Korean Journal of Microbiology (2019) Vol. 55, No. 3, pp. 220-225
DOI https://doi.org/10.7845/kjm.2019.9058

pISSN 0440-2413
eISSN 2383-9902

Comparison between DNA- and cDNA-based gut microbial community analyses using 16S rRNA gene sequences§

Hyejun Jo†, Jiwan Hong†, and Tatsuya Unno*

Faculty of Biotechnology, College of Applied Life Sciences, Jeju National University, Jeju 63243, Republic of Korea

16S rRNA 유전자 서열 분석을 이용한 DNA 및 cDNA 기반 장내 미생물 군집 분석의 비교§

조혜준† · 홍지완† · 운노타쓰야*

제주대학교 생명자원과학대학 생명공학부

(Received June 14, 2019; Revised July 3, 2019; Accepted July 5, 2019)

Studies based on microbial community analyses have increased in the recent decade since the development of next generation sequencing technology. Associations of gut microbiota with host's health are one of the major outcomes of microbial ecology filed. The major approach for microbial community analysis includes the sequencing of variable regions of 16S rRNA genes, which does not provide the information of bacterial activities. Here, we conducted RNA-based microbial community analysis and compared results obtained from DNA- and its cDNA-based microbial community analyses. Our results indicated that these two approaches differed in the ratio of Firmicutes and Bacteroidetes, known as an obesity indicator, as well as abundance of some key bacteria in gut metabolisms such as butyrate producers and probiotics strains. Therefore, cDNA-based microbial community may provide different insights regarding roles of gut microbiota compared to the previous studies where DNA-based microbial community analyses were performed.

Keywords: cDNA, gut microbiota, microbial community analysis, miSeq

Since the development of the next generation sequencing (NGS) technology, studies involving unculturable microbiota has dramatically increased, especially related to intestinal microbiota. More than a decade ago, Turnbaugh *et al*. (2006) reported that differences in the gut microbiota affect host's energy harvesting capacity. In addition, other studies indicated significant associations of gut microbiota with intestinal diseases such as diabetes (Houghton *et al.*, 2018), inflammatory bowel disease (Eom *et al.*, 2018), and colon cancer (Jahani-Sherafat *et al.*, 2018). Furthermore, recent studies have found that gut microbiota is also associated with liver disease (Vassallo *et al.*, 2015) and brain-related diseases such as autism (Liu *et al.*, 2019) and Parkinson's disease (Hopfner *et al.*, 2017).

In general, gut microbiota analysis is done by sequencing one or two of the variable regions of 16S rRNA genes using NGS sequencers (i.e., MiSeq). In many cases, feces are considered as a sink of gut microbes, thus fecal DNA is used to amplify 16S rRNA gene for microbial community analyses. Obtained sequences are trimmed and filtered using bioinformatics software such as QIIME (Caporaso *et al.*, 2010) and MOTHUR (Schloss *et al.*, 2009). Clean reads were clustered to assign operational taxonomic units (OTUs), then the distribution of OTUs are used for diversity analyses to investigate microbial

† These authors contributed equally to the study.
***For correspondence.** E-mail: tatsu@jejunu.ac.kr;
Tel.: +82-64-754-3354; Fax: +82-64-756-3351
§Supplemental material for this article may be found at http://www.kjom.org.

community shifts and differentially abundant OTUs.

Probiotics such as *Bifidobacterium* and *Lactobacillus* are beneficial bacteria to hosts' health. NGS-based microbial community studies have revealed that other beneficial bacteria such as *Roseburia*, *Anaerstipes*, *Ruminococcus*, *Butyricicoccus*, etc. (Vital *et al.*, 2017) known as butyrate-producing bacteria. Moreover, some bacteria that reside in the mucosal layer also play important key roles in regulating intestinal immune function (Eom *et al.*, 2018). Although the importance of these bacteria has been well documented, these bacteria are often present with low abundance and only detected the significant differential abundance through bioinformatics tools developed for biomarker identification. For examples, effects of anti-obese functional food are often evaluated by being applied to animals, followed by differential abundance tests.

