# Introduction

Obesity is defined as excessive fat accumulation which may present additional health risks that could lead complications like metabolic syndrome[1, 2]. Obesity is commonly classified by body mass index (BMI), a crude measure of weight status[1]. Obesity is the second leading cause of preventable death in the United states, after tobacco use, according to the National Institute of Health[3]. Obesity is estimated to cause 300, 000 deaths annually[3]. Obesity has doubled worldwide since the 1980’s, according to the 2014 Global Status Report from the World Health Organization[4]. About 2 billion adults worldwide were overweight (BMI≥25kg/m2) and more than 600 million were obese (BMI ≥30kg/m2) in 2014, while 42 million children under the age of 5 years were overweight or obese in 2013[4]. The United States has the highest prevalence of obesity with about 70% of adults being considered overweight, including 36% of obese adults[5, 6]. Obesity is also one of the most important global public health concerns due to not only its prevalence but also due to numerous studies having reported excess body weight as a known risk factor associated with mortality from cancer and chronic disease such as diabetes, fatter liver disease and cardiovascular disease[7, 8]. China has the lowest prevalence of obesity (5.2% in 2010) however, the absolute number of obese people in China exceeds that of the United States[9].

Aside for its health risks, obesity presents with major costs to healthcare. In 2008, the annual medical cost was estimated to be 147 billion dollars[10]. Those who were obese had medical costs that averaged 1,429 dollars more than those of normal weight[10]. In addition to healthcare costs, the cost of job absenteeism associated with obesity costed 4.3 million dollars annually along with lower productivity at work costing employers 506 dollars per obese person per year[10].

The fundamental cause of obesity is excessive caloric intake, that is also accompanied by a sedentary lifestyle[11]; a trend that is apparent in global societies that are moving towards “the Western model”[2]. Other contributing risk factors include a genetic predisposition to obesity[12], epigenetics[13], increasing maternal age[14], sleep deprivation[15], endocrine disruptors[16] and pharmaceutical iatrogenesis[16]. Many studies have also linked microbiota to obesity. Gut microbiota, which constitutes a collection of microorganisms that inhabits human intestines, play an intricate and vital role in the physiological process of the human body, including digestion and metabolism[17]. Microbiota increase energy production from the diet, induce low-grade inflammation, regulate the fatty acid tissue composition as well as control the appetite through the gut-brain axis[17]. Hence, they are key players in the development of obesity and metabolic disorders.

The literature contains an abundance of studies that relate the contribution of microbiota to obesity that have been conducted on animal models. Data collected from animal models have shown that after germfree mice that are genetically resistant to obesity, when introduced to certain gut microbes, increased their calorie uptake and accumulated fat and developed insulin resistance[18]. Another study has shown when mice were colonized with gut microbiota from obese mice, they were able to harvest energy more efficiently and increased the rate of body fat accumulation[19].

The studies that have been conducted on the human populations, through the use of 16 rRNA sequencing, have found statistically significant associations between gut microbiota and obesity. A strong and consistent taxonomic signature of obesity was identified in a large cross-sectional study of American Adults[11]. Studies have also found alterations in the gut microbiota in adolescent[20, 21] and obese children[22]. It has also been found that the western diet influences the growth of microbial organisms that contribute to obesity[23]. It was found that a population in the midst of westernization compared to current industrialize countries, had differences in gut microbiota that are associated with obesity; suggesting that the gut microbiota has evolved along with changing environments[23].

Currently, there are limited data conducted from studies using the human population[17]. Also, currently, researchers are producing a vast amount of knowledge regarding microbes and their association with disease such as obesity. The difficulty is that a standard of demonstrating results of these microbes does not exists, and often illustrations are not intuitive and difficult to comprehend. Through the process of manually curating the metadata such as age, BMI, disease status, sex, country of residence alongside microbial “signatures” onto a standardized database, this study will attempt to discover patterns of obesity in relation to the gut microbiota. **The knowledge that treating patients with the right microbiome may cure a multitude of disease such as cancer alongside the significant drop in cost in genome sequencing has dramatically expanded the efforts in microbiome research**. The Waldron lab at City University of New York, School of Public Health (CUNY SPH) in collaboration with colleagues at the University of Trento in Italy, are further expanding on building the body of knowledge in relation to microbiome research through the contributions of CuratedMegatenomic Data, a curated database of processed whole-metagenome shot gun sequencing data of the human microbiome. The purpose of this project is to contribute to the effort of unveiling multiple links to the microbiome and human health as the microbiome field establishes itself. The outcome of this will identify common microbiome signatures that have been reported to be differentially abundant in various studies in relation to the obesity global epidemic.

Researchers are attempting to change the way medicine is practiced by investigating if bacteria can help treat human diseases ranging from inflammatory bowel disease, diabetes, autism, cancer, and aids

**Methods**

*Types of Studies*

This systematic review consisted of observational studies including cross-sectional and case-control studies. Longitudinal studies were not included because they were studies of dietary and surgical interventions which resulted in lean status. The differential abundance of the microbiome in the stool collections after the interventions are not representative of obesity vs. non-obese status but rather increased caloric intake; a confounder of obesity. The interventions removed the confounder by including a very low-calorie diet or a decreased caloric intake due to bariatric surgery, laparoscopic sleeve gastrectomy or gastric bypass surgery.

