Different analysis strategies of 16S rRNA gene data generate contrasting views of gut bacterial communities associated with diet, health and obesity

Jose F Garcia-Mazcorro [Corresp., 1] , Jorge R. Kawas [2] , Cuauhtemoc Licona-Cassani [3] , Susanne Talcott [4] , Giuliana Noratto [4]

[1] Research and Development, MNA de Mexico, San Nicolas de los Garza, Nuevo Leon, Mexico

[2] Faculty of Agronomy, Universidad Autónoma de Nuevo León, General Escobedo, Nuevo Leon, Mexico

[3] Department of Bioengineering, Biosystems and Synthetic Biology, Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Monterrey, Nuevo Leon, Mexico

[4] Department of Nutrition and Food Sciences, Texas A&M University, College Station, Texas, United States

Corresponding Author: Jose F Garcia-Mazcorro
Email address: josegarcia_mex@hotmail.com

Background. A main function of diet is to nurture the gut microbiota and this relationship affects host health. However, different analysis strategies can generate different views on the relative abundance of each microbial taxon, which can affect our conclusions about the significance of diet to gut health in lean and obese subjects. Here we explored the impact of using different analysis strategies to study the gut microbiota in a context of diet, health and obesity.

Methods. Over 15 million 16S rRNA gene sequences from published studies involving dietary interventions in obese laboratory rodents were analyzed. Three strategies were used to assign the 16S sequences to Operational Taxonomic Units (OTUs) based on the GreenGenes reference OTU sequence files clustered at 97% and 99% similarity.

Results. Different strategies to select OTUs influenced the relative abundance of all bacterial taxa, but the magnitude of this phenomenon showed a strong study effect. Different taxa showed up to 20% difference in relative abundance within the same study, depending on the analysis strategy. Very few OTUs were shared among the samples. ANOSIM test on unweighted UniFrac distances showed that study, animal model, region of the 16S gene, and dietary treatment (in that order) were the most important factors explaining the differences in bacterial communities. Except for obesity status, the contribution of diet and other factors to explain the variability in bacterial communities was lower when using weighted UniFrac distances. Predicted functional profile and high-level phenotypes of the microbiota showed that each study was associated with unique features and patterns.

Conclusions. The results confirm previous findings showing a strong study effect on gut microbial composition and raise concerns about the impact of analysis strategies on the membership and composition of the gut microbiota. This study may be helpful to guide future research aiming to investigate the relationship between diet, health, and the gut microbiota.

1 Different analysis strategies of 16S rRNA gene data
2 generate contrasting views of gut bacterial
3 communities associated with diet, health and obesity

4

5

6 Jose F. Garcia-Mazcorro[1,*], Jorge R. Kawas[2], Cuauhtemoc Licona-Cassani[3], Susanne U.
7 Mertens-Talcott[4] and Giuliana Noratto[4]

8

9 [1] Research and Development, MNA de Mexico, San Nicolas de los Garza, Nuevo Leon, Mexico
10 [2] Faculty of Agronomy, Universidad Autonoma de Nuevo Leon, General Escobedo, Nuevo
11 Leon, Mexico
12 [3] Department of Bioengineering, Biosystems and Synthetic Biology, Instituto Tecnologico y de
13 Estudios Superiores de Monterrey (ITESM), Monterrey, Nuevo Leon, Mexico
14 [4] Department of Nutrition and Food Science, Texas A&M University, College Station, Texas,
15 USA

16

17 Corresponding Author:
18 Jose Garcia-Mazcorro[1]
19 Avenida Acapulco 770, San Nicolas de los Garza, Nuevo Leon, 66477, Mexico
20 Email address: josegarcia_mex@hotmail.com

21

22 Abstract

23 **Background.** A main function of diet is to nurture the gut microbiota and this relationship
24 affects host health. However, different analysis strategies can generate different views on the
25 relative abundance of each microbial taxon, which can affect our conclusions about the
26 significance of diet to gut health in lean and obese subjects. Here we explored the impact of
27 using different analysis strategies to study the gut microbiota in a context of diet, health and
28 obesity.
29 **Methods.** Over 15 million 16S rRNA gene sequences from published studies involving dietary
30 interventions in obese laboratory rodents were analyzed. Three strategies were used to assign the
31 16S sequences to Operational Taxonomic Units (OTUs) based on the GreenGenes reference
32 OTU sequence files clustered at 97% and 99% similarity.
33 **Results.** Different strategies to select OTUs influenced the relative abundance of all bacterial
34 taxa, but the magnitude of this phenomenon showed a strong study effect. Different taxa showed
35 up to 20% difference in relative abundance within the same study, depending on the analysis
36 strategy. Very few OTUs were shared among the samples. ANOSIM test on unweighted UniFrac
37 distances showed that study, animal model, region of the 16S gene, and dietary treatment (in that
38 order) were the most important factors explaining the differences in bacterial communities.
39 Except for obesity status, the contribution of diet and other factors to explain the variability in

40   bacterial communities was lower when using weighted UniFrac distances. Predicted functional
41   profile and high-level phenotypes of the microbiota showed that each study was associated with
42   unique features and patterns.
43   **Conclusions.** The results confirm previous findings showing a strong study effect on gut
44   microbial composition and raise concerns about the impact of analysis strategies on the
45   membership and composition of the gut microbiota. This study may be helpful to guide future
46   research aiming to investigate the relationship between diet, health, and the gut microbiota.
47
48   Introduction
49   The digestive tract of humans and other animals is inhabited by trillions of microbes and viruses
50   that have evolved with their host as a unit throughout millennia. Host genetics is one of the most
51   important factors shaping the gut microbiota (*Bonder et al. 2016; Daḇrowska & Witkiewicz,*
52   *2016; Zhao et al. 2016; Knowles et al. 2019; Suzuki et al. 2019*) but environmental factors may
53   dominate over genetics in some circumstances (*Muegge et al. 2011; Rothschild et al. 2018*).
54   Among all environmental factors that can modulate the gut microbiota, diet and dietary patterns
55   have the strongest potential to do so. For example, diet can output similar microbiota functions
56   across mammalian phylogeny (*Muegge et al. 2011*) and certain diets can induce and perpetuate
57   obesity, a phenomenon that is closely interlinked with the gut microbiota (*Bäckhed et al. 2004;*
58   *Turnbaugh et al. 2008*).
59   Several food ingredients such as polysaccharides and polyphenols have been reported to
60   influence lipid metabolism by altering gut microbiota composition. However, there are still many
61   unknowns about the effect of diet and other factors on the gut microbial ecosystem and host'
62   health. For instance, microbes display high levels of cell-to-cell variability, even between
63   members of the same strain under controlled homogeneous environments (*Davidson et al. 2008*).
64   This individuality strengthens even more the notion of a highly personalized microbiome in each
65   human or animal host, which ultimately affects host response to diet. Moreover, we often do not
66   deal with the microbes per se; instead, we habitually deal with DNA nucleotide sequences
67   obtained from these unique microbes at discrete time-points. In order to efficiently classify
68   microbes based on molecular data (e.g. DNA sequences), scientists have developed
69   categorization (or grouping) items and rules besides the idea of species, a term that is obviously
70   ambiguous for organisms with eccentric reproductive strategies (*Angert 2005*). The term
71   Operational Taxonomic Unit (OTU) was invented in a context of numerical taxonomy (*Sneath*
72   *1957; Sokal & Sneath 1963*) and nowadays is mostly used to catalogue genetic sequences from
73   marker genes (e.g. the 16S gene) based on nucleotide similarities. In short, similar sequences are
74   catalogued within the same OTU, and therefore each OTU is thought to represent similar
75   microorganisms. Alternatives to OTUs have been suggested (*Callahan, McMurdie & Holmes*
76   *2017*) but the idea of using nucleotide similarities to catalogue microbes prevails. Note that the
77   supposition that similar 16S sequences come from similar organisms is far from being true
78   (*Jaspers et al. 2004*).

79   Several papers have investigated the gut microbiota in relationship to diet, health and
80 obesity. However, only few have looked at data from multiple studies to expose the impact of
81 analysis strategies on this relationship. Lozupone *et al.* (2013) performed a meta-analysis of
82 studies of the human microbiota (gut, oral, vaginal, skin and other) and showed that samples of
83 the Western adult fecal microbiota clustered strongly by study. However, the authors did not
84 discuss the potential contribution of diet in the differences in the microbiota and only used one
85 approach to select OTUs. Walters *et al.* (2014) performed a similar meta-analysis of human gut
86 microbes associated with obesity and Inflammatory Bowel Disease and showed that specific
87 microbial signatures of obesity were not consistent between studies. The authors suggested that
88 this was due to differences in effect sizes and explained that some conditions such as
89 inflammatory bowel disease are associated with more obvious differences in microbiota
90 compared to other conditions such as obesity, whose association with the microbiota,
91 accordingly to the authors, is less clear. The authors of this study also did not look at the
92 contribution of diet to the observed differences and did not explore the use of different analysis
93 strategies. A recent meta-analysis of samples from rodents and humans confirmed that diet can
94 induce reproducible microbiome alterations but only focused on high-fat diets and did not assess
95 the impact of analysis strategies on the relative abundance of taxa (*Bisanz et al. 2019*). Others
96 have used a different approach looking only at the –published– abundance of a few selected
97 bacterial groups (e.g. *Bifidobacterium*) and their relationship with dietary components (*So et al.*
98 *2018; Wilson et al. 2019*).
99   The mucin-degrader bacterium *Akkermansia muciniphila* is a good example to emphasize
100 the relevance of these issues. This bacterium has anti-inflammatory and anti-obesity effects that
101 are mediated by a close biochemical interaction with the host through the colonic mucus (*Cani &*
102 *de Vos 2017*). The relative abundance of this bacterium in feces ranges from 0.1% to up to 85%,
103 and this variation is often considered to reflect a response to dietary components and health
104 status (*Garcia-Mazcorro et al. 2020*). However, some of this variation may also be derived from
105 the specific analysis strategy used during the analysis, although this issue has received little
106 attention (*Garcia-Mazcorro et al. 2020*). The aim of this study was to explore how different
107 analysis strategies impact the results from gut microbiological studies in a context of diet, health
108 and obesity.
109
Materials & Methods
111 **Ethical Considerations**
112 This study used 16S rRNA gene sequence data from public datasets published by our research
113 group (Table 1), all of which received approval for the use of the animals.
114
115 **16S gene sequencing data**
116 We used 16S gene sequencing data from seven of our previous publications dealing with
117 modulation of the gut microbiota using dietary interventions in animal obese models (Table 1).
118 Performing a comparative analysis of data generated by the same research group is advantageous

