RESEARCH PAPER/REPORT

Analysis of gut microbiota – An ever changing landscape

Michael J. Sadowsky a,b, Christopher Staley a, Cheryl Heiner c, Richard Hall c, Colleen R. Kelly d,e, Lawrence Brandt f, and Alexander Khorut g,h

a BioTechnology Institute, University of Minnesota, St. Paul, Minnesota, USA; b Department of Soil, Water & Climate, University of Minnesota, St. Paul, Minnesota, USA; c Pacific Biosciences, Menlo Park, CA USA; d Department of Medicine, The Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA; e Women’s Medicine Collaborative, The Miriam Hospital, Providence, Rhode Island, USA; f Department of Medicine, Division of Gastroenterology, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY; USA; g Department of Medicine, Division of Gastroenterology, University of Minnesota, Minneapolis, Minnesota, USA; h Center for Immunology, University of Minnesota, Minneapolis, Minnesota, USA

ABSTRACT

In the last two decades, the field of metagenomics has greatly expanded due to improvement in sequencing technologies allowing for a more comprehensive characterization of microbial communities. The use of these technologies has led to an unprecedented understanding of human, animal, and environmental microbiomes and have shown that the gut microbiota are comparable to an organ that is intrinsically linked with a variety of diseases. Characterization of microbial communities using next-generation sequencing-by-synthesis approaches have revealed important shifts in microbiota associated with debilitating diseases such as Clostridium difficile infection. But due to limitations in sequence read length, primer biases, and the quality of databases, genus- and species-level classification have been difficult. Third-generation technologies, such as Pacific Biosciences’ single molecule, real-time (SMRT) approach, allow for unbiased, more specific identification of species that are likely clinically relevant. Comparison of Illumina next-generation characterization and SMRT sequencing of samples from patients treated for C. difficile infection revealed similarities in community composition at the phylum and family levels, but SMRT sequencing further allowed for species-level characterization - permitting a better understanding of the microbial ecology of this disease. Thus, as sequencing technologies continue to advance, new species-level insights can be gained in the study of complex and clinically-relevant microbial communities.

KEYWORDS

Clostridium difficile; fecal microbial transplant; gut microbiota; next-generation sequencing; SMRT sequencing; PacBio

Over the last decade great advances have been made in our understanding of the structure, function, and ecology of microorganisms in natural systems, microbial community structure in a variety of habitats, and host- and microbe-microbe interactions. These advances have been achieved, in large part, by the development of molecular tools, chiefly next-generation DNA sequencing technologies, that allow for the determination of microbial taxa without the need for culture, as well as bioinformatic tools to process the billions of DNA sequences generated from these technologies. These technologies not only overcome issues related to the “great plate count anomaly,” in which <1% of microbial species in some matrices are able to be cultured in vitro, but provide unprecedented understanding of the complexity of numerous microbial systems, many of which were previously passed-over as overly complicated or simplified because of methodological limitations. Recently, however, more gut microbiota have been cultivated and there have been reports of the use of a variety of laboratory media and growth conditions, both aerobic and anaerobic, to culture a large number of microbial species identified by DNA-based methods.

These recent advances in sequencing technologies have also led to the creation of a new field, commonly referred to as metagenomics, where the DNA in any sample (e.g., from human body sites or the environment) can be sequenced and the data assembled into overlapping sequence reads called contigs that allow for description of microbial taxa (from assembly of 16S rRNA genes), metabolic pathways, and, in less complex communities, assembly of whole microbial genomes. This revolution is akin to Antonie van
Leeuwenhoek’s development of the microscope about 300 y ago, as it is also opening our eyes to a whole new microbial world.\(^1\) The development of metagenomic technologies and 16S rRNA amplicon analyses has also led to the unprecedented understanding of human, animal and environmental microbiomes and is paving the way for development of microbiota-based therapeutics to correct dysbiosis and correct a variety of ailments,\(^11\) as well as a better understanding of host-microbe genomic interactions leading to symbioses, plant and animal disease and restoration of function or enhancement of plant growth.\(^12\) However, a tradeoff currently exists between whole genomic shotgun (WGS) metagenomic and 16S-rRNA-amplicon sequencing methodologies.\(^1\) The WGS methodology allows for a less biased, albeit lower-coverage, characterization of all microbial taxa, independent of a single taxonomic marker gene (i.e., 16S rRNA), as well as functional genes. In contrast, 16S-rRNA-amplicon sequencing allows for greater taxonomic coverage, but is limited to biases associated with primer sequences and PCR, often limiting the taxa characterized to a single kingdom and the exclusion of many taxa.\(^13,14\)