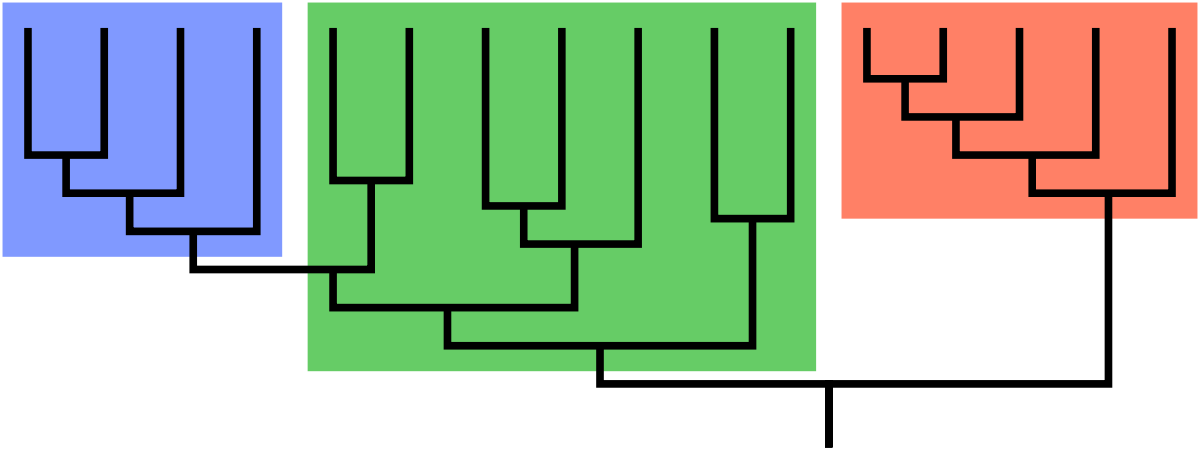
*Participants/Population*

Participants included in the study consisted of adult, pediatric and adolescent populations. The range in age was from 3yrs up to 75 yrs. Studies of infants or participants under the age of 3yrs were excluded due to the difference of the microbiome composition in comparison to adults. By the age of three years the microbiome stabilizes and is comparable to the adult microbiome. In order to assess microbiota associated with obesity status, studies eligible for inclusion assessed differential microbiota from categories of Obese vs. normal weight and/or lean individuals. An example of a cladogram (family tree) is shown in figure 1.

*Microbiome Measures*

The microbial measures in this systematic review includes a summary of the taxa, groups of one or more populations of an organism or organisms that form a unit. More specifically, the reported taxa in microbiome studies are monophyletic taxa or clades, which are lineal descendents of a common ancestor. There are 8 main hierarchical levels of naming the microbiota or taxa in the order of Kingdom, phyla, class, order, family, genus and species.

Figure 1:



In this figure of a Cladogram (family tree), the blue and red subgroups are clades or monophyletic (complete) groups, each showing its common ancestor stemmed at the bottom of the subgroup branch. The green subgroup is not a clade; it is a paraphyletic group, an incomplete clade because it excludes the blue branch despite that it has also descended from the common ancestor stem at the bottom of the green branch. However, the blue and green subgroup together form a clade again.

Relative abundance describes the proportion of a bacteria of a particular species that exists relative to the total number of bacteria in the same community. This tells us how common or rare a species is in the population relative to other species. Species richness is used to quantify the number of species within a biological community, and species abundance is the number of individuals per species. Relative abundance is a significant measure because two different communities may be equal in richness of species that are present but one community may contain species that are equally common and another community one species may significantly outnumber the other species. This difference in relative abundance of one species among two communities or sample groups is known as differential abundance.

This is a systematic review comparing studies that have reported differentially abundant bacterial clades at any taxonomic level in obese vs. non-obese populations. The relative abundance of microbiota are measured in obese and non-obese individuals and compared. Various statistical methods were applied to detect any statistically significant difference in the differential abundance of microbiota. Any taxa that were found to have been significantly increased or decreased in obese groups were curated onto a google spreadsheet. The google spreadsheet validated all manual entries and was used to help assist in organizing and preserving the metagenomic data relating to the studies from the studies that have reported taxa to have been significantly differentially abundant.

Relative abundance is a significant outcome measure in microbiome studies because it is theorized that the richness or diversity of gut bacterial clades conforms to various distinct patterns seen in different medical/psychiatric conditions. Thus, a description of relative abundance is used to quantify how evenly distributed species are in different study groups. Thus, differential abundance is significant in being able to distinguish the distinct patterns of the microbiota that exists in the human gut under various medical conditions. This review summarizes the distinct patterns of bacterial taxa which have been more frequently reported UP or DOWN in studies of differential abundance of the microbiome in obesity.

Alpha diversity is a measure that expresses the average number of species in a specific body site or subsite. This systematic review includes this measure to account for the qualitative differences in alpha diversity observed in obese populations in comparison to normal weight or lean individuals. Alpha diversity was reported to be either decreased, increased or remained the same. Alpha diversity is accounted for in this systematic review because human disease is not only associated with dysbiosis but also an overall loss of microbial diversity in the gut microbiota. Alpha diversity was measured in various studies with indices calculated from the use of Pielou, Shannon, Chao1, Simpson, inverse Simpson or richness.

As already mentioned, species richness describes how diverse the population is in terms of the number of species present within a community. The pielou index is a derivative of the Shannon diversity index, and it describes evenness of a community. Both the Shannon and Simpson index account for both abundance and evenness of the species present and is used to characterize species diversity in a community. However, the Simpson index gives more weight to common or dominant species present, whereas the Shannon index assumes all species are represented in a sample and is assumed that they are randomly sampled.

