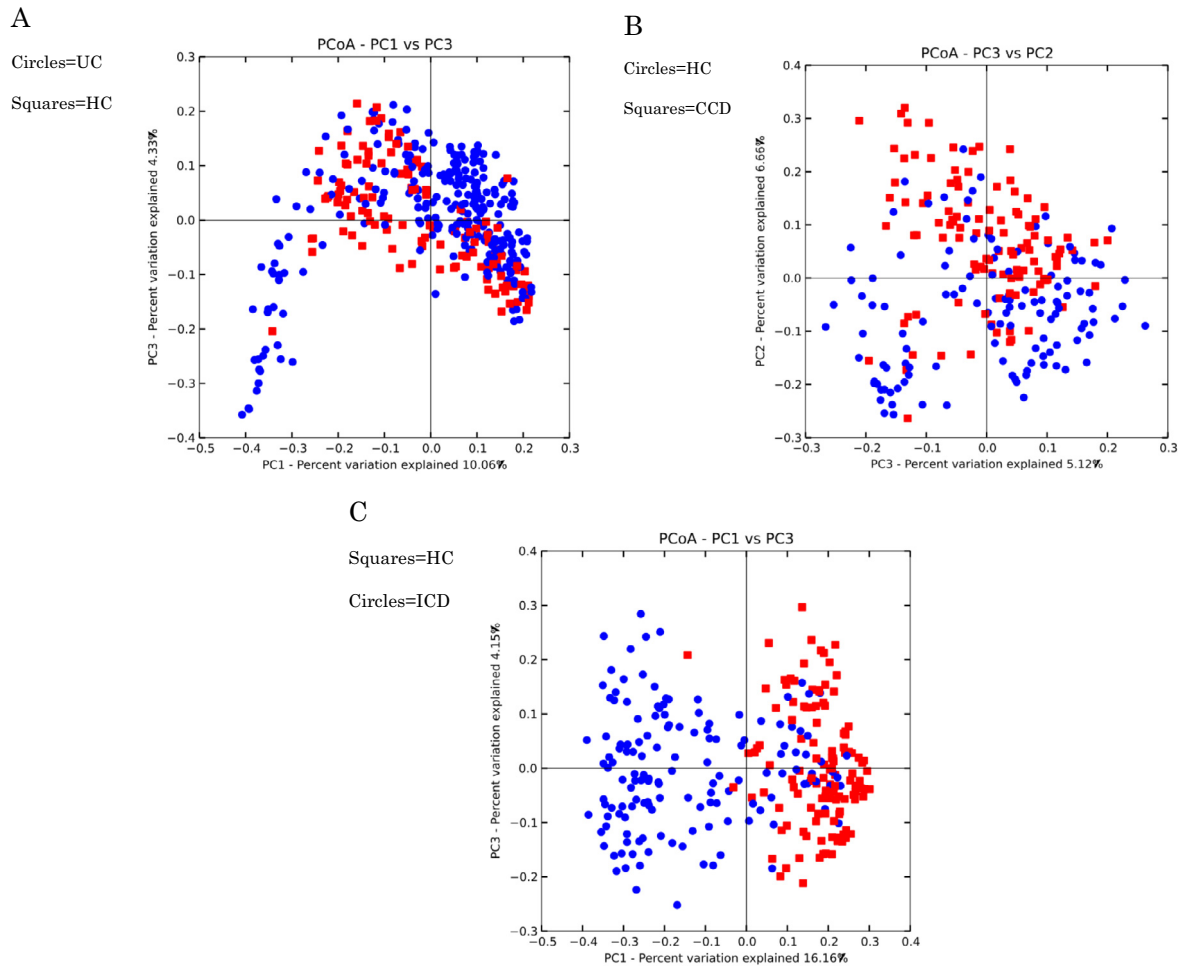


Fig. 10. PCoA plots of healthy controls versus subjects with IBD. Distances were calculated with unweighted UniFrac. (A) HC vs UC samples, (B) HC vs CCD samples, (C) HC vs ICD samples. Distances between healthy controls and all IBD categories are significantly different (P -value < 0.050 with PERMANOVA tests (999 permutations)). Data were evenly sampled at 1004 sequences per sample.

