# A Taxonomic Signature of Obesity in the Microbiome? Getting to the Guts of the Matter

**Mariel M. Finucane[1], Thomas J. Sharpton[1¤], Timothy J. Laurent[1], Katherine S. Pollard[1,2]***

1 The J. David Gladstone Institutes, San Francisco, California, United States of America, 2 Institute for Human Genetics and Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States of America

## Abstract

Obesity is an important and intractable public health problem. In addition to the well-known risk factors of behavior, diet, and genetics, gut microbial communities were recently identified as another possible source of risk and a potential therapeutic target. However, human and animal-model studies have yielded conflicting results about the precise nature of associations between microbiome composition and obesity. In this paper, we use publicly available data from the Human Microbiome Project (HMP) and MetaHIT, both surveys of healthy adults that include obese individuals, plus two smaller studies that specifically examined lean versus obese adults. We find that inter-study variability in the taxonomic composition of stool microbiomes far exceeds differences between lean and obese individuals within studies. Our analyses further reveal a high degree of variability in stool microbiome composition and diversity across individuals. While we confirm the previously published small, but statistically significant, differences in phylum-level taxonomic composition between lean and obese individuals in several cohorts, we find no association between BMI and taxonomic composition of stool microbiomes in the larger HMP and MetaHIT datasets. We explore a range of different statistical techniques and show that this result is robust to the choice of methodology. Differences between studies are likely due to a combination of technical and clinical factors. We conclude that there is no simple taxonomic signature of obesity in the microbiota of the human gut.

## Introduction

Obesity is among the defining public health challenges of our time, with an estimated 3.4 million annual deaths attributable to high BMI [1]. Dietary and lifestyle interventions have only modest effects, and it is unclear whether these benefits persist over time [2]. Thus, there is substantial interest in alternative approaches to weight loss.

A tantalizing new theory has emerged in recent years, suggesting that the gut microbiome may offer a therapeutic target. Supporting a causal role of gut microbes in obesity, studies in mice showed that obesity can be induced in lean individuals via fecal transplants from obese individuals [3,4]. While the mechanisms through which gut microbes influence BMI are unknown, multiple investigations of gut microbiome composition in both mice and humans have shown that obese individuals have a lower ratio of bacteria from the phylum *Bacteroidetes* to bacteria from the phylum *Firmicutes* than lean individuals [3,5–7]. Obese individuals have also been shown to harbor less diverse bacterial communities [7,8].

Both the scientific literature [9,10] and the popular press [11] have heralded the association of obesity and the relative abundance of *Bacteroidetes* vs. *Firmicutes* as a robust finding. However, several recent reports question the strength of this association. Two large studies found no association between

obesity and the *Bacteroidetes*: *Firmicutes* ratio [12,13]. Furthermore, several publications actually report a *higher* ratio of *Bacteroidetes* to *Firmicutes* among obese individuals [14], in direct contradiction with the original findings.

The Human Microbiome Project (HMP) Consortium has collated the largest existing dataset describing the microbiota of healthy individuals, with sequences curated using stringent quality control. The cohort includes 16S rRNA sequencing of stool microbiomes from more than 200 adults living in Houston and Saint Louis [15], and it contains subjects with a comprehensive range of BMI values, including 24 obese ($BMI \geq 30$) and 123 lean ($BMI \leq 25$) individuals. These data provide an opportunity to investigate the conflicting findings about taxonomic composition of the gut microbiome and obesity.

To this end, we conducted an extensive assessment of the relationship between BMI and the taxonomic composition of the gut microbiome in the HMP dataset and compared our results to trends in the MetaHIT data [16], which is another large survey of healthy obese and non-obese adults, as well as to two earlier studies that specifically sampled lean and obese individuals [6,7]. Our analysis expands upon the work of the HMP Consortium, which explored a large set of candidate relationships between host phenotypic data (e.g., BMI, age, blood pressure) and microbial, enzymatic and pathway abundance and did not find a significant

association between BMI and the relative abundance of *Bacteroidetes* or *Firmicutes* [13]. That analysis was designed to search automatedly over a large number of candidate relationships by a single pre-specified approach. To ensure that an association between BMI and gut community composition was not missed, we employed a range of graphical and statistical modeling techniques, quantified community composition in a variety of ways, and performed power calculations. This thorough interrogation of the data confirms that there is no association between BMI and stool microbiome taxonomic composition or diversity in the HMP cohort. When we examined the relationship between stool microbiome composition and BMI across different studies, we found that inter-study variability far exceeds differences in composition between lean and obese individuals within each study. Our results suggest that there is no simple relationship between BMI and gut microbiota and that significant technical and clinical differences exist between published studies.