Mathematically the Shannon index is represented by the following formula:

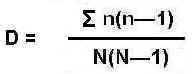
**s**

**H = ∑ - (Pi \* ln Pi)**

**i=1**

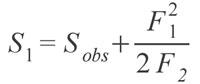
Where Pi is the fraction of the entire population made up of species i, S is the number of species encountered, ∑ is the sum from species 1 to species S. A high H value, would represent a more diverse community, and a value of 0, would represent a community with only one species. If species are evenly distributed then the H value would be high, therefore also allowing us to know how the abundance of the species is distributed among all the species and at the same time telling us the number of species of the community.

Mathematically the Simpson index is represented by the following formula:

,

Where *n* is the total number of organisms of a particular species, and *N* is the total number of organisms of all species. The value D ranges between 0 1nd 1; 0 represents infinite diversity and 1, no diversity. A larger D value represents a lower diversity. Diversity is then calculated giving the *Simpsons Index of diversity*, obtained from *1-D*, where the value ranges from 0 and 1, the greater the value the greater the sample diversity. The reciprocal of Simpson's index, 1/D gives a value that starts with 1 as the lowest possible figure, representing a community with only one species. The higher the value the greater the diversity. A sample with 10 species would give the maximum value of 10.

Chao1 is an alpha diversity measure which measures the true species diversity by considering the number of rare species that are found in a sample to calculate how likely it is that there are more undiscovered species. In this way, it can estimate the true species diversity of a sample. This is done by the following formula:

,

where *Sobs* is the number of species in the sample, *F1* is the number of singletons (i.e., the number of species with only a single occurrence in the sample) and *F2* is the number of doubletons (i.e., the number of species with exactly two occurrences in the sample). The number of rare species that are found in a sample determine how likely that there are more undiscovered species. This formula allows to include species that are rare (singletons) still being discovered because as soon as all species have been recovered at least twice (doubletons), there is likely no more species to be found.

**Outcome Measures**

The frequency of taxonomic signatures were measured amongst the outcome obesity. Traditionally, a person is considered obese if they are more than 20% over their ideal body weight. However, there are additional factors that must be taking into consideration such as a person’s height, age, sex and build. The National Institute of Health (N.I. H,) has redefined obesity using a Body Mass Index (B.M.I.). This why obesity was defined by the included studies by their BMI, however, rarely did studies include waist size, a more accurate measure of obesity status. BMI is an index which relates the body weight in kilograms (kg) to height, in meters (m) divided by the height in meters (m) squared. The conventional BMI measure of obesity is a BMI that is approximately 30 and above. Studies in some of the countries, particularly Asian countries had a different BMI measure of obesity. Children of included studies were classified as obese according to a percentile of body weight distribution in the population.

**Measures of Association**

The measures of association were based on fold change, Spearman’s correlation and Pearson’s correlation.

Fold change in microbiome analysis is the ratio in the change of microbiota abundance of individuals with obese BMI in respect to individuals with non-obese BMI.

Fold change is a measure commonly used in the analysis of gene expression data for measuring the change in the expression of a gene, a measure suitable for microbiome analysis, which makes use of 16s gene sequencing with the use of various sequencing platforms.

Spearman’s correlation coefficient is a nonparametric measure of rank correlation and was used to measure the degree of similarity between BMI and microbiota. In this way, the test was able to determine to what degree microbiota was significantly related to BMI using a monotonic function.

Pearson’s correlation coefficient between two variables is equal to the Spearman correlation. The difference is that while Spearman’s correlation assesses monotonic relationships, Pearson’s correlation assesses linear relationships.

**Searches**

For my search inquiry of studies potentially eligible for inclusion in the systematic review, Pubmed was searched through February 2019. Through the use of keywords. Boolean operators (AND) were used to combine searches. The search consisted of one search inquiry presented on PubMed: Obesity AND microbiota AND abundance. From this search inquiry, I selected only those that were relevant based on title or abstract. I the discarded any studies that did not meet inclusion or exclusion criteria. I also searched the bibliographies of the selected studies and read through relevant reviews.

**Data Extraction**

Relevant metagenomic and methodological data on studies based differentially abundant microbes in obesity were curated and extracted onto a google spreadsheet, provided by Professor Levi Waldron at CUNY SPH. The data on the microbial clades extracted were checked by using an NCBI taxonomy. An example of a microorganism identified by NCBI taxonomy would be the following: k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Faecalibacterium|s\_\_Faecalibacterium\_prausnitzii.

Where each taxonomical level is represented by the letters k, p, c, o, f, g, and s, where K =Kingdom, p=phylum, c=class, o=order, f=family g=genus, and s=species.

The intensive curation procedure was designed in a way that prevented errors of manual entry. This strengthen the validity of the curation process. The curation sheet contained predefined columns and within most of the columns, there were predefined values, that are commonly used in microbiome research. This structure allowed to check for valid manual entries for each cell in every column.

The curation sheet was built to included searches only through PubMed, with a column to enter the Pubmed ID for every retrieved article. To assist in identifying relevant articles and results of a research topic, the tables and figures of results were entered within the column ‘Source within paper’ and I provided a free-form description of the study in a column adjacent to the source. The significant threshold was entered into a separate column, to ensure that only statically significant results were used from the literature. It was also noted if the threshold was corrected for multiple hypothesis testing (M.H.T.). There are two separate columns to which the size of the population was noted. The Control (unexposed) column and the Case (exposed) columns accounted for the number of individuals in each comparison groups. In reference to the study design, there are predefined columns with predefined values which included case-control, longitudinal, cross-sectional, laboratory experiment and randomized controlled trial (R.C.T.). For each study entered, I listed the confounders that were controlled for, including the exclusion criteria onto a separate predefined column.