It may be difficult to reveal the changes caused by functional food if it only activates certain species of beneficial bacteria whose abundance is extremely low. DNA based microbial community analysis may not be sensitive enough to capture the small quantitative differences likely due to the PCR-bias. RNA-seq is a method to investigate differentially expressed genes (DEGs), which is basically the sequencing of cDNA. We hypothesized using cDNA for 16S rDNA microbial community analysis may help us in comparing activeness of bacteria. In this study, we compare how results of the DNA-based and cDNA-based microbial community analyses turned out differently.

Materials and Methods

Animal experiment

C57BL/6J female mice were raised at the Laboratory Animal Center Jeju National University under conditions of light and shade with 12-h cycle, 23 ± 2°C of temperature and 50 ± 5% of relative humidity, which was approved by the Jeju National University Institutional Animal Care and Use Committee (JNU-IACUC2016-0042). Animals were acclimated for a week and housed in cages with food (normal diet) and water *ad libitum* for 12 weeks.

Fecal DNA extraction and 16S rRNA gene based microbial community analysis

We collected feces of mice at 12 weeks and froze them immediately using liquid nitrogen. Feces were kept under -80°C until used for DNA and RNA extraction. Fecal DNA was extracted using QIAamp PowerFecal DNA Kit (Qiagen) and fecal RNA was extracted using QIAamp RNeasy Power Microbiome Kit (Qiagen). We synthesized cDNA from RNA using PrimeScript™ 1st strand cDNA Synthesis Kit (TaKaRa). For 16S rRNA gene sequencing library preparation, the V4 region was amplified by PCR. PCR was performed using a KAPA HiFi HotStart ReadyMix PCR Kit (KAPABIOSYSTEMS) as follows: 95°C for 2 min, 25 cycles of 95°C for 20 sec, 55°C for 15 sec, and 72°C for 1 min, and 72°C for 5 min. The obtained PCR products were purified using an Agencourt AMPure XP (Beckman Coulter, Inc.). Illumina index was added by another PCR, and final PCR products were purified. After measurement DNA concentration using Qubit fluorometer (Invitrogen Inc.), we pooled equi-molar PCR amplicons and stored -20°C until sent for sequencing with MiSeq at Macrogen Inc.

Sequence processing and microbial community analysis

MiSeq output were processed according to the standard operational protocol defined (https://www.mothur.org/wiki/MiSeq_SOP) using MOTHUR (Schloss *et al.*, 2009). Briefly, paired-end assembly was done with make.contigs, aligned it with the SILVA Database (Quast *et al.*, 2013). After eliminating the singleton, we used the pre.cluster subroutine to correct the error for rare sequences, and then we removed the chimeric sequences using VSEARCH (Rognes *et al.*, 2016). Taxonomic classification was done using Greengenes_13_5 database (McDonald *et al.*, 2012). Sequences classified to undesired taxa (i.e., Chloroplast and Mitochondria) were removed using remove.lineages MOTHUR subroutine and clustered using OptiClust (Westcott and Schloss, 2017) to assign operational taxonomic units (OTU) at dissimilarity 0.03. Distribution of OTUs were subsequently used to calculate ecological indices (i.e., richness and evenness). Distance between each sample was calculated using Bray-curtis coefficient. Non-metric multidimensional scaling (NMDS) analysis and analysis of molecular variance (AMOVA) were performed using nmds and amova MOTHUR subroutines, respectively. Differential

abundance analysis was performed using the linear discriminant analysis effect size (LEfSe) (Segata *et al.*, 2011).

Results and Discussion

Sequencing data quality analysis

In this study, we have obtained 1,161,820 clean reads (min 16,280, max 105,502 reads per sample). To normalize the number of reads per sample, 10,000 reads were randomly sub-sampled from each. Rarefaction curve analysis showed 10,000 reads per sample should be enough to capture a large portion of OTUs in each sample (Supplementary data Fig. S1A). In addition, Good's coverage was more than 99% in all samples (Supplementary data Fig. S1B). These results suggest that sequencing effort performed in this study was enough to estimate majorities of species in each sample.