## Results

### The *Bacteroidetes:Firmicutes* ratio is not associated with obesity or BMI

We began by attempting to reproduce the best-known result supporting the theory that obese individuals have a lower ratio of *Bacteroidetes* to *Firmicutes* in their guts. We found no difference between obese versus lean individuals in their relative abundance of *Bacteroidetes* or *Firmicutes* ($p = 0.30$ and $0.86$, respectively, by *t*-test).

Importantly, our failure to detect these differences was not due to insufficient statistical power. With our sample sizes and previously reported effect sizes, it is very unlikely that we would have found no association in the HMP data if an association did exist in the St. Louis and Houston populations. For example, using effect sizes from Turnbaugh et al. [7] for the V6 region of the 16S rRNA gene in European Americans, we would have had 96% power to detect a difference in the relative abundance of *Bacteroidetes* and 80% power for *Firmicutes*. These power calculations account for the fact that the proportion of obese individuals in the HMP cohort is lower than in [7].

Because the detrimental health effects of being overweight occur along a continuum of BMI values and not just above the obesity cutoff (BMI $\geq 30$), we next looked for a quantitative association between the continuous BMI variable and the ratio of *Bacteroidetes* to *Firmicutes*. There was no association (Figure 1; linear regression $p = 0.41$), and this ratio varied greatly between individuals regardless of BMI.

### Alternative quantifications of taxonomic composition are also not associated with BMI

Next, we investigated the possibility that – for the purposes of detecting an association with BMI – the ratio of *Bacteroidetes* to *Firmicutes* did not adequately summarize the taxonomic composition of the gut microbiome at the phylum level. Specifically, we quantified the relative abundance of the five major bacterial phyla in each sample and constructed a phylum-level compositional profile for each individual. We then visualized these compositional profiles as a function of BMI. No signal was apparent (Figure 2).

Then, to ensure that we had not missed a subtle pattern in this plot, we used a statistical model to isolate BMI effects from residual variance due to sampling and measurement error. Specifically, we modeled the isometric log ratio transform [17] of the relative abundance of each major phylum in each sample using a linear model, including a fixed effect of phylum plus a phylum-specific effect of BMI plus a random error term. We again found no
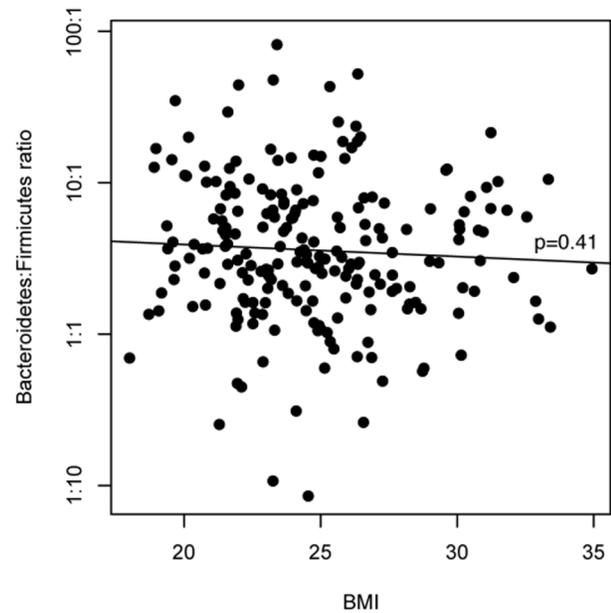


**Figure 1. There is no association between BMI and the *Bacteroidetes:Firmicutes* ratio in HMP stool microbiomes.**
doi:10.1371/journal.pone.0084689.g001

significant association between BMI and taxonomic composition at the phylum level.