A column for Country included all Countries and this helped account for the diversity of a research topic. Columns for body site, body-subsite, and Condition included in each of the cells an extensive list to choose from, however, entries were not limited to this, and I was allowed for free entry of any kind. A column for the sequencing type included the values 16S or whole metagenomic sequencing (WMS), however other sequencing types can be entered. Predefined values were also listed in cells pertaining to the column of the 16S variable region. Regions included V1-V2, V3-V5, V1, V2, V3, V5, V5-V6 or NA (non-applicable) if it was not mentioned in the study. Variable regions are sequences due to its great variability in amino acid sequence among different immunoglobulins of the same class. In other words each individual has their own characteristic amino acid sequence encoding for bacteria. In this way, the human body is no longer defined merely by DNA but also by the millions of microbes that we harbor and co-exist with. Our microbiome is commonly referred to by scientists as our “second genome,” as it uniquely maps each segment of our body.[1] In fact, the microbiome community outnumbers the total number of unique genes by 150times. This “new too” allows for the current systematic review to attempt in summarizing and finding potentially associated of microbial taxons in obesity.  [1]

16S rDNA sequencing is particularly preferred sequencing method used in microbiome research for its sensitivity to the presence of rare bacterial species. It has the capacity to detect rare bacteria, slow growing bacteria, uncultivable bacteria and culture-negative infections in an 0.1% of a mixture. However, WMS is a good alternative to 16S DNA for the analysis of the microbiome. It offers several advantages such that it allows direct inference of the metabolic capacity and physiological features of the metagenome without having to know the genotypes and phenotypes of the bacterial community. It also makes it possible to overcome the problems of 16S sequencing such as unknown copy number of the 16S gene and lack of sufficient sequence similarity of the universal 16S primers to some of the target 16S genes.

A column for sequencing platform included values for Illumina, Roche454, Ion Torrent, and Qiagen and again results were not limited to these types. A column was also dedicated for the date of curation and curator name to keep track of the curation process and progress of each curator.

**Assessment of Validity**

The google spreadsheet used to enter the curated metagenomic data contained columns which were used to collect information regarding the validity of the studies used in this systematic review. A column was used to indicate the significance threshold for which results only with p-values of less than .05 were included. As already mentioned, a column was dedicated to representing studies that have included M.H.T. in their statistical analysis. Many of my studies chosen did, in fact, include M.H.T. which is crucial to the validity of the results that are reported. M.H.T. is performed to correct or adjust for type 1 errors by computing for the probability of obtaining at least 1 false positive in the results. Such MHTs include False Discovery Rate (FDR), Benjamini-Hochberg procedure and Bonferroni tests. Adjusted p-values of less than .05 were included in the spreadsheet, denoting the significance threshold with q, to indicate the adjustment. However, due to the limited research on the microbiome and obesity conducted on humans, I choose to also include studies in this systematic review that were not corrected for M.H.T.

Various statistical methods were used in the studies, for which a column with predefined values for curating only the statistical methods that are commonly used in microbiome research. In addition to Spearman’s and Pearson’s correlation coefficient already mentioned, other statistical tests used in the included studies were the T-test, Kruskall-Wallis, Mann-Whitney (Wilcoxon), DESeq2, and, LEfSe, GLM ANOVA.

The Mann-Whitney (Wilcoxon) test, a non- parametric test of the null hypothesis stating it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample.[2] What makes it efficient is that unlike the t-test it does not require the assumption of a normal distribution, for which microbiome samples do not follow, each has a unique microbiome signature. [2] The results of this test on samples that do not follow a normal distribution are nearly as efficient as a t-test on normal distributions. [2]

The LEfSe (linear discriminant analysis effect size) test, which uses relative abundances to find biomarkers between 2 or more groups. [3] Once results are obtained, they are plotted with bars which represent the effect size for particular taxa in a specific group. The use of the colors red or green represent which group the taxa were found to be more abundant compared to the other group. The absolute values of the effect size are used to get an interpretation of the scale of the difference between 2 groups for a particular taxa. [3]

\*\*\*I plan on expanding on each statistical test in final draft\*\*\*

\*\*\* Will expand on the importance of Sample size, Confounders/mediators controlled for, antibiotics criteria\*\*\*

**Results**

**Literature Search**

General characteristics of the studies are as follows: from the 436 articles retrieved, 54 were selected based on tittle and abstract to be read. After reading the selected articles in depth 16 studies were included. In addition to these 16 articles, in order to add additional valuable data to this review 3 articles were identified from previous systematic review, 1 article was found from a bibliography of the selected articles and 1 parameters

Table I.