Comparison of DNA- and cDNA-based diversity analyses

Results in Fig. 1 show that species richness (Chao) was not significantly different between DNA and cDNA groups, while species evenness (Shannon) was significantly higher in cDNA groups compared to that of DNA. These results suggest that species abundance in cDNA groups are more evenly distributed. In other words, the abundance of rare species in DNA groups was increased, while the abundance of the major species was decreased. Since the abundance of cDNA is linked to the bacterial activity, our results may suggest that rare species might be more active than we expected from the DNA abundance and vice versa. In addition, results from NMDS analysis indicate significant difference between DNA and cDNA groups ($P < 0.001$) (Supplementary data Fig. S2).

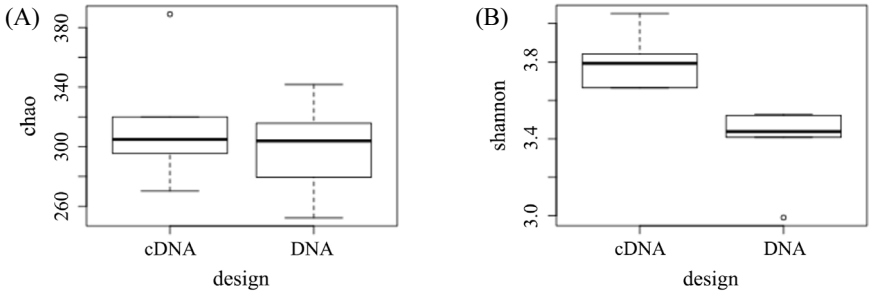

Fig. 1. Analysis of ecological indices for species richness (A) and evenness (B).

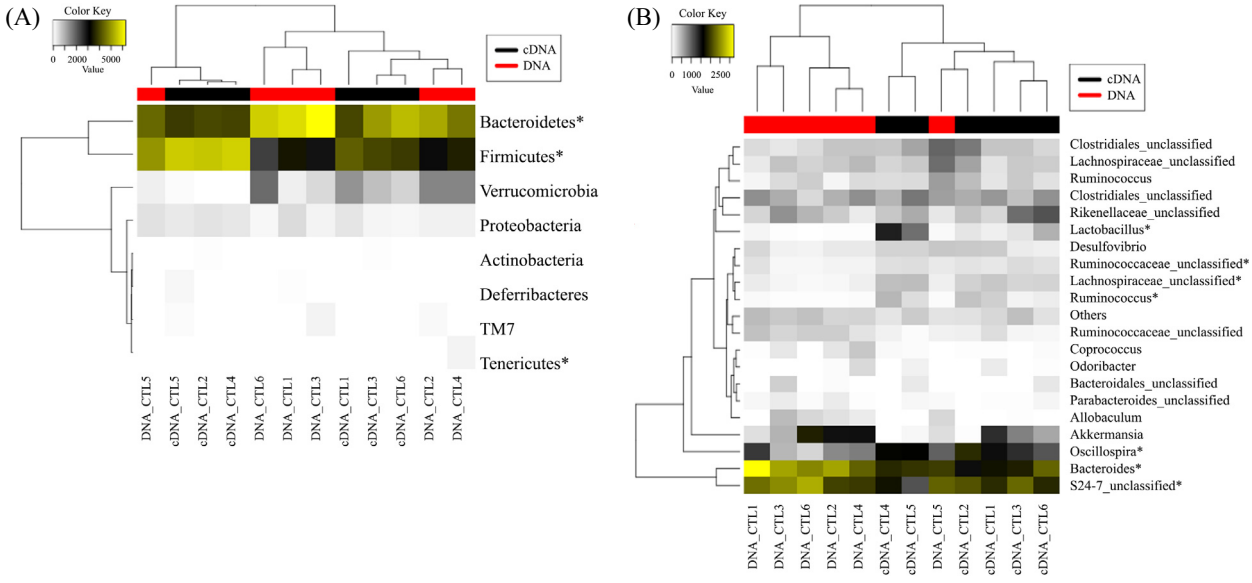

Fig. 2. Taxonomic composition analyses at the phylum level (A) and the genus level (B).
*indicates significant difference between DNA and cDNA groups