119 because technical variation is likely to be less compared to the variation obtained from multiple
120 research groups. The sequencing procedure was performed at UC Davis for three studies, while
121 the remaining four studies were sequenced at the Molecular Research LP. There were slight
122 variations in the DNA extraction procedures, but we did not use FastPrep for bead beating in any
123 of these studies. This is an important clarification because some people consider the use of
124 FastPrep essential for optimal lysis.
125   Six factors were studied across the samples: 'dietary treatment', with 11 levels; 'study',
126 with seven levels, one for each study; 'animal model', with two levels: mice and rats;
127 'sequencing technique', with two levels: pyrosequencing and MiSeq; 'obesity status' at the time
128 of sampling, with two levels: lean and obese; and 'anatomical site', with two levels: colon and
129 feces. It is important to keep in mind that in the case that one factor is biologically significantly
130 associated with a different microbiota in nature (e.g. obesity status), the existence of other
131 interactive factors (e.g. dietary treatment) may mask the differences we observe during analysis.
132 Unless otherwise stated, the data was analysed with QIIME (*Caporaso et al. 2010*) v.1.8.0. After
133 demultiplexing and quality filtering, we used the sequence file to assign 16S reads to OTUs
134 based on the GreenGenes reference OTU sequence files (v.13.8, August 2013 release) clustered
135 at 97% (99,322 sequences) and 99% similarity (203,452 sequences).
136
137 **Assignment of 16S sequences to OTUs**
138 Three strategies were used to assign the 16S sequences to OTUs. First, the conservative closed
139 approach that discards sequences for not matching any sequence in the sequence reference data.
140 Second, a de novo approach, which does not use a reference data set to cluster the sequences
141 (*Westcott & Schloss 2015*), thus being relatively free to depict the variety of sequences in a
142 sample. Finally, an open approach that combines these two approaches (*Rideout et al. 2014*), first
143 performing a closed approach followed by clustering of remaining sequences de novo. From this
144 point on, we will refer to these strategies as the name of the clustering method (e.g. closed) and
145 the reference file utilized (e.g. OTUs clustered at 97%). For example, 'closed97' and 'closed99'
146 will refer to a closed approach using the 97% and the 99% OTU reference files, respectively. We
147 only used the default 97% in the similarity option of the pick_otus.py script but changing this
148 parameter can also drastically affect the results (more on this in "Similarity percentage between
149 16S rRNA gene sequences" in Supplemental Information).
150
151 **Taxonomic and diversity analyses**
152 We combined all OTU tables (one for each of the seven studies) from the closed97 and the
153 closed99 approaches into two separate OTU tables (one for each approach) and used these OTU
154 tables for taxonomic classification and diversity analyses. The unique fraction metric, or
155 UniFrac, is a phylogenetic method for comparing microbial communities based on the
156 phylogenetic distance between sets of taxa in a phylogenetic tree as the fraction of the branch
157 length of the tree that leads to descendants from one environment or the other, but not both
158 (*Lozupone et al. 2005*). Both weighted and unweighted UniFrac distances were used for

159   comparing microbial communities because they can lead to different insights into factors that
160   structure microbial communities (*Lozupone et al. 2007*). Principal Coordinate Analyses (PCoA)
161   using these UniFrac distances were performed in QIIME and visualized using Emperor.
162

163   **Prediction of functional profiling and phenotypes**
164   PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states,
165   *Langille et al. 2013*) was used to predict the functional profiling based on the 16S sequences.
166   PICRUSt results were analysed in STAMP (*Parks et al. 2010*). Additionally, we used BugBase
167   (*Ward et al. 2017*) to predict organism-level microbiome phenotypes for each study separately.
168

169   **QIIME2**
170   QIIME2 was introduced in 2017 on the basis of a plugin architecture that allows third parties to
171   contribute functionality (*Bolyen et al. 2019*). Based on the observations made by Thompson *et*
172   *al.* (2017) in a context of OTUs and sequence variants, DADA2 (*Callahan et al. 2016*) and
173   Deblur (*Amir et al. 2017*) were also used to select OTUs for a few selected studies in QIIME2.
174

175   **Statistical analysis**
176   The non-parametric analysis of similarities (ANOSIM) and the Adonis tests were used to
177   determine whether the clustering of samples by a given factor (e.g. study) is statistically
178   significant based on UniFrac distances, using the compare_categories.py script in QIIME with
179   default number of permutations (999). In our experience, these tests usually have low sensitivity
180   (they usually yield low p values even for weak clustering of samples), therefore it is informative
181   to look at both the p values and the percentage of variation explained by the factor. We used the
182   non-parametric Kruskal-Wallis to compare the number of OTUs between the different levels of
183   any given factor, and also to compare the results from BugBase.
184

Results
186   We analysed >15 million 16S reads from 164 samples from seven studies dealing with dietary
187   interventions in obese laboratory animals (Table 1). In general, each study investigated the effect
188   of different diets on the gut microbiota of obese laboratory animals and compared the results
189   with data from control obese and/or lean animals (Supplemental Table S1).
190   We detected 5,246 OTUs using the combined OTU table from the closed97 approach
191   (n=162), and 8,898 OTUs from the closed99 approach (n=163). This difference of ~3,600 OTUs
192   is substantial (70% more OTUs compared to the closed97 approach) and likely reflects the
193   higher number of reference OTUs available for clustering the unknown 16S gene sequences. The
194   number of detected OTUs from the closed97 approach was always lower, and with the exception
195   of the peach study, the number of OTUs from the de novo approaches was always higher (Table
196   2). No singletons were found in any analysis using the open approach because the default script
197   prevents it, but the de novo approach always showed ~2 times higher percentage of singletons
198   compared to the closed approach, except for the peach study (Table 2). The factor 'animal

199 model' generated the highest number of significantly different OTUs, and 'obesity status' the
200 lowest (Table 3). The analysis of the combined OTU tables from the closed97 and the closed99
201 approach showed that only 161 and 165 OTUs (~3% of all OTUs detected) were present in 50%
202 of the samples, respectively. Very few OTUs (3-4) were present in >80% of samples using either
203 approach, and no OTU was present in >90% of samples.
204
205 **Community membership structure**
206 Table 2 shows an accurate numerical impression of relative abundances of all bacterial phyla
207 across different OTU picking strategies, with up to 20% difference in relative abundance of some
208 taxa, depending on the strategy analysis. However, Table 2 lacks a general view of how each
209 taxon is represented across the different levels of each factor. A visual analysis of membership
210 data revealed interesting patterns, for instance, the relative abundances of Firmicutes and
211 Bacteroidetes were equally represented in each level of all six factors studied, while
212 Cyanobacteria was poorly represented in samples from mice (Figure 1). However, note that these
213 results may also be misleading when considering the very few OTUs that were shared across all
214 samples.
215
216 **Peach study**
217 The peach study was the only one using 454 pyrosequencing and a different primer set. As
218 expected, an increase in quality threshold reduced the number of sequences that passed quality
219 filtering. This is important because it has been suggested that a more stringent quality filtering
220 helps to "reduce the number of spurious OTUs" (*Buza et al. 2019*), although in practice it is
221 difficult to determine the exact quality threshold to differentiate "true" vs "false" OTUs (*Edgar*
222 *2017*). We will not discuss this issue for Illumina platforms because defaults have already been
223 established (*Bokulich et al. 2013*).
224   Overall the different taxa remained similar in relative abundance but a higher base quality
225 score (qual) threshold of 34 (default: qual 25) had a strong effect on the number of sequences
226 available for OTU picking (62,971 sequences with qual 25 vs. 11,553 with qual 34) and
227 consequently on the number of the OTUs discovered (758 OTUs with qual 25 vs 442 OTUs with
228 qual 34 in the closed97 approach). The difference in qual threshold also had an effect in the
229 proportions of Bacteroidetes that went down from ~37% (qual 25) to 15% (qual 34), and
230 Firmicutes that went up from ~50% to 70%. This discrepancy was not related to the lower
231 number of sequences available for OTU picking because a lower rarefaction in all other analyses
232 (e.g. with qual 25) revealed similar relative proportions compared to the analysis with higher
233 rarefaction depth. A higher base quality score threshold also affected the presence of some low
234 abundant groups (e.g. at qual 34 Deferribacteres and Fusobacteria were not detected using all
235 approaches).
236   In the peach study, Bacteroidetes displayed the highest standard deviation (SD) and
237 showed the biggest difference (~5%) in relative abundances, particularly between the open97 and
238 the open99 approaches (Table 2). The difference between the lowest and the highest value was

239   minimal for Firmicutes (~2%), Proteobacteria (~2%) and others (Table 2). Tenericutes showed
240   the highest SD/average ratio (68.7), which implies that the variability was proportionally higher
241   in this taxon, and Firmicutes the lowest (1.7), which implies that the variability was
242   proportionally lower in this taxon. The detected phyla varied from 10 (closed97) to 13 (de
243   novo97 and open97).
244
245   **Wheat study**
246   In the wheat study, Firmicutes displayed the highest SD and showed the biggest difference
247   (~20%) in relative abundances, particularly between the closed97 and the closed99 approaches
248   (Table 2). This contrasts heavily with the biggest difference of ~5% observed in the peach study.
249   This difference in closed approaches was also noticeable in Bacteroidetes (~11% difference),
250   Proteobacteria (~10% difference), and others (Table 2) and this was not related to rarefaction
251   depth. This difference was not noticeable for the open and the de novo approaches where
252   Firmicutes, Bacteroidetes, Proteobacteria and others showed very similar proportions in all cases
253   (Table 2). Verrucomicrobia showed the highest SD/average ratio (107.9) and Firmicutes the
254   lowest (11.3). The number of detected phyla varied between 10 (closed97) and 14 (de novo).
255
256   **Quinoa study**
257   Similar to the wheat study, in the quinoa study Firmicutes displayed the highest SD and showed
258   the biggest difference (~15%) in relative abundances, particularly in the closed approaches
259   (Table 2). This difference in closed approaches was also noticeable in Bacteroidetes (~11%
260   difference) and minimal for Proteobacteria (~2% difference) and others, and this was again not
261   related to rarefaction depth. Also similar to the wheat study, this difference was not noticeable
262   for the open and the de novo approaches where Firmicutes, Bacteroidetes, Proteobacteria and
263   others showed very similar proportions in all cases (Table 2). Verrucomicrobia showed the
264   highest SD/average ratio (101.8) and Firmicutes the lowest (7.3). The number of detected phyla
265   varied from 8 (closed97) to 12 (de novo).
266
267   **Barley study**
268   In the barley study, Bacteroidetes displayed the highest SD but Firmicutes showed the biggest
269   difference (~9%) in relative abundances, particularly in the closed approach (Table 2). This
270   difference in closed approaches was minimal in Bacteroidetes (~3% difference), Proteobacteria
271   (~2% difference) and others and this was again not related to rarefaction depth. Also similar to
272   the other studies, this difference was not noticeable in the de novo and open approaches (Table
273   2). Verrucomicrobia showed the highest SD/average ratio (91.9) and Firmicutes the lowest (5.5).
274   The number of detected phyla varied from 8 (closed97) to 12 (de novo).
275
276   **Cherry study**
277   In contrast to the wheat, quinoa and barley studies, where the closed97 and the closed99
278   approaches showed different abundance of taxa, in the cherry study there was good agreement