sible for causing inappropriate immune responses in the host [47–49], as well as opening up niches for occupation by invasive or pro-inflammatory species. Increased Proteobacteria and Fusobacteria (and a related decrease of less oxygen tolerate taxa, such as many of the Firmicutes) could be due to increased available oxygen in the intestinal lumen of subjects with IBD. These increased oxygen-tolerant taxa have the potential to produce pro-inflammatory responses in the host through flagellin or lipopolysaccharides [50,51]. *Bifidobacterium* may be increased in abundance due to oxygen tolerance relative to other taxa [52]. Previous proteomic data suggests that certain opportunistic *Bacteroides* sp. pathogens are increased in IBD subjects, at the expense of *Prevotella* species, which could explain the shifts within Bacteroidetes in our observations [53].

Alpha diversity between IBD categories was also consistent with prior observations (i.e. lower in subjects with IBD relative to healthy controls), and is shown in Fig. 11.

3.3. Predicted disease states via fecal sequencing

The question of whether the combined IBD microbiota data can provide an accurate training dataset for predicting disease states was tested using supervised learning (with the same ROC based approach described above for BMI samples). All levels of taxonomy (phylum to species) were tested, with 10-fold cross validation for error rate prediction. The species level table performed best for

almost all comparisons between healthy controls and individual disease states and all disease states combined, which are shown in Table 7, with one exception—the family level had a slightly higher AUC ROC value (only 0.00326 higher) for healthy controls versus ICD samples, but this is much smaller than the accuracy that would be lost for each of the other categories when using the family level versus the species level. When all four classes of samples are combined, the supervised learning accuracy drops (68.6%). The confusion matrix for the combined four class data is shown in Table S2.

These tests indicate that the combined inflammatory bowel disease sequence sets are accurate at placing individuals into an IBD/healthy category. These tests could therefore serve as a complement to other, non-invasive diagnoses of symptoms (serological markers, bowel movement frequency, bloody diarrhea, rectal lesions, and a variety of non-gastrointestinal tract symptoms, which by themselves also suffer from a lack of sensitivity/specificity in diagnoses) and minimize risks (e.g. nosocomial infections) to patients. The AUC ROC approach requires a two class comparison, so applications of this would be most optimal for an initial diagnosis to determine if the subject likely has IBD or not, or to complement other diagnostics that suggest a particular IBD type.

4. Conclusions

Our meta-analyses show that studies of inflammatory bowel disease reveal clear, consistent differences between healthy and

Table 6

Taxa associated with IBD. + or – indicates that taxa were significantly enriched or depleted relative to healthy controls with a Kruskal–Wallis test and Bonferroni multiple comparisons correction. Only taxa that were present in at least 0.5% of one of the categories (HC, UC, CCD, or ICD) were included. Data were rarefied to 1004 sequences per sample. The ‘Matches Study’ column includes a reference with consistent observations to the abundance results reported in this study.

Taxonomy	UC	CCD	ICD	Matches study
Actinobacteria		+	+	[34,37,40]
<i>Bifidobacterium</i> (sp)		+	+	[40]
<i>Bifidobacterium adolescentis</i>		+	+	
Bacteroidetes				
<i>Bacteroides eggerthii</i>			–	
<i>Parabacteroides distasonis</i>		–		
<i>Parabacteroides</i> (sp)		–	–	
<i>Bacteroides uniformis</i>		–		
<i>Bacteroides</i> (sp)	+	+		
<i>Paraprevotella</i> (sp)		+		
<i>Prevotella</i> (sp)	–	–	–	[40]
<i>Prevotella copri</i>	–		–	
Rikenellaceae			–	
Barnesiellaceae (unknown genus)			–	
Firmicutes				
<i>Faecalibacterium prausnitzii</i>	–	+	–	[37,39] [34,40]
<i>Oscillospira</i> (sp)		–	–	
Clostridiales (unknown family)			–	[37,41]
<i>Clostridium</i> (sp)	–	–	–	
<i>Roseburia</i> (sp)			–	[39,40]
Lachnospiraceae (unknown genus)			–	[34]
<i>Coprococcus</i> (sp)			–	[37]
<i>Ruminococcus gnavus</i>		+		
<i>Ruminococcus</i> (sp)		–	–	[37,39]
<i>Lachnospira</i> (sp)			–	[34,37]
<i>Blautia</i> (sp)		+		
<i>Blautia producta</i>			+	
<i>Dialister</i> (sp)		+		
<i>Veillonella dispar</i>			+	[40]
<i>Phascolarctobacterium</i> (sp)		–	–	[39]
<i>Lactobacillus</i> (sp)			+	[37,40]
Fusobacteria			+	[41]
<i>Fusobacterium</i> (sp)			+	[41]
Gamma-Proteobacteria			+	[34,37,39–41]
Enterobacteriaceae			+	[37,39–41]
Verrucomicrobia		–	–	
<i>Akkermansia muciniphila</i>		–		

