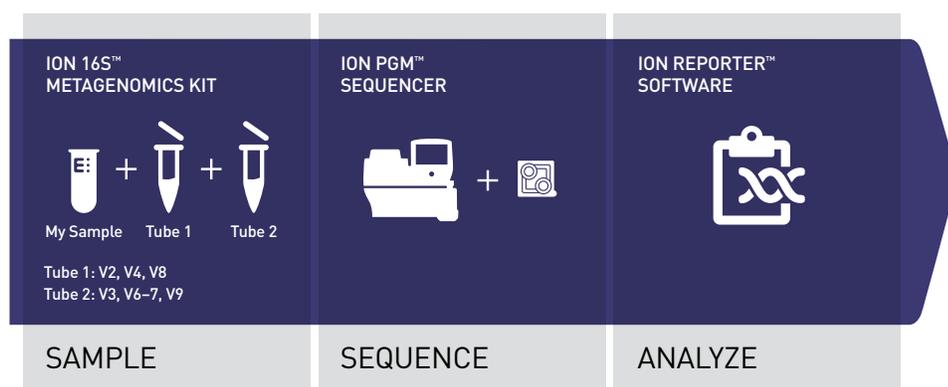


# 16S rRNA Sequencing

An integrated research solution for bacterial identification using 16S rRNA sequencing on the Ion PGM™ System with Ion Reporter™ Software

## Summary

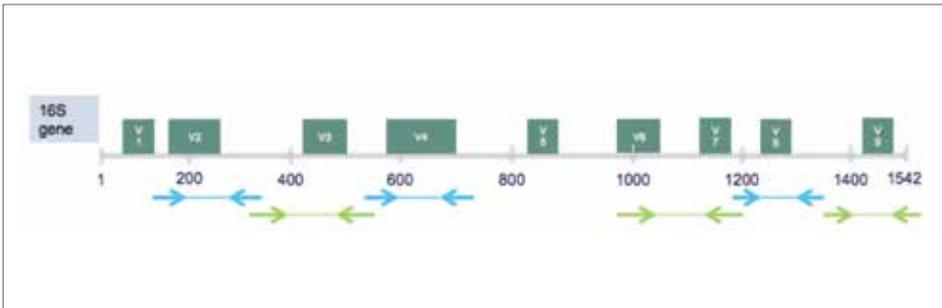
- A novel approach that simultaneously examines 7 of the 9 hypervariable regions in the bacterial 16S rRNA gene is presented, with PCR primers designed to target >80% of the sequences found in the Greengenes database
- Detection sensitivity of 1:1,000 was determined for all organisms in a mock bacterial community sample, with a range of  $10^3$ – $10^6$  16S rDNA gene copies per reaction
- Direct integration between Torrent Suite™ Software and Ion Reporter™ Software allows simple run setup with automated and user-defined analysis—in as few as five clicks—that simplifies data interpretation and enables species-level identification (Figure 1)
- Three reference query options classify individual reads via alignment to either the premium curated MicroSEQ® ID or curated Greengenes databases as well as an optimal two-step alignment that utilizes both databases



**Figure 1. Integrated, cost-effective, scalable 16S rRNA sequencing workflow for the Ion PGM™ System, with a rapid workflow from library to analyzed results.** Following library construction and template preparation, 400 bp sequencing runs are completed in just 2.3 hours for the Ion 314™ Chip v2 and 4.4 hours for the Ion 318™ Chip v2. BAM files produced with Torrent Suite™ Software are automatically analyzed, annotated, and taxonomically assigned in the Ion Reporter™ Software 16S metagenomics workflow.

## Introduction

Understanding the diversity of bacterial communities, in almost any environment, has been greatly facilitated by rapid advances in next-generation sequencing (NGS) technologies and associated bioinformatics approaches. Research using 16S ribosomal RNA (rRNA) sequencing is a fast, inexpensive profiling technique based on variation in the bacterial 16S rRNA gene. This method has a wide range of uses, including the characterization of bacterial populations, taxonomical analysis, and species identification. For example, the Ion PGM™ System was used to characterize bacterial communities important for research in human health using archived samples from a diversity of sites including the mouth [1, 2], airway [3, 4], diabetic mycetoma (or “Madura foot”) [5], meconium [6], the hands of healthcare workers [7], brain abscesses [8], and the human and murine intestinal tract [9, 10, 11, 12].