279   between the closed97 and the closed99 approaches for Firmicutes, Bacteroidetes, Proteobacteria
280   and other groups, but these approaches showed dissimilar proportions compared to the open and
281   the de novo approaches (Table 2). The cherry study was also interesting because the de novo
282   approach showed the presence of bacterial groups (e.g. Nitrospirae, Chlorobi, Planctomycetes)
283   that were not detectable using the closed and the open approaches, and this was likely not related
284   to rarefaction depth because the open approach used similar thresholds. Bacteroidetes displayed
285   the highest SD and showed the biggest difference (~7%) in relative abundances. Tenericutes
286   showed the highest SD/average ratio (100.9) and Firmicutes the lowest (3.4). The number of
287   detected phyla varied from 14 (open99) to 22 (de novo). The use of DADA2 and Deblur in
288   QIIME2 showed 1,329 and 1,263 OTUs, respectively. This is about half the lower number of
289   OTUs (2,439 with closed97) obtained with the other approaches (Table 2).
290
291   **Raspberry study**
292   Unlike the other studies discussed above, the raspberry study was very interesting because it had
293   the lowest variation among the different OTU picking approaches (0.56% difference in
294   Bacteroidetes and 0.55% difference in Firmicutes between the lowest and the highest results),
295   which was also reflected in the SD/average ratio (highest: 32.8 for Cyanobacteria, lowest: 1.2 for
296   Firmicutes) (Table 2). Moreover, both the open97 and the open99 approaches detected as many
297   as 41 different taxa at the phylum level, while the number of taxa in all other studies using the
298   same approach only varied from 8 to 22. This strongly suggests that the microbiota data
299   contained within each study is unique and may sometimes contain a high proportion of taxa that
300   go unnoticed in other studies. This is very important in a context of the role of rare taxa in
301   maintaining the stability of ecosystems (*Jousset et al. 2017*). DADA2 and Deblur showed 791
302   and 721 OTUs, respectively. This contrasts heavily with the numbers (2,751-92,486 OTUs)
303   obtained from all approaches (Table 2).
304
305   **Apple study**
306   In the apple study, Bacteroidetes showed the highest variation and also the biggest difference
307   (~11%) especially between the closed and the de novo approaches (Table 2). This difference
308   between the two approaches was also noticeable in Firmicutes and Tenericutes (Table 2).
309   Tenericutes showed the highest SD/average ratio (77.3) and Proteobacteria the lowest (4.9). The
310   number of detected phyla varied from 15 to 35, thus making the apple study the second study
311   more variable after the raspberry study. The use of DADA2 and Deblur in QIIME2 showed
312   1,357 and 1,149 OTUs, respectively. This contrasts heavily with the numbers (2,095-153,681
313   OTUs) obtained from all approaches (Table 2).
314
315   **Firmicutes/Bacteroidetes ratio**
316   The first paper that discussed about the possible usefulness of the Firmicutes/Bacteroidetes ratio
317   was published by Ley *et al.* (2005), where the showed that obese mice had 50% more Firmicutes,
318   with a proportionally lower abundance of Bacteroidetes, thus leading to a higher

319   Firmicutes/Bacteroidetes ratio compared to lean mice. However, each phylum is composed by
320   hundreds of different species, and therefore the ratio between the two has little significance as
321   discussed elsewhere (*Delzenne & Cani 2011*). In this current study, this ratio varied from 1.1 to
322   3.3 due to the differences in relative abundance in the two phyla (Table 2).
323
324   **The phylum Verrucomicrobia**
325   Firmicutes and Bacteroidetes are usually the most abundant phyla in the gut microbiota
326   and they have attracted most of the attention. However, other taxa deserve attention to better
327   comprehend the functioning of the gut microbial ecosystem. *A. muciniphila* is a taxon that has
328   generated interest as a new generation probiotic candidate to help obese patients and is
329   considered to be a member of Verrucomicrobia based on 16S gene analysis. *Akkermansia* was
330   detected in ~15% the samples (25/164) with an average of 3% and it was more represented in
331   colon samples (Figure 1). Interestingly, the whole Verrucomicrobia phylum also had an average
332   of 3%, which implies that most or all Verrucomicrobia was represented by *Akkermansia*. If we
333   translate the proportions of 16S reads into numbers, a 3% would represent approximately 300
334   million cells (3×10^8) in a hypothetical environment of 1×10^{10} cells/g of intestinal contents. While
335   this conversion is not necessarily accurate, the high numbers may bear some ecological relevance
336   if one considers that as low as 1,000 cells of other microbes are enough to thrive (*Nilsson et al.*
337   *2019*). In three studies, the abundance of Verrucomicrobia was >5 times higher using a closed
338   approach compared to the other approaches (Table 2).
339
340   **The phylum Cyanobacteria**
341   Another group of interest for gut health is Cyanobacteria (*Bárcena et al. 2019*), which is often
342   considered a contaminant and removed from 16S gene analysis. However, Ley *et al.* (2005)
343   showed a deep-branching clade of Cyanobacteria in the guts of animals and mentioned that they
344   may represent descendants of non-photosynthetic ancestral Cyanobacteria that have become
345   adapted to life in the mammalian gut. Here, Cyanobacteria was detected in ~50% the samples
346   (74/164), was always low in relative abundance (<0.5%) and showed the lowest SD among the
347   OTU picking strategies (Table 2). While 0.5% may be considered low, this percentage implies
348   approximately 50 million (5×10^7) cells in the same hypothetical environment of 1×10^{10} cells/g of
349   intestinal contents as discussed above.
350
351   **Analyses of UniFrac distances**
352   **UniFrac analyses from the closed97 approach**
353   The first two coordinates explained 18% of the variation using unweighted distances (Figure 2),
354   which is similar to other studies that have shown that the first two coordinates explained 23% of
355   the variation in a data set of unweighted UniFrac distances from samples of the human
356   microbiota (*Lozupone et al. 2013*). Study, animal model, sequencing procedure, and dietary
357   treatment were the factors that explained the highest proportion of variability in the data (Figure

358   2, Table 4). This is interesting because others have also shown a strong study effect (*Lozupone et*
359   *al. 2013*).
360   In this study, the first 2 coordinates explained 34% of the variation using the weighted
361   UniFrac distances (Table 4), which is about twice the variation explained by the first two axes
362   using unweighted UniFrac. Accordingly to ANOSIM tests, the effect of study, animal model,
363   and sequencing procedure explained much smaller proportion of the variability in the data using
364   weighted distances, while obesity status explained a higher proportion of the variation (23.6% vs
365   9.5% using unweighted distances, Table 4).
366
367   **UniFrac analyses from the closed99 approach**
368   The analysis of unweighted UniFrac distances from the closed99 approach revealed similar
369   results compared to the closed97 regarding the relative contribution of each factor in explaining
370   the variability in the data (Figure 3 and Table 5). The first two coordinates explained 25% of the
371   variation (Figure 3), which is even closer to the results of Lozupone et al. (2013). Interestingly,
372   *Akkermansia* was not part of the ten most abundant groups (Figure 3) and the factor 'study'
373   explained 80% of the variability in the unweighted UniFrac data accordingly to the ANOSIM
374   test (Table 5). In our experience it is uncommon that one single factor explains that much of the
375   variability between microbial communities. Similar to the analysis using the closed97 approach,
376   the effect of 'study', 'animal model' and 'sequencing procedure' at explaining the variability in
377   the data was lower using weighted UniFrac, while obesity status explained much more of the
378   variability in the data (Table 5).
379
380   **UniFrac analyses from closed97 approach on mice samples**
381   To discover any additional pattern or association between the microbial communities, we
382   performed a separate analysis of mice samples only (n=120). Briefly, unweighted UniFrac
383   analyses using closed97 showed that 'treatment' and 'study' were the most important factors,
384   each explaining almost half of the variability in the data accordingly to ANOSIM tests, while the
385   use of weighted UniFrac distances revealed a lower contribution of 'study' and 'treatment' and a
386   higher contribution of obesity status at differentiating microbial communities (more about this in
387   "UniFrac analyses from closed97 approach on mice samples" in Supplemental Information).
388
389   **Prediction of functional profiling**
390   PICRUSt revealed a total of 217 features that were significantly different between different
391   studies, 211 between different treatments, 145 between sequencing techniques, 136 between
392   mice and rats, 69 between lean and obese, and 32 between feces and colon contents (adjusted P <
393   0.05, see 'PICRUSt results' in Supplemental Information). This is interesting because the factor
394   'study' was also associated with the highest amount of variability. The weighted Nearest
395   Sequenced Taxon Index (weighted NSTI) score was developed to summarize the extent to which
396   microorganisms in a given sample are related to sequenced genomes (e.g. a NSTI score of 0.03
397   indicates that the genomes of the microbes were well represented and that the average microbe in

398   the sample can be predicted using a relative from the same "species"). Higher NSTI values
399   indicate a lower accuracy of the predictions. Here, NSTI scores varied widely, from a minimum
400   of 0.03 (sample W11, feces from lean mice sequenced with MiSeq, wheat study) to a maximum
401   of 0.34 (sample W32, feces from obese mice on quinoa sequences with MiSeq, quinoa study).
402   There was no indication to suggest that any group of sequences were associated with lower or
403   higher NSTI scores.
404

405   **Prediction of organism-level microbiome phenotypes**
406   BugBase revealed interesting differences in phenotypes within each study. For example, obese
407   rats in a high-fat diet had higher proportions of bacteria with the potential of forming biofilms
408   among the different treatments in the apple study, lean mice had higher proportions of bacteria
409   containing mobile elements in the barley study, and obese mice had higher proportions of stress-
410   tolerant bacteria in the cherry study (more on this in 'BugBase results' in Supplemental
411   Information). Overall, these results also indicate that each study showed unique peculiarities,
412   ultimately derived from the molecular composition of the 16S gene sequences at any given time-
413   point within each particular study.
414

Discussion

416   Bacteria and other microorganisms are artificially classified for various reasons, for example to
417   understand their relationship with other organisms. In most scientific publications, it is implicitly
418   assumed that this artificial classification yields biologically relevant groups that exist in nature at
419   real, normal abundances in a state of "normobiosis". This cannot be better exemplified by the
420   widespread use of the opposite term "dysbiosis", a term that is not only inaccurate but also
421   misleading (*Brüssow 2019*). However, the true abundance of any bacterial group in nature is very
422   difficult to determine with accuracy, due to both biological facts (*Davidson & Surette 2008;*
423   *Jaspers & Overmann 2004*), methodological biases, and the many remaining unknowns in
424   microbiome research (*Thomas & Segata 2019*). This work provides clear evidence of this issue
425   by showing that different analysis strategies generate up to 20% difference in relative abundance
426   of some taxa from the same samples, depending on the strategy.
427   Our work also sheds light into the issue of the number of bacterial species in the gut.
428   Considering the lowest sequence abundance threshold of one, it has been suggested that our
429   planet harbor about 10,000 bacterial species (*Caporaso et al. 2011*) but others have suggested
430   5.6 million OTUs as the lower bound of the microbial diversity on Earth (*Rideout et al. 2014*).
431   On the other hand, the human microbiome is thought to harbor 15,000 to 36,000 species (*Frank*
432   *et al. 2007*) yet others have suggested only 4,930 species of bacteria in that environment (*Pasolli*
433   *et al. 2019*). In technical terms, the number of OTUs is dependent on the number of sequences
434   that pass quality filtering (i.e. the higher the quality threshold, the lower the number of sequences
435   available to catalogue). However, the number of OTUs is also dependent on the specific strategy
436   used to catalogue sequences into OTUs. For instance, in this study the number of OTUs from the
437   closed97 approach was always lower, and with one exception, the number of OTUs from the de

438   novo approaches was always higher. While the higher number of OTUs detected by the de novo
439   approach may not be biologically relevant (*Edgar 2017*), this may deserve more thought if we
440   consider the unexplored variety of microbes in nature.