Table 7

Predicting IBD state via supervised learning. Values listed are average values of 5× repeated ROC analyses, using random forest method, with 10-fold cross validation. Healthy controls versus each individual category of disease are shown as well as healthy controls versus all IBD categories combined. A ROC AUC value of 0.5 is no better than random guess, whereas 1.0 indicates perfect specificity and sensitivity.

Categories compared	ROC AUC value	ROC AUC stdev
Healthy controls versus ulcerative colitis	0.92258803	0.04239816
Healthy controls versus colonic Crohn's disease	0.87879176	0.07375804
Healthy controls versus ileal Crohn's disease	0.96996245	0.0378
Healthy controls versus all IBD categories	0.92404109	0.04297096

diseased individuals that replicate across studies and where the biological effects between clinical states exceed per-study effects. In contrast, the differences between obese and lean individuals, even though these differences can be transmitted to recipient gnotobiotic mice [32], are much less consistent. The ratio of Firmicutes:Bacteroidetes is not a consistent feature distinguishing lean from obese human microbiota generally, and previous studies did not find this pattern consistently when comparing human obese and lean fecal 16S data. However, in these previous studies, assessment of taxon abundance was not performed consistently with respect to the clustering and taxonomic assignment of sequences between studies. The approach to clustering and quality control (i.e., ‘‘closed reference’’) used here is a conservative one, as it limits the potential for sequencing noise and chimeras to interfere with the results at the cost of perhaps discarding real, novel reads. However, as human-associated microbiota have been sequenced with considerable coverage, and the most up to date Greengenes reference database was used, the potential for large numbers of novel reads being discarded is minimal. As shown in the gnotobiotic mouse transplant systems, microbes can have an impact on obesity, so the limited significant differences observed in the taxa (Kruskal–Wallis tests) are surprising. It is possible that the differences between the lean and obese microbiota are present in the pan-genome of the microbes (not detectable by amplicon sequencing), that low-abundance taxa in the fecal samples are colonizing other locations in the colon (e.g. the proximal colon or the cecum) of the recipient animals and driving the differences observed in the hosts' physiology, and hence are not apparent in the fecal communities of the human subjects being studied, or that the differences may be multiple, small shifts in the community that escape detection with significance tests, but when combined can differentiate lean and obese communities with an approach like supervised learning.

To determine whether a consistent pattern was observable across multiple studies with IBD and normal versus obese individuals, all available datasets were reprocessed using the latest reference 16S rRNA gene database. For clinical utility, such as evaluating a particular form of Crohn's disease, a taxonomic signature needs to be consistent across populations. In the BMI data analysis, there was no significant signal in the Firmicutes:Bacteroidetes ratio or in overall diversity of the samples to differentiate obese and normal weight subjects. Indicator taxa were found within studies to be significant, although these were not found to be consistently significant across studies (a trait also frequently seen with host genetic markers in GWAS). The taxa found to be significantly different in the BMI analyses may be linked to increased inflammation in individuals with high BMI (decreased *F. prausnitzii*) or increased extraction of energy from dietary polysaccharides (*Megasphaera* sp). A large number of significant shifts were