We then considered the possibility that a BMI association exists at a finer taxonomic resolution, despite the lack of association at the phylum level. For each individual, we generated a new taxonomic composition profile that quantified the relative abundance of each bacterial genus in the individuals stool microbiome. We applied principal components analysis to these genus-level profiles to reduce their dimensionality, as in Smith et al.
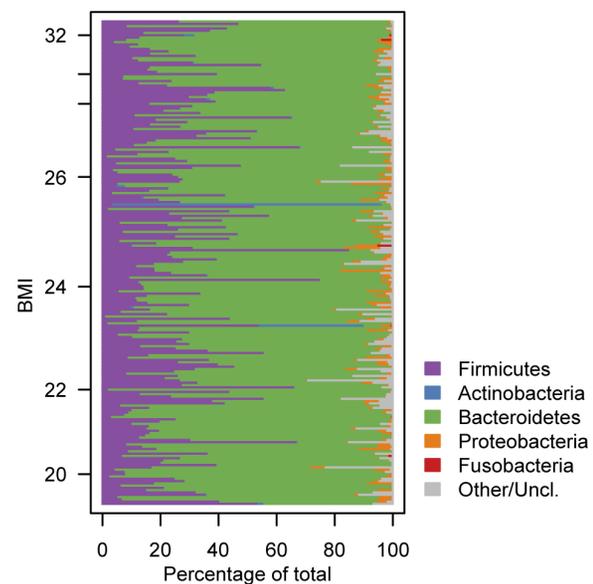


**Figure 2. There is no relationship between BMI and the phylum-level composition of the microbiome.** Each row shows the relative abundance of major gut bacterial phyla in an individual. Individuals are ordered according to their BMI.
doi:10.1371/journal.pone.0084689.g002

[18]. We then tested for an association between BMI and any of the first six principal components (which explain 96% of the variance in genus-level profiles). We found no significant associations.

Additionally, to ensure that the principal components reduction did not obscure an association, we used logistic regression to model the probability of observing each major genus as a function of BMI. We again found no associations.

## Gut microbiome community diversity is not associated with BMI

Finally, we investigated the hypothesis that BMI is associated not with the relative abundance of particular taxa but rather the diversity of taxa present. This possibility was supported by Turnbaugh et al. [7] and by Le Chatelier et al. [8], who both concluded that a diverse gut microbiome can have a protective effect against obesity. Following the approach of Turnbaugh et al. [7], we used the 97% identity operational taxonomic unit (OTU) calls on the sequencing reads for each HMP sample to calculate rarefaction curves. We then used these curves to compare richness levels (i.e., total number of OTUs) between obese and lean individuals. In contrast to the results in [7] and [8], we found no relationship between richness and obesity, but rather observed a high degree of residual variability in OTU richness across individuals.

To ensure that these surprising results were not an artifact of diversity measure or calculation procedure, we validated our findings using the Shannon entropy measure as well as a variety of microbial ecology analysis software packages (mothur[19], QIIME [20], and MetaPhlAn[21]). This sensitivity analysis confirmed that our null diversity results were robust.

## Obesity effects are not consistent across studies

To ensure that we could reproduce the significant findings of previous analyses [6,7], and to assess whether another large, recent study with a wider range of BMI values (MetaHIT [16]) could help clarify the contradictory results, we reanalyzed data from [6], [7], and from the Danish subjects in [16]. Although the primary analyses of the present manuscript were restricted to the V35 region of the 16S rRNA gene, we also included HMP data from the V13 region to assess the possibility that a 16S-region-specific bias could be obscuring a true underlying relationship.

In Figure 3, we show that variation in the relative abundance of *Firmicutes* and *Bacteroidetes* is much larger among studies than between lean and obese individuals within any study. Not only do the MetaHIT and HMP results fail to recapitulate the findings of Ley and Turnbaugh, but they actually go in the opposite direction. In the HMP data, this finding is consistent for the V13 and V35 regions of the 16S locus, both of which were sequenced in the same individuals on the same platform (Roche 454). The substantial between-study variability could be due to an unmeasured factor such as diet [22] or due to technical factors such as DNA extraction technique, region of the 16S locus targeted, or sequencing platform [23]. However, the consistent lack of BMI association for the V13 and V35 regions in the HMP data suggest that 16S region is not a major confounder, at least in this cohort and for these two variable regions.

## Discussion

Fecal transplant studies in mice have shown conclusively that the microbiome has a causal effect on obesity [3,4], and a number of high-profile papers have found that obese individuals have lower ratios of *Bacteroidetes* to *Firmicutes* [3,5–7]. Yet we and others

have found no relationship between BMI and the gut microbiome's phylum-level composition in large-scale analyses [12,13]. These contradictory observations suggest that no simple taxonomic signature of obesity exists in the gut microbiome.