Results in Fig. 2A suggest that cDNA samples showed a significantly higher abundance of Firmicutes and lower abundance of Bacteroidetes. Previous studies had reported that the higher ratio of Firmicutes/Bacteroidetes indicate obese-type microbiota (Ley *et al.*, 2006), while there are contradicting reports regarding the use of Firmicutes/Bacteroidetes ratio as an obese indicator (Duncan *et al.*, 2008). Perhaps, this discrepancy could be solved by considering the fact that the simple ratio of these two phyla does not consider "bacterial activities." Applying cDNA-based microbial community analysis to investigate the effects of diet on gut microbiota and the host's physiological functions, we may find new roles of gut microbiota.

Results in Fig. 2B show that DNA and cDNA groups are separately clustered at the genus level. In cDNA samples, significantly higher abundance was observed for two unclassified genera in the family Ruminococcoaceae and Lachnospiraceae, as well as the genus *Lactobacillus*, *Ruminococcus*, and *Oscillospira* ($P < 0.05$). On the other hand, one unclassified genus in the family S24-7 and the genus *Bacteroides* were found significantly more abundant in DNA samples ($P < 0.05$). As suggested in the

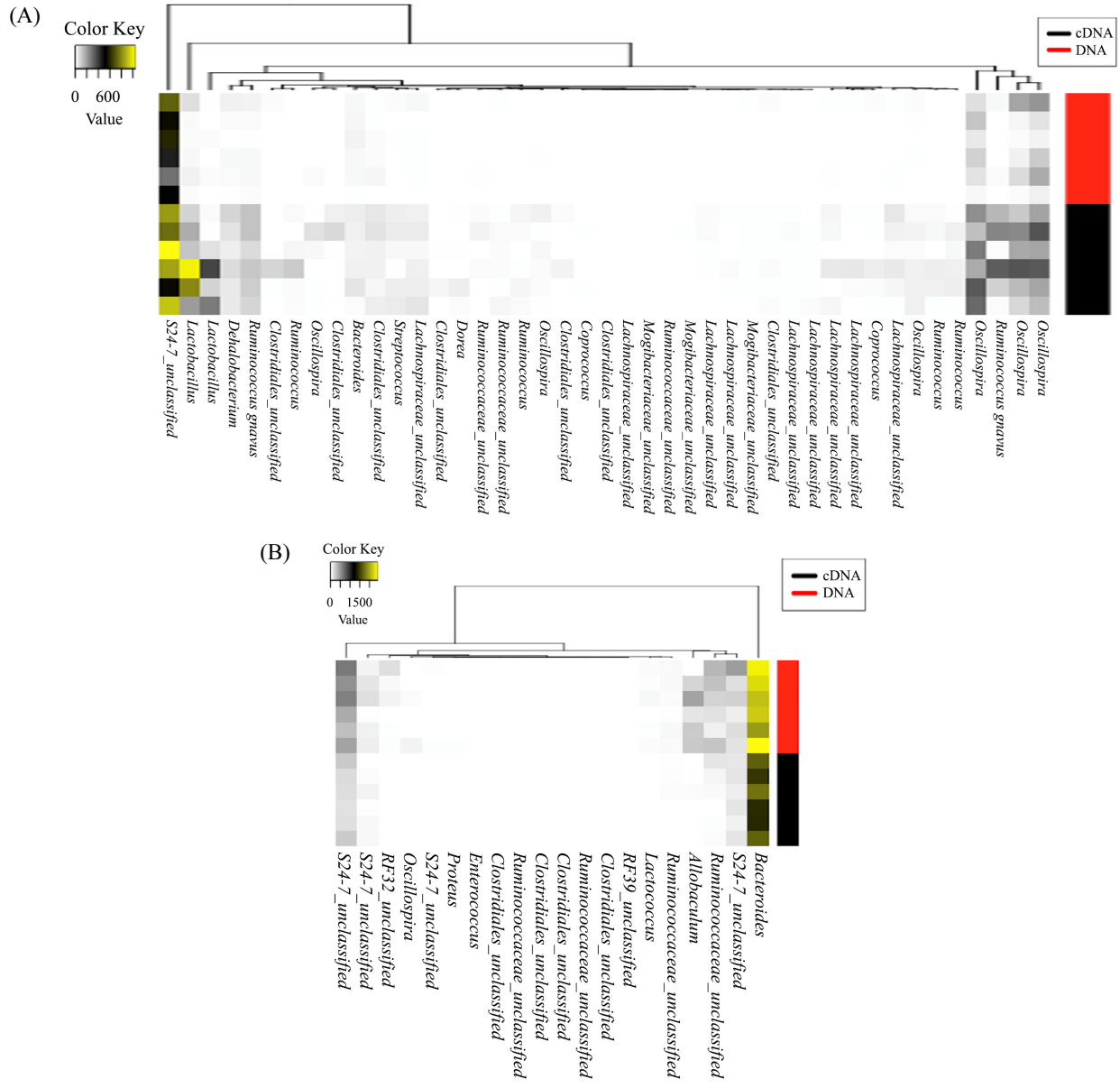

Fig. 3. Differentially more abundant OTUs in cDNA (A) and DNA (B).

Shannon diversity index (Fig. 1B), the abundant bacteria S24-7 and Bacteroides in DNA samples were shown less abundant in cDNA groups, likewise, rare species in the DNA samples were shown more abundant in cDNA groups. It should be noted that three genera found more abundant in cDNA samples are often reported as key bacteria that regulate host's metabolic functions. *Lactobacillus* is a well-known probiotic bacteria, *Oscillospira* has been reported to be associated host's physiological functions and leanness (Konikoff and Gophna, 2016), and *Ruminococcus* are butyrate-producing bacteria (Vital *et al.*, 2017). Those bacteria that are associated with the hosts' physiological functions are expected to be active although low in abundance, therefore, the use of cDNA may enable allows us to find those important rare bacterial species.

