The characterization of novel tissue microbiota using an optimized 16S metagenomic sequencing pipeline

Jérôme J. Lluch, Florence Servant, Sandrine Païssé, Carine Valle, Sophie Valiere, Claire Kuchly, Gaëlle Vilchez, Cecile Donnadieu, Michael Courtney, Remy Burcelin, et al.

HAL Id: hal-02637487

https://hal.inrae.fr/hal-02637487v1

Submitted on 27 May 2020

The Characterization of Novel Tissue Microbiota Using an Optimized 16S Metagenomic Sequencing Pipeline

Jérôme Lluch[1,2☯], Florence Servant[1☯], Sandrine Païssé[1], Carine Valle[1], Sophie Valière[2,3], Claire Kuchly[2,3], Gaëlle Vilchez[2,3], Cécile Donnadieu[2,4], Michael Courtney[1], Rémy Burcelin[5], Jacques Amar[5,6], Olivier Bouchez[2,4], Benjamin Lelouvier[1]*

1 Vaiomer SAS, Labège, France, 2 INRA, GeT-PlaGe, Genotoul, Castanet-Tolosan, France, 3 INRA, UAR1209, Castanet-Tolosan, France, 4 INRA, UMR1388, GenPhySE, Castanet-Tolosan, France, 5 INSERM U1048, I2MC, Toulouse, France, 6 Rangueil Hospital, Department of Therapeutics, Toulouse, France

☯ These authors contributed equally to this work.
* benjamin.lelouvier@vaiomer.com

Abstract

Background

Substantial progress in high-throughput metagenomic sequencing methodologies has enabled the characterisation of bacteria from various origins (for example gut and skin). However, the recently-discovered bacterial microbiota present within animal internal tissues has remained unexplored due to technical difficulties associated with these challenging samples.

Results

We have optimized a specific 16S rDNA-targeted metagenomics sequencing (16S metabarcoding) pipeline based on the Illumina MiSeq technology for the analysis of bacterial DNA in human and animal tissues. This was successfully achieved in various mouse tissues despite the high abundance of eukaryotic DNA and PCR inhibitors in these samples. We extensively tested this pipeline on mock communities, negative controls, positive controls and tissues and demonstrated the presence of novel tissue specific bacterial DNA profiles in a variety of organs (including brain, muscle, adipose tissue, liver and heart).

Conclusion

The high throughput and excellent reproducibility of the method ensured exhaustive and precise coverage of the 16S rDNA bacterial variants present in mouse tissues. This optimized 16S metagenomic sequencing pipeline will allow the scientific community to catalogue the bacterial DNA profiles of different tissues and will provide a database to analyse host/bacterial interactions in relation to homeostasis and disease.

Abbreviations: BEI, biodefense and emerging infections; CPU, central processing unit; DNA, deoxyribonucleic acid; dNTP, deoxyribonucleotide triphosphate; EDTA, ethylene diamine tetra acetic; FLASH, fast length adjustment of short reads; HPLC, high performance liquid chromatography; INRA, institut national de la recherche agronomique; INSERM, institut national de la santé et de la recherche médicale; LB, lysogeny broth; LED, light emitting diode; MAT, mesenteric adipose tissue; NIAID, national institute of allergy and infectious diseases; NIH, national institute of health; NGS, next generation sequencing; OTU, operational taxonomic unit; PCoA, principal coordinates analysis; PCR, polymerase chain reaction; QC, quality control; RAM, random access memory; rDNA, ribosomal deoxyribonucleic acid; RDP, ribosomal database projet; RFLP, restriction fragment length polymorphism; RNA, ribonucleic acid; SEM, standard error of the mean; SOP, standard operating procedure; TBE, tris borate ethylene diamine tetra acetic; UV, ultraviolet.

Introduction

Animal cells coexist with a complex ecosystem of bacteria and archaea. This microbiota, which outnumbers eukaryotic cells at least tenfold [1], is mostly present in the gastrointestinal tract and at other epithelial surfaces such as the skin, oral cavity, lung mucosa and vagina [1–3]. A large body of evidence demonstrates the importance of epithelial bacteria in the maintenance of health [4,5]. Recent studies are consistent with the existence of microbiota in diverse tissues and organs such as the liver, adipose tissue, blood and atheroma plaque and these bacteria may play a role in non-infectious pathologies [6–10]. Importantly, the function of this microbiota could impact the physiology of the tissue. For example, gram-negative bacteria in adipose tissue from obese patients [11] are responsible for the triggering of pre-adipocyte precursors and macrophage proliferation [12].

Identifying the bacterial taxa (living bacteria or bacterial DNA) present within tissues will aid in elucidating the molecular mechanisms implicated in the control of cellular and physiological functions of the host. The exhaustive study of tissue microbiota requires culture-independent methods such as metagenomic sequencing. 16S rDNA-targeted metagenomic sequencing (also referred to as 16S metagenomics or 16S metabarcoding) allows the analysis of the relative proportion of bacterial taxa in a sample using specific amplification by PCR of the 16S ribosomal RNA gene (16S) coupled to next generation high throughput sequencing (NGS). Whereas for a number of years Roche 454 pyrosequencing has been the gold standard for 16S metagenomics [13,14], the release of the MiSeq kit reagents v2 (2x250 bp pair ended reads) and v3 (2 x300 bp pair ended reads) by Illumina, permitted for the first time the use of the MiSeq technology to reach an amplicon length compatible with 16S metagenomics. MiSeq technology combines several major advantages compared to 454 technology: i) higher output (8.5 Gb for kit v2 and 15 Gb for kit v3) allowing more exhaustive analysis of complex microbiota and/or more samples per sequencing run ii) lower cost per read and iii) a simplified procedure for library construction. Technical limitations exist that hamper the metagenomic analysis of tissue microbiota, including high abundance of PCR inhibitors and other eukaryotic products, which complicate dramatically the extraction and sequencing of bacterial DNA present within the samples [15,16]. This study describes the design, and validation of an optimized 16S metagenomics pipeline to investigate taxonomic diversity in tissue microbiota using MiSeq reagent kits v2 and v3 and presents its application in the analysis of microbiota in liver, muscle, heart, brain and adipose tissue. In addition to protocol optimization for tissue sample, we designed the pipeline with several specificities to reduce cost and complexity, and to facilitate the adaptation of the method to new primers and future technical improvements from Illumina.

Deciphering the tissue microbiota will help to identify the molecular crosstalk between the host and the bacteria and will thus lay the groundwork for the understanding of homeostatic and pathological mechanisms and the identification of novel therapeutic strategies.

Materials and Methods

Sample preparation and DNA extraction

BEI mock communities. Genomic DNA from microbial mock communities B, HM-782D (v5.1L, even, low concentration) and HM-783D (v5.2L staggered, low concentration) were obtained from BEI Resources (NIAID, NIH as part of the Human Microbiome Project, Manassas, VA, USA). HM-783D contains genomic DNA mixture from 20 bacterial strains containing staggered ribosomal RNA operon counts (1,000 to 1,000,000 copies per organism per μl). HM-782D contains genomic DNA from the same 20 bacterial strains with equimolar (even)

ribosomal RNA operon counts (100,000 copies per organism per μl). See S1 Table for bacterial strain list.

Designed mock community. The designed mock community was prepared by cloning the complete 16S rDNA gene of 14 different bacterial species. Genomic DNA from *Acinetobacter johnsonii* (NCIMB 8154), *Aminobacter aminovorans* (NCIMB 9039), *Devosia riboflavin* (NCIMB 8177), *Eubacterium barkeri* (NCIMB 10623), *Geofilum rubicundum* (NCIMB 14482), *Paracoccus denitrificans* (NCIMB 8944), *Prosthecobacter fusiformis* (NCIMB 12777) and *Xanthomonas sp.* (NCIMB 592) were obtained from NCIMB (Aberdeen, Scotland). Genomic DNA from *Lactococcus lactis* (CIRMBP-611) was obtained from CIRM-BP (INRA UMR 1282 ISP, Nouzilly, France). Bacterial strains *Bifidobacterium animalis subsp. lactis*, *Cupriavidus necator (Ralstonia eutropha)*, *Escherichia coli*, *Ralstonia mannitolilytica*, and *Ralstonia pickettii* were provided by Dr Remy Burcelin (Inserm/UPS UMR 1048—I2MC, Toulouse France). The genomic DNA of these 5 bacterial strains was extracted using the Trizol method following the protocol recommended by the manufacturer (Life Technologies, Grand Island, NY, USA). The complete 16S rRNA gene of the 14 bacterial strains was amplified using the 8F (AGAGTTTG ATCCTGGCTCAG) [17] and 1489R (TACCTTGTTACGACTTCA) [18] primers. The PCR products were purified using the NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) and cloned into *Escherichia coli* TOP10F using the pCR2.1 TOPO TA cloning kit (Life Technologies). Recombinant clones were verified by checking the insert size by PCR with the M13F (GTAAAACGACGGCCAG) and M13R (CAGGAAACAGCTATGAC) primers, by restriction fragment length polymorphism with NotI and SpeI restriction enzymes (New England Biolabs, Ipswich, MA, USA), and by Sanger sequencing (GATC Biotech, Constance, Germany) of both strands (1480 bp, using the 16S rDNA gene of *E. coli* as a reference, although the length varies depending on the organisms [19,20]). The validated recombinant clones were then cultured in liquid LB medium supplemented with ampicillin (65 μg/ml) overnight at 37°C. After centrifugation (5 min 8000 g at 4°C), plasmids were extracted using the PureLink® Quick Plasmid Miniprep Kit (Life Technologies). Following the plasmid purification, a second verification of the inserted DNA sequence (including PCR, RFLP and Sanger sequencing as described above) was performed for final plasmid validation. For the designed mock community mixture, the concentration of each plasmid extract was defined using Nano-Drop 2000 UV spectrophotometer (Thermo Scientific, Waltham, MA, USA). See Table 1 for the complete bacterial strain list. The genomic DNA of the strains of *Bifidobacterium animalis subsp. lactis*, *Cupriavidus necator* (*Ralstonia eutropha*), *Escherichia coli*, *Ralstonia mannitolily-tica*, and *Ralstonia pickettii*, extracted for the designed mock community mixture preparation, were also used to prepare several mixes of genomic bacterial DNA, using NanoDrop 2000 UV spectrophotometer to set the proportions of each strains in the mix.