There are a number of possible reasons for these conflicting results. One possibility is that unmeasured confounders obscure a true underlying relationship. For example, diabetes status [24], diet [22], total caloric intake [25], or the duration of fasting periods [26] may be associated both with BMI and with microbiome composition. Thorough and standardized collection of host physiological data is needed to evaluate the contribution of these variables to the relationship between BMI and the composition of the intestinal microbiome. Another possibility is that technological issues may have differential effects across studies [14,23]. Finally, it may be the case that the microbiome's effect on obesity is not mediated through its taxonomic composition but rather its function, since closely related taxa can have widely varying functions and distantly related taxa can have similar functions. This theory is supported by data from Turnbaugh et al. [7] (who found an enrichment of genes involved in carbohydrate and lipid metabolism in obese individuals), but not by results from the HMP Consortium [13] (who found no association between BMI and any pathway abundance). Additional functional metagenomic investigations are needed to determine whether a robust relationship exists between BMI and microbiome function. Once the nature of this relationship is better understood, further epidemiological work will be needed to estimate the proportion of human obesity attributable to microbial factors.

A limitation of our primary analyses is that they were restricted to the healthy subjects of the HMP cohort, none of whom had BMI > 35. For example, we hypothesize that the HMP's health screen may have excluded low-diversity individuals, thus confounding our ability to discern an association between obesity and richness. We note, however, that the Danish MetaHIT cohort included $n = 12$ (17%) subjects with BMI > 35, and it nonetheless revealed no association between obesity and the taxonomic composition of the gut microbiome.

## Methods

We downloaded high-quality, taxonomically annotated Roche V35 16S rRNA gene sequences from the HMP Data Analysis and Coordination Center (www.hmpdacc.org). These are PCR-amplified V35 regions sequenced *en masse* on a Roche 454 instrument. These sequences were previously subject to extensive quality control analyses [15], including trimming, denoising, and chimera filtering. For each sequence, we extracted phylum-level taxonomic annotations and bootstrap statistics via RDP classifier 2.2 [27] using the default 032010 training set and taxonomy and the 'allrank' output format option, as per the HMP SOP. Sequences with annotations having a bootstrap statistic less than 80% were treated as "unclassified". Sequences were then mapped to their corresponding HMP sample identifier and used to calculate phylum-level relative abundance for each sample. For subjects with multiple stool samples, since longitudinal BMI trajectory data were not available, we analyzed the sample with the largest number of reads. Out of a total of 217 available samples, we excluded five samples with < 1000 reads, bringing our total sample size down to 212 samples. The same approach was used to obtain HMP V13 16S rRNA gene sequences.