**Figure 2.** *E. coli* 16S rRNA gene with variable regions illustrated (green squares). Amplicons V2, V4, and V8 are targeted in multiplex PCR reaction tube 1. Amplicons V3, V6-7, and V9 are targeted in multiplex PCR reaction tube 2.

Semiconductor-based sequencing has also revealed the composition of bacterial populations in a wide variety of environmental sources such as: waterways in response to oil sands mining [13]; uranium mine tailings [14, 15]; contaminated arctic soils [16]; the northern Gulf of Mexico [17] and English Channel [18]; soil and rhizosphere of a cactus [19] and willow [20]. For a complete list of 16S rRNA profiling research publications using Ion Torrent™ technology, visit [lifetechnologies.com/ionmetagenomicspublications](http://lifetechnologies.com/ionmetagenomicspublications)

### The bacterial 16S rRNA gene

Assessing microbial diversity, identifying bacterial species, and carrying out grouping by shared sequence characteristics (taxonomical assignment) is possible due to the high degree of conservation of the 16S rRNA gene across the domains of Bacteria and Archaea. Taxonomical assignment is possible due to the presence of 9 hypervariable regions (V1–V9) that contain sufficient sequence diversity to classify microbes (Figure 2). Moreover, since conserved regions flank these variable regions, PCR amplification using universal primers is possible.

When used in NGS workflows, universal primers typically allow targeting of a single, or a few, 16S

rRNA gene variable regions in a single PCR reaction, which is performed on a bacterial population as a whole. Due to the clonal nature of massively parallel NGS, every PCR amplification product and subsequent sequencing read can be considered to be representative of a single bacterium within a mixed population. This approach has allowed characterization of bacterial communities without isolation and culturing. Through a deep sequencing approach, 16S rRNA profiling has fundamentally changed our understanding of countless microbial communities and has proven to be an important discovery tool, revealing the “microbial dark matter” of our planet due to the difficulty of culturing most bacterial species.

However, no single hypervariable region is sufficiently diverse to differentiate among all bacterial species. This fact, coupled with read length restrictions on most NGS platforms, requires concessions to be made in terms of the discriminatory power of 16S rRNA profiling and the hypervariable regions investigated when planning an NGS project. To address this, we describe a novel approach that simultaneously surveys 7 out of 9 hypervariable regions in the 16S rRNA gene (Figure 2). With integrated analysis in Ion Reporter™

Software, we tested primer pools targeting the 16S rRNA gene on two mock bacterial community control samples, achieving excellent sensitivity and species-level discrimination.

### Experimental design

To increase the resolving power of 16S rRNA profiling, primers were designed to amplify variable regions 2, 4, and 8 in a single tube with the resulting amplicon fragments of ~250 base pairs (bp), ~288 bp, and ~295 bp, respectively (Figure 2). In a second single tube, a multiplex PCR reaction targets variable regions 3, 6-7, and 9 with resulting amplicon fragments of ~215 bp, ~260 bp, and ~209 bp, respectively (Figure 2). The primer pools were designed to target >80% of sequences found in the Greengenes database with 100% identity for a primer pair amplifying at least one variable region.

Two mock bacterial community samples were amplified using both primer pools: a balanced community [21] with equal representation of ribosomal DNA (rDNA) copies for 20 bacterial species, and a staggered community [22] with variable amounts of rDNA copies for 20 bacterial species ( $10^3$ – $10^6$  copies per organism per  $\mu\text{L}$ ) (Table 1, supplementary information).

The PCR amplification products were used to create a library via the Ion Plus Fragment Library Kit (Cat. No. 4471252) with sample indexing using the Ion Xpress™ Barcode Adapters 1-96 Kit (Cat. No. 4474517). Template preparation was performed using the Ion OneTouch™ 2 System and the Ion PGM™ Template OT2 400 Kit (Cat. No. 4479878). Sequencing was conducted using the Ion PGM™ Sequencing 400 Kit (Cat. No. 4482002) on the Ion PGM™ System using the Ion

**Table 1. Sample multiplexing guideline for 16S rRNA profiling.**

| Chip type     | Sample complexity* |             |            |
|---------------|--------------------|-------------|------------|
|               | Low (10)           | Medium (20) | High (>30) |
| Ion 314™ Chip | 4                  | 2           | 1          |
| Ion 316™ Chip | 20                 | 10          | 2          |
| Ion 318™ Chip | 40                 | 20          | 4          |

\*Targeting about 350,000 reads per sample; the complexity and sensitivity required will impact the true number of samples that can be multiplexed per chip.