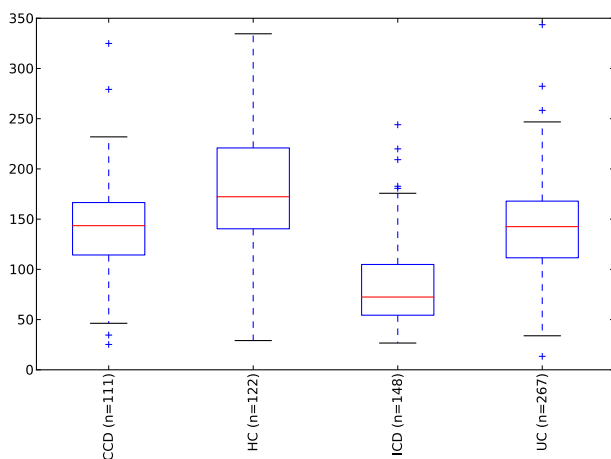


Fig. 11. Alpha diversity for IBD subjects and healthy controls. Y-axis indicates observed species value. Healthy control samples are significantly different from all inflammatory bowel disease categories with a P -value of <0.05 (Monte Carlo permutation test, permutations = 999). The samples were repeatedly sampled (10×) at 1000 sequences/sample.

detected in IBD subjects, which may be related to a dysregulated immune system or altered environment (luminal oxygen content or antibiotics). Judicious use of supervised learning, optimized to maximize sensitivity and minimize noise, can be used to distinguish lean and obese individuals with some degree of accuracy. The inflammatory bowel disease studies showed clearer patterns: many taxa are significantly enriched or depleted in subjects with IBD, significantly reduced overall diversity is present in subjects with inflammatory bowel disease, and significant differences are detecting when clustering IBD samples and healthy controls. Using the combined IBD data as a training set, a reasonably accurate assignment of subjects between healthy and particular IBD state is possible, but would need to complement other diagnostics in a clinical setting.

Acknowledgements

This work was supported in part by the National Institutes of Health, the Crohns and Colitis Foundation of America, and the Howard Hughes Medical Institute.

We would also like to thank Dan Knights for his insight and assistance with the supervised learning analyses.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.febslet.2014.09.039>.