We tested for associations between arcsin square root genus relative abundance and all additional quantitative phenotypes from the HMP. We found no significant associations (FDR-adjusted $p < 0.2$) in any body site, consistent with the low number
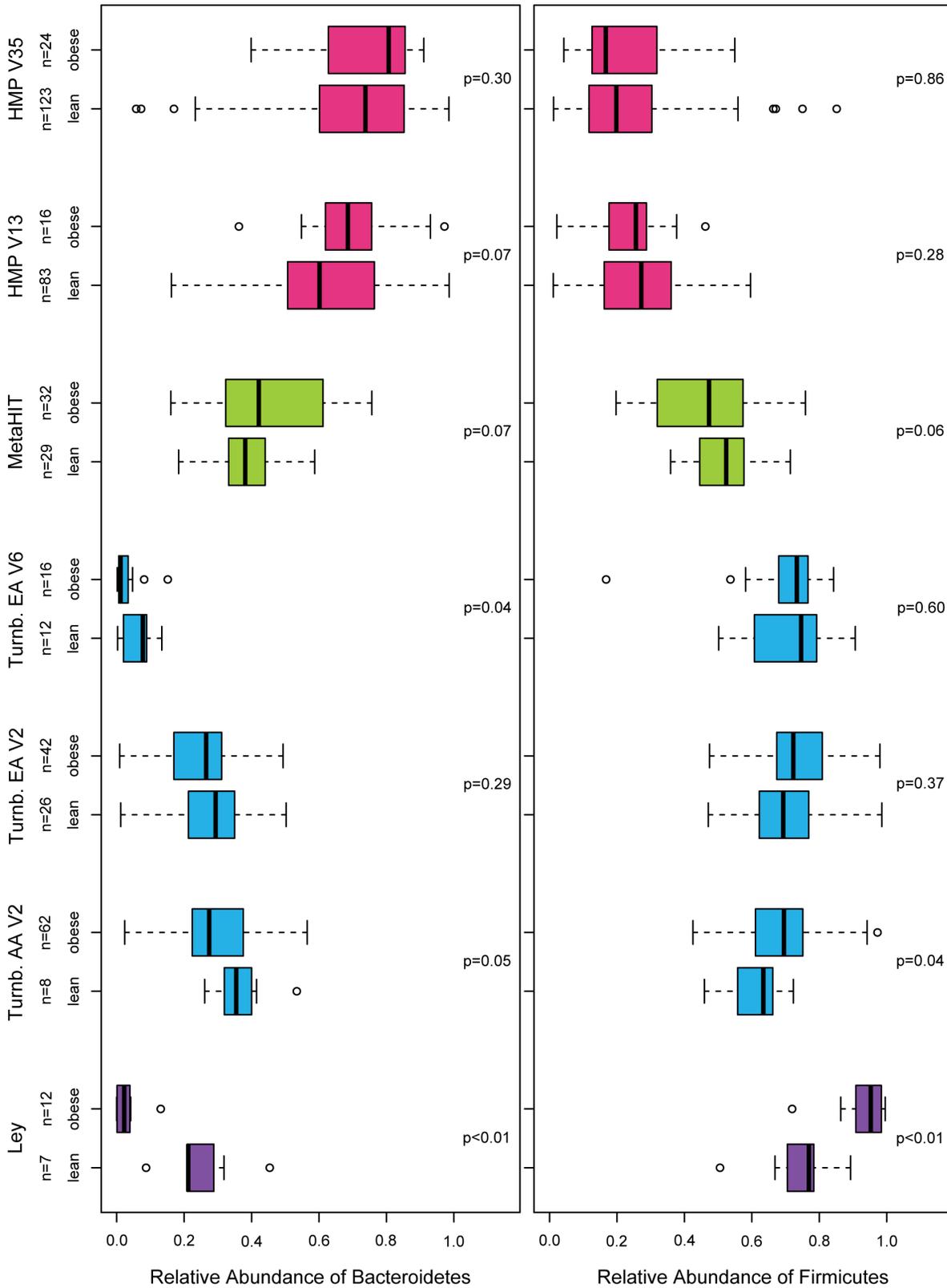
**Figure 3. The between-study variability in the relative abundance of *Bacteroidetes* and *Firmicutes* is greater than the within-study differences between lean and obese individuals.** The Ley data are from [6]. The "Turnb." data are from Turnbaugh et al. [7], from African Americans (AA) and European Americans (EA), from variable regions (V) 2 and 6. The MetaHIT data are from the Danish subjects in [16] who do not have inflammatory bowel disease. The HMP data are from V13 and V35. We note that the primary results from this manuscript were generated using data from HMP V35. All p-values by *t*-test.

doi:10.1371/journal.pone.0084689.g003

of microbe-phenotype associations found in [13] and supportive of the validity of our primary findings regarding BMI.

The MetaHIT data was composed of shotgun metagenomes, which we downloaded from The European Bioinformatics Institute (www.ebi.ac.uk). We restricted our analysis to the Danish samples that contained read-lengths of at least 75 base pairs (bp) ($n = 70$), and we randomly analyzed 20 M reads from each sample. We quantified microbiome phylum-level diversity from these samples by using the STAP database [28] and the GreenGenes database [29] to identify metagenomic homologs of the 16 S locus and the RDP classifier to taxonomically annotate these sequences. We conducted a statistical simulation to identify the optimal bootstrap statistic thresholds for classifying 16 S RNA metagenomic reads into phyla using the RDP classifier. Briefly, we used Grinder [30] to simulate 10,000 75-bp 16 S reads from the STAP database and the GreenGenes database, plus 90,000 75-bp reads from the coding sequences (CDS) of 11 bacterial genomes randomly selected from the J. Craig Venter Institute?s Comprehensive Microbial Resource database [31]. All reads were subject

to classification using the RDP classifier. We found that a bootstrap threshold of 70% captured 83% of the simulated 16 S sequences and correctly classified 99% of these reads while filtering out all but 0.001% of CDS reads.