Tissue samples. In addition to fecal samples, we choose six tissue samples representing diverse organs: ileum, liver, skeletal muscle, heart, brain and mesenteric adipose tissue (MAT). Tissue samples were collected from C57-BL/6J mice (Charles River, Wilmington, MA, USA) raised in a specific pathogen-free animal facility. Samples were directly frozen in liquid nitrogen and stored at -80°C until DNA extraction.

For DNA extraction, to access most of the bacterial DNA present in the tissues (free DNA or in living /dormant/degraded bacteria, circulating or inside eukaryotic cells) without damaging the DNA, we tested empirically numerous protocols of lysis (mechanical and/or enzymatic). The lysis protocol that gave the best yield of bacterial DNA extraction from fecal and tissue samples consisted of a mechanical disruption step for 5 seconds using Turax (KA, Germany) followed by a lysis step using acid-washed glass beads (Sigma, Saint-Louis, MO, USA) and Tissue Lyser (Qiagen, Venlo, Netherlands) for 2x 3 min at 30 Hz. After this lysis step, the total genomic DNA was extracted using the QIAamp DNA Stool kit (Qiagen) for fecal

Table 1. Designed mock community.

Species	Family	Phylum	Copies of16S rDNA
Acinetobacter johnsonii	Moraxellaceae	Proteobacteria (Gamma)	10^7
Aminobacter aminovorans	Phyllobacteriaceae	Proteobacteria (alpha)	10^7
Bifidobacterium animalis subsp. lactis	Bifidobacteriaceae	Actinobacteria	10^7
Cupriavidus necator (= Ralstonia eutropha)	Burkholderiaceae	Proteobacteria (Beta)	10^7
Devosia riboflavina	Hyphomicrobiaceae	Proteobacteria (alpha)	10^7
Escherichia coli	Enterobacteriaceae	Proteobacteria (Gamma)	10^7
Eubacterium barkeri	Eubacteriaceae	Firmicutes	10^7
Geofilum rubicundum	Marinilabiliaceae	Bacteroidetes/Chlorobi	10^7
Lactococcus lactis	Streptococcaceae	Firmicutes	10^7
Paracoccus denitrificans	Rhodobacteraceae	Proteobacteria (alpha)	10^7
Prosthecobacter fusiformis	Verrucomicrobiaceae	Chlamydiae/Verrucomicrobia	10^7
Ralstonia mannitolilytica	Burkholderiaceae	Proteobacteria (Beta)	10^7
Ralstonia pickettii	Burkholderiaceae	Proteobacteria (Beta)	10^7
Xanthomonas sp.	Xanthomonadaceae	Proteobacteria (Gamma)	10^7

doi:10.1371/journal.pone.0142334.t001

samples and Trizol (Life Technologies) for all other tissues according to the manufacturer's instructions. The quality and quantity of DNA extracts were analyzed by agarose gel electrophoresis (1% agarose in TBE 0.5X) and NanoDrop 2000 UV spectrophotometer (Thermo Scientific).

Primer design and library preparation

Except mentioned otherwise, all samples (Figs 1–4, S1 and S2 Figs) were analyzed in triplicate starting from the extracted DNA. The replicates presented were technical except in Fig 3B which displays both technical and biological replicates (3 mice with three technical replicates each) and Fig 4 which only presents biological replicates (three mice for each tissue). Negative controls to assess technical background were performed using Nuclease-free water (Ambion, LifeTechnologies) either in place of the tissue sample during the extraction step (lysis + trizol protocol), or in place of the extracted DNA during the library preparation. Each triplicate underwent all library preparation steps, sequencing and bioinformatics analysis, as described below. The V3-V4 hyper-variable regions of the 16S rDNA gene were amplified from the DNA extracts during the first PCR step using universal primer Vaiomer 1F (CTTTCCCTACACG ACGCTCTTCCGATCT-TCCTACGGGAGGCAGCAGT, partial P5 adapter–primer) and universal primer Vaiomer 1R (GGAGTTCAGACGTGTGCTCTTCCGATCT-GGACTACC AGGGTATCTAATCCTGTT, partial P7 adapter–primer) which are fusion primers based on the qPCR primers designed by Nadkarni et al. [21]. Primers Vaiomer 1F and 1R include specificity for the 16S rDNA gene of 95% of the bacteria in the Ribosomal Database Project and part of the P5/P7 adapter targeted by the second PCR step (CTTTCCCTACACGAC and GGAGTTCAGACGTGT). Our primer design and 2 step PCR strategy allow shorter primers which are more suitable for the amplification of bacterial DNA extracted from tissue samples carrying large amounts of PCR inhibitors and eukaryotic DNA. This PCR was performed using 2 U of a DNA-free Taq DNA Polymerase and 1x Taq DNA polymerase buffer (MTP Taq DNA Polymerase, Sigma). The buffer was complemented with 10 nmol of dNTP mixture (Eurodex, Souffelweyersheim, France), 15 nmol of each primer (Sigma) and Nuclease-free water (Ambion, Life Technologies) in a final volume of 50 µl. The PCR reaction was carried out on a Veriti Thermal Cycler (Life Technologies) as follows: an initial denaturation step (94°C for
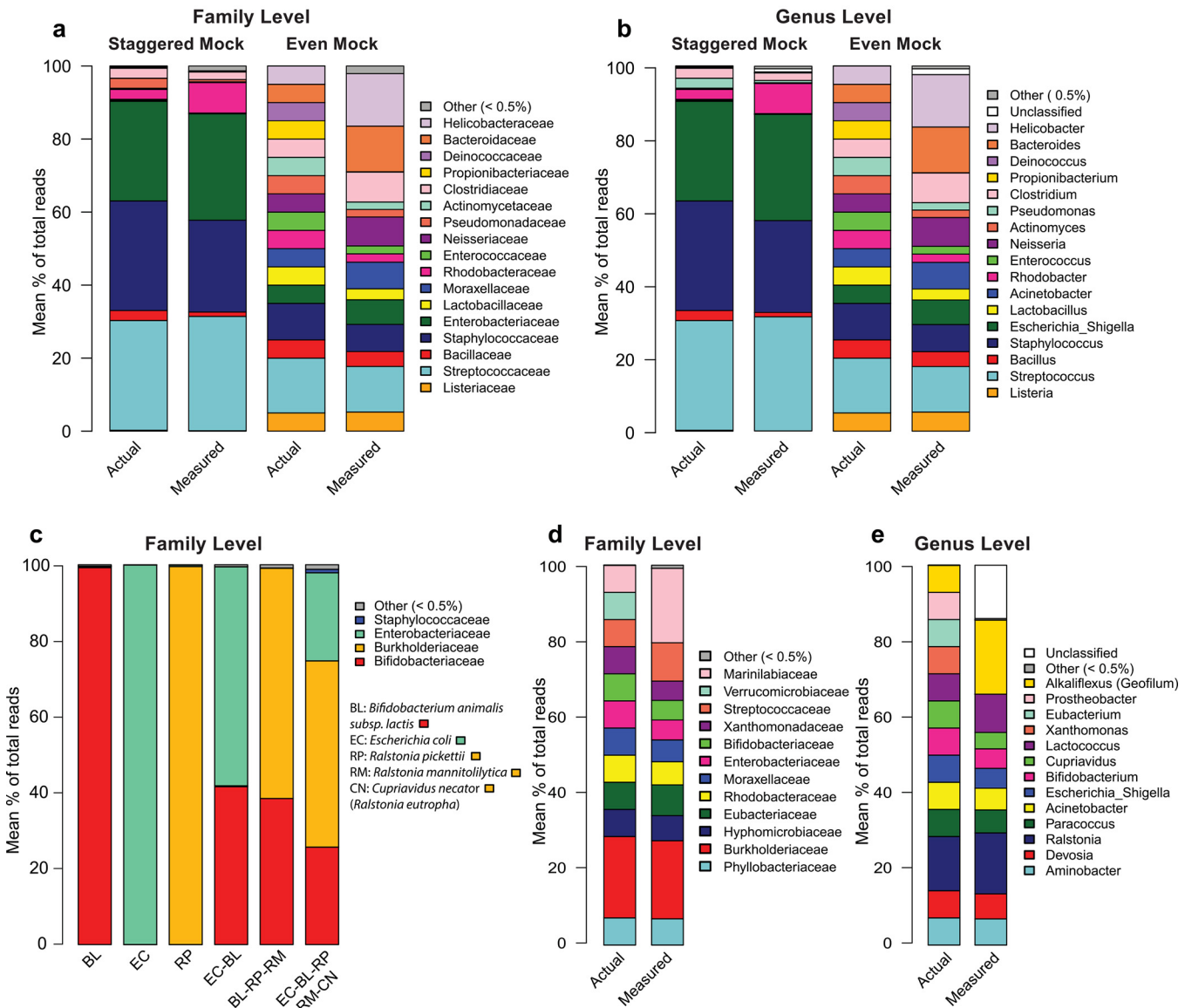
Fig 1. Pipeline validation against mock communities. (a) Stacked bar charts showing the actual relative abundance of bacterial families in the staggered and even BEI genomic mock communities and the measured relative abundance of the families obtained with the MiSeq sequencing pipeline. **(b)** The relative abundance as in **a**, but at the genus taxonomic level. **(c)** Relative abundance of the families obtained by sequencing genomic DNA of a single strain of bacteria or simple mix of bacterial genomic DNA. **(d)** Stacked bar charts showing the actual relative abundance of bacterial families in the plasmid based mock community and the measured relative abundance of the families obtained with the MiSeq sequencing pipeline. **(e)** The relative abundance as in **d**, but at the genus taxonomic level. The sequencing was performed in triplicate for all the samples (starting from the extracted DNA); the means of the triplicates are shown on the stacked bar charts.