Studying the human and environmental microbiome has led to the development of distinct subfields revolving around the human microbiome project (https://commonfund.nih.gov/hmp/overview)\(^15\) and the earth microbiome project (http://www.earthmicrobiome.org/)\(^16\) that specialize in understanding the microbiota in various habitats, e.g., the gut, ear, vagina, skin, soils, rivers, sand, etc. While these projects have rapidly advanced our understanding of various niches, they often suffer from overstating taxonomic significance and relevance as well as a lack of the ability to definitively equate structure/function relationships. These limitations result, in large part, due to low taxonomic resolution resulting from methodological limitations, chiefly read length, but also sequencing depth, sequence read quality and the quality of taxonomic databases.

Despite these shortcomings, we have come a long way in under a decade, owing to rapid advances in first- and second-generation high-throughput sequencing technologies. First-generation technologies, such as the Roche 454 pyrosequencing platform,\(^3\) were able to produce 100 Mb paired-end runs in 7 hours with a read length of 500 bp at a cost of about $80 per Mb. This emulsion-PCR-based approach was most limited by the number of samples that could be run at the same time and relatively high cost per sample, although high-resolution taxonomic classification was possible owing to the relatively long read length (~700 bp). While this read length seemed promising among other early platforms i.e., the Illumina HiSeq platform (up to 200 bp from paired-end reads) or the Applied Biosystems SOLiD system (~80 bp),\(^17\) declining costs and improved sequencing chemistries favoring longer read lengths have since led to a departure from the emulsion-based approach.

Metagenomics and 16S rRNA amplicon analyses, in contrast, were rapidly advanced by the development of second-generation technologies (often referred to as “next”-generation) revolving around Illumina-based PCR and sequencing-by-synthesis approaches.\(^18\) The Illumina MiSeq and HiSeq platforms can produce 1300 Mb in a single run, with read lengths up to 300 bp, in 3 d with a cost approaching $6.00 per Mb. New instruments (e.g., the HiSeq2500 platform), improved chemistries, and the ability to barcode and multiplex samples for later bioinformatics-binning and independent analyses continue to improve the amount and quality of sequence data generated. This technology will ultimately be limited by read length maxima and long processing times, due in part to the large number of sample reads produced per run.

Key to understanding microbiomes, and thus the underlying microbial ecology of any system, is the taxonomic resolution possible using a given technique. Such resolution in 16S rRNA amplicon sequencing is due to primer choice, which induces some degree of bias and may not target archaea, as well as the reference databases that can be used to assign taxonomic status to operational taxonomic units (OTUs).\(^13,14,19,20\) Most studies use hypervariable regions V5/V6, V3/V4, or V4, which produce the best accuracy relative to mock communities,\(^14\) and the most commonly used V4 region primers are also inclusive of archaea.\(^21\) The accuracy and resolution (taxonomic level to which OTUs can be assigned) of taxonomic assignment is limited by the length of the hypervariable region(s) of the 16S rRNA gene and the overall quality of the database. For example, characterization of the gut microbiota associated with fecal microbiota transplantation (FMT) to treat \textit{Clostridium difficile} infection in 2013, using Ribosomal Database Project (RDP) ver. 7.0 to classify V6 sequence data,
resulted in many more unclassified bacteria at the phylum level than did similar characterization in 2016 using V5-V6 sequence data and the RDP ver. 14 database.\textsuperscript{22,23} The improved database and longer read length also resulted in the ability to assign OTUs to a more specific taxonomic level, from phylum characterizations previously to family and genus level assignments currently. Early on, however, microarrays or phylorichs were the only reliable technologies available to assign sequence data to the genus and species level.\textsuperscript{24} This technology was limited to uncovering those taxa known to be present and species covered by the array, rather than also encompassing unknown microbes present in the sample.