441   As mentioned above, bacteria and other microbes display high levels of cell-to-cell
442   variability; in other words, these organisms show high levels of individuality, which implies that
443   every host carries a unique set of microbes. In support of this hypothesis, the analysis of the
444   combined OTU tables from the closed97 and the closed99 approach showed that only very few
445   OTUs were shared among the samples (3-4 OTUs were present in >80% of samples using either
446   approach, and no OTU was present in >90% of samples). These results strongly suggest a unique
447   microbial profile (i.e. variety of 16S sequences) in each sample. More importantly, the fact that
448   each bacterial cell displays a high level of individuality raises doubts about the relevance of
449   nucleotide similarities to estimate or predict similarities in nature. In other words, it is feasible
450   and very likely that even those same OTUs that were apparently shared by multiple samples
451   represent different microorganisms.

452   The analysis of 16S data from each study revealed interesting patterns. However, there
453   were two studies that attracted our attention. The raspberry study was very interesting because it
454   had the lowest variation in relative abundance of taxa among the different OTU picking
455   approaches, which was also reflected in the SD/average ratio. Also, in the raspberry study both
456   the open97 and the open99 approaches detected as many as 41 different taxa at the phylum level,
457   while the number of taxa in all other studies using the same approach only varied from 8 to 22.
458   This strongly suggests that the microbiota data contained within each study is unique and may
459   sometimes contain a high proportion of taxa that go unnoticed in other studies. This is very
460   important in a context of the role of rare taxa in maintaining the stability of ecosystems (*Jousset*
461   *et al. 2017*).

462   This paper shows interesting results about the impact of the analysis strategies in the
463   relative abundance of Verrucomicrobia and Cyanobacteria and these taxa are also important in a
464   context of low abundant or rare taxa (or OTUs in datasets). Several strategies have been used to
465   remove the so-called rare sequences. Needham *et al.* (2017) used 99% de novo OTUs (i.e. OTUs
466   generated using 99% similarity without the use of a reference database) that exceeded a threshold
467   of 0.4% in relative abundance without providing any rationale for using this threshold. In
468   contrast, Navas-Molina *et al.* (2013) filtered the data based on the proportions of sequences
469   represented by the OTUs (the authors suggested to remove OTUs that represents less than
470   0.005% of all sequence reads based on the analyses by *Bokulich et al. 2013*) and Lozupone *et al.*
471   (2013) discarded whole samples that did not have at least 100 sequences after quality filtering
472   and OTU assignment. Although the decision of removing rare sequences is often based on
473   rational arguments (*Bokulich et al. 2013*) it is still arbitrary to choose one strategy or another.
474   This is important because the so-called rare microbes may in fact be keystone species that can
475   regulate entire microbial environments, including host-associated microbiomes (*Jousset et al.*
476   *2017*).

477   The factor(s) that better explain the differences in gut microbial communities are of
478   interest for various disciplines in biomedical sciences. For instance, the results of this current
479   work showed that the factor 'animal model' generated the highest number of significantly
480   different OTUs, while 'obesity status' the lowest. Unlike other studies suggesting that microbial
481   signatures of obesity are not consistent between studies because of effect sizes (*Walters et al.*
482   *2014*), in this study the lack of statistical difference in gut microbiota between obese and lean is
483   most likely due to the fact that the obese rodents were subject to different dietary treatments, thus
484   making difficult to dissect between obesity status and dietary treatment. In this regard, the results
485   of the analysis of UniFrac distances are particularly important considering the thoughts discussed
486   by Lozupone *et al.* (2007) that explained that unweighted UniFrac is better suited to detect
487   effects of different "founding" populations (e.g. the source of bacteria that first colonize the gut
488   of newborn individuals and factors that are restrictive for microbial growth such as temperature).
489   In contrast, they suggested that weighted UniFrac can more accurately reveal the effects of more
490   transient factors, for example nutrient availability. Host genetics can be considered as a founding
491   effect that since the moment of birth imposes certain restrictions to growth of some microbes,
492   and in our analyses the factor 'animal model' explained high proportion of the variation when
493   using unweighted distances. In contrast, the analysis of weighted distances revealed that the
494   effect of obesity status at sampling explained much more variability in the data compared to
495   unweighted analyses, a result that also makes sense if we consider obesity as a more transient
496   condition that suddenly appeared in the host-microbiota relationship. More research is needed to
497   better investigate the relationship of the different factors as well as their interactions in the gut
498   microbiota.
499

Conclusions

501   In summary, diet and the gut microbiota are strongly related to host health and their interactions
502   are "convoluted and multi-faceted" (*Xu & Knight 2015*). This paper aimed to contribute to
503   current knowledge about the impact of using different analysis strategies in the relationship
504   between the gut microbiota, diet and obesity. The results show that the use of different strategies
505   to select OTUs have an impact on the relative proportions of bacterial taxa, particularly when
506   using a closed OTU picking approach. We also demonstrated the impact of using OTU reference
507   sequences clustered at different similarity percent and confirmed previous observations regarding
508   a strong study effect. We invite others to consider and further expand the feasible possibility that
509   the variations among studies are related to the individuality of bacteria. Overall, the results are
510   useful to guide future research and meta-analyses aiming to investigate the complex relationship
511   between diet, health, and the gut microbiota.
512

Acknowledgements

515

References

517   **Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, Kightley EP,**
518   **Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017.** Deblur rapidly resolves single-
519   nucleotide community sequence patterns. *mSystems* **2:**e00191-16 DOI
520   10.1128/mSystems.00191-16.
521   **Angert ER. 2005.** Alternatives to binary fission in Bacteria. *Nature Reviews Microbiology*
522   **3:**214-224 DOI 10.1038/nrmicro1096.
523   **Bäckhed F, Ding H, Wang T, Hooper LV, Young Koh G, Nagy A, Semenkovich CF,**
524   **Gordon JI. 2004.** The gut microbiota as an environmental factor that regulates fat
525   storage. *Proceedings of the Natural Academy of Sciences USA* **101:**15718-15723 DOI
526   10.1073/pnas.0407076101.
527   **Bárcena C, Valdés-Mas R, Mayoral P, Garabaya C, Durand S, Rodríguez F, Fernández-**
528   **García MT, Salazar N, Nogacka AM, Garatachea N, Bossut N, Aprahamian F,**
529   **Lucia A, Kroemer G, Freije JMP, Quirós PM, López-Otín C. 2019.** Healthspan and
530   lifespan extension by fecal microbiota transplantation into progeroid mice. *Nature*
531   *Medicine* **25:**1234-1242. DOI 10.1038/s41591-019-0504-5.
532   **Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ. 2019.** Meta-analysis reveals
533   reproducible gut microbiome alterations in response to a high-fat diet. *Cell Host &*
534   *Microbe* **26:**265-272 DOI 10.1016/j.chom.2019.06.013.
535   **Bonder MJ, Kurilshikov A, Tigchelaar EF, Mujagic Z, Imhann F, Vich Vila A, Deelen P,**
536   **Vatanen T, Schirmer M, Smeekens SP, Zhernakova DV, Jankipersadsing SA,**
537   **Jaeger M, Oosting M, Cenit MC, Masclee AAM, Swertz MA, Li Y, KKumar V,**
538   **Joosten L, Harmsen H, Weersma RK, Franke L, Hofker MH, Xavier RJ, Jonkers D,**
539   **Netea MG, Wijmenga C, Fu J, Zhernakova A. 2016.** The effect of host genetics on the
540   gut microbiome. *Nature Genetics* **48:**1407-1412 DOI 10.1038/ng.3663.
541   **Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA,**
542   **Caporaso JG. 2013.** Quality-filtering vastly improves diversity estimates from Illumina
543   amplicon sequencing. *Nature Methods* **10:**57-59 DOI 10.1038/nmeth.2276.
544   **Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H,**
545   **Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A,**
546   **Brislawn CJ, Titus Brown C, Callahan BJ, Caraballo-Rodríguez AM, Chase J,**
547   **Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet**
548   **C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM,**
549   **Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H,**
550   **Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Bin**
551   **Kang K, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J,**
552   **Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz**
553   **C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC,**
554   **Morton JT, Turan Naimey A, Navas-Molina JA, Felix Nothias L, Orchanian SB,**
555   **Pearson T, Peoples SL, Petras D, Lai Preuss M, Pruesse E, Buur Rasmussen L,**
556   **Rivers A, Robeson II MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Jin**

557   **Song S, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A,**
558   **Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y,**
559   **Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC,**
560   **Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R,**
561   **Caporaso JG.** Reproducible, interactive, scalable and extensible microbiome data
562   science using QIIME 2. *Nature Biotechnology* **37:**852-857 DOI 10.1038/s41587-019-
563   0209-9.
564   **Brüssow H. 2020.** Problems with the concept of gut microbiota dysbiosis. *Microbial*
565   *Biotechnology* **13:**423-434 DOI 10.1111/1751-7915.13479.
566   **Buza TM, Tonui T, Stomeo F, Tiambo C, Katani R, Schilling M, Lyimo B, Gwakisa P,**
567   **Cattadori IM, Buza J, Kapur V. 2019.** iMAP: an integrated bioinformatics and
568   visualization pipeline for microbiome data analysis. *BMC Bioinformatics* **20:**374 DOI
569   10.1186/s12859-019-2965-4.
570   **Callahan BJ, McMurdie PJ, Holmes SP. 2017.** Exact sequence variants should replace
571   operational taxonomic units in marker-gene data analysis. *The ISME Journal* **11:**2639-
572   2643 DOI 10.1038/ismej.2017.119.
573   **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016.** DADA2:
574   high-resolution sample inference from Illumina amplicon data. *Nature Methods* **13:**581-
575   583 DOI 10.1038/nmeth.3869.
576   **Cani PD, de Vos WM. 2017.** Next-generation beneficial microbes: the case of *Akkermansia*
577   *muciniphila*. *Frontiers in Microbiology* **8:**1765 DOI 10.3389/fmicb.2017.01765.
578   **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer**
579   **N, Gonzalez Peña A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D,**
580   **Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J,**
581   **Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J,**
582   **Knight R. 2010.** QIIME allows analysis of high-throughput community sequencing data.
583   *Nature Methods* **7:**335-336 DOI 10.1038/nmeth.f.303.
584   **Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ,**
585   **Fierer N, Knight R. 2011.** Global patterns of 16S rRNA diversity at a depth of millions
586   of sequences per sample. *Proceedings of the Natural Academy of Sciences USA*
587   **108:**4516-4522 DOI 10.1073/pnas.1000080107.
588   **Dąbrowska K, Witkiewicz W. 2016.** Correlations of host genetics and gut microbiome
589   composition. *Frontiers in Microbiology* **7:**1357 DOI 10.3389/fmicb.2016.01357.
590   **Davidson CJ, Surette MG. 2008.** Individuality in Bacteria. *Annual Review of Genetics* **42:**253-
591   268 DOI 10.1146/annurev.genet.42.110807.091601.
592   **Delzenne NM, Cani PD. 2011.** Interaction between obesity and the gut microbiota: relevance in
593   nutrition. *Annual Review of Nutrition* **31:**15-31 DOI 10.1146/annurev-nutr-072610-
594   145146.
595   **Edgar RC. 2017.** Accuracy of microbial community diversity estimated by closed- and open-
596   reference OTUs. *PeerJ* **5:**e3889 DOI 10.7717/peerj.3889.