doi:10.1371/journal.pone.0142334.g001

10 min), 35 cycles of amplification (94°C for 1 min, 68°C for 1 min and 72°C for 1 min) and a final elongation step at 72°C for 10 min. Amplicons were then purified using the magnetic beads Agencourt AMPure XP—PCR Purification (Beckman Coulter, Brea, CA, USA) following the 96 well format procedure modified as follow: beads/PCR reactional volume ratio of 0.8 x and final elution volume of 32 µl using Elution Buffer EB (Qiagen). The concentration of the purified amplicons was controlled using Nanodrop 8000 spectrophotometry (Thermo Scientific).
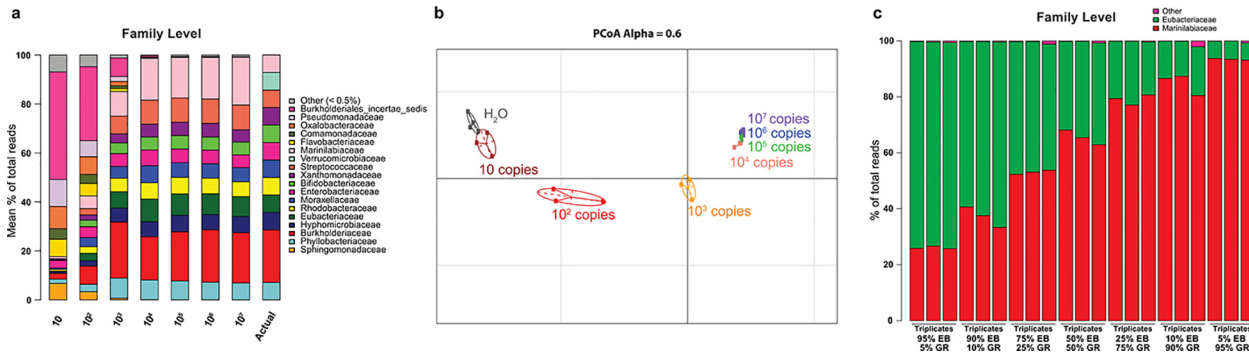
Fig 2. Pipeline sensitivity and assessment of the relative abundance accuracy. (a) Stacked bar charts showing the relative abundance of bacterial families obtained by sequencing of a serial dilution of the plasmid-based mock community (from 10 and 10^7 copies of 16S gene for each of the 14 bacterial strains) compared to the actual relative abundance (right stacked bar). **(b)** Generalized UniFrac distance based PCoA analysis of the sequencing of a serial dilution of the plasmid-based mock community shown in a compared with the negative control generated by sequencing molecular biology-grade water with the same pipeline (H_2O). UniFrac weight parameter (Alpha) was set to 0.6 for this analysis. **(c)** Stacked bar charts showing the relative abundance of bacterial families obtained by sequencing 7 mixes with different ratios of two of the bacterial plasmids containing the 16S gene of *Geofilum rubicundum* (GR) and *Eubacterium barkeri* (EB). The sequencing was performed in triplicate for all the samples (starting from the extracted DNA); the means of the triplicates are shown in **a** and individual triplicates are shown in **b** and **c**.

doi:10.1371/journal.pone.0142334.g002

Sample multiplexing was performed using tailor-made 6 bp unique indexes, which were added during the second PCR step at the same time as the second part of the P5/P7 adapters used for the sequencing step on the MiSeq flow cells with the forward primer Vaiomer 2F (AATGATACGGCGACCACCGAGATCTACACT-CTTTCCCTACACGAC, partial P5 adapter–primer targeting primer 1F) and reverse primer Vaiomer 2R (CAAGCAGAAG ACGGCATACGAGAT-index-GTGACT-GGAGTTCAGACGTGT, partial P7 adapter including index–primer targeting primer 1R). This second PCR step was performed on 50–200 ng of purified amplicons from the first PCR using 2.5 U of a DNA free Taq DNA Polymerase and 1x Taq DNA polymerase buffer. The buffer was complemented with 10 nmol of dNTP mixture (Euromedex), 25 nmol of each primer (Eurogentec, HPLC grade) and Nuclease-free water (Ambion, Life Technologies) up to a final volume of 50 μl. The PCR reaction was carried out on a Veriti Thermal Cycler (Life Technologies) and ran as follow: an initial denaturation step (94°C for 10 min), 12 cycles of amplification (94°C for 1 min, 65°C for 1 min and 72°C for 1 min) and a final elongation step at 72°C for 10 min. Amplicons were purified as described for the first PCR round. The concentration of the purified amplicons was measured using Nano-drop 8000 spectrophotometry (Thermo Scientific) and the quality of a set of amplicons (12 samples per sequencing run) was tested using Agilent DNA 7500 chips using the Bioanaly-zer 2100 (Agilent Technologies, Santa Clara, CA, USA). Controls were carried out to ensure that the high number of PCR cycles (35 cycles for PCR 1 + 12 cycles for PCR2) did not create significant amounts of PCR chimera or other artifacts. The region of the 16S rDNA gene to be sequenced has a length of 467 bp for a total amplicon length of 522 bp after PCR 1 and of 588 bp after PCR 2 (using the 16S rDNA gene of *E. coli* as a reference). All libraries were pooled in the same quantity in order to generate equivalent number of raw reads with each library. The DNA concentration in the pool (no dilution, diluted 10x and 20x in EB + Tween 0.5% buffer) was quantified by qPCR using the 7900HT Fast Real-Time PCR System (Life Technologies) and KAPA Library Quantification Kits for Illumina Platform (Kapa Biosystems, Inc., Wilming-ton, MA, USA) as recommended by the manufacturer (Illumina, San Diego, CA, USA). The final pool, at a concentration after dilution between 5 and 20 nM, was used for sequencing.
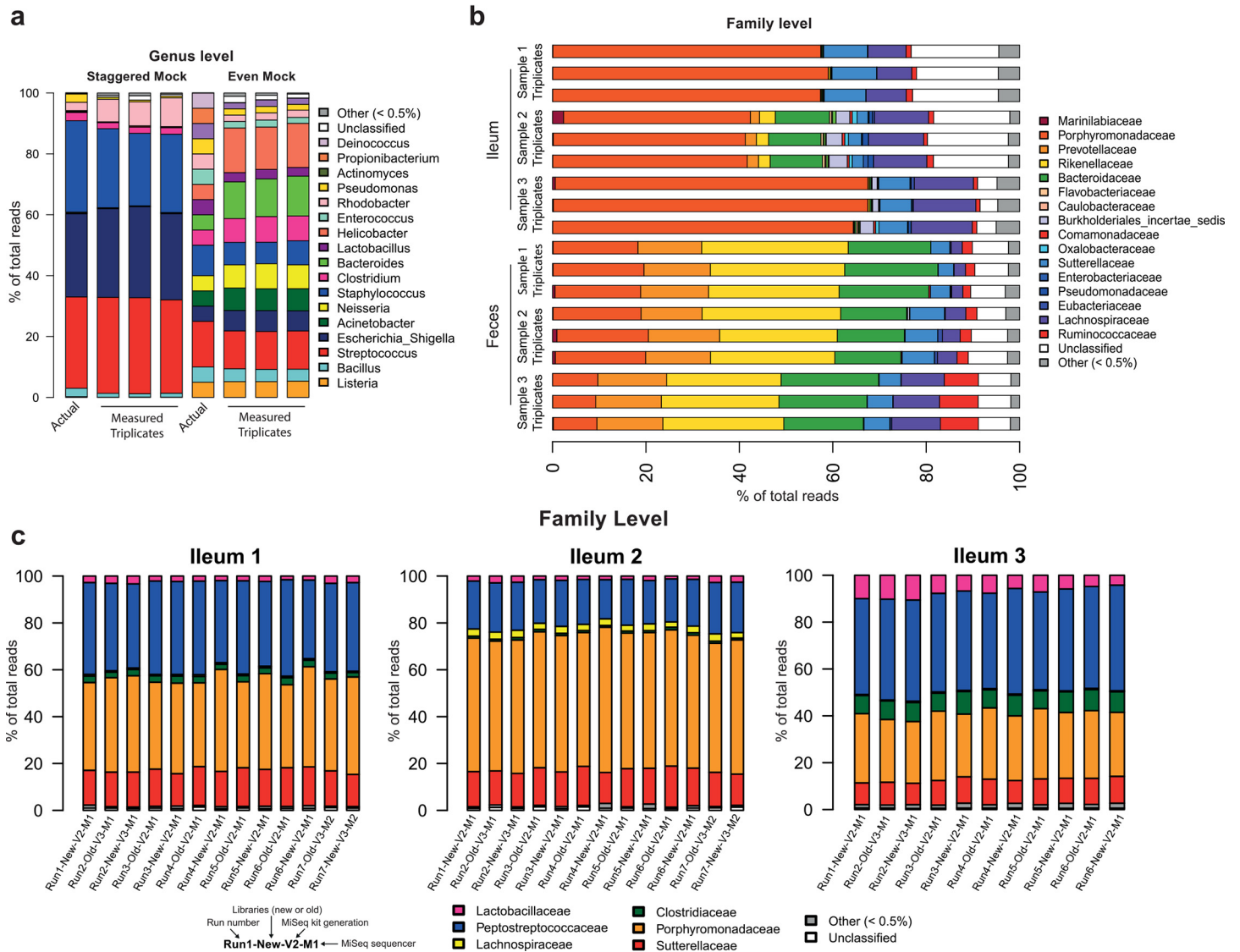
Fig 3. Replicability and reproducibility. (a) Stacked bar charts showing the actual relative abundance of bacterial families in the staggered and even BEI genomic mock communities and the measured relative abundance in triplicates of the families obtained with the MiSeq sequencing pipeline. **(b)** Stacked bar charts showing the relative abundance of bacterial families obtained by sequencing of triplicates of fecal samples and ileum mucosa samples collected from three different mice (Sample 1, 2 and 3). **(c)** Stacked bar charts showing the relative abundance of bacterial families obtained by sequencing of three samples of mouse ileum mucosa in six to seven runs each with different parameters described in the legend at the bottom left: different runs, new libraries from same extracted DNA or the same libraries already prepared, different MiSeq kit generation and different sequencers. Two different experimenters performed the different runs and reagent batch numbers (including Taq polymerase) varied from run to run.