Publicly available, high-quality 16 S amplicon sequences generated as part of the Ley [6] and Turnbaugh [7] studies were downloaded and taxonomically annotated. Specifically, we downloaded 18,348 full-length 16 S assembled from shotgun sequences generated in [6] from gordonlab.wustl.edu/microbial_ecology_human_obesity/and 817,942 V6 hypervariable regions and 1,119,519 V2 variable regions of 16 S RNA pyrosequencing sequences generated in [7] from gordonlab.wustl.edu/NatureTwins_2008/. All sequences were classified into phyla using the RDP classifier as described above.

## Author Contributions

Conceived and designed the experiments: MMF TJS KSP. Analyzed the data: MMF TJS TL. Wrote the paper: MMF TJS TL KSP.

## References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, et al. (2013) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. The Lancet 380: 2224–2260.
2. Douketis J, Macie C, Thabane L, Williamson D (2005) Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. International Journal of Obesity 29: 1153–1167.
3. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, et al. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nature 444: 1027–131.
4. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI, et al. (2008) Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host & Microbe 3: 213–223.
5. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, et al. (2005) Obesity alters gut microbial ecology. Proceedings of the National Academy of Sciences, USA 102: 11070–11075.
6. Ley RE, Turnbaugh PJ, Klein S, Gordon JI, et al. (2006) Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022–1023.
7. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. (2009) A core gut microbiome in obese and lean twins. Nature 457: 480–484.
8. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. (2013) Richness of human gut microbiome correlates with metabolic markers. Nature 500: 541–546.
9. Knights D, Costello EK, Knight R (2011) Supervised classification of human microbiota. FEMS Microbiology Reviews 35: 343–359.
10. Sweeney TE, Morton JM (2013) The human gut microbiome: A review of the effect of obesity and surgically induced weight loss. JAMA Surgery: 1–7.
11. Zimmer C (2010) How microbes defend and define us. The New York Times, 13 July, Page D1.
12. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. (2011) Enterotypes of the human gut microbiome. Nature 473: 174–180.
13. The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486: 207–214.
14. Ley RE (2010) Obesity and the human microbiome. Current Opinion in Gastroenterology 26: 5–11.
15. The Human Microbiome Project Consortium (2012) A framework for human microbiome research. Nature 486: 215–221.
16. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464: 59–65.
17. Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, Ltd.
18. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, et al. (2013) Gut microbiomes of malawian twin pairs discordant for kwashiorkor. Science 339: 548–554.
19. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75: 7537–7541.
20. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) Qiime allows analysis of high-throughput community sequencing data. Nature methods 7: 335–336.
21. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, et al. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. Nature Methods 9: 811–814.
22. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, et al. (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Science Translational Medicine 1: 6ra14.
23. Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, et al. (2013) Meta-analyses of studies of the human microbiota. Genome research 23: 1704–1714.
24. Graessler J, Qin Y, Zhong H, Zhang J, Licinio J, et al. (2012) Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inammatory and metabolic parameters. The Pharmacogenomics Journal: 1–9.
25. Schwiertz A, Taras D, Schafer K, Beijer S, Bos NA, et al. (2010) Microbiota and SCFA in lean and overweight healthy subjects. Obesity 18: 190–195.
26. Crawford PA, Crowley JR, Sambandam N, Muegge BD, Costello EK, et al. (2009) Regulation of myocardial ketone body metabolism by the gut microbiota during nutrient deprivation. Proceedings of the National Academy of Sciences, USA 106: 11276–11281.
27. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. Applied and environmental microbiology 73: 5261–5267.
28. Wu D, Hartman A, Ward N, Eisen JA (2008) An automated phylogenetic tree-based small subunit rrna taxonomy and alignment pipeline (stap). PloS one 3: e2566.
29. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimerachecked 16s rrna gene database and workbench compatible with arb. Applied and environmental microbiology 72: 5069–5072.
30. Angly FE, Willner D, Rohwer F, Hugenholtz P, Tyson GW (2012) Grinder: a versatile amplicon and shotgun sequence simulator. Nucleic acids research 40: e94–e94.
31. Davidsen T, Beck E, Ganapathy A, Montgomery R, Zafar N, et al. (2010) The comprehensive microbial resource. Nucleic acids research 38: D340–D345.