597 **Frank DN, St. Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007.**
598   Molecular-phylogenetic characterization of microbial community imbalances in human
599   inflammatory bowel diseases. *Proceedings of the Natural Academy of Sciences USA*
600   **104:**13780-13785 DOI 10.1073/pnas.0706625104.
601 **Garcia-Mazcorro JF, Ivanov I, Mills DA, Noratto G. 2016.** Influence of whole-wheat
602   consumption on fecal microbial community structure of obese diabetic mice. *PeerJ*
603   **4:**e1702 DOI 10.7717/peerj.1702.
604 **Garcia-Mazcorro JF, Lage NN, Mertens-Talcott S, Talcott S, Chew B, Dowd SE, Kawas**
605   **JR, Noratto GD. 2018.** Effect of dark sweet cherry powder consumption on the gut
606   microbiota, short-chain fatty acids, and biomarkers of gut health in obese db/db mice.
607   *PeerJ* **6:**e4195 DOI 10.7717/peerj.4195.
608 **Garcia-Mazcorro JF, Mills DA, Murphy K, Noratto G. 2017.** Effect of barley
609   supplementation on the fecal microbiota, caecal biochemistry and key biomarkers of
610   obesity and inflammation in obese db/db mice. *European Journal of Nutrition* **57:**2513-
611   2528 DOI 10.1007/s00394-017-1523-y.
612 **Garcia-Mazcorro JF, Mills DA, Noratto G. 2016.** Molecular exploration of fecal microbiome
613   in quinoa-supplemented obese mice. *FEMS Microbiology Ecology* **92:**fiw089 DOI
614   10.1093/femsec/fiw089.
615 **Garcia-Mazcorro JF, Minamoto Y, Kawas JR, Suchodolski JS, de Vos WM. 2020.**
616   *Akkermansia* and microbial degradation of mucus in cats and dogs: implications to the
617   growing worldwide epidemic of pet obesity. *Veterinary Sciences* **7:**44. DOI
618   10.3390/vetsci7020044.
619 **Garcia-Mazcorro JF, Pedreschi R, Chew B, Dowd SE, Kawas JR, Noratto G. 2018.** Dietary
620   supplementation with raspberry extracts modifies the fecal microbiota in obese diabetic
621   db/db mice. *Journal of Microbiology and Biotechnology* **28:**1247-1259 DOI
622   10.4014/jmb.1803.03020.
623 **Garcia-Mazcorro JF, Pedreschi R, Yuan J, Kawas JR, Chew B, Dowd SE, Noratto G. 2019.**
624   Dietary supplementation with whole apples for six weeks is associated with a distinctive
625   microbiota, proteomics and metabolomics profile in the gut of Dawley Sprague rats fed a
626   high-fat diet. *PLoS ONE* **14:**e0212586 DOI 10.1371/journal.pone.0212586.
627 **Jaspers E, Overmann J. 2004.** Ecological significance of microdiversity: identical 16S rRNA
628   gene sequences can be found in Bacteria with highly divergent genomes and
629   ecophysiologies. *Applied Environmental Microbiology* **70:**4831-4839 DOI
630   10.1128/AEM.70.8.4831-4839.2004.
631 **Jousset A, Bienhold C, Chatzinotas A, Gallien L, GObet A, Kurm V, Küsel K, Rilling MC,**
632   **Rivett DW, Salles JF, van der Heijden MGA, Youssef NH, Zhang X, Wei Z, WH**
633   **Gera Hol. 2017.** Where less may be more: How the rare biosphere pulls ecosystems
634   strings. *The ISME Journal* **11:**853-862 DOI 10.1038/ismej.2016.174.

635   **Knowles SCL, Eccles RM, Baltrūnaitė L. 2019.** Species identity dominates over environment
636   in shaping the microbiota of small mammals. *Ecology Letters* **22:**826-837 DOI
637   10.1111/ele.13240.
638   **Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC,**
639   **Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013.**
640   Predictive functional profiling of microbial communities using 16S rRNA marker gene
641   sequences. *Nature Biotechnology* **31:**814-823 DOI 10.1038/nbt.2676.
642   **Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. 2005.** Obesity
643   alters gut microbial ecology. *Proceedings of the Natural Academy of Sciences USA*
644   **102:**11070-11075 DOI 10.1073/pnas.0504978102.
645   **Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vázquez-Baeza Y,**
646   **Jansson JK, Gordon JI, Knight R. 2013.** Meta-analyses of studies of the human
647   microbiota. *Genome Research* **23:**1704-1714. DOI 10.1101/gr.151803.112.
648   **Lozupone C, Knight R. 2005.** UniFrac: a new phylogenetic method for comparing microbial
649   communities. *Applied Environmental Microbiology* **71:**8228-8235 DOI
650   10.1128/AEM.71.12.8228-8235.2005.
651   **Lozupone CA, Hamady M, Kelley ST, Knight R. 2007.** Quantitative and qualitative β
652   diversity measures lead to different insights into factors that structure microbial
653   communities. *Applied Environmental Microbiology* **73:**1576-1585 DOI
654   10.1128/AEM.01996-06.
655   **Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B,**
656   **Knight R, Gordon JI. 2011.** Diet drives convergence in gut microbiome functions
657   across mammalian phylogeny and within humans. *Science* **332:**970-974 DOI
658   10.1126/science.1198719.
659   **Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vázquez-Baeza Y, Xu**
660   **Z, Ursell LK, Lauber C, Zhou H, Song SJ, Huntley J, Ackermann GL, Berg-Lyons**
661   **D, Holmes S, Caporaso JG, Knight R. 2013.** Advancing our understanding of the
662   human microbiome using QIIME. *Methods Enzymology* **531:**371-444 DOI
663   10.1016/B978-0-12-407863-5.00019-8.
664   **Needham DM, Sachdeva R, Fuhrman JA. 2017.** Ecological dynamics and co-occurrence
665   among marine phytoplankton, bacteria and myoviruses shows microdiversity matters.
666   *The ISME Journal* **11:**1614-1629 DOI 10.1038/ismej.2017.29.
667   **Nilsson OR, Kari L, Steele-Mortimer O. 2019.** Foodborne infection of mice with *Salmonella*
668   *Typhimurium*. *PLoS ONE* **14:**e0215190 DOI 10.1371/journal.pone.0215190.
669   **Noratto GD, Garcia-Mazcorro JF, Markel M, Martino HS, Minamoto Y, Steiner JM,**
670   **Byrne D, Suchodolski JS, Mertens-Talcott SU. 2014.** Carbohydrate-free peach (*Prunus*
671   *persica*) and plum (*Prunus salicina*) juice affects fecal microbial ecology in an obese
672   animal model. *PLoS ONE* **9:**e101723 DOI 10.1371/journal.pone.0101723.
673   **Parks DH, Beiko RG. 2010.** Identifying biologically relevant differences between metagenomic
674   communities. *Bioinformatics* **26:**715-721 DOI 10.1093/bioinformatics/btq041.

675 **Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P,**
676   **Tett A, Ghensi P, Collado MC, Rice BL, DuLong C, Morgan XC, Golden CD,**
677   **Quince C, Huttenhower C, Segata N. 2019.** Extensive unexplored human microbiome
678   diversity revealed by over 150,000 genomes from metagenomes spanning age,
679   geography, and lifestyle. *Cell* **176:**649-662 DOI 10.1016/j.cell.2019.01.001.
680 **Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, Chase J,**
681   **McDonald D, Gonzalez A, Robbins-Pianka A, Clemente JC, Gilbert JA, Huse SM,**
682   **Zhou H-W, Knight R, Caporaso JG. 2014.** Subsampled open-reference clustering
683   creates consistent, comprehensive OTU definitions and scales to billions of sequences.
684   *PeerJ* **2:**e545 DOI 10.7717/peerj.545.
685 **Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI,**
686   **Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vich Vila A, Zmora N, Pevsner-**
687   **Fisher M, Israeli D, Kosower N, Malka G, Chen Wolf B, Avnit.Sagi T, Lotan-**
688   **Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A,**
689   **Elinav E, Segal E. 2018.** Environment dominates over host genetics in shaping human
690   gut microbiota. *Nature* **555:**210-215 DOI 10.1038/nature25973.
691 **Sneath PHA. 1957.** The application of computers to taxonomy. *Journal of General*
692   *Microbiology* **17:**201-226.
693 **So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT, Shanahan ER, Staudacher**
694   **HM, Campbell KL. 2018.** Dietary fiber intervention on gut microbiota composition in
695   healthy adults: a systematic review and meta-analysis. *The American Journal of Clinical*
696   *Nutrition* **107:**965-983 DOI 10.1093/ajcn/nqy041.
697 Sokal RR, Sneath PHA. Principles of numerical taxonomy. San Francisco: W.H. Freeman, 1963.
698 **Suzuki TA, Phifer-Rixey M, Mack KL, Sheehan MJ, Lin D, Bi K, Nachman MW. 2019.**
699   Host genetic determinants of the gut microbiota of wild mice. *Molecular Ecology*
700   **28:**3197-3207 DOI 10.1111/mec.15139.
701 **Thomas AM, Segata N. 2019.** Multiple levels of the unknown in microbiome research. *BMC*
702   *Biology* **17:**48. DOI 10.1186/s12915-019-0667-z.
703 **Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi**
704   **A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-**
705   **Baeza Y, González A, Morton JT, Mirarab S, Xu ZZ, Jian L, Haroon MF, Kanbar**
706   **J, Zhu Q, Song SJ, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G,**
707   **Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman**
708   **JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert**
709   **JA, Knight R, The Earth Microbiome Project Consortium. 2017.** A communal
710   catalogue reveals Earth's multiscale microbial diversity. *Nature* **551:**457-463 DOI
711   10.1038/nature24621.
712 **Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI. 2008.** Diet-induced obesity is linked to
713   marked but reversible alterations in the mouse distal gut microbiome. *Cell Host &*
714   *Microbe* **3:**213-223 DOI 10.1016/j.chom.2008.02.015.