doi:10.1371/journal.pone.0142334.g003

Sequencing

The pool was denatured (NaOH 0.1N) and diluted to 7 pM. The PhiX Control v3 (Illumina) was added to the pool at 15% of the final concentration as described in the Illumina procedure. 600 µl of this pool and PhiX mixture were loaded onto the Illumina MiSeq cartridge according to the manufacturer's instructions using MiSeq Reagent Kit v2 (2x250 bp Paired-End Reads, 8.5 Gb output) or MiSeq Reagent Kit v3 (2x300 bp Paired-End Reads, 15 Gb output, only for samples marked as V3 in Fig 3C and S2E Fig). FastQ files were generated at the end of the run
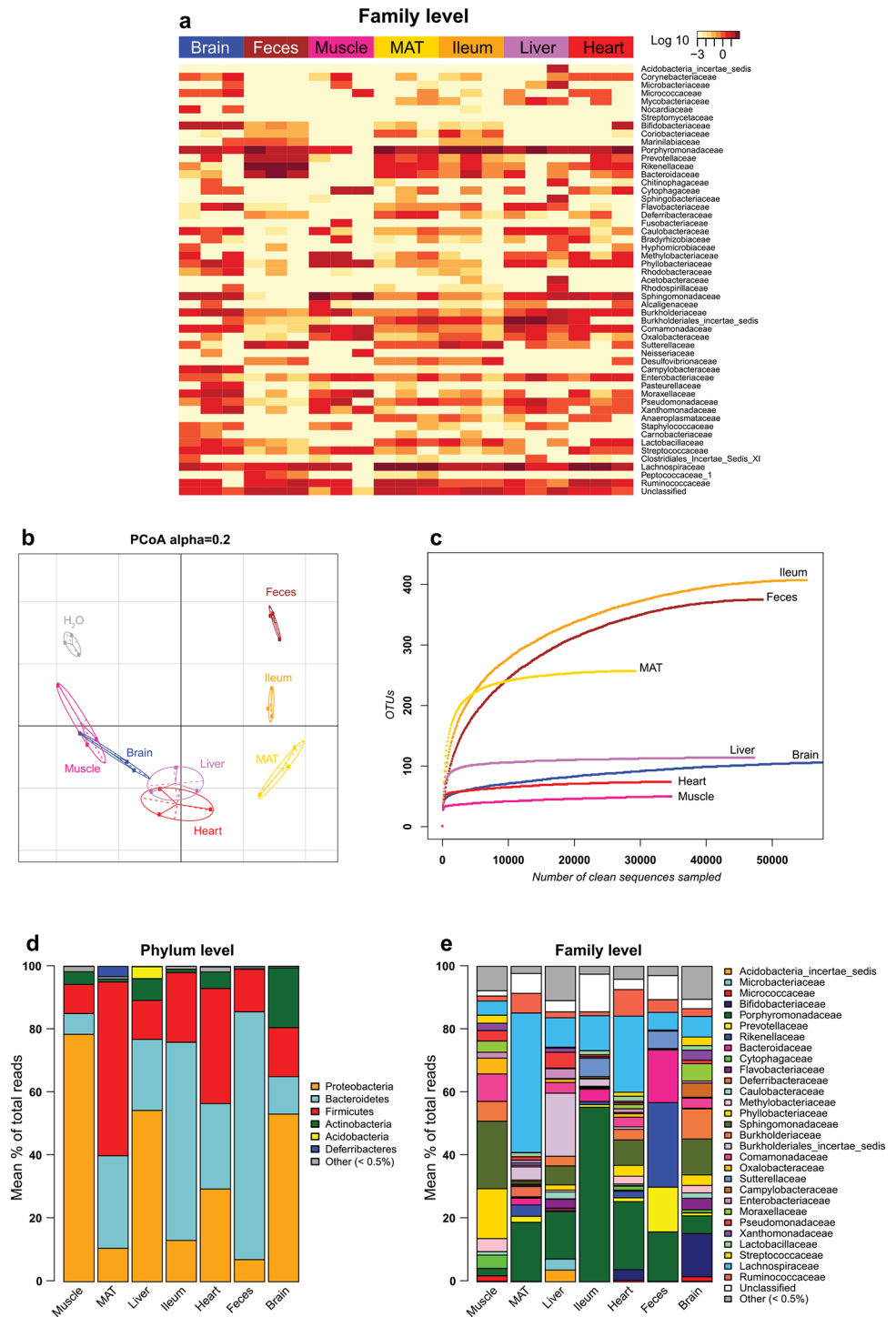
Fig 4. 16S metagenomics on diverse tissue samples. (a) Heatmap of the relative abundance of each bacterial family from sequencing of different mouse tissue samples performed in triplicate (three different mice for each tissue). Each line corresponds to a bacterial family; each one of the three columns for a tissue corresponds to a different mouse. **(b)** Generalized UniFrac distance-based PCoA analysis of sequencing data from the samples shown in **a** compared with the negative control generated by sequencing molecular biology-grade water with the same pipeline (H₂O). UniFrac weight parameter (Alpha) was set to 0.2 for this analysis. **(c)** Rarefaction curve of the sequencing of the samples shown in **a** and **b**. For each tissue, only the sample with the median number of OTU is displayed. **(d)** Stacked bar charts showing the relative abundance

of bacterial phyla obtained by sequencing of the mouse samples shown in **a**, **b** and **c**. **(e)** The relative abundance as in **d**, but at the family taxonomic level. MAT: Mesenteric adipose tissue. OTU: Operational Taxonomic Unit.

doi:10.1371/journal.pone.0142334.g004

to perform the quality control. The quality of the run was checked internally using PhiX Control and then each pair-end sequence was assigned to its sample using the multiplexing index.