314™, 316™, or 318™ Chips (Cat. No. 4482261, 4483324, 4484355) with various barcoded samples per chip. Sample multiplexing and chip choice is dependent on the degree of sensitivity required and the complexity of the sample under investigation. Table 1 provides a rough guideline for planning a 16S rRNA profiling project. Primary data analysis was performed with Torrent Suite™ Software v4.0 with automated secondary analysis using Ion Reporter™ Software v4.0.

### **Ion Reporter™ Software and 16S metagenomics analysis**

Ion Reporter™ Software comprises a suite of bioinformatics tools that streamline and simplify analysis of semiconductor-based sequencing data. Inclusion of an Ion Reporter™ Software instance in your metagenomics sequencing template during run setup using Torrent Suite™ Software results in 16S rRNA sequencing data to be automatically uploaded to a local or hosted version of Ion Reporter™ Software for read mapping, annotation, and reporting. Further, data analysis can be shared with collaborators—all with version control and audit traceability. Additional sequencing data storage and compute power is available through the Ion Reporter™ Server System for a combined hardware and software solution.

### **Three reference mapping options for flexible bacterial identification**

The 16S rRNA workflow module in Ion Reporter™ Software can classify individual reads via three reference library options: (i) Basic Local Alignment Search Tool (BLAST) alignment to the curated Greengenes database, which contains >400,000 records that were curated for content; (ii) BLAST alignment to the premium curated MicroSEQ® ID database, a high-quality library of full-length 16S rRNA sequences for >15,000 organisms that has been manually curated for sequence quality, length, annotation, and phylogeny, with frequent taxonomical updates; (iii) an optimal two-step BLAST alignment to both reference libraries for access to manually curated and public content. In the first step, reads are aligned to the MicroSEQ® ID library with any unaligned reads subject to a second alignment to the Greengenes database to achieve rapid and exhaustive bacterial identification.

### **Analysis results output, downloads, and visualization**

Following successful completion, clicking on an analysis will open the Analyses view in Ion Reporter™ Software. There are three main sections in this tab: (i) Visualization/Downloads, (ii) Summary, and (iii) Results (Figure 3). Clicking the links under Visualization/Downloads enables the download of results files for

all samples as a compressed file containing text files in tab-delimited format for results by (i) primer, (ii) primer with slash calls, and (iii) consensus.

Taxonomical assignments are reported as a consensus of the results from all of the primers and by each primer, with the option to report multiple taxonomical assignments (slash call). Slash calls can result for a particular variable region when a sequence identifies multiple taxa within a set percentage range. By default, alignment at various taxonomical levels follows the Clinical and Laboratory Standards Institute (CLSI) guidelines requiring the family level to have <97% identity, with genus >97% identity and species >99% identity.

Note that below the genus-level identification, taxonomical assignment should be considered presumptive family level. Family assignment can be the consequence of an assignment to the best match, but this assignment may not capture other nearly equivalent assignments that could be equally valid.

Biological analysis and identification of microbes is enhanced through interactive graphs (powered by Krona) to enable data exploration with visualization of results by primer or consensus that can be graphically viewed at six taxonomical ranks (species, genus, family, order, class, phylum) (Figure 4). Additionally, images of the bacterial diversity at any level in the taxonomy can be exported for publications and reports.

The Summary section provides results including total and valid reads for the experiment as well as the number of low-copy reads and unmapped reads for each



variable region (Figure 3A). Summarized for forward and reverse reads are the number of reads found (i.e., the number of reads where the forward or reverse primer was found), the number of full-coverage reads (number of reads in which both forward and reverse primer were found), the number of short reads (number of reads in which the forward or reverse primer was found, but the trimmed read did not meet the defined length requirement), and the number of valid reads or the number of reads after trimming that—depending on the specified analysis—contained only the forward or reverse primer or both.