Characteristics of Included Studies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Author | Country | Study Design | Groups | N (Participants/obese) | Age Group |
| Haro et al. | United States | Cross-sectional | BMI <30  BMI ≤30≥33  BMI >33 | 75/49 | Adult 20< age <75yrs |
| Chierico et al. | Italy | Case-control | Obese  Normal weight | 69/45 | Adult and Adolescent |
| Riva et al. | Italy | Cross-sectional | Obese  Normal weight | 78/42 | Pediatric |
| Hu et al. | Korea | Cross-sectional | Obese  Normal weight | 134/67 | Adolescent |
| de la Cuesta-Zuluaga et al. | Colombia | Cross-sectional | Obese  Overweight  Normal weight | 441/132 | Adult 18-62yrs |
| Peters et al. | United States | Cross-sectional | Obese  Overweight  Normal weight | 599/142 | Adult 18-86yrs |
| Gao et al. | China | Cross-sectional | Gender stratified  Underweight: BMI <18.5  Normal weight: BMI 18.5-23  Overweight: BMI 23-27.5  Obese: BMI ≥27.5 | 489/58 | Adult |
| Lopez-Contreras et al. | Mexico | Cross-sectional | Obese  Normal weight | 138/71 | Children 6-12yrs |
| Chavez-Carbajal et al. | Mexico | Case-control | Obese  Obese + metabolic syndrome (OMS)  Normal Weight | No Obesity-MS: 42/17  Obesity +MS: 50/25 | Adult Women |
| Verdam et al. | Netherlands | Cross-sectional | Obese  Overweight  Normal weight | 23/15 | Adult |
| Million et al. | France | Case-Control | Obese  Normal Weight | 115/68 | Adult |
| Borgo et al. | Italy | Case-Control | Obese  Normal Weight | 61/28 | Pediatric |
| Wu Y et al. | China | Case-Control | Obese  Normal Weight | 62/33 | Adult |
| Graessler et al. | Germany | Case-control | Morbidly obese BMI >40w/comorbidity | 12/6? | Adult <60 yrs. |
| Kasai et al. | Japan | Cross-sectional | Lean BMI<18.5kg/m2  Normal Weight BMI18.5-25kg/m2  Obese BMI ≥25 kg/m2 | 108/33 | Adults <65 yrs. |
| Michail et al. | United States | Case-control | Obese  Normal Weight | 37/11 | Pediatric |
| Karvonen et al. | United States | Cross-sectional | Overweight/obese  Normal weight | 502/146 | 3 yrs. Old |
| Andoh et al. | Japan | Case-Control | Obese  Lean | 20/10 | Adult |
| Wang et al. | China | Case-control | Obese before SG  Obese before RYBG  Normal Weight | 39/19  27/7 | Adult (18-65 yrs.) |
| Liu et al. | China | Case-control | Obese  Lean | 151/72 | Adult (18-30yrs) |
| Kashtanova et al. | Russia | Cross-sectional | Obese  Abdominal Obese  Normal Weight | 92/23  92/53 | Adult (25-76 yrs) |

Table II.

Descriptive Results of Studies Comparing Obesity to Control groups

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Author | **Case definition of Obesity** | **Confounders Control** | **Sequence Platform** | **Sequencing type** | **Statistical Test** | | **Corrected for MHT** | | **16S Variable Region** | | **Change in Alpha Diversity in Overall Obese Population vs. Controls** | | **Number of Differential Microbes Reported in Overall Obese vs. Controls by Statistical Method** | |
| Haro et al. | BMI ≥ 30 kg/m2 | Diet , age, gender | Illumina | 16S | T-test | | YES | | V4 | | No | | 14 | |
| Chierico et al. | 30≥BMI≤60 kg/m2 | Age matched;  Antibiotics; cases and controls were subjected to an extensive list of exclusions Cases: antibiotics, pre-biotics, helicobacter pylori, corticosteroids, vitamin E, fish oil, Chronic G.I. disease, bariatric surgery. Controls: anti-biotic, pre-biotic, Chronic disease, G.I. infection, omnivore diet | Roche454 | 16S | Kruskall-Wallis | YES | | V1-3 | | Decreased | | 54 | |
| Riva et al. | BMI z-score: -2.14-5; p<0.0001 | Antibiotics, probiotics, neonatal disease, congenital malformation, chronic or acute intestinal and obesity-related co morbidity conditions | Illumina | 16S | Pearson Correlation | | Yes | | NA | | No Change | | 12 | |
| Hu et al. | BMI ≥30 kg/m2 or ≥99th BMI percentile | gender and age checked at baseline; Antibiotics | Qiagen | 16S | Mann-Whitney (Wilcoxon), | | Yes | | V1-3 | | No change | | 11 | |
| de la Cuesta-Zuluaga et al. | BMI >30kg/m2 | Enrolled in similar proportions by city, sex and age; excluded BMI < 18kg/m2, pregnant women, neurodegenerative disease, current or recent cancer and G.I. disease | Illumina | 16S | Spearman Correlation | | Yes | | V4 | | NA | | 41 | |
| Peters et al. | BMI ≥ 30 kg/m2 | Race, age, polyp status ad study adjusted; antibiotics in NYU study; NYU and CDC study included extensive list of exclusions\* | Illumina | 16S | DESeq2 | | Yes | | V4 | | Decreased | | 34 | |
| Gao et al. | BMI ≥ 27.5 kg/m2 | adjusted for sex and age; antibiotics, diabetes, chronic diarrhea or constipation, long medication use | Illuminan | 16S | Mann-Whitney (Wilcoxon) | | Yes | | V3-V4 | | No change | | 1 | |
| Lopez-Contreras et al. | BMI ≥95th perentile | Antibiotics, <10% body weight loss, diarrhoea, acute G.I. illness | Illumina | 16S | DESeq2 | | Yes | | V4 | | No change | | 5 | |
| Chavez-Carbajal et al. | BMI ≥30 kg/m2 | Antibiotics, chronic diseases, smoking, pregnancy, allergies, thyroid disease, eating disorders, supplementations, ACD | Ion Torrent | 16S | Kruskall-Wallis | | Yes | | V3 | | Increased | | 24 | |
| Verdam et al. | BMI ≥ 30.5 kg/m2 | Antibiotics, anti-inflammatory drugs, alcohol consumption >80 ml/week | NA | 16S | Spearman | | yes | | NA | | decreased | | 27 | |
| Million et al. | BMI ≥30 kg/m2 | Age, antibiotics; Controls subjected to an extensive list of exclusions including antibiotics | Qiagen | qPCR | Mann-Whitney  (Wilcoxon) | | No | | Tuf gene | | NA | | 5 | |
| Borgo et al. | WHO criteria BMI >30kg/m2 | Age & sex matched. Participants subjected to an extensive list of exclusions | Ingeny | 16S | Mann-Whitney  (Wilcoxon) | | No | | V2-V3 | | NA | | 7 | |
| Wu Y et al. |  | Antibiotics & BMI | Illumina | 16S | Mann-Whitney (Wilcoxon) | | Yes | | V3-V4 | | Incrased | | 32 | |
| Graessler et al. | Criteria for bariatric surgery (BMI>40kg/m2) | Fulfilled criteria for bariatric surgery | Illumina | WMS | ANOVA | | Yes | | NA | | NA | | 3 | |
| Kasai et al. | BMI ≥25 kg/m2 | Antibiotics; correlation with microbiota checked for age and HBA1c at baseline | Illumina | 16S | Mann-Whitney (Wilcoxon) | | No | | V3-V4 | | Increased | | 11 | |
| Michail et al. | BMI > 95% for age | Antibiotics caloric intake, carbohydrate, protein or fat intake checked at baseline | Ion Torrent | 16S shotgun | ANOVA | | Yes | | NA | | NA | | 7 | |
| Karvonen et al. | BMI >85th percentile | Antibiotics  models adjusted for maternal education, diversity, age | Illumina | 16S | Mann-Whitney (Wilcoxon) | | Yes | | V4 | | No change | | 4 | |
| Andoh et al. | BMI >30Kg/m2 | Antibiotics  medical treatment, drugs and supplement use | Illumina | 16S | t-test | | No | | V3-V4 | | Decreased | | 28 | |
| Wang et al. | BMI >28 Kg.m2 | Antibiotics  Age and sex | Ion S5 XL | 16S | LEfSe | | No | | V4 | | No change | | 10 | |
| Liu et al. | BMI >28 Kg/m2 | Antibiotics Age and sex matched; no food with probiotics 7 days before stool collections | Illumina | Metagenomic | Mann-Whitney (Wilcoxon | | Yes | | NA | | Decreased | | 26 | |
| Kashtanova et al. | BMI >30 Kg/m2 | Antibiotics  participants were subjected to an extensive list of exclusions | Illumina | 16S | Generalized linear model | | Yes | | V3-V4 | | NA | | 3 | |
|  |  |  |  |  |  | |  | |  | |  | |  | |