Undoubtedly, modern microbiome analyses owe their successes to next-generation sequencing technologies that can rapidly resolve complex microbial communities, regardless of sample composition, to the taxonomic levels of families and often genera for relatively inexpensive single sample costs. However, third-generation (next-next-generation) technologies are now available, and these offer even greater taxonomic resolution than second-generation sequencing methodologies.\textsuperscript{25,26} While some of these technologies have mostly been used for sequencing of whole animal, plant, and microbial genomes, they are only beginning to be evaluated to a degree required to ascertain all their pros and cons in resolving complex microbiomes,\textsuperscript{27–29} such as those found in the gut or soil. One such technology is Pacific Biosciences’ (PacBios’) single molecule, real-time (SMRT) sequencing technology found in the RSII and the Sequel platforms.\textsuperscript{30,31} These sequencing platforms have advantages over previous technologies as they produce very long read lengths, allowing for potential species-level taxonomic resolution and unbiased sequencing results due to a lack of target amplification. Initial issues with high sequence error have been mostly overcome by oversampling until consensus sequences were obtained. Currently, both RSII and the Sequel platforms are in use, with the most recent Sequel platform producing read lengths of 8,000–13,000 bp with throughput at approximately 4 to 7 Gb per SMRT cell. The Sequel system has scalable throughput and produces roughly 7 times the reads as the RSII. Both systems allow for barcoded multiplex sequencing. Another newer technology, Oxford Nanopore Technologies’ MinION nanopore sequencer,\textsuperscript{32} has been tested for metagenomic detection of bacterial and viral pathogens, but extensive data on a variety of matrices are currently not available.

While there is little doubt that PacBios’ technology has been a great boon for sequencing the complex genomes present in eukaryotes, and for high-resolution, full genome sequencing of prokaryotes, little is known about the use and resolution of this technology for understanding the intestinal microbiota in patients having debilitating diseases, such as that caused \textit{C. difficile} infection (CDI). This bacterium produces a devastating, and often fatal, colitis that affects nearly 500,000 people per year in the US.\textsuperscript{33} Traditional antibiotic treatment using antibiotic therapies may be effective, but recurrence is seen in approximately 20–30% of cases due to the disruption of the native microbiota which may lead to continual infection.\textsuperscript{11} We have pioneered the use of FMT to treat this disease with success rates of \textasciitilde 96%.\textsuperscript{11} We now understand that restoration of normal intestinal function after FMT is due to replacement of healthful gut microbiota (primarily those species within the \textit{Firmicutes} and \textit{Bacteroidetes}) that fill all available niches, as well as an increase in secondary bile acids that limit \textit{C. difficile} spore germination and vegetative growth.\textsuperscript{34} Despite these advances, however, specific species that may be important for this functional restoration are unknown, in large part due to the limited taxonomic resolution of the Illumina-based sequencing technologies we have used to examine gut recolonization.\textsuperscript{22,24,35} Such limitation has led to the generation of often limited correlations and a lack of functionally based hypotheses, aside from those based on data obtained using predictive software such as PICRUSt.\textsuperscript{36}