The Results section has three data views, with presentation possible by primer, by consensus, and by primer with slash calls (Figure 3B). The number reads (count) is indicated for each taxonomical level (from phylum to species); additionally, the percentage of the total, valid, and mapped reads is indicated for each taxonomical level. At the species level, the percent identification range is reported as well as the forward:reverse read percentage (F:R%). The Results view also reports the number of reads that map to each reference database (DB counters) if the two-step alignment method is used, with the first number indicating the reads aligning to the MicroSEQ® ID library and the second number reporting the reads aligning to the Greengenes database.

### **Staggered 16S mock bacterial community analysis**

Mock community DNA samples ([www.beiresources.org](http://www.beiresources.org)) are standards created as part of the Human Microbiome Project (Table 1, supplementary information). The staggered microbial mock

community consists of genomic DNA from 20 bacterial strains with staggered rDNA gene counts ranging from  $10^3$  to  $10^6$  copies per organism per  $\mu\text{L}$  [22]. Using the staggered mock community, all species were identified using the two multiplexed primer pools targeting 7 of 9 hypervariable regions with the detection sensitivity dependent on the number of analyzed reads. With 360,000 analyzed reads (the equivalent of a run on the Ion 314™ Chip), a 1:100 detection sensitivity was achieved (detection of all organisms with  $10^4$ – $10^6$  rDNA copies/multiplexed reaction). A sensitivity of 1:1000 was achieved through an increase in the number of sequencing reads to >2.8 million valid reads; for example, 3.7 million and 6.0 million reads using the Ion 316™ Chip and Ion 318™ Chip, respectively.

### **Balanced 16S mock bacterial community analysis**

The balanced microbial mock community consists of genomic DNA from 20 bacterial strains with equimolar rDNA gene counts of  $10^6$  copies per organism per  $\mu\text{L}$  [21]. By using the balanced mock community, the impact of adjusting the metagenomics workflow analysis parameters for 16S rRNA profiling in Ion Reporter™ Software can be illustrated.

User-defined analysis parameters include: detection of primers (both ends, single end, and none); read length filter (filters out all reads shorter than the specified length after trimming primers; this filter is disabled when primer detection is set to both ends); minimum alignment coverage (percentage coverage required between a hit and a query); read abundance filter (number of unique reads needed for a particular hit to be a valid); genus

and species cutoff (percentage identity required to make a genus or species identification; default values are 97.0% and 99.0%, respectively); and slash ID reporting percentage (difference in percent match between the top hit and the next hit; if more than one species or genus is found within this range of percentage difference from each other, all hits will be reported as “slash ID”). Note that the default metagenomics workflow analysis parameters have been optimized for family-level identification.

To illustrate how to improve confidence in genus/species-level identification, the default metagenomics workflow analysis parameters were changed to require the detection of primers at both ends (the default requires single-end detection) of a read. In Ion Reporter™ Software, 16S rRNA reads can be trimmed by primer, with a tolerance of three errors, and/or by length. Users can specify whether both primers are required to be present in a read, or just one, or no primers at all. If a single primer or no primer is selected, trimmed reads will require attainment of a user-specified length. Requiring both primers to be present increases the stringency of the reads used for taxonomical assignment, with a downside that non-inclusion of trimmed reads with a single primer present may underrepresent the diversity in a bacterial population.