Table III

|  |  |
| --- | --- |
| Most frequently Identified Taxa |  |
| Taxon | **Frequency** |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Faecalibacterium|s\_\_Faecalibacterium\_prausnitzii | 11 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides | 7 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Prevotellaceae|g\_\_Prevotella | 7 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Lachnospiraceae | 7 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Clostridiaceae | 6 |
| k\_\_Bacteria|p\_\_Actinobacteria | 5 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_caccae | 5 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Bacilli|o\_\_Lactobacillales|f\_\_Streptococcaceae|g\_\_Streptococcus | 5 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Oscillospira | 5 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Actinomycetales|f\_\_Actinomycetaceae|g\_\_Actinomyces | 4 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Actinomycetales|f\_\_Propionibacteriaceae|g\_\_Propionibacterium|s\_\_Propionibacterium\_acnes | 4 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_ovatus | 4 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Rikenellaceae|g\_\_Alistipes | 4 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_Barnesiellaceae | 4 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Bacilli|o\_\_Bacillales|f\_\_Gemellaceae | 4 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Faecalibacterium | 4 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Negativicutes|o\_\_Selenomonadales|f\_\_Veillonellaceae|g\_\_Veillonella|s\_\_Veillonella\_parvula | 4 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Coriobacteriales|f\_\_Coriobacteriaceae|g\_\_Adlercreutzia | 3 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Coriobacteriales|f\_\_Coriobacteriaceae|g\_\_Collinsella|s\_\_Collinsella\_aerofaciens | 3 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_uniformis | 3 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_vulgatus | 3 |
| k\_\_Bacteria|p\_\_Firmicutes | 3 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia | 3 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Lachnospiraceae|g\_\_Coprococcus | 3 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Lachnospiraceae|g\_\_Dorea|s\_\_Dorea\_formicigenerans | 3 |

Table IV. Taxa most frequently found UP vs. DOWN

|  |  |
| --- | --- |
| Taxon | UP |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Prevotellaceae|g\_\_Prevotella | 6 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Faecalibacterium|s\_\_Faecalibacterium\_prausnitzii | 6 |
| k\_\_Bacteria|p\_\_Actinobacteria | 5 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Bacilli|o\_\_Lactobacillales|f\_\_Streptococcaceae|g\_\_Streptococcus | 5 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Lachnospiraceae | 4 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Actinomycetales|f\_\_Propionibacteriaceae|g\_\_Propionibacterium|s\_\_Propionibacterium\_acnes | 4 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Clostridiaceae | 4 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Negativicutes|o\_\_Selenomonadales|f\_\_Veillonellaceae|g\_\_Veillonella|s\_\_Veillonella\_parvula | 4 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Actinomycetales|f\_\_Actinomycetaceae|g\_\_Actinomyces | 3 |
| k\_\_Bacteria|p\_\_Actinobacteria|c\_\_Actinobacteria|o\_\_Coriobacteriales|f\_\_Coriobacteriaceae|g\_\_Collinsella|s\_\_Collinsella\_aerofaciens | 3 |

Table V.