715   **Walters WA, Xu Z, Knight R. 2014.** Meta-analyses of human gut microbes associated with
716   obesity and IBD. *FEBS Letters* **588:**4223-4233 DOI 10.1016/j.febslet.2014.09.039.
717   **Ward T, Larson J, Meulemans J, Hillmann B, Lynch J, Sidiropoulos D, Spear JR,**
718   **Caporaso G, Blekhman R, Knight R, Fink R, Knights D. 2017.** BugBase predicts
719   organism-level microbiome phenotypes. bioRxiv DOI 10.1101/133462.
720   **Westcott SL, Schloss PD. 2015.** De novo clustering methods outperform reference-based
721   methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*
722   **3:**e1487 DOI 10.7717/peerj.1487.
723   **Wilson B, Rossi M, Dimidi E, Whelan K. 2019.** Prebiotics in irritable bowel syndrome and
724   other functional bowel disorders in adults: a systematic review and meta-analysis of
725   randomized controlled trials. *American Journal of Clinical Nutrition* **109:**1098-1111.
726   DOI 10.1093/ajcn/nqy376.
727   **Xu Z, Knight R. 2015.** Dietary effects on human gut microbiome diversity. *British Journal of*
728   *Nutrition* **113:**S1-S5 DOI 10.1017/S0007114514004127.
729   **Zhao P, Irwin DM, Dong D. 2016.** Host genetics is associated with the gut microbial
730   community membership rather than the structure. *Molecular Biosystems* **12:**1676-1686
731   DOI 10.1039/C5MB00850F.

Figure 1

Proportions of relative abundances (percentages) of 16S reads.

The data is shown accordingly to the factors (A) animal model, (B) obesity status, (C) sequencing, (D) anatomical site, (E) study, and (F) treatment. Letters in bars represent different taxa (A: Firmicutes, B: Bacteroidetes, C: Proteobacteria, D: Verrucomicrobia, E: Deferribacteres, F: Actinobacteria, G: Cyanobacteria, and H: others). The bars do not represent the relative abundance of each taxon; instead, they show the proportion of 16S relative abundances from each level of the factors investigated. A balanced stacked bar denotes that the taxon was equally represented in each level of the factor (e.g. Firmicutes and Bacteroidetes between levels of all factors). An unbalanced stacked bar denotes that the taxon was more or less represented between the levels of the factor, for example Cyanobacteria between mice and rats, or between colon and feces.
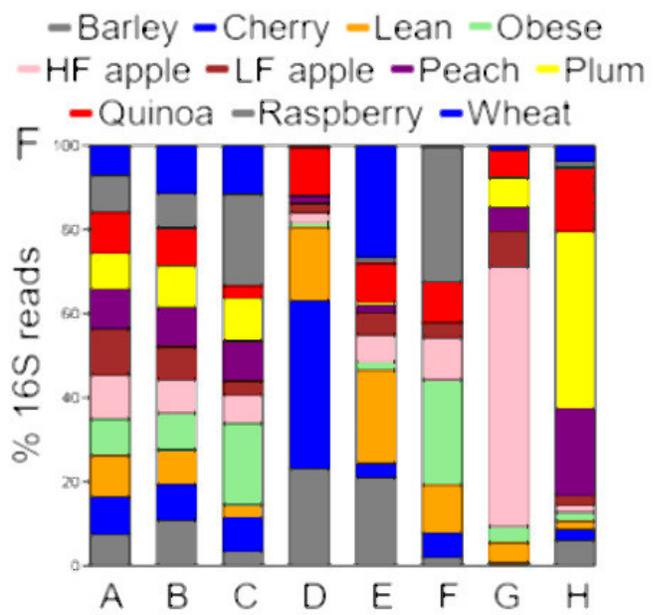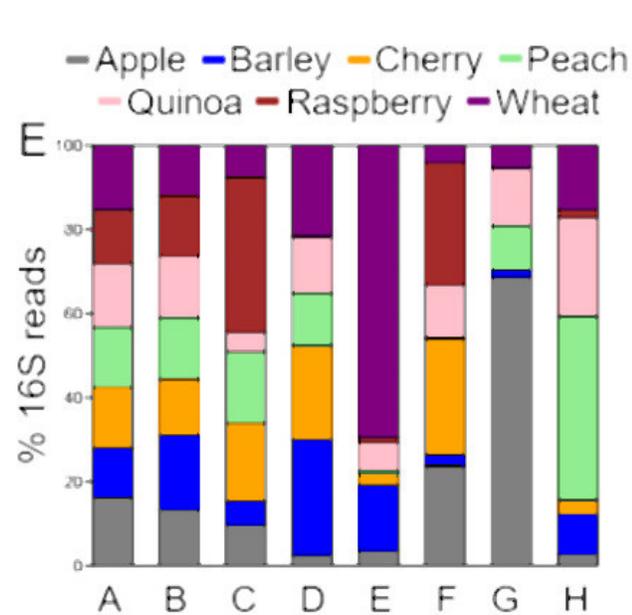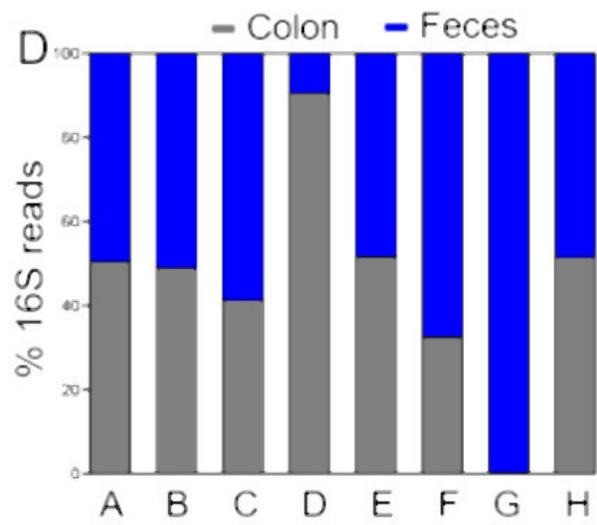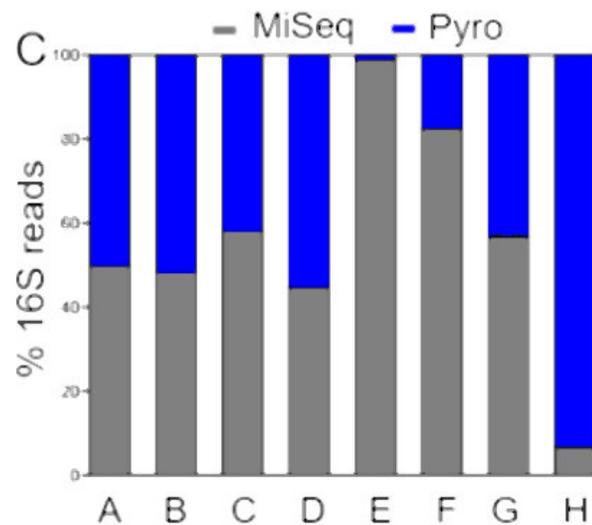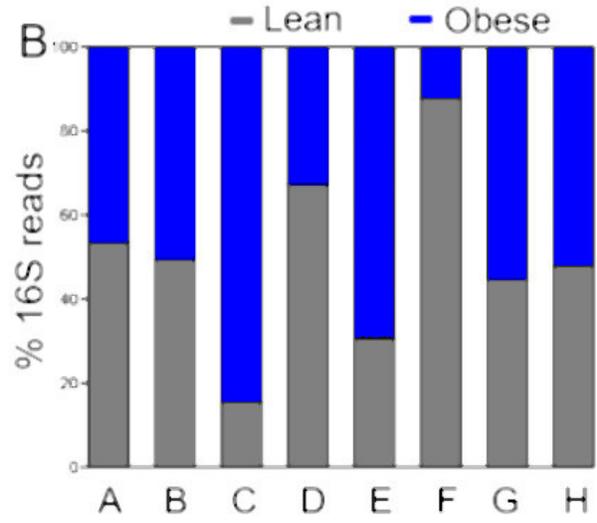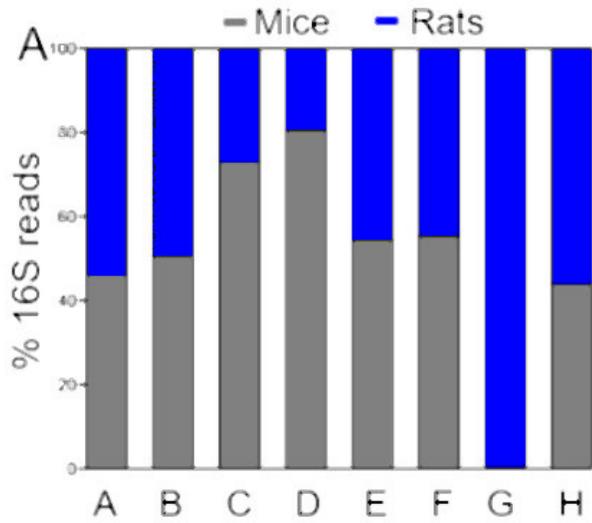
Figure 2

PCoA plots of unweighted UniFrac distances using data from the closed approach using the reference OTUs sequence file at 97% similarity (closed97 approach).