To overcome these limitations, we examined the potential use of third-generation PacBio RSII- and Sequel-based sequence analyses of the microbiota from feces of patients receiving FMT to cure CDI. We hypothesized that PacBio sequence analysis would provide greater taxonomic resolution needed to better understand the role of specific microbiota in the cure of FMT. Both pre- and post-FMT samples were examined using the RSII platform, and Sequel results were compared against Illumina sequencing data from samples collected at 2 and 8 weeks after FMT was administered. Samples characterized by Sequel were taken from Rhode Island subject RS05 (female, age 60), further described in a recent randomized clinical trial.\textsuperscript{23} Briefly, for PacBio sequencing SMRTbell libraries were prepared using SMRTbell Template Prep Kit 1.0 as described previously (http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-2-kb-Template-Preparation-and-
The 2 and 3 kb pooled libraries were sequenced on the Sequel system, using Sequel Binding Kit 1.0, and Sequel Sequencing Kit v1.1. Libraries for RSII sequencing were prepared in the same manner as above, but were done on a different sample set. Libraries were sequenced using DNA/Polymerase Binding Kit P6 v2 and DNA Sequencing Reagent Kit 4.0 v2. A total of we 12 RS II SMRT Cells were run for the pre-FMT sample and 45 RS II SMRT Cells were run for the post-FMT samples. CCS (Circular Consensus) reads were generated using default parameters in the SMRT Link 3.1.1 application (Pacific Biosciences, Menlo Park, CA) and reads were aligned against the RefSeq genomic database using blastn. Blast results were imported into MEGAN (http://www-ab.informatic.uni-tuebingen.de/software/megan6/) and a lowest common ancestor (LCA) algorithm was used to assign a taxonomy to each sequence.

For pre-FMT samples 12 SMRT cells were used on the RSII system, producing 73,237 reads and 93,055,558 bases with a N50 CCS read length of 1,373. Post-FMT samples were sequenced on the RSII system using 45 SMRT cells and generated 905,267 reads comprising 662,421,070 bases and at N50 CCS read length of 739. In contrast, the Sequel system was used for Rhode Island clinical samples. Each sample was run on 1 SMRT cell and generated, on average, 80,540 (range 59,674–113,489) CCS reads, composed of 162,687,835 bases (range 125,466,241 to 153,874,067) and a N50 CCS read length of 2,171 (range 1,803 to 2,471).

When grouped broadly to class, sequencing and characterization of the fecal microbiota before FMT using third-generation PacBio RSII technology revealed similar taxonomic composition to that characterized by next-generation sequencing of the 16S rRNA gene and fluorescent in situ hybridization. By all methods, pre-FMT communities showed high relative abundances of Gammaproteobacteria, determined by microarray to consist primarily of the genera Enterobacter, Escherichia, Lactobacillus, Raoultella, and Veillonella among 3 patient samples. In a separate sample, characterized using RS II technology, the genera Klebsiella, Citrobacter, and Veillonella predominated in the pre-FMT community, and species-level classification revealed high abundances of Klebsiella oxytoca, Pseudomonas aeruginosa, and E. coli (Fig. 1).

In contrast, the community composition post-FMT appeared to depend on the technology used: sequencing (next-generation or PacBio) revealed expansions of both Firmicutes and Bacteroidetes whereas microarray characterization showed greater expansion of the clostridia with lower abundances of Bacteroidetes. At more specific resolution, the post-FMT community characterized by RS II technology showed expansion of the Bacteroides (primarily B. uniformis and B. cacao) and Alistipes (A. shahii and A. onderdonkii), while species within the Firmicutes were identified by microarray. Although different samples were characterized in each study, results suggest that, broadly, the sequencing results are similar across platforms, although it is not possible to determine whether more specific (genus and species) differences in taxonomic characterization result from sequencing method or individual sample.