Taxa most frequently found DOWN vs. UP

|  |  |
| --- | --- |
| Taxon | DOWN |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides | 7 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Faecalibacterium|s\_\_Faecalibacterium\_prausnitzii | 5 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_caccae | 4 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_Barnesiellaceae | 3 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Clostridia|o\_\_Clostridiales|f\_\_Ruminococcaceae|g\_\_Oscillospira | 3 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae|g\_\_Bacteroides|s\_\_Bacteroides\_uniformis | 3 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Rikenellaceae|g\_\_Alistipes | 3 |
| k\_\_Bacteria|p\_\_Firmicutes|c\_\_Erysipelotrichia|o\_\_Erysipelotrichales|f\_\_Erysipelotrichaceae | 3 |
| k\_\_Bacteria|p\_\_Verrucomicrobia|c\_\_Verrucomicrobiae|o\_\_Verrucomicrobiales|f\_\_Akkermansiaceae|g\_\_Akkermansia|s\_\_Akkermansia\_muciniphila | 2 |
| k\_\_Bacteria|p\_\_Bacteroidetes|c\_\_Bacteroidia|o\_\_Bacteroidales|f\_\_Bacteroidaceae | 2 |

**Figures**

Figure 2: PRISMA figure



Citations identified from electronic searches on PubMed

N=436

Citations clearly not relevant based on abstract and/or tittle

(n=382)

Potentially relevant articles retrieved for detailed examination

(n=54)

Articles excluded (systematic review, irrelevant case comparison, pregnant women, scarce representation of bacteria)

(n=

Articles with events identified from previous systematic review

(n=3)

Articles found from Microbial Signatures curation spreadsheet

(n=1)

Articles found from bibliography of selected articles

(n=1)

**Hypothesis test**: taxa identified more frequently than expected by chance? Have to think about how.

**Hypothesis test**: chi-squared test or odds ratio w/ confidence interval for commonly identified taxon identified (yes/no) vs:

* outcome measure (BMI/obesity)
* study quality measures

**Hypothesis test**: Wilcoxon rank-sum test between # of taxa identified and study quality measures

The text of the results should describe the tables and figures in more detail. Do not discuss the implications of the results yet (that is part of the conclusions and discussions).

**Conclusion**

In this systematic review, 17 studies on the differential abundance of the microbiome in relation to obesity were recovered from Pubmed with the use of various Boolean operators. The study conducted was able to distinguish 10 taxa that were most frequently found UP vs. DOWN and 10 taxa that were most frequently found DOWN vs UP. According to the literature there was both consistencies and discrepancies of the most frequently identified taxa in terms of the direction of their relative abundance identified in obese populations.

“*Bacteroidetes*(Gram negative), *Firmicutes* (Gram positive) and *Actinobacteria* (Gram positive), are most abundant and have been found to play a dominant role in the pathophysiology of metabolic disorders - specifically, obesity”

Amongst the taxa that were most frequently identified in the UP direction in comparison to the DOWN direction in obesity groups include at the species level Faecalibacterium prausnitzii, Propionibacterium acnes, Veillonella parvula and Collinsella aerofacins. Taxa frequently found UP at the genus level include Streptococcus, Prevotella, and Actenomyces. At the phylum level only Actinobacteria was frequently found to be increased in obese populations.

Amongst the taxa that were most frequently identified in the DOWN direction in comparison to the Up direction in obesity groups include at the species level Bacteroides caccae, Bacteroides plebeius, and Faecalibacterium prausnitzii. In the genus level, Oscillospira and Bacteroides were most frequently found DOWN. Taxa frequently found DOWN at the family level include Barnesiellaceae, Rikenellaceae, Erysipelotrichaceae, and Bacteroidaceae. At the phylum level, Bacteroidetes was most frequently found down.

There is inconsistency that exists in the microbiome literature regarding the relative abundance of firmicutes to bacteroidets phyla ratio upon comparing the gut microbiota of the obese to lean subjects. population. Numerous studies described an increase firmicutes and a decreased Bacteroidetes contributing to a high ratio of firmicutes to Bacteroidetes. However other studies have found a decreased Bacteroidetes/firmicutes ratio while others have reported no difference. In fact a meta-analysis that was previously conducted across multiple studies concluded that there were no statistically significant differences of the Firmicutes/Bacteroidetes ration between obese and normal-weight adults.

Despite the inconsistency that is observed in the literature regarding these two phyla, it is interesting to observe that the curation of the metagenomic data of the microbiome in obese subjects revealed that half (5 out 10) of bacterial clades that were reported UP within the obese populations were descendants of the Firmicutes phyla. In particular, Faecalibacterium prausnitzii, was found to be the most frequent taxa that differentially increased at a frequency of 7 times in the obese groups. This is consistent to what has been described of this taxa in the literature as it is the most common species in the gastrointestinal tract of adults consuming a western diet and it has been found to be associated with failure in dietary weight loss. Streptococcus, a descendant of the firmicutes phyla was also found to appear to have been frequently increased in obese populations and this observation is also in line with what has been reported in the literature. It was found that obese pregnant women were significantly more likely to be colonized by rectovaginal group B streptococcus when compared to with nonobese pregnant women (28.4% vs. 22.2%, P<0.001) and were also 35% more likely than nonobese women to test positive for group B streptococcus after adjusted for race, smoking and diabetes.