Bioinformatics analysis

New MiSeq Indexes.   In order to increase the number of samples multiplexed on the same MiSeq run, new indexes were designed using two steps implemented in Perl scripts. The first step had the following rules: i) index length must be 6 bases in the alphabet ['A', 'C', 'G', 'T'], ii) each index pair must exhibit at least 2 mismatches and iii) the sequence of each index pair must not be reverse complement nor complement. The first step resulted in 466 new indexes. To avoid any mismatch and allow a proper calibration of the MiSeq camera, we selected before each run in the second step up to 320 optimal indexes (S2 Table) having i) minimum similarities between them, ii) equivalent proportions of A/C and T/G (corresponding to the two excitation LED of the MiSeq sequencer) and iii) proportions of each base at each position close to 25% among all the indexes. To reduce even more the risk of miss-assignment of reads to samples, we did not allow during demultiplexing any mismatch in the index sequences.

Sequencing Data Preprocessing.   The raw MiSeq sequencing data were processed using the NG6 application [22] for demultiplexing without mismatches on the index sequence and FASTQ generation using Illumina CASAVA v1.8.2 on the GenoToul computer cluster (68 nodes, each node consists of 40 core/80 threads CPU and 256 GB of RAM). The determination of a quality threshold to trim the reads based on the Illumina quality scores is difficult due to overestimation of the error probability for low quality base calls [23]. Therefore, we did not apply any *a priori* filters based on the Illumina quality scores to avoid unneeded loss of good sequence bases. The quality filters implemented in this approach are mainly based on the pair joining quality and the random nature of sequencing errors. Moreover, the quality trimming based on Illumina scores can reduce the length of the reads to the extent that overlap is lost, rendering them unusable for the rest of the analysis. To overcome this difficulty and to filter out poor quality reads, read pair joining was performed using FLASH (v1.2.7) with default values except for the following constraints i) reduce base mismatches between the 2 reads in the overlapping region by setting the maximum mismatch density to 0.1 ii) limit the minimum length of the overlap to 10 for 2x250bp reads and to 110 for 2x300 bp reads iii) set the maximum overlap length to 70 for 2x250bp reads and to 170 for 2x300 bp reads. The resulting FASTQ files for successfully joined pairs were converted to FASTA format using FASTX Toolkit (v0.0.14) and merged into a single file. This file constitutes our sequence working set.

16S Metagenomic Analysis Pipeline.   The bioinformatics pipeline used for the 16S metagenomics studies in tissues was based on the protocol published by Kozich et al. [24] and has similarities with the protocol developed in parallel by Unno [25], with adjustments for specific difficulties presented by the analysis of tissue microbiota and to adapt to the short overlap of our paired-end reads. The pipeline, run on an Intel Xeon server (16 cores/32 Threads CPU and 208 GB of RAM), is composed of the following steps:

Step 1: *Sequence Filtering and Trimming*. The previous step produces sequences of different lengths, which is not suited for sequence clustering into Operational Taxonomy Units (OTU). Non-specific PCR amplification could also have occurred especially with tissue samples containing high concentrations of both eukaryotic and non-eukaryotic DNA. In order to address

these issues, the Mothur analysis environment v1.33.0 [26] was used first to de-replicate (i.e. to cluster redundant sequences) the sequence working set obtained in the previous step and align the non-redundant sequences to the SILVA bacteria reference 16S alignment (v102) distributed with Mothur. The sequences that do not align to the reference 16S alignment are considered as non-specific PCR amplification and therefore culled from the sequence working set. The remaining aligned sequences were trimmed to obtain blunt ends to allow sequence clustering.

Step 2: *Sequencing Error Reduction.* As the OTU clustering could be dramatically affected by sequencing errors, and although the Illumina MiSeq technology has a low error rate, quality filtering is required to obtain accurate amplicon sequences. Contrary to the Roche 454 technology, the insertion-deletion errors with MiSeq are negligible (<0.001%). Substitution errors are more frequent (0.1%) and have to be tackled. The substitutions occur at random positions and at extremely low frequencies, making them relatively easy to detect when sequence coverage is sufficiently deep. We took advantage of this feature to aggregate rare sequences with more abundant and closely related groups of sequences. The Mothur pre.cluster command achieves this by merging low frequency sequences with very close higher frequency sequences using a modified single linkage algorithm. Three differences were allowed during this clustering step. At the end of this pre-clustering step, the singletons (unclustered sequences) were considered as either uncaught sequencing errors or extremely rare taxa with abundances under the background level of the method and were withdrawn from the sequence working set.

Step 3: *PCR Chimera Removal.* The chimeras were screened *de-novo* as described in the MiSeq SOP (August 2013, http://www.mothur.org/wiki/MiSeq_SOP), i.e. considering the most abundant sequences to build a reference of non-chimeric sequences, using UCHIME v4.2 [27] and removed from the sequence working set.

Step 4: *Taxonomic Assignment of Sequences.* The Mothur implementation of the Naive Bayesian Classifier [28] was run against the RDP rRNA training set (v9) to provide a taxonomic assignment to every sequence with a minimum bootstrap confidence score of 80%. The sequences that were not assigned to the Bacteria domain were filtered out for the rest of the analysis (including mitochondrial and chloroplast DNA). The primers specificity ensured that no sequences corresponding to mitochondria or chloroplast were found during this filtering.

Step 5: *Clustering into OTU.* The clustering of sequences was performed using the average neighbor algorithm on our clean sequence working set as described in the MiSeq SOP at a threshold of 0.03% identity. Finally, a consensus taxonomy is provided for each OTU based on the taxonomic assignment of individual reads using the default cutoff (51%).

For the samples sequenced in this study, we obtained on average 146,987 raw sequences (raw read pairs) per sample. 119,413 sequences (81.24% of the raw sequences) were kept on average after the manufacturer passing filters, we obtain 88,868 sequences on average (74.4% of the manufacturer passing filters pairs) after pair joining and related quality filters and finally 60,979 sequences on average (51.1% of the manufacturer passing filters pairs) after OTU clustering and related quality filters.

Statistics, data analysis and figures preparation

Custom R and Perl scripts were used to perform the data acquisition from Mothur output files, the statistical data analysis and generate the figures.

The 16S metagenomics profiles are displayed using heatmaps and barplots. Heatmaps were generated using the gplots CRAN library (v2.11.0) to plot the relative abundance of sequences for each sample assigned to the taxa at the different taxonomic levels. For better visualization of low abundant taxa, the data were log10-transformed. Barplots represent the proportion of sequences classified to taxa at the different taxonomic levels. The proportions presented are

either those calculated on single samples or averages calculated over different sequencing triplicates. For clarity, the low abundant taxa (mean proportion <0.5%) have been grouped into a single category called "other (<0.5%)". The 'unclassified' category represents the sequences that could not be assigned to a taxon at a given taxonomic level with at least 80% confidence score.

The dissimilarities between the 16S profiles of samples were analyzed using the principal coordinate analysis (PCoA). The Generalized UniFrac distance [29] was calculated using the GUniFrac CRAN library (v1.0) to generate the distance matrix. The alpha parameter, controlling the weight assigned to the most abundant taxa, is specified on each figure. For 16S rich samples such as stool or intestinal tract samples, the potential 16S contaminants contributed by the PCR mix is negligible. However, for more challenging samples such as tissue samples, negative target controls (molecular biology grade water sequenced with the same metagenomic pipeline) were added to the PCoA to ensure that the potential amplification of contaminants do not impact the sequencing data of our samples.

To determine the suitable sequencing depth for each tissue, a rarefaction curve was plotted for each sample. The rarefaction curve plots the number of OTU as a function of the number of sequences sampled in the clean sequence working set. We used a step of 100 sequences for the X axis and sampled each step 1000 times to calculate the average number of OTU plotted on the Y axis. For better readability, only the median curve for each tissue is shown.

Sequencing data are available on the European Nucleotide Archive (ENA) under the study ID "PRJEB10949" at http://www.ebi.ac.uk/ena/data/view/PRJEB10949. Custom R and Perl scripts are available on sourceforge at http://sourceforge.net/projects/tissue-microbiota/files/

Ethics Statement

Studies on mice were performed in accordance with the article R-214-89 of the French "Code rural et de la pêche maritime" section 6 "Use of living animal for scientific research" and approved by the ethical committee CEEA-122 of the SICOVAL Prologue Biotech Institute.

Results

Overall strategy and primer choice

The metagenomics pipeline comprises: 1) a DNA extraction step optimized to maximize the recovery of bacterial DNA from tissue samples; 2) a sequence library construction method based on a two-step PCR using short primers (the first PCR to target the 16S sequence, the second PCR to add a single index and the Illumina adapters; 3) the sequencing step on MiSeq and 4) a tailor-made bioinformatics processing of the data. We sequenced amplicons of 467 bp encompassing the V3 and V4 variable regions of the 16S gene since these variable regions are reported to have the broadest phylogenetic information for studying the microbiota of human and other mammalian species [30–32].