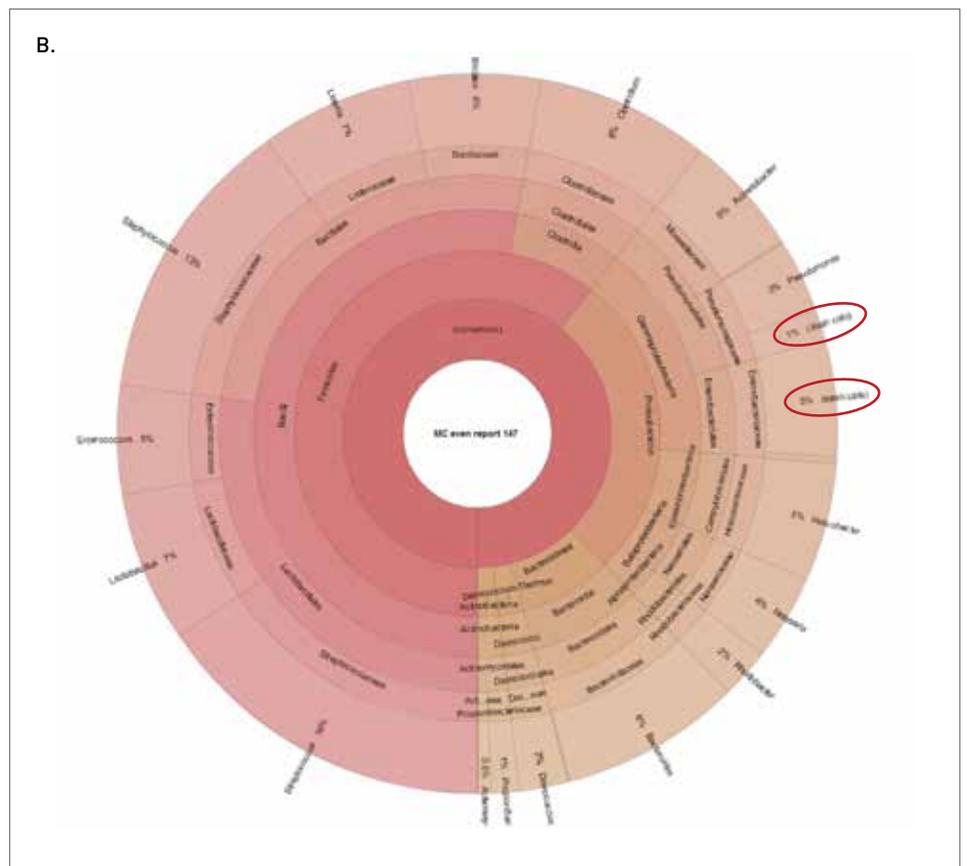
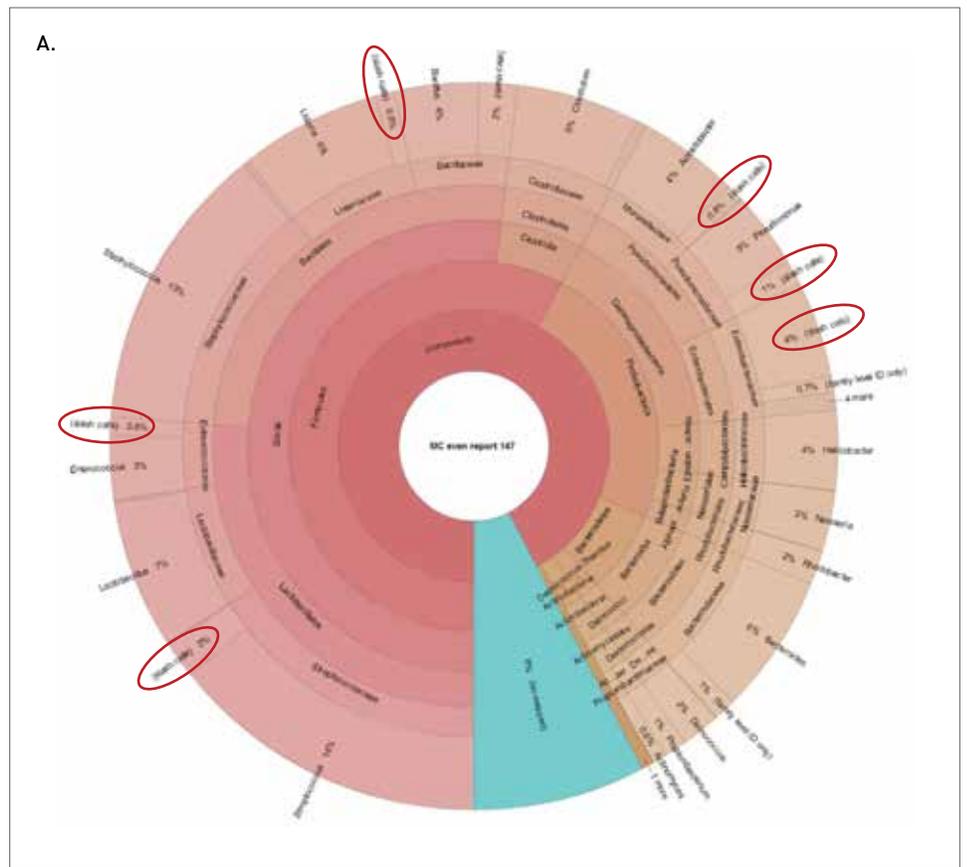
To further improve genus/species-level resolution, the default metagenomics workflow analysis parameters for slash ID reporting percentage was altered from 0.8 to 0.2. For example, a species-level slash call could result for a 99.9% match and another hit with 99.1%