Each point represents a sample from one of the studies detailed in Table 1, and the plots highlight the effect of (A) animal model, (B) obesity status, (C) sequencing technique, (D) anatomical site, (E) study and (F) treatment. The ten most abundant bacterial groups are superimposed in all plots (the bigger the circle, the bigger the relative abundance of each taxa) and labelled with numbers on the first plot (1: Ruminococcaceae, 2: *Bacteroides*, 3: Clostridiales, 4: Lachnospiraceae, 5: *Oscillospira*, 6: Enterobacteriaceae, 7: S24-7, 8: *Lactobacillus*, 9: *Akkermansia*, 10: *Allobaculum*) to show that clustering of samples is driven by specific bacterial groups that have previously been shown to influence (or be influenced by) health status, such as *Akkermansia*. The values for each axis are only shown in A to facilitate viewing. These plots were built using a rarefaction depth of 100 sequences per sample to account for as many samples as possible (only two samples were left out using this rarefaction depth). The UniFrac data was obtained from the closed97 approach but other approaches may reveal other patterns (see Table 4 and Table 5).
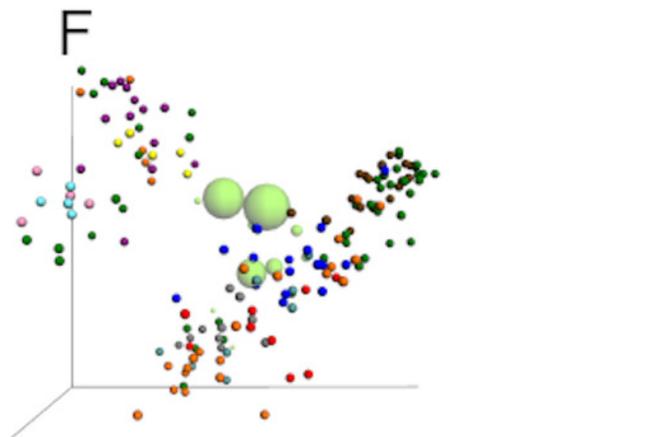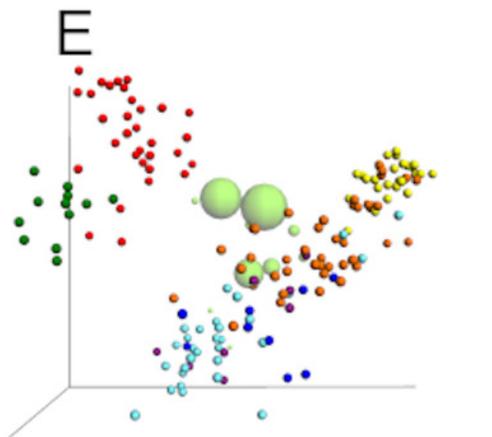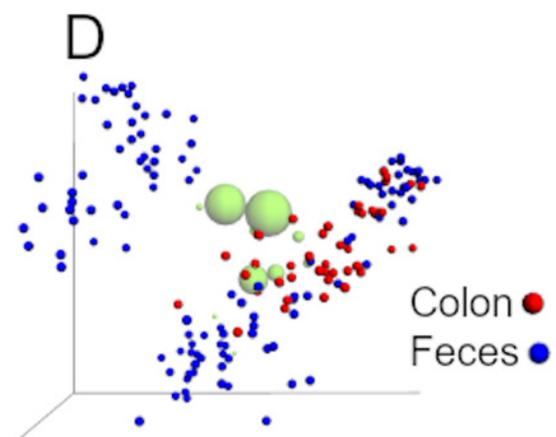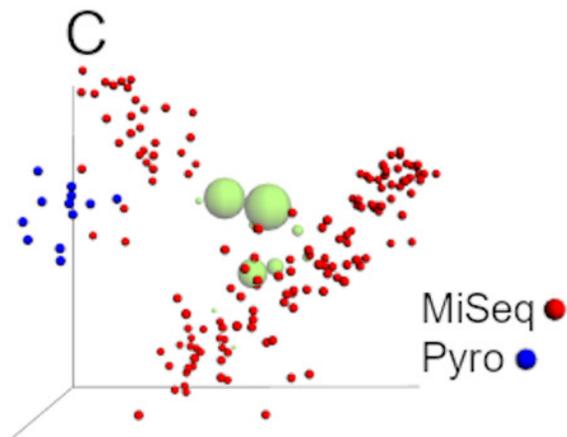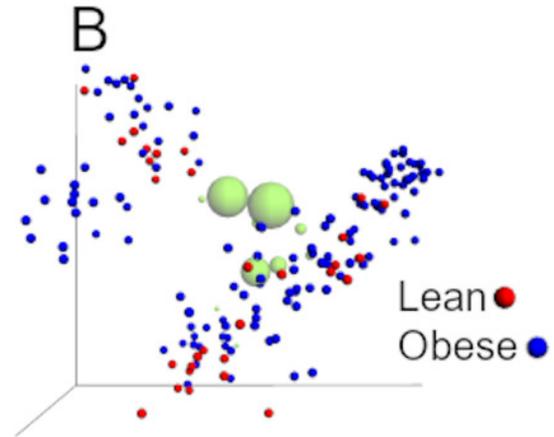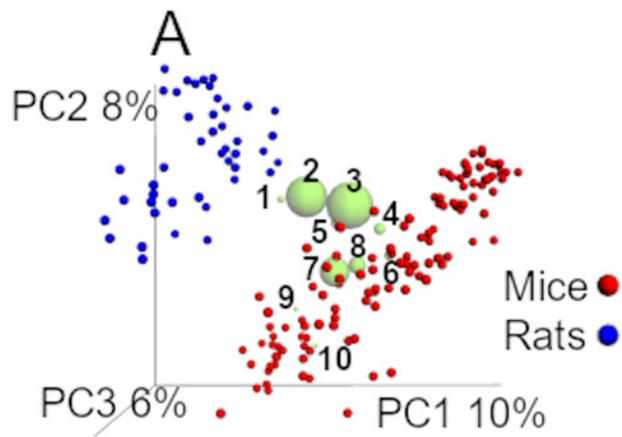
Figure 3

PCoA plots of unweighted UniFrac distances using data from the closed approach using the reference OTUs sequence file at 99% similarity (closed99 approach).

Each point represents a sample from one of the studies detailed in Table 1, and the plots highlight the effect of (A) animal model, (B) obesity status, (C) sequencing technique, (D) anatomical site, (E) study and (F) treatment. The ten most abundant bacterial groups are superimposed in all plots (the bigger the circle, the bigger the relative abundance of each taxa) and labelled with numbers on the first plot (1: Ruminococcaceae, 2: *Bacteroides*, 3: Clostridiales, 4: Lachnospiraceae, 5: *Oscillospira*, 6: Enterobacteriaceae, 7: S24-7, 8: *Lactobacillus*, 9: *Parabacteroides*, 10: *Allobaculum*) to show that clustering of samples is driven by specific bacterial groups that have previously been shown to influence (or be influenced by) health status, such as Ruminococcaceae. The values for each axis are only shown in A to facilitate viewing. These plots were built using a rarefaction depth of 440 sequences per sample to account for as many samples as possible (only two samples were left out using this rarefaction depth). The UniFrac data was obtained from the closed99 approach but other approaches may reveal other patterns (see Table 4 and Table 5).
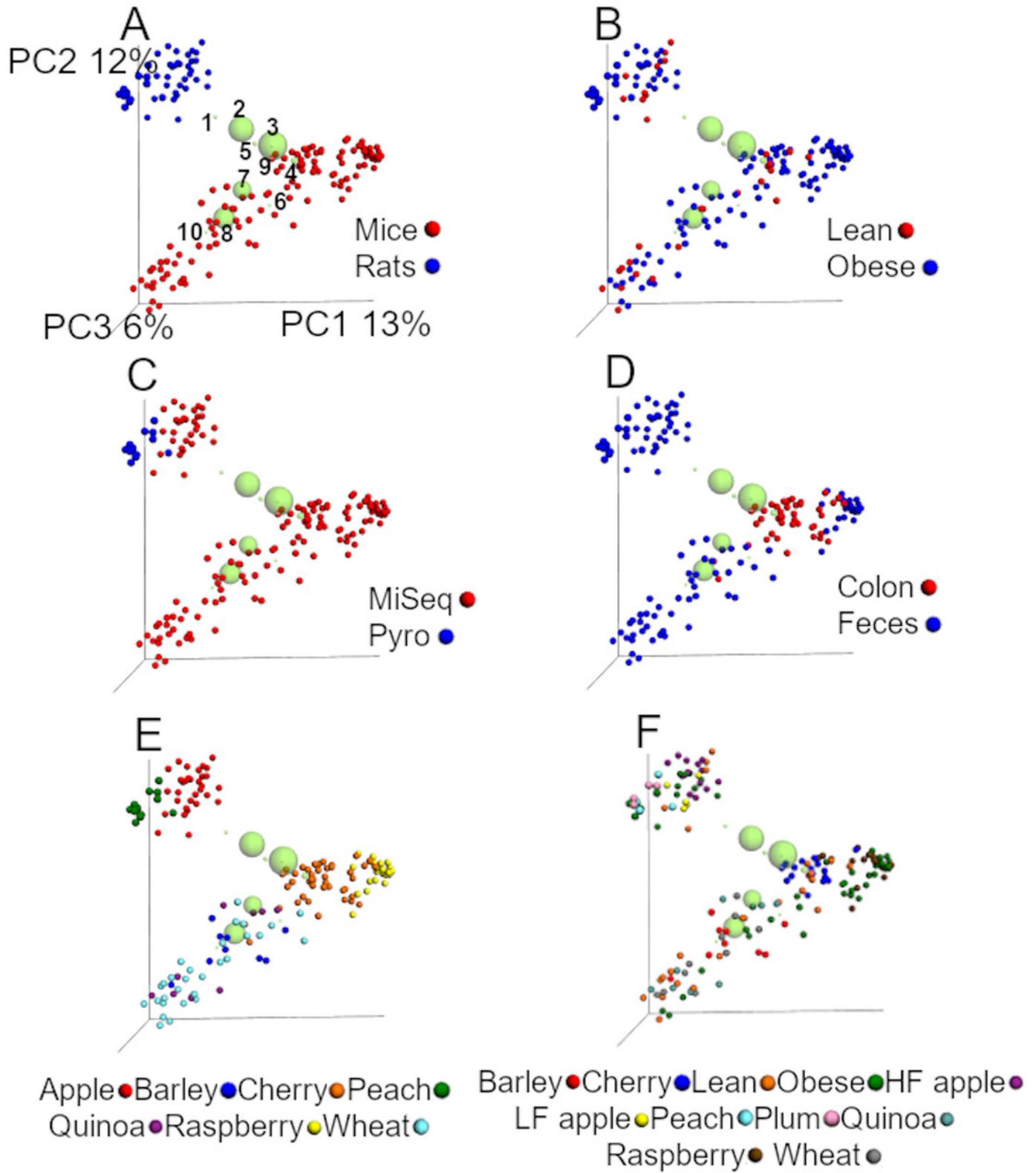
Table 1(on next page)

List of publications from which the data for this study came from.