We further compared fecal microbial communities characterized by Illumina 16S rRNA sequencing of the V5-V6 hypervariable regions, processed as described previously, with PacBio Sequel characterization performed as described above. For characterization on both platforms, DNA was extracted from the same fecal sample using the MoBio PowerSoil DNA extraction kit (MoBio Laboratories, Inc., Carlsbad, CA, USA). DNA was sequenced on both platforms and samples included a pre-FMT time point as well as 2- and 8-week time points after FMT using autologous stool (A-FMT) and donor stool (H-FMT; Table 1). The patient was treated for recurrent Clostridium difficile infection as documented in a previous study, had a recurrence of infection following the 8-week time point after A-FMT, and was subsequently treated by H-FMT.
Community diversity, characterized using the Shannon index (Table 1), tended to be higher among samples sequenced using PacBio technology, but differences between platform were not significant (ANOVA $P = 0.082$). Similarly, when A-FMT and H-FMT samples were grouped, differences in diversity between treatment groups did not differ ($P = 0.064$ and $0.149$ for Illumina and PacBio data, respectively). PacBio characterization of the pre-FMT sample revealed that *Klebsiella* and *Veillonella* were the predominant genera, while insufficient sequence data were obtained for characterization of this sample by Illumina. Furthermore, stark differences in community composition, when classified to genera (Fig. 2), were observed between platforms: there were greater relative abundances of *Bacteroides* and *Alistipes* observed in samples characterized by PacBio, while those characterized by Illumina showed greater abundances of families found at lower abundances. Notably, very low abundances of genera within the *Firmicutes* (a phylum typically prevalent in fecal communities) were observed when samples were characterized by PacBio.

Ordination of samples by principal coordinate analysis revealed similar separation of autologous FMT and heterologous FMT samples between platforms (Fig. 3), in agreement with results described previously for the larger study cohort. However, post-heterologous-FMT samples analyzed using PacBio were considerably more alike than when analyzed using Illumina, similar to taxonomic characterization.

**Figure 2.** Distribution of genera in samples characterized using PacBio Sequel (A) and Illumina (B) platforms. A-FMT: patients received autologous FMT; H-FMT: patients received heterologous FMT.

**Figure 3.** Principal coordinate analyses of genus compositions of samples characterized using PacBio Sequel (A) or Illumina (B) platforms. A-FMT: patients received autologous FMT; H-FMT: patients received heterologous FMT.

| Time point (platform) | PacBio | Illumina 16S rRNA amplicons |
|-----------------------|--------|-----------------------------|
| pre-FMT (RSII)        | 2.95   | ND                          |
| post-FMT (RSII)       | 4.65   | ND                          |
| pre-FMT (Sequel)      | 4.27   | ND                          |
| A-FMT 2 weeks (Sequel)| 3.14   | 1.46                        |
| A-FMT 8-weeks (Sequel)| 4.78   | 2.48                        |
| H-FMT 2-weeks (Sequel)| 5.97   | 4.25                        |
| H-FMT 8-weeks (Sequel)| 5.74   | 3.85                        |

**Note.**  
*A-FMT* – patients received autologous FMT; *H-FMT* – patients received heterologous FMT with donor material.  
$^1$ND: not determined.
Resolution of communities at the species level revealed a greater shift in community composition among autologous FMT samples than was observed in heterologous FMT samples (Fig. 4), where autologous FMT samples were characterized by fluctuations in abundance of species predominantly within the genus *Bacteroides*. In contrast, heterologous FMT communities appeared more taxonomically stable and were comprised of a highly abundant species of *Alistipes* and more consistent distribution of *Bacteroides* spp. than was observed in autologous FMT samples.