identity, and with the parameter change the next hit would need to be >99.7% to be reported as a slash call. A drawback to restricting slash calls is that sequencing errors could result in erroneous species identification.

Requiring the presence of both primers in a read and restricting the slash ID reporting percentage improved genus-level identification of organisms using sequence data from the balanced microbial mock community (Figure 4). Compared to the default parameters, more stringent parameters resulted in considerably fewer slash calls at the genus level and no unmapped reads (Figure 4B).

### Conclusions

16S rRNA profiling is a powerful method for characterizing bacterial populations; however, read length restrictions on most NGS platforms require compromises in terms of the discriminatory power of 16S sequencing due to the targeting of only 1–3 hypervariable regions. The novel approach described here targets 7 of the 9 hypervariable regions in the 16S rRNA gene for optimal characterization of bacterial species. When combined with the scalable Ion PGM™ System and Ion Reporter™ Software, the multiplexed primer pools achieve excellent detection sensitivity with a choice of reference database query options for optimal taxonomical assignment. Further, direct integration with Torrent Suite™ Software and simple stepwise informatics available using the metagenomics workflow in Ion Reporter™ Software enables any laboratory to easily and rapidly perform bacterial identification.



**Figure 4. Automated analysis, annotation, and taxonomical assignment implemented in Ion Reporter™ Software.** (A) Consensus view at the genus level of the balanced mock bacterial sample using the default analysis parameters. Unmapped reads are shown in blue with slash calls at the genus level circled in red. (B) Consensus view at the genus level of the balanced mock bacterial sample using more stringent analysis parameters. Slash calls at the genus level are circled in red.