According to the differential abundance test, there were 43 and 20 OTUs significantly abundant in cDNA and DNA groups, respectively (Fig. 3). Significantly more abundant OTUs in cDNA group include unclassified OTUs in the order Clostridiales (n = 7) and the family Lachnospiraceae (n = 9), while including 5 OTUs in the genus *Oscillospira*, 5 OTUs in the genus *Ruminococcus*, and 2 OTUs in the genus *Lactobacillus* (Fig. 3A). Interestingly, a total of the unclassified genus in the family S24-7 was less abundant in cDNA, but the abundance of one of the S24-7 OTUs was significantly higher in cDNA. On the other hand, the abundant OTUs in DNA group include 4 unclassified OTUs in the family S24-7 and 4 unclassified OTUs in Ruminococcaceae, and other 6 genera, where the highest abundance difference was observed for the genus *Bacteroides* followed by the minor abundance differences for *Allobaculum*, *Lactococcus*, *Enterococcus*, *Proteus*, and one of the *Oscillospira* OTUs (Fig. 3B). *Bacteroides* are rich in polysaccharide utilization loci (PULs) so that they can digest fibers and provide substrates to other bacteria in the gut (Wexler and Goodman, 2017), therefore, they are likely get activated when undigestible food such as dietary fibers are injected. On the other hand, *Bacteroides* is also known as a pathogen outside the intestinal environment due to their virulence factors (Wexler, 2007), suggesting that *Bacteroides* may not be active in the gut. In addition, *Bacteroides* can degrade O-glycan, a major component of mucin, and resides within the mucosal layer, thus activation of this bacterium may cause intestinal inflammation. For these reasons, we assume *Bacteroides* may not be that active than we expect from the abundance data. Nonetheless, it is worth mentioning that *Bacteroides* is the key player in gut metabolic functions and their abundance may significantly differ between DNA- and cDNA-based microbial community analyses, suggesting that the use of cDNA may provide important information of the metabolic changes derived from gut microbiota shifts.

In summary, our results suggest that cDNA based microbial community analysis showed different results from that of DNA-based. The discrepancies included abundance of major fecal phyla (i.e., Firmicutes and Bacteroidetes) and several key bacteria such as SCFA producers and probiotic strains. While cDNA-based microbial community analysis is not an accepted protocol for the studies regarding gut microbiota, the study suggest it is worth investing further, which in future may provide new aspects of roles of gut microbiota.