We used diverse mouse tissues and mock bacterial communities to test the pipeline and analyze the potential of the MiSeq technology for 16S metagenomics analysis of animal tissues. We performed numerous controls both *in vitro* and *in silico* to ensure the absence of artefacts such as cross-contamination between tissue samples or amplification of bacterial DNA contaminants from reagents. Indeed, many reagents required in the sequencing pipeline, including most of the TAQ polymerases available on the market, contain non negligible amounts of bacterial DNA [33–35]. A significant part of the pipeline optimization involved testing combinations of reagents to minimize bacterial contaminants and adapting the protocol to increase the yield of extraction and amplification of the bacterial DNA present in the tissue. The protocols

of tissue dissection and DNA extraction were also carefully designed to minimize any risk of contamination between different tissue samples.

Pipeline validation using BEI mock communities

First, we assessed the ability of the sequencing pipeline to detect bacterial taxa in the mock communities from BEI Resources (S1 Table) and to assign the corresponding OTU at the genus and family levels. We tested both the staggered (1,000 to 1,000,000 copies of RNA operon per organism per µl depending on the organisms) and even mock communities (100,000 copies of operon counts per organism per µl for each organism). All the bacteria strains present in the mock communities were sequenced and assigned to the family and genus levels (Fig 1A and 1B), except *Deinococcus radiodurans* strain R1: an unusual extremophile environmental bacteria of the phylum *Deinoccus-Thermus* [36] and for which the 16S gene cannot be amplified by the primer pair we designed (as predicted by *in silico* validation). *Propionibacterium acnes* was sequenced and assigned to the family and genus levels, but reads assigned to *Propionibacterium acnes* represented less than 0.5% of the total reads and therefore fell into the "Other" category on the barplots.

The PCR amplification, sequencing steps and bioinformatics processing create a biased distribution of the observed community in comparison to the actual distribution of the mock communities provided by BEI. This biased distribution depends on the 16S region amplified, the PCR primers and the proportion of each strain in the mock community [37–39]. The measured abundance of each strain at the genus level differs from the actual abundance by 0.4% to 9.7% (Median = 3.0) and by 0.02% to 6.7% (Median = 0.2) in the even and the staggered mock communities respectively (Fig 1A and 1B).

Pipeline validation using designed mock communities

The mock communities from BEI are used as standards by many research groups for molecular microbiology and metagenomic sequencing. This type of mock community is a useful tool to compare performance with published studies. Most of the organisms of the BEI mock communities are particularly common bacterial strains that are well represented in the taxonomic databases used in bioinformatics. However, biological tissue samples contain mostly less common bacteria that are poorly represented or absent from the available databases. For this reason, we also tested the pipeline using mock communities composed of random bacteria representative of different phyla of interest (Table 1). We purposely included strains of bacteria absent from the RDP database (*Geofilum rubicundum*, a genus and species of bacteria discovered in 2012 [40]) and strains of Verrucomicrobia (*Prosthecobacter fusiformis*), a phylum that cannot be amplified with our primers according to *in silico* simulations. We sequenced several mixes of these bacterial DNAs, either genomic DNA (Fig 1C) or plasmids incorporating the 16S genes of different strains of bacteria (Fig 1D and 1E). Genomic DNA is more similar to tissue samples and the use of plasmid-based mock communities allows precise quantitation of the number of 16S gene copies included in the mix.

The separate sequencing of the genomes of *Bifidobacterium animalis subsp. lactis*, *Ralstonia pickettii* and *Escherichia coli* resulted in assignment to the corresponding bacterial families (Bifidobacteriaceae, Burkholderiaceae, and Enterobacteriaceae) for 99.3%, 99.5%, and 99.9% of the reads respectively (Fig 1C).

A 1:1 genomic DNA mix of *Escherichia coli* and *Bifidobacterium animalis subsp. lactis* resulted in an assignment of respectively 41.6% and 57.7% to the corresponding family (Enterobacteriaceae and Bifidobacteriaceae) (Fig 1C). The deviation from 50% is explained by i) a

degree of unavoidable bias introduced by the sequencing pipeline and ii) the variation in 16S copy number per genome and genome size between bacteria [20,41].

A 1:1:1 genomic DNA mix of *Bifidobacterium animalis subsp. lactis*, *Ralstonia picketti* and *Ralstonia mannitolilytica* resulted in assignments of 37.0% to the *Bifidobacterium animalis subsp. lactis* family (Bifidobacteriaceae) and 59.5% to the family of the *Ralstonia* species (Burkholderiaceae) (Fig 1C).

Finally, we mixed and sequenced the same genomic DNA proportion (20%) of *Escherichia coli*, *Bifidobacterium animalis subsp. lactis* and three Burkholderiaceae: *Ralstonia picketti*, *Ralstonia mannitolilytica* and *Cupriavidus necator*. The resulting assignments were: 23.2% Enterobacteriaceae, 25.7% Bifidobacteriaceae and 49.0% Burkholderiaceae (Fig 1C).

We then tested a mix of 14 plasmids each of which carried a copy of the 16S gene of different strains (Table 1). Except for *Prosthecobacter fusiformis*, which, as confirmed by *in silico* analysis, cannot be amplified by the primers used, we identified all strains at the family level, including the recently discovered *Geofilum rubicundum* (Fig 1D), and 11 strains at the genus level (Fig 1E). The *Geofilum* genus is absent from the RDP database and is considered by the Naive Bayesian classifier as a bacterium from the *Alkaliflexus* genus, the genus closest to *Geofilum*[40]. The assigned proportions of the different bacteria were generally very close to the real copy numbers present in the samples (Fig 1D and 1E).

Pipeline versatility: example of primer optimization

We took advantage of the versatility of the pipeline to design and incorporate a new set of primers to amplify the Verrucomicrobia phylum in addition to the bacterial taxa recognized by the former set of primers ($\geq$ 95% of Ribosomal Database Project). With the new set of primers, all 14 strains of the designed mock community, including *Prosthecobacter fusiformis*, were sequenced and assigned to the correct family and 12 of them to the genus level (S1 Fig).

Pipeline sensitivity

We investigated the bias of the MiSeq pipeline by assessing the impact of sample dilution and variation of the relative proportions of bacterial 16S genes. We sequenced serial dilutions of the plasmid-based mock community to obtain between 10 and 10^7 copies of 16S gene for each of the 14 bacterial strains (Fig 2A and 2B). Between 10^7 copies and 10^4 copies of the 16S gene, the relative proportions were consistent and close to the actual values as shown on the bar plot (Fig 2A) and PCoA analysis (Fig 2B). However, at 10^3 copies or lower, the measured proportions progressively divert from the actual composition and the technical background (mainly arising from bacterial DNA contaminants present in the TAQ polymerase) represented an increasing fraction of the reads (Fig 2A) and became closer to the profile of the negative control (H_2O) (Fig 2B).

Accuracy of assessment of relative abundance. We sequenced seven mixes with different ratios of bacterial plasmids containing the 16S genes of *Geofilum rubicundum* and *Eubacterium barkeri*. The observed ratios corresponded consistently to the actual ratio hierarchy (Fig 2C) with some variations depending on the strains and the ratio of the plasmids. With *Geofilum rubicundum*, low proportions were overestimated: 5% *Geofilum rubicundum* generated 26.1% of the assigned reads, whereas 95% of the same plasmid produced a much more accurate ratio (93.4%) (Fig 2C). The bar plots showed that triplicates of each mix were consistent for all ratios (between 0.3% and 7.3% of variability, median = 3.8) (Fig 2C).

Replicability and reproducibility

We analyzed the replicability of the pipeline during the same sequencing run (Fig 3A and 3B and S2A–S2D Fig) and between runs (Fig 3C, S2E Fig). For assessing the variability within the

same run, all samples were treated in triplicate (whole metagenomic pipeline starting from extracted DNA). As shown in Fig 3A and 3B, the variability of the pipeline between triplicates is very low (median difference = 0.01 read % ± SEM 0.01) for all sample types (mock communities or tissues samples) and is independent of the bacterial taxa considered, their relative abundance or the taxonomy level. We also tested the reproducibility between several MiSeq runs on the same samples of ileum mucosa. Three samples of mouse ileum mucosa were sequenced on six to seven runs each using either the same library preparations (named "old") or by generating new libraries (named "new") from the extracted DNA samples (Fig 3C). We performed the different sequencing runs over a period of three months using either MiSeq Kit V2 (2x250bp) or V3 (2x300bp) and a different set of indexes. Moreover, the last run was performed on a different MiSeq desktop sequencer from the six previous runs. As shown on the bar plots (Fig 3C, S2E Fig) and the PCoA analysis (S3 Fig), variability between runs was very low (median difference = 0.01% of read ± SEM 0.02).