## References

1. Junemann S, Prior K, Szczepanowski R, et al. (2012) Bacterial community shift in treated periodontitis patients revealed by Ion Torrent 16S rRNA gene amplicon sequencing. *PLoS ONE* 7: e41606.
2. Dassi E, Ballarini A, Covello G, et al. (2014) Enhanced microbial diversity in the saliva microbiome induced by short-term probiotic intake revealed by 16S rRNA sequencing on the Ion Torrent PGM platform. *J Biotech* In Press, Accepted Manuscript.
3. Blainey PC, Milla CE, Cornfield DN, and Quake SR (2012) Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Sci Transl Med* 4: 153ra130.
4. Salipante S, Sengupta DJ, Rosenthal C et al. (2013) Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS ONE* 8:e65226.
5. Salipante SJ, Sengupta DJ, Hoogestraat DR, et al. (2013) Molecular diagnosis of *Actinomyces madurae* infection by 16S rRNA deep sequencing. *J Clin Microbiol* 51(12):4262-4265
6. Ardisson AN, de la Cruz DM, Davis-Richardson AG, et al. (2014) Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS ONE* 9:e90784.
7. Rosenthal M, Aiello AE, Chenoweth C, et al. (2014) Impact of technical sources of variation on the hand microbiome dynamics of healthcare workers. *PLoS ONE* 9:e88999.
8. Kommedal Ø, Wilhelmsen MT, Skrede S, et al. (2014) Massive parallel sequencing provides new perspectives on bacterial brain abscesses. *J Clin Microbiol* 52(6):1990-1997.
9. Petrof EO, Gloor GB, Vanner SJ, et al. (2013) Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. *Microbiome* 1:3.
10. Milani C, Hevia A, Foroni E, et al. (2013) Assessing the fecal microbiota: an optimized Ion Torrent 16S rRNA gene-based analysis protocol. *PLoS ONE* 8:e68739.
11. Mir SA, Nagy-Szakal D, Dowd SE, et al. (2013) Prenatal methyl-donor supplementation augments colitis in young adult mice. *PLoS ONE* 8:e73162.
12. Bel S, Elkis Y, Elifantz H, et al. (2014) Reprogrammed and transmissible intestinal microbiota confer diminished susceptibility to induced colitis in TMF-/- mice. *Proc Natl Acad Sci USA* 111(13):4964-4969.
13. Yergeau E, Lawrence JR, Sanschagrin S, et al. (2012) Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl Environ Microbiol* 78:7626-7637.
14. Bondici VF, Lawrence JR, Khan NH, et al. (2013) Microbial communities in low-permeability, high pH uranium mine tailings: characterization and potential effects. *J Appl Microbiol* 114:1671-1686.
15. Khan N, Bondici V, Medihala P, et al. (2013) Bacterial diversity and composition of an alkaline uranium mine tailings-water interface. *J Microbiol* 51(5):558-569.
16. Bell TH, Yergeau E, Maynard C, et al. (2013) Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. *ISME J* 7:1200-12109.
17. Ortmann A and Ortell N (2013) Changes in free-living bacterial community diversity reflect the magnitude of environmental variability. *FEMS Microbiol Ecol* 87(1):291-301.
18. Taylor JD, Cottingham SD, Billinge J, Cunliffe M (2013) Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. *ISME J* 8: 245-248.
19. Nessner Kavamura V, Taketani RG, Lançon MD, et al. (2013) Water regime influences bulk soil and rhizosphere of *Cereus jamacaru* bacterial communities in the Brazilian Caatinga Biome. *PLoS ONE* 8:e73606.
20. Yergeau E, Sanschagrin S, Maynard C, et al. (2013) Microbial expression profiles in the rhizosphere of willows depend on soil contamination. *ISME J* 8: 344-358.
21. The following reagent was obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community B (Even, High Concentration), v5.2H, for *Whole Genome Shotgun Sequencing*, HM-277D.
22. The following reagent was obtained through BEI Resources, NIAID, NIH as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community B (Staggered, High Concentration), v5.2H, for *Whole Genome Shotgun Sequencing*, HM-277D.

## Supplementary information

Table 1. Microbial mock community B

| Organism   | NCBI reference sequence | Organism  | NCBI reference sequence |
|--|-------------------------|---|-------------------------|
| <i>Acinetobacter baumannii</i> , strain 5377           | NC_009085               | <i>Listeria monocytogenes</i> , strain EGDe             | NC_003210               |
| <i>Actinomyces odontolyticus</i> , strain 1A.21        | NZ_AAYI02000000         | <i>Neisseria meningitidis</i> , strain MC58             | NC_003112               |
| <i>Bacillus cereus</i> , strain NRS 248                | NC_003909               | <i>Propionibacterium acnes</i> , strain KPA171202       | NC_006085               |
| <i>Bacteroides vulgatus</i> , strain NCTC 11154        | NC_009614               | <i>Pseudomonas aeruginosa</i> , strain PAO1-LAC         | NC_002516               |
| <i>Clostridium beijerinckii</i> , strain NCIMB 8052    | NC_009617               | <i>Rhodobacter sphaeroides</i> , strain ATH 2.4.1       | NC_007493,<br>NC_007494 |
| <i>Deinococcus radiodurans</i> , strain R1 (smooth)    | NC_001263, NC_001264    | <i>Staphylococcus aureus</i> , strain TCH1516           | NC_010079               |
| <i>Enterococcus faecalis</i> , strain OG1RF            | NC_017316               | <i>Staphylococcus epidermidis</i> , FDA strain PCI 1200 | NC_004461               |
| <i>Escherichia coli</i> , strain K12, substrain MG1655 | NC_000913               | <i>Streptococcus agalactiae</i> , strain 2603 V/R       | NC_004116               |
| <i>Helicobacter pylori</i> , strain 26695              | NC_000915               | <i>Streptococcus mutans</i> , strain UA159              | NC_004350               |
| <i>Lactobacillus gasseri</i> , strain 63 AM            | NC_008530               | <i>Streptococcus pneumoniae</i> , strain TIGR4          | NC_003028               |

Find out more at [lifetechnologies.com](http://lifetechnologies.com)

 **life** technologies

A Thermo Fisher Scientific Brand