Taken together, these results suggest that PacBio has the sensitivity and accuracy to determine species level resolution of microbial taxa present in fecal samples, to a degree not possible using other technologies. Fecal microbiomes are relatively complex and are reported to contain 700–1000 species of bacteria and archaea. Despite this complexity, however, adequate species-level resolution was obtained in these preliminary studies, and a previous study has reported the utility of third-generation sequencing technology to characterize changes in the virome, as well.37 While these studies do not show significant difference in α diversity between the Illumina and PacBio platforms, there is a clear trend toward greater diversity of near significance being uncovered by the PacBio platform. Moreover, there was more variability in ordination of communities characterized using 16S amplicons, perhaps due to primer biases. Future studies may focus on cross-platform comparisons sequencing 16S rRNA amplicons alone (in which primer bias would be inherent) or comparison of Illumina shotgun sequencing data with results from third-generation platforms to better resolve the extent to which primer biases versus sequencing platforms influence discrepancies in taxonomic characterization. Nevertheless, new bioinformatic tools coupled with the PacBio Sequel platform with multiplexed samples hold promise in providing unparalleled species-level resolution of microbiota in complex communities.

**Disclosure of potential conflicts of interest**
No potential conflicts of interest were disclosed.

**ORCID**
Michael J. Sadowsky [http://orcid.org/0000-0001-8779-2781](http://orcid.org/0000-0001-8779-2781)

**References**

[1] Gilbert JA, Dupont CL. Microbial metagenomics: beyond the genome. Ann Rev Mar Sci 2011; 3:347-71.

[2] Staley C, Sadowsky MJ. Application of metagenomics to assess microbial communities in water and other environmental matrices. J Mar Biol Assoc United Kingdom 2016; 96:121-9.

[3] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bember LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 2005; 437:376-80; PMID:16056220

[4] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009; 75:7537-41; PMID:19801464

[5] Coporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010; 7:335-6; PMID:20383131

[6] Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol Rev 1995; 59:143-69; PMID:7535888

[7] Sommer MO. Advancing gut microbiome research using cultivation. Curr Opin Microbiol 2015; 27:127-32; PMID:26401902

[8] Lau JT, Whelan FJ, Herath I, Lee CH, Collins SM, Bercik P, Surette MG. Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. Genome Med 2016; 8:72; PMID:27363992

[9] Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 1998; 5:R245-9; PMID:9818143

[10] Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. Microbial diversity in...
the deep sea and the underexplored “rare biosphere.” Proc Natl Acad Sci U S A 2006; 103:12115-20; PMID:16880384

[11] Sadowsky MJ, Khoruts A. Faecal microbiota transplantation is promising but not a panacea. Nat Microbiol 2016; 1:16015; PMID:27572174

[12] Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Raup F, Huettel B, Reinhardt R, Schmelzer E, et al. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature 2012; 488:91-5; PMID:22859207

[13] Youssef N, Sheik CS, Krumholz LR, Najaf NZ, Roe BA, Elshahed MS. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl Environ Microbiol 2009; 75:5227-36; PMID:19561178

[14] Claesson MJ, Wang QO, O’Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O’Toole PW. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res 2010; 38:e200; PMID:20880993

[15] The Human Microbiome Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012; 486:207-14; PMID:22699609

[16] Gilbert JA, Meyer F, Jannson J, Gordon J, Jansson J, Tiedje J, Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law MJ. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012; 2012:251364.

[17] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 2012; 6:1621-4; PMID:22402401; http://dx.doi.org/10.1038/ismej.2012.8

[18] Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. Ribosomal Database Project: data and tools for high-throughput rRNA analysis. Nucleic Acids Res 2014; 42:D633-42; PMID:24288368; http://dx.doi.org/10.1093/nar/gkt1244

[19] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72:5069-72; PMID:16820507; http://dx.doi.org/10.1128/AEM.03006-05

[20] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 2012; 6:1621-4; PMID:22402401; http://dx.doi.org/10.1038/ismej.2012.8

[21] Hamilton MJ, Weingarden AR, Unno T, Khoruts A, Sadowsky MJ. High-throughput DNA sequence analysis reveals stable engraftment of gut microbiota following transplantation of previously frozen fecal bacteria. Gut Microbes 2013; 4:125-35; PMID:23333862; http://dx.doi.org/10.4161/gmic.23571