The same ileum samples (Fig 3C, S2E Fig), with the same 467 bp amplicon size, sequenced either with the MiSeq Kit v2 (2x250 bp) or Miseq Kit V3 (2x300 bp) gave the same results, despite the difference of overlap between the read pairs (33 bp for MiSeq V2 and 133 bp for MiSeq V3, using the 16S rDNA gene of E. coli as a reference). These show that the small overlap with the MiSeq V2 kit did not affect the quality of the results.

16S metagenomics on diverse tissue samples

Finally, we tested the pipeline by sequencing mouse samples from different organs and tissues: feces, ileum, liver, skeletal muscle, heart, brain and mesenteric adipose tissue (MAT). Each tissue was isolated from three different mice. Heat map representation (Fig 4A) and PCoA analysis (Fig 4B) showed that each sample was successfully sequenced and displayed a 16S metagenomic profile distinct from the technical background: H₂O added during the library preparation (Fig 2B, Fig 4B and S4 Fig) or at the beginning of the extraction step (S4 Fig). Thus, the tissue profiles observed were not the result of an artefact due to the amplification of the bacterial DNA contaminating the reagents. Each tissue has a distinct taxonomic profile (which excludes the presence of a significant contamination between tissues) with similarities between certain tissues (e.g. liver and heart) (Fig 4A and 4B). The high output of MiSeq sequencing, 20–45 million raw reads per run and between 30, 000 and 100, 000 sequences per sample (after filtering, cleaning, and assignation), generated high levels of relevant information and consequently the rarefaction analyses (Fig 4C) displayed an exhaustive description of the diversity of all tested samples. Some tissues, (e.g. feces, Ileum and adipose tissue), displayed a very high diversity in terms of bacterial content which could not have been assessed correctly with other lower throughput techniques. Bar plot representations of the results (Fig 4D and 4E) confirmed the previous observations and showed that the pipeline allowed effective assignment for this type of sample: 99.3% - 100% of assigned reads at the phylum level (Fig 4D) and 88.0% - 98.3% at the family level (Fig 4E). Assignments were more variable at the genus level, between 34.4% and 88.8%, probably reflecting a higher number of unknown genera in the reference taxonomy database for those samples (data not shown).

Discussion

By employing an optimized MiSeq-based pipeline for the analysis of bacterial 16S rDNA, we have characterized the diversity of bacterial taxa present in mouse tissues.

This study opens new avenues for the understanding of host to microbiota interactions in relation to homeostasis and disease. 16S metagenomic sequencing techniques, first used in samples extremely rich in bacterial content such as environmental samples or feces, need

specific optimizations to study more challenging samples like animal tissues that have lower levels of bacterial sequences and contain molecules that interfere with the DNA amplification methods [15,16]. Other groups have explored the potential of MiSeq-based methods for analyzing environmental and gut samples [24,42–44], but to our knowledge these have not previously been used to study microbiota in animal tissues. We have adapted and optimized the procedures provided by Illumina by introducing numerous unique modifications that are critical in order to obtain the sensitivity and robustness required to sequence the bacterial DNA present in animal tissues.

This optimized MiSeq pipeline is based on two PCR steps and a novel indexing strategy. The two-step PCR protocol allowed the use of shorter primers to target the 16S genes and to add the MiSeq adapters necessary for the sequencing steps (approximately 50 bases instead of up to 90 bases with a one-step PCR protocol). The shorter primers allow more efficient amplification of the samples, since long primers complicate the PCR cycle (higher annealing temperatures and increase of secondary structure formation) especially for challenging tissue samples. This strategy also allows the possibility of changing the specificity of the primer pairs without the need to redesign, resynthesize and re-optimize the primer pairs of the second PCR step. Indeed, the method allowed rapid modification of the primers to add the capability of sequencing the Verrucomicrobia phylum. With the same strategy, new primers can be efficiently designed to increase the length of the amplicon to exploit the increased read length offered by the new MiSeq Kit V3 (2x300 bp), target other regions of the 16S gene, other bacterial genes, or genomes from other kingdoms such as fungi. The second PCR step, which adds the indexes and sequencing adapters to the amplicons, can also be modified independently of the first step. Indeed we have replaced the original 180 indexes by 320 new indexes to take advantage of the higher throughput of the MiSeq v3 kit, without any need to modify or optimize the pipeline.

We extensively validated the pipeline with mock communities and negative controls to ensure the validity of the obtained results. We observed moderate differences between the actual proportions of 16S genes in the mock communities and the measured proportions assessed by sequencing. Indeed, as already reported, bias in the library preparation (PCR and cleaning), the sequencing steps and the bioinformatics analysis frequently introduce an over- or under-estimation of the proportions of a bacterial taxon, which depends on the individual taxa and their relative proportions in the samples [37–39,45]. However, the overall proportions were well respected, triplicates were consistent even between dilutions and progression was clearly observed when the proportion of a bacterial taxon was increased in the sample.

Because of the high diversity of the human microbiota and its variation among individuals, technologies with higher throughput than Roche 454 pyrosequencing are necessary to correlate microbiome composition with clinical parameters or disease states. The MiSeq technology, as assessed by the rarefaction analysis, allowed an exhaustive description of even the rich microbial population present in samples like feces and intestinal mucosa.

Another advantage of the MiSeq pipeline described here over 454-based sequencing protocols was the much higher reproducibility and replicability. The poor reproducibility of 454 technology between replicates and between runs has been noted in the literature [46–48]. Our MiSeq-based pipeline displayed a high reproducibility between replicates both within the same run and across runs. This reproducibility was manifest over broad time intervals and included sample triplicates from separate library preparations. This offers the important advantage of allowing robust comparisons of results between runs, for example for the analysis of large cohorts of samples in clinical trials or population studies.

In terms of read length and quality of assignment, studies have shown that despite higher read length, 454-based analyses do not exhibit a better depth of assignment than MiSeq [13,49]. Some studies point out that the type of mutations occurring with MiSeq (mainly

substitutions), combined with a suitable quality filtering step allowed by read pairing makes this technology more reliable in terms of taxonomic assignment [24,50,51]. In our studies, MiSeq reached the required criteria in terms of quality and read depth with a taxonomic assignment at both the family and genus levels. We did not attempt to assign the reads and OTU to the species level, since NGS technologies that have an output sequence length of only several hundred base pairs cannot be used to assign with confidence the reads and OTU to species level except in a handful of specific cases [14,52–54].

Tests of sensitivity demonstrated the reliability and reproducibility of the pipeline down to 10^3 copies of 16S rDNA in our samples. This corresponds to around 100–500 bacterial genomes in the sample, since each bacterium has about 1–15 copies of the 16S gene per genome [20,41]. In our experience, this level of sensitivity is sufficient for the analysis of human and animal tissue samples. With samples of low bacterial content, it is necessary to include negative controls in order to take into account the technical background during the analysis. All our negative controls (as illustrated in the different figures) demonstrate that the bacterial contaminants do not affect the quality of our results. In addition, the fact that each tissue sample has a specific and reproducible bacterial taxa profile supports the fact that the observed profile cannot be attributed to reagent contaminants or contaminants from other samples.

Because of technical difficulties and the belief that tissues and organs of healthy individuals do not contain bacteria, few studies have been undertaken to examine tissue bacteria except from epithelial surfaces [1,2,6–9]. We describe here the successful sequencing of microbiota from a variety of internal organs and identified reproducibly a bacterial DNA profile specific for each tissue. These findings raise important questions concerning the role of tissue microbiota in animal physiology. The presence of bacterial DNA in tissues does not necessarily imply the presence of living bacteria. Free bacterial DNA could be present in the intercellular space or most likely within cells, in particular immune cells. On the contrary, the reported absence of cultivable bacteria in healthy tissue is not synonymous with the absence of living bacteria in these tissues. Indeed, several technical limitations as well as biological features of living or dormant bacteria explain why the vast majority of microbial species from both environmental or tissue samples remain uncultivated [10,55,56]. Consequently, we cannot ascertain today what proportion of the bacterial DNA sequenced in tissues corresponds to living bacteria.

In addition to the tissue-specificity of the bacterial DNA profiles, a striking observation in our study was the alpha diversity richness in mesenteric adipose tissue. This supports the conclusions of several published studies concerning the involvement of host microbiota in obesity and diabetes [8,11,12,57–59]. The presence of a specific bacterial DNA profile in the brain is much more surprising since brain is assumed to be a sterile organ in the absence of disease. However, this observation, which was not possible before the development of next generation sequencing, is in accordance with previous studies that analyzed brain tissue from immunodepressive humans and rodents [60]. It would be interesting to investigate the potential role of bacteria or bacterial DNA present in the central nervous system in the control of the brain immune system by host microbiota [61,62] and in the development of neurodegenerative diseases [63–65].