Table 1 List of publications from which the data for this study came from†.					
Publication	Number of samples	SRA Bioproject	Animal species	Sequencing Technique	Reported OTUs
Peach (*Noratto et al. 2014*)	12	PRJNA217444	Zucker rats (obese)	Pyrosequencing	1,549 (approach not mentioned)
Wheat (*Garcia-Mazcorro et al. 2016*)	30	PRJNA281761	db/db mice	MiSeq (Illumina)	8,686 (open)/1,302 (closed)
Quinoa (*Garcia-Mazcorro, Mills & Noratto 2016*)	9[a]	PRJNA299688	db/db mice	MiSeq	5,774 (open)
Barley (*Garcia-Mazcorro et al. 2017*)	10[a]	PRJNA314690	db/db mice	MiSeq	5,366 (open)
Cherry (*Garcia-Mazcorro et al. 2018*)	44[b]	PRJNA407462	db/db mice (colon microbiota)	MiSeq	Not mentioned[c]
Raspberry (*Garcia-Mazcorro et al. 2018*)	27	PRJNA415476	db/db mice	MiSeq	675[d] (open)
Apple (*Garcia-Mazcorro et al. 2019*)	32	PRJNA504388	Dawley Sprague rats	MiSeq	69,010 (1,339[d]) (open)

†In the original published articles, all of these studies used the 97% OTU files and a 97% similarity threshold. The compositional information about all diets is provided as Supplementary Information.
[a]These studies (quinoa and barley) used the lean and obese controls from the paper published about whole-wheat by Garcia-Mazcorro et al. (2016). [b]The associated BioProject does not contain all of these samples. [c]The original cherry study did not report the number of OTUs. In this work, we report 2,439 to 138,203 OTUs for the cherry study, depending on the approach (see Table 2). [d]After removal of very low abundant OTUs (i.e. OTUs with <0.005% of all reads). By default, the open approach in QIIME discards singletons (i.e. OTUs that appear only once).

1

Table 2(on next page)

Number of OTUs and relative proportions (i.e. percentages of 16S rRNA gene sequences) of the most abundant phyla.

Table 2 Number of OTUs and relative proportions (i.e. percentages of 16S rRNA gene sequences) of the most abundant phyla.

Peach study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	758 ↓	114	50.39%	40.84% ↑	6.09% ↑	2.37% ↑	0.072% ↑	0.128% ↓	0.065% ↓
Closed99	1,074	192	52.35% ↑	39.31%	5.77%	2.23%	0.069%	0.148%	0.077%
De novo	1,549	147	50.14% ↓	35.95%	4.28% ↓	1.67% ↓	0.042%	1.766%	0.112%
Open97	1,603	NA	50.18%	35.15% ↓	4.29%	1.73%	0.040% ↓	1.766%	0.113% ↑
Open99	1,680 ↑	NA	50.15%	39.95%	4.29%	1.73%	0.042%	1.775% ↑	0.112%
SD	NA	NA	0.88	2.45 ↑	0.85	0.31	0.02 ↓	0.85	0.02

Wheat study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	1,302 ↓	414	45.81% ↓	32.52%	13.41% ↑	5.89% ↑	2.25% ↓	0.0249%	0.0282% ↓
Closed99	2,008	734	65.97% ↑	21.29% ↓	7.64%	2.06%	2.90%	0.0161% ↓	0.0456%
De novo	37,474 ↑	28,466	57.96%	31.82%	3.18% ↓	0.81% ↓	3.46%	0.3405%	0.0927%
Open97	8,686	NA	58.73%	32.78%	3.22%	0.85%	3.53% ↑	0.3471%	0.0912%
Open99	9,013	NA	58.73%	32.79% ↑	3.21%	0.85%	3.53% ↑	0.3476% ↑	0.0908%
SD	NA	NA	6.51 ↑	4.53	4.19	2.03	0.52	0.17	0.03 ↓

Quinoa study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	1,062 ↓	333	55.34% ↓	30.29% ↑	4.68% ↑	7.20% ↑	2.20% ↓	0.051%	0.050%
Closed99	1,606	578	69.77% ↑	21.32% ↓	2.73%	3.29%	2.61%	0.045% ↓	0.070% ↑
De novo	17,046 ↑	11,500	63.70%	27.53%	2.45% ↓	1.04% ↓	3.30%	0.280% ↑	0.041%
Open97	5,774	NA	64.08%	28.18%	2.46%	1.07%	3.35%	0.275%	0.037%
Open99	5,976	NA	64.07%	28.17 %	2.46%	1.07%	3.36% ↑	0.275%	0.035% ↓
SD	NA	NA	4.62 ↑	3.04	0.89	2.49	0.49	0.12	0.01 ↓

Barley study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	1,078 ↓	342	53.50% ↓	32.38%	5.09% ↑	6.63% ↑	2.29% ↓	0.0157%	0.022% ↓
Closed99	1,586	572	62.54% ↑	29.04% ↓	3.01%	2.67%	2.60%	0.0131% ↓	0.053%
De novo	15,599 ↑	10,429	55.65%	36.03%	2.08%	1.20% ↓	3.25%	0.2859%	0.106% ↑
Open97	5,366	NA	55.86%	36.75%	2.07%	1.23%	3.32%	0.2809%	0.104%
Open99	5,594	NA	55.85%	36.77% ↑	2.06% ↓	1.23%	3.32% ↑	0.2812%	0.102%
SD	NA	NA	3.09	3.14 ↑	1.21	2.17	0.45	0.14	0.04 ↓

Cherry study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	2,439 ↓	1,942	50.89%	36.60%	6.75% ↑	4.67%	0.180%	0.000% ↓	0.000% ↓
Closed99	4,217	3,226	51.68% ↑	36.00% ↓	6.18%	5.11% ↑	0.172% ↓	0.012%	0.001%
De novo	138,203 ↑	64,000	47.27% ↓	43.20% ↑	4.80%	3.46%	0.208%	0.079%	0.002%
Open97	69,658	NA	48.53%	41.71%	4.73% ↓	2.90% ↓	0.208%	0.096% ↑	0.002% ↑
Open99	70,886	NA	48.53%	41.75%	4.73% ↓	2.90% ↓	0.209% ↑	0.001%	0.002%
SD	NA	NA	1.68	3.12 ↑	0.90	0.98	0.02	0.05	0.00 ↓

Raspberry study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	2,751 ↓	722	44.99% ↓	40.66% ↑	13.33% ↑	0.052% ↑	0.092%	0.002% ↓	0.001% ↓
Closed99	4,433	1,231	45.13%	40.53%	13.32%	0.052% ↑	0.092%	0.002% ↓	0.001% ↓
De novo	92,486 ↑	69,306	46.13% ↑	39.40% ↓	12.90% ↓	0.049% ↓	0.097% ↑	0.005% ↑	0.002% ↑
Open97	21,243	NA	46.10%	39.73%	12.99%	0.049% ↓	0.088% ↓	0.004%	0.001% ↓
Open99	21,834	NA	46.10%	39.72%	12.99%	0.049% ↓	0.088% ↓	0.004%	0.001% ↓
SD	NA	NA	0.55	0.56 ↑	0.19	0.002	0.004	0.001	0.0004 ↓

Apple study

Approach	OTUs	Singletons	Firmicutes	Bacteroidetes	Proteobacteria	Verrucomicrobia	Actinobacteria	Tenericutes	Cyanobacteria
Closed97	2,095 ↓	510	55.32% ↓	39.17% ↑	3.48% ↓	0.499%	0.270%	0.002% ↓	0.542% ↑
Closed99	3,363	907	55.43%	38.74%	3.97% ↑	0.510% ↑	0.246% ↓	0.002% ↓	0.448%
De novo	153,681 ↑	81,032	61.48%	28.63% ↓	3.89%	0.279% ↓	0.432%	1.517%	0.361%
Open97	69,010	NA	61.84% ↑	29.28%	3.96%	0.282%	0.435%	1.578% ↑	0.364%
Open99	70,056	NA	61.83%	29.61%	3.96%	0.282%	0.436%	1.577%	0.364%
SD	NA	NA	3.25	5.14 ↑	0.19	0.12	0.09	0.79	0.07 ↓

The symbols (↑) and (↓) are used for the highest and the lowest value of each dataset, respectively. This data was obtained from the analysis of sequences using a 97% similarity against the reference OTU files.

1

Table 3(on next page)

Number of OTUs that were significantly different accordingly to the different factors.

1

Table 3 Number of OTUs that were significantly different* accordingly to the different factors.		
Factor	**FDR**	**Bonferroni**
Study (7 levels)	309	174
Treatment (11 levels)	278	105
Sequencing (two levels)	225	112
Animal model (two levels)	134	53
Anatomical site (two levels)	31	14
Obesity (two levels)	4	3
*Defined as an adjusted p<0.01 in non-parametric Kruskal-Wallis (adjusted for False Discovery Rate and Bonferroni).		

Table 4(on next page)

Summary of results for all samples (n=162) from the Adonis and ANOSIM tests for comparing categories using UniFrac data from the closed97 approach.

Table 4. Summary of results for all samples (n=162) from the Adonis and ANOSIM tests for comparing categories using UniFrac data from the closed97 approach.				
	Adonis		ANOSIM	
	Unweighted	Weighted	Unweighted	Weighted
Study	$P < 0.001$ 24.8%	$P < 0.001$ 19.8%	$P = 0.001$ 66.3%	$P = 0.001$ 31.5%
Model	$P < 0.001$ 8.8%	$P < 0.001$ 7.4%	$P = 0.001$ 56.7%	$P = 0.145$ 4.5%
Sequencing	$P < 0.001$ 5.5%	$P < 0.01$ 3.2%	$P = 0.001$ 57.8%	$P = 0.612$ -3.2%
Treatment	$P < 0.001$ 20.4%	$P < 0.001$ 21.1%	$P = 0.001$ 34.2%	$P = 0.001$ 20.1%
Obesity	$P < 0.001$ 1.9%	$P < 0.001$ 4.4%	$P = 0.023$ 9.5%	$P = 0.001$ 23.6%
Site	$P < 0.001$ 3.8%	$P < 0.01$ 2.2%	$P = 0.702$ -1.9%	$P = 0.868$ -4.7%
A rarefaction depth of 100 sequences per sample to account for as many samples as possible (only two samples were left out using this rarefaction depth). A total of 999 permutations were used to calculate the statistics.				

1

Table 5(on next page)

Summary of results for all samples (n=163) from the Adonis and ANOSIM tests for comparing categories using UniFrac data from the closed99 approach.

Table 5. Summary of results for all samples (n=163) from the Adonis and ANOSIM tests for comparing categories using UniFrac data from the closed99 approach.				
	Adonis		ANOSIM	
	Unweighted	Weighted	Unweighted	Weighted
Study	$P < 0.001$ 33.2%	$P < 0.001$ 33.4%	$P = 0.001$ 80.1%	$P = 0.001$ 43.5%
Model	$P < 0.01$ 11.9%	$P < 0.001$ 10.4%	$P = 0.001$ 53.9%	$P = 0.016$ 9.8%
Sequencing	$P < 0.001$ 6.7%	$P < 0.001$ 3.7%	$P = 0.001$ 61.6%	$P = 0.529$ -2%
Treatment	$P < 0.001$ 24.6%	$P < 0.001$ 28.2%	$P = 0.001$ 33.8%	$P = 0.001$ 25%
Obesity	$P < 0.001$ 1.8%	$P < 0.001$ 5.5%	$P = 0.237$ 3.1%	$P = 0.001$ 20.6%
Site	$P < 0.001$ 5.1%	$P < 0.01$ 2.4%	$P = 0.960$ -5.3%	$P = 0.953$ -6.3%
A rarefaction depth of 440 sequences per sample to account for as many samples as possible (only one sample was left out using this rarefaction depth). A total of 999 permutations were used to calculate the statistics.				

1