적 요

최근 10년간 미생물생태분석 기반의 연구는 차세대염기서열분석법이 개발된 이래로 지속적으로 증가하고 있다. 장내미생물생태와 건강의 연관성은 미생물 생태학 분야에 있어서 중요한 결과로 여겨지고 있다. 미생물 군집 분석은 주로 16S rRNA 유전자 가변 영역의 염기서열을 통해 분석되지만 이는 미생물의 활성 정보를 제공하지 않는다. 본 연구에서는 cDNA 기반의 미생물 생태분석을 수행하고 DNA 및 cDNA기반의 미생물생태분석 결과를 비교하였다. 두 가지의 서로 다른 접근법이 Butyrate producer와 probiotics와 같이 장내 대사과정에서 중요한 미생물의 abundance 뿐만 아니라 비만 지표로 알려진 Firmicutes 와 Bacteroidetes의 비율에 있어서 차이가 있음을 나타내었다. 따라서, cDNA 기반 미생물 군집은 이전에 수행된DNA 기반 미생물 군집 분석과 비교하여 장내미생물생태의 역할과 관련된 또 다른 분석 방향성을 제공한다.

Acknowledgements

This work was supported by the 2019 education, research and student guidance grant funded by Jeju National University.

References

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336.

Duncan SH, Lobley GE, Holtrop G, Ince J, Johnstone AM, Louis P, and Flint HJ. 2008. Human colonic microbiota associated with diet, obesity and weight loss. *Int. J. Obes. (Lond)* **32**, 1720–1724.

Eom T, Kim YS, Choi CH, Sadowsky MJ, and Unno T. 2018. Current understanding of microbiota- and dietary-therapies for treating inflammatory bowel disease. *J. Microbiol.* **56**, 189–198.

Hopfner F, Kunstner A, Muller SH, Kunzel S, Zeuner KE, Margraf NG, Deuschl G, Baines JF, and Kuhlenbaumer G. 2017. Gut microbiota in Parkinson disease in a northern German cohort. *Brain Res.* **1667**, 41–45.

Houghton D, Hardy T, Stewart C, Errington L, Day CP, Trenell MI, and Avery L. 2018. Systematic review assessing the effectiveness of dietary intervention on gut microbiota in adults with type 2 diabetes. *Diabetologia* **61**, 1700–1711.

Jahani-Sherafat S, Alebouyeh M, Moghim S, Ahmadi Amoli H, and Ghasemian-Safaei H. 2018. Role of gut microbiota in the pathogenesis of colorectal cancer; a review article. *Gastroenterol. Hepatol. Bed Bench* **11**, 101–109.

Konikoff T and Gophna U. 2016. *Oscillospira*: a central, enigmatic component of the human gut microbiota. *Trends Microbiol.* **24**, 523–524.

Ley RE, Turnbaugh PJ, Klein S, and Gordon JI. 2006. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023.

Liu F, Li J, Wu F, Zheng H, Peng Q, and Zhou H. 2019. Altered composition and function of intestinal microbiota in autism spectrum disorders: a systematic review. *Transl. Psychiatry* **9**, 43.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, and Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, and Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–596.

Rognes T, Flouri T, Nichols B, Quince C, and Mahe F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, *et al.* 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, and Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, and Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031.

Vassallo G, Mirijello A, Ferrulli A, Antonelli M, Landolfi R, Gasbarrini A, and Addolorato G. 2015. Review article: Alcohol and gut microbiota - the possible role of gut microbiota modulation in the treatment of alcoholic liver disease. *Aliment. Pharmacol. Ther.* **41**, 917–927.

Vital M, Karch A, and Pieper DH. 2017. Colonic butyrate-producing communities in humans: an overview using omics data. *mSystems* **2**, e00130-17.

Westcott SL and Schloss PD. 2017. OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. *mSphere* **2**, e00073-17.

Wexler HM. 2007. Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621.

Wexler AG and Goodman AL. 2017. An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 17026.