References

- [1] Lozupone, C.A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vazquez-Baeza, Y., Jansson, J.K., Gordon, J.I. and Knight, R. (2013) Meta-analyses of studies of the human microbiota. *Genome Res.* 23, 1704–1714.
- [2] Liu, Z., DeSantis, T.Z., Andersen, G.L. and Knight, R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencing. *Nucleic Acids Res.* 36, e120.
- [3] DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L. (2006) Greengenes, a chimeric-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072.
- [4] Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461.
- [5] Lozupone, C. and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235.
- [6] Kelly, T., Yang, W., Chen, C.S., Reynolds, K. and He, J. (2008) Global burden of obesity in 2005 and projections to 2030. *Int. J. Obes.* 32, 1431–1437.
- [7] Ley, R.E., Turnbaugh, P.J., Klein, S. and Gordon, J.I. (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* 444, 1022–1023.
- [8] Turnbaugh, P.J., Hamady, M., Yatsunenok, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R. and Gordon, J.I. (2009) A core gut microbiome in obese and lean twins. *Nature* 457, 480–484.
- [9] Knights, D., Parfrey, L.W., Zaneveld, J., Lozupone, C. and Knight, R. (2011) Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe* 10, 292–296.
- [10] Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jørgensen, T., Brandslund, I., Nielsen, H.B., Juncker, A.S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E.G., Brunak, S., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W.M., Zucker, J.D., Raes, J., Hansen, T., Meta, H.I.T.C., Bork, P., Wang, J., Ehrlich, S.D. and Pedersen, O. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–546.
- [11] Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.
- [12] Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E. and Gewirtz, A.T. (2010) Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* 328, 228–231.
- [13] Bervoets, L., Van Hoorenbeeck, K., Kortleven, I., Van Noten, C., Hens, N., Vael, C., Goossens, H., Desager, K.N. and Vankerckhoven, V. (2013) Differences in gut microbiota composition between obese and lean children: a cross-sectional study. *Gut Pathog.* 5, 10.
- [14] Colson, P., Fancello, L., Gimenez, G., Armougom, F., Desnues, C., Fournous, G., Yoosuf, N., Million, M., La Scola, B. and Raoult, D. (2013) Evidence of the megavirome in humans. *J. Clin. Virol.* 57, 191–200.
- [15] Ferrer, M., Ruiz, A., Lanza, F., Haange, S.B., Oberbach, A., Till, H., Bargiela, R., Campoy, C., Segura, M.T., Richter, M., von Bergen, M., Seifert, J. and Suarez, A. (2013) Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environ. Microbiol.* 15, 211–226.
- [16] Furet, J.P., Kong, L.C., Tap, J., Poitou, C., Basdevant, A., Bouillot, J.L., Mariat, D., Corthier, G., Dore, J., Henegar, C., Rizkalla, S. and Clement, K. (2010) Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* 59, 3049–3057.
- [17] Ley, R.E., Peterson, D.A. and Gordon, J.I. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848.
- [18] Verdum, F.J., Fuentes, S., de Jonge, C., Zoetendal, E.G., Erbil, R., Greve, J.W., Buurman, W.A., de Vos, W.M. and Rensen, S.S. (2013) Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. *Obesity* 21, E607–E615.
- [19] Xu, P., Li, M., Zhang, J. and Zhang, T. (2012) Correlation of intestinal microbiota with overweight and obesity in Kazakh school children. *BMC Microbiol.* 12, 283.
- [20] Duncan, S.H., Lobley, G.E., Holtrop, G., Ince, J., Johnstone, A.M., Louis, P. and Flint, H.J. (2008) Human colonic microbiota associated with diet, obesity and weight loss. *Int. J. Obes.* 32, 1720–1724.
- [21] Jumpertz, R., Le, D.S., Turnbaugh, P.J., Trinidad, C., Bogardus, C., Gordon, J.I. and Krakoff, J. (2011) Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am. J. Clin. Nutr.* 94, 58–65.
- [22] Patil, D.P., Dhotre, D.P., Chavan, S.G., Sultan, A., Jain, D.S., Lanjekar, V.B., Gangawani, J., Shah, P.S., Todkar, J.S., Shah, S., Ranade, D.R., Patole, M.S. and Shouche, Y.S. (2012) Molecular analysis of gut microbiota in obesity among Indian individuals. *J. Biosci.* 37, 647–657.
- [23] Schwirtz, A., Taras, D., Schafer, K., Beijer, S., Bos, N.A., Donus, C. and Hardt, P.D. (2010) Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* 18, 190–195.
- [24] Tims, S., Derom, C., Jonkers, D.M., Vlietinck, R., Saris, W.H., Kleerebezem, M., de Vos, W.M. and Zoetendal, E.G. (2013) Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J.* 7, 707–717.
- [25] Zupancic, M.L., Cantarel, B.L., Liu, Z., Drabek, E.F., Ryan, K.A., Cirimotich, S., Jones, C., Knight, R., Walters, W.A., Knights, D., Mongodin, E.F., Horenstein, R.B., Mitchell, B.D., Steinle, N., Snitker, S., Shuldiner, A.R. and Fraser, C.M. (2012) Analysis of the gut microbiota in the old order Amish and its relation to the metabolic syndrome. *PLoS One* 7, e43052.
- [26] Team, R. D. C. (2008) R: A language and environment for statistical computing. in: Vienna, Austria.
- [27] Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.
- [28] Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D. and Lewis, J.D. (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108.
- [29] Shannon, C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–473.
- [30] Holt, J.G. (1994) *Bergey's Manual of Determinative Bacteriology*, Lippincott Williams & Wilkins, London.
- [31] Breimann, L. (2001) Random forests. *Machine Learn.* 45, 5–32.
- [32] Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffn, N.W., Lombard, V., Henrissat, B., Bain, J.R., Muehlbauer, M.J., Ilkayeva, O., Semenkovich, C.F., Funai, K., Hayashi, D.K., Lyle, B.J., Martini, M.C., Ursell, L.K., Clemente, J.C., Van Treuren, W., Walters, W.A., Knight, R., Newgard, C.B., Heath, A.C. and Gordon, J.I. (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341, 1241214.
- [33] Baumgard, D.C. and Carding, S.R. (2007) Inflammatory bowel disease: cause and immunobiology. *Lancet* 369, 1627–1640.
- [34] Mazmanian, S.K. and Sartor, R.B. (2012) Intestinal microbes in inflammatory bowel diseases. *Am. J. Gastroenterol. Suppl.* 1, 15–21.
- [35] Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. and Relman, D.A. (2005) Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.
- [36] Frank, D.N., St Amand, A.L., Feldman, R.A., Boedeker, E.C., Harpaz, N. and Pace, N.R. (2007) Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13780–13785.
- [37] Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S.P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., Schauer, D.B., Ward, D.V., Korzenik, J.R., Xavier, R.J., Bousvaros, A. and Alm, E.J. (2012) Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS One* 7, e39242.
- [38] Momozawa, Y., Defontaine, V., Louis, E. and Medrano, J.F. (2011) Characterization of bacteria in biopsies of colon and stools by high throughput sequencing of the V2 region of bacterial 16S rRNA gene in human. *PLoS One* 6, e16952.
- [39] Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., Bousvaros, A., Korzenik, J., Sands, B.E.,