Actinobacteria is also another dominant phylum known to contribute to the pathophysiology of metabolic disorders, and the results of this study do reveal that Actinobacteria was frequently found to increase in the obese population. A higher level of Actinobacteria has been demonstrated to be more prevalent obese persons compared to healthy individuals. Its species descendant, Propionibacterium acnes was also found to be frequently increased in the obese population. This is no unexpected as this bacterium is known to induce an immunological reaction that causes inflammation and is known to induce the expression of proinflammatory cytokines from peripheral blood monocytes. Inflammation is also directly related to diet intake of omga-6 ad omega 3, polyunsaturated fatty acids (PUFA). The western diet is a contributing factor to the obesity epidemic as it maintains a significantly higher concentration of omega 6 PUFA and a lower omega 3 PUFA because of the predominance of omega 6 that appears in most vegetable oils and processed foods of the Western civilizations. In fact the ration has been found to have risen to 10:1, whereas in nonwestern diets it has been to be between 3 to 2:1.

**Discussion**

* discuss causality
* discuss study quality and possible relationship with results (e.g. if study quality measures are associated with identifying certain taxa or overall # of taxa)
* Discuss limitations of the current review (do **not** discuss limitations of individual studies unless those limitations are present across many of the studies included in the review)
  + Some common potential limitations:
    - Narrow definitions of the outcome (e.g. obesity can be measured using a variety of methods and not just BMI or obesity/overweight categories)
    - Missing populations of interest (e.g. all papers reviewed were only from European studies and so Asian/African/American populations are missing)
    - Poor study design (no study is perfect--which is why you’re doing the review--but if you feel that many of the studies included have serious design flaws you should mention how they could be improved)

**Conclusions**

* frequently identified taxa
* consistency / discrepancy by outcome measure

Citations

* *Obesity*. Available from: <https://www.who.int/topics/obesity/en/>.
* 2. Chavez-Carbajal, A., et al., *Gut Microbiota and Predicted Metabolic Pathways in a Sample of Mexican Women Affected by Obesity and Obesity Plus Metabolic Syndrome.* Int J Mol Sci, 2019. **20**(2).
* 3. *Obesity: Facts, Figures, Guidlines* Available from: <https://www.wvdhhr.org/bph/oehp/obesity/mortality.htm>.
* 4. Organization, W.H., *Global Status Report on Noncommunicable Diseases*. 2014. p. 79-95.
* 5. Arnold, M., et al., *Duration of Adulthood Overweight, Obesity, and Cancer Risk in the Women's Health Initiative: A Longitudinal Study from the United States.* PLoS Med, 2016. **13**(8): p. e1002081.
* 6. Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014. **384**(9945): p. 766-81.
* 7. Arnold, M., et al., *Global burden of cancer attributable to high body-mass index in 2012: a population-based study.* Lancet Oncol, 2015. **16**(1): p. 36-46.
* 8. Bhaskaran, K., et al., *Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults.* Lancet, 2014. **384**(9945): p. 755-65.
* 9. Stevens, G.A., et al., *National, regional, and global trends in adult overweight and obesity prevalences.* Popul Health Metr, 2012. **10**(1): p. 22.
* 10. *The Healthcare Costs of Obesity*. Available from: <https://stateofobesity.org/healthcare-costs-obesity/>.
* 11. Peters, B.A., et al., *A taxonomic signature of obesity in a large study of American adults.* Sci Rep, 2018. **8**(1): p. 9749.
* 12. Lu, Y. and R.J. Loos, *Obesity genomics: assessing the transferability of susceptibility loci across diverse populations.* Genome Med, 2013. **5**(6): p. 55.
* 13. Waterland, R.A., *Epigenetic mechanisms affecting regulation of energy balance: many questions, few answers.* Annu Rev Nutr, 2014. **34**: p. 337-55.
* 14. We, J.S., et al., *Effect of Maternal Age at Childbirth on Obesity in Postmenopausal Women: A Nationwide Population-Based Study in Korea.* Medicine (Baltimore), 2016. **95**(19): p. e3584.
* 15. Patel, S.R., et al., *The association between sleep patterns and obesity in older adults.* Int J Obes (Lond), 2014. **38**(9): p. 1159-64.
* 16. Petrakis, D., et al., *Endocrine Disruptors Leading to Obesity and Related Diseases.* Int J Environ Res Public Health, 2017. **14**(10).
* 17. Verdam, F.J., et al., *Human intestinal microbiota composition is associated with local and systemic inflammation in obesity.* Obesity (Silver Spring), 2013. **21**(12): p. E607-15.
* 18. Xiao, L., et al., *High-fat feeding rather than obesity drives taxonomical and functional changes in the gut microbiota in mice.* Microbiome, 2017. **5**(1): p. 43.
* 19. Ussar, S., et al., *Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome.* Cell Metab, 2015. **22**(3): p. 516-530.
* 20. Del Chierico, F., et al., *Gut Microbiota Markers in Obese Adolescent and Adult Patients: Age-Dependent Differential Patterns.* Front Microbiol, 2018. **9**: p. 1210.
* 21. Hu, H.J., et al., *Obesity Alters the Microbial Community Profile in Korean Adolescents.* PLoS One, 2015. **10**(7): p. e0134333.
* 22. Lopez-Contreras, B.E., et al., *Composition of gut microbiota in obese and normal-weight Mexican school-age children and its association with metabolic traits.* Pediatr Obes, 2018. **13**(6): p. 381-388.
* 23. de la Cuesta-Zuluaga, J., et al., *Gut microbiota is associated with obesity and cardiometabolic disease in a population in the midst of Westernization.* Sci Rep, 2018. **8**(1): p. 11356.