[22] Kelly C, Khoruts A, Staley C, Sadowsky M, Abd M, Alani M, Bakow B, Curran P, McKenney J, Tisch A, et al. Fecal microbiota transplant prevents recurrence in multiply recurrent C. difficile. Ann Intern Med 2016; 165:609-16; PMID:27547925; http://dx.doi.org/10.7326/M16-0271

[23] Shankar V, Hamilton MJ, Khoruts A, Kilburn A, Unno T, Paliy O, Sadowsky MJ. Species and genus level resolution analysis of gut microbiota in Clostridium difficile patients following fecal microbiota transplantation. Microbiome 2014; 2:13; PMID: 24855561; http://dx.doi.org/10.1186/2049-2618-2-13

[24] Moore E, Yu KF. PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 2015; 13:278-89; PMID: 26542840; http://dx.doi.org/10.1016/j.gpb.2015.08.002

[25] Singer E, Bushnell B, Coleman-Derr D, Bowman B, Bowers RM, Levy A, Gies EA, Cheng J-F, Copeland A, Klenk H-P, et al. High-resolution phylogenetic microbial community profiling. ISME J 2016; 10:2020-32; PMID:26859772; http://dx.doi.org/10.1038/ismej.2015.249

[26] Marshall CW, Ross DE, Fichot EB, Norman RS, May HD. Electrosynthesis of commodity chemicals by an autotrophic microbial community. Appl Environ Microbiol 2012; 78:8412-20; PMID:23001672; http://dx.doi.org/10.1128/AEM.02401-12

[27] Fichot EB, Norman RS, Loman N, Constantinidou C, Chan J, Halachev M, Sergeant M, Penn C, Robinson E, Pallen M, et al. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. Microbiome 2013; 1:10; PMID:24450498; http://dx.doi.org/10.1186/2049-2618-1-10

[28] Frank JA, Pan Y, Tooming-Klunderud A, Ejisink VG, McHardy AC, Nederbragt AJ, Pope PB. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci Rep 2016; 6:25373; PMID:27156482; http://dx.doi.org/10.1038/srep25373

[29] Mosher JJ, Bowman B, Bernberg EL, Shevchenko O, Kan J, Korlach J, Kaplan LA. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. 2014.

[30] Willemsen R, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci Rep 2016; 6:25373; PMID:27156482; http://dx.doi.org/10.1038/srep25373

[31] WJ A. Next generation DNA sequencing (II): techniques, applications. J Next Gener Seq Appl 2016; S1:005.

[32] Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 2016; 14:265-79; PMID:27646134; http://dx.doi.org/10.4161/gmic.23571

[33] Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JJ, Phipps EC, et al. Burden of Clostridium difficile infection in the United States. N Engl J Med 2015; 372:825-34; PMID:25714160; http://dx.doi.org/10.1056/NEJMoa1408913
[34] Weingarden AR, Chen C, Bobr A, Yao D, Lu Y, Nelson VM, Sadowsky MJ, Khoruts A. Microbiota transplantation restores normal fecal bile acid composition in recurrent _Clostridium difficile_ infection. Am J Physiol - Gastrointest Liver Physiol 2014; 306:G310-9; PMID: 24284963; http://dx.doi.org/10.1152/ajpgi.00282.2013

[35] Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, Knights D, Unno T, Bobr A, Kang J, et al. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent _Clostridium difficile_ infection. Microbiome 2015; 3:10; PMID: 25825673; http://dx.doi.org/10.1186/s40168-015-0070-0

[36] Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 2013; 31:814-21; PMID: 23975157; http://dx.doi.org/10.1038/nbt.2676

[37] Broecker F, Klumpp J, Schuppler M, Russo G, Biedermann L, Hombach M, Rogler G, Moelling K. Long-term changes of bacterial and viral compositions in the intestine of a recovered _Clostridium difficile_ patient after fecal microbiota transplantation. Cold Spring Harb Mol case Stud 2016; 2:a000448; PMID: 27148577; http://dx.doi.org/10.1101/mcs.a000448