- Xavier, N. and Huttenhower, C. (2014) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79.
- [40] Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J.K. and Engstrand, L. (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139 (1844–1854), e1.
- [41] Gevers, D., Kugathasan, S., Denson, L.A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., Morgan, X.C., Kotic, A.D., Luo, C., Gonzalez, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R., Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R. and Xavier, R.J. (2014) The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392.
- [42] Sheil, B., Shanahan, F. and O'Mahony, L. (2007) Probiotic effects on inflammatory bowel disease. *J. Nutr.* 137, 819S–824S.
- [43] Pierce, E.S. (2010) Ulcerative colitis and Crohn's disease: is *Mycobacterium avium* subspecies paratuberculosis the common villain? *Gut Pathog.* 2, 21.
- [44] Human Microbiome Project, C. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214.
- [45] Yatsunenkov, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R. and Gordon, J.I. (2012) Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.
- [46] Brandt, L.J. and Aroniadis, O.C. (2013) An overview of fecal microbiota transplantation: techniques, indications, and outcomes. *Gastrointest. Endosc.* 78, 240–249.
- [47] Collado, M.C., Derrien, M., Isolauri, E., de Vos, W.M. and Salminen, S. (2007) Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl. Environ. Microbiol.* 73, 7767–7770.
- [48] Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L.G., Chatel, J.M., Sokol, H., Thomas, M., Wells, J.M. and Langella, P. (2013) Faecalibacterium prausnitzii and human intestinal health. *Curr. Opin. Microbiol.* 16, 255–261.
- [49] Gauffin Cano, P., Santacruz, A., Moya, Á. and Sanz, Y. (2012) Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. *PLoS One* 7, e41079.
- [50] Strauss, J., Kaplan, G.G., Beck, P.L., Rioux, K., Panaccione, R., Devinney, R., Lynch, T. and Allen-Vercoe, E. (2011) Invasive potential of gut mucosa-derived *Fusobacterium nucleatum* positively correlates with IBD status of the host. *Inflamm. Bowel Dis.* 17, 1971–1978.
- [51] Vijay-Kumar, M. and Gewirtz, A.T. (2009) Flagellin: key target of mucosal innate immunity. *Mucosal Immunol.* 2, 197–205.
- [52] Shimamura, S., Abe, F., Ishibashi, N., Miyakawa, H., Yaeshima, T., Araya, T. and Tomita, M. (1992) Relationship between oxygen sensitivity and oxygen metabolism of *Bifidobacterium* species. *J. Dairy Sci.* 75, 3296–3306.
- [53] Juste, C., Kreil, D.P., Beauvallet, C., Guillot, A., Vaca, S., Carapito, C., Mondot, S., Sykacek, P., Sokol, H., Blon, F., Lepercq, P., Levenez, F., Valot, B., Carré, W., Loux, V., Pons, N., David, O., Schaeffer, B., Lepage, P., Martin, P., Monnet, V., Seksik, P., Beaugerie, L., Ehrlich, S.D., Gibrat, J.-F., Van Dorsselaer, A. and Doré, J. (2014) Bacterial protein signals are associated with Crohn's disease. *Gut*.