In terms of bacterial DNA profile, we found in the gut (fecal and ileum samples) DNA belonging mostly to the Bacteroidetes and Firmicutes phyla, as previously shown by other groups in both mice and Humans [66–69]. Mesenteric adipose tissue also contains DNA mostly from Bacteroidetes and Firmicutes but not in the same proportion (more Firmicutes than Bacteroidetes). In addition, the mesenteric adipose tissue is the only tissue sample in our study shown to contain significant amounts of DNA belonging to the Deferribacters phylum.

Muscle, liver, heart and brain samples contain more DNA from the Proteobacteria and Actinobacteria phyla than the gut or adipose tissue. Liver is the only tissue to have DNA belonging to the Acidobacteria phylum. Overall, the bacterial DNA profiles of these non-intestinal tissues differ dramatically from those of the fecal or ileum samples, despite the fact that their microbiome derives at least partially (and probably mostly) from the gut as a result of bacterial translocation [70–73]. The difference in bacterial profiles between the gut and other tissues could be explained by the role of filter played by the intestinal and immune cells. This mechanism limits the translocation of a specific portion of the gut microbiota to the periphery. In addition, each tissue has its own specific bacterial DNA profile, but which is shared at least partially among individuals. The tissue specificity of the bacterial DNA profile could be explained both in terms of microbiology (living/dormant bacteria in a specific tissue environment, which is already demonstrated for example for bacterial species in blood [74–79]) and in terms of immunology (eukaryotic cells which could carry specific bacterial DNA depending on their location [80,81]). These mechanisms, as well as the physiological role of the bacteria and bacterial DNA present in the tissue, remain to be studied in detail.

Conclusions

This enhanced method of 16S tissue metagenomic sequencing has permitted the first characterisation of bacterial DNA profiles in several animal tissues and will allow the scientific community to address the nature and role of tissue microbiota in human physiology and disease. These observations introduce the possibility of designing new approaches for the discovery of novel biomarkers and therapeutic targets.

Supporting Information

S1 Fig. Validation of the new primers against mock communities. (a) Stacked bar charts showing the actual relative abundance of bacterial families in the plasmid based mock community and the measured relative abundance of the families obtained with the MiSeq sequencing pipeline using either the original primers (described in the methods), or the new primers (designed to amplify also the Verrucomicrobia phylum). **(b)** The relative abundance as in **a**, but at the genus taxonomic level. The sequencing was performed in triplicate for all the samples (starting from the extracted DNA); the means of the triplicates are shown on the stacked bar charts.
(PDF)

S2 Fig. Replicability and reproducibility. (a-d) Stacked bar charts showing in triplicates the actual relative abundance and the measured relative abundance of the families **(a, c)** and genus **(b, d)** of the BEI mock communities **(a, b)** and our own designed mock communities **(c, d)**. **(e)** Stacked bar charts showing the relative abundance of bacterial families obtained by sequencing of three samples of mouse ileum mucosa in six to seven runs each with different parameters described in the legend at the bottom left: different runs, new libraries from same extracted DNA or the same libraries already prepared, different MiSeq kit generations and different sequencers. Two experimenters performed the different runs and reagent batch numbers (including Taq polymerase) varied from run to run.
(PDF)

S3 Fig. PCoA analysis of the reproducibility between runs. Generalized UniFrac distance based PCoA analysis of the sequencing of six samples of mouse ileum mucosa in six to seven runs each with different parameters as described in the Fig 3A and S2E Fig. UniFrac weight

parameter (Alpha) was set to 0.6 for this analysis.
(PDF)

S4 Fig. PCoA analysis of the negative controls. Generalized UniFrac distance based PCoA
analysis of the sequencing of 10 samples of mouse mesenteric adipose tissue and 2 x 5 negative
controls technical replicates. H20 ext: negative control performed by replacing the tissue sam-
ple by molecular grade water in lysis/extraction step of the pipeline. H20: negative control per-
formed by replacing the extracted DNA by molecular grade water in the first step of library
preparation. UniFrac weight parameter (Alpha) was set to 0.6 for this analysis.
(PDF)

S1 Table. BEI Resources mock communities. Bacterial stain list of the genomic DNA mixture
with concentration (16S rDNA copies/μl) of the even and staggered Mixture.
(PDF)

S2 Table. Optimal multiplexing indexes. 320 indexes designed to allow a proper calibration
of the MiSeq camera and accurate demultiplexing of samples.
(PDF)

Author Contributions

Conceived and designed the experiments: BL FS SP OB. Performed the experiments: JL SP CV.
Analyzed the data: BL FS SP. Contributed reagents/materials/analysis tools: FS SV CK GV CD
OB RB. Wrote the paper: BL FS MC. Gave conceptual advice and revised the manuscript: MC
RB JA.

References

1. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science (80-). 2009; 326: 1694–7. doi: 10.1126/science. 1177486

2. Segal LN, Blaser MJ. A brave new world: the lung microbiota in an era of change. Ann Am Thorac Soc. 2014; 11 Suppl 1: S21–7. doi: 10.1513/AnnalsATS.201306-189MG PMID: 24437400

3. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature. 2014; 509: 357–60. doi: 10.1038/nature13178 PMID: 24739969

4. Holmes E, Li J V, Marchesi JR, Nicholson JK. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. Cell Metab. Elsevier Inc.; 2012; 16: 559–64. doi: 10.1016/j. cmet.2012.10.007

5. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. Science (80-). 2012; 336: 1262–7. doi: 10.1126/science.1223813

6. Yan AW, Schnabl B. Bacterial translocation and changes in the intestinal microbiome associated with alcoholic liver disease. World J Hepatol. 2012; 4: 110–8. doi: 10.4254/wjh.v4.i4.110 PMID: 22567183

7. Koren O, Spor A, Felin J, Fåk F, Stombaugh J, Tremaroli V, et al. Human oral, gut, and plaque micro-biota in patients with atherosclerosis. Proc Natl Acad Sci. 2011; 108 Suppl: 4592–8. doi: 10.1073/pnas. 1011383107

8. Amar J, Serino M, Lange C, Chabo C, Iacovoni J, Mondot S, et al. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. Diabetologia. 2011; 54: 3055–61. doi: 10.1007/ s00125-011-2329-8 PMID: 21976140

9. Amar J, Lange C, Payros G, Garret C, Chabo C, Lantieri O, et al. Blood microbiota dysbiosis is associ-ated with the onset of cardiovascular events in a large general population: the D.E.S.I.R. Study. PLoS One. 2013; 8: e54461. doi: 10.1371/journal.pone.0054461 PMID: 23372728

10. Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory dis-eases. FEMS Microbiol Rev. 2015; 1–25. doi: 10.1093/femsre/fuv013 PMID: 25940667

11. Burcelin R, Serino M, Chabo C, Garidou L, Pomié C, Courtney M, et al. Metagenome and metabolism: the tissue microbiota hypothesis. Diabetes Obes Metab. 2013; 15 Suppl 3: 61–70. doi: 10.1111/dom. 12157 PMID: 24003922

12. Luche E, Cousin B, Garidou L, Serino M, Waget A, Barreau C, et al. Metabolic endotoxemia directly increases the proliferation of adipocyte precursors at the onset of metabolic diseases through a CD14-dependent mechanism. Mol Metab. 2013; 2: 281–91. doi: 10.1016/j.molmet.2013.06.005 PMID: 24049740

13. Loman NJ, Misra R V, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012; 30: 434–9. doi: 10.1038/nbt.2198 PMID: 22522955

14. Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a Microbiome Study. Cell. 2014; 158: 250–262. doi: 10.1016/j.cell.2014.06.037 PMID: 25036628

15. Tichopad A, Didier A, Pfaffl MW. Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants. Mol Cell Probes. 2004; 18: 45–50. doi: 10.1016/j.mcp.2003.09.001 PMID: 15036369

16. Opel KL, Chung D, McCord BR. A study of PCR inhibition mechanisms using real time PCR. J Forensic Sci. 2010; 55: 25–33. doi: 10.1111/j.1556-4029.2009.01245.x PMID: 20015162

17. Lane DJ. 16S/23S rRNA sequencing. Nucleic Acid Techniques in Bacterial Systematics. 1991. pp. 115–175.

18. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991; 173: 697–703. PMID: 1987160

19. Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. Annu Rev Genet. 2004; 38: 771–92. doi: 10.1146/annurev.genet.38.072902.094318 PMID: 15568993

20. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS One. 2013; 8: e57923. doi: 10.1371/journal.pone.0057923 PMID: 23460914

21. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology. 2002; 148: 257–66. PMID: 11782518

22. Mariette J, Escudié F, Allias N, Salin G, Noirot C, Thomas S, et al. NG6: Integrated next generation sequencing storage and processing environment. BMC Genomics. 2012; 13: 462. doi: 10.1186/1471-2164-13-462 PMID: 22958229

23. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008; 36: e105. doi: 10.1093/nar/gkn425 PMID: 18660515

24. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. American Society for Microbiology; 2013; 79: 5112–20. doi: 10.1128/AEM.01043-13 PMID: 23793624

25. Unno T. Bioinformatic Suggestions on MiSeq-based Microbial Community Analysis. J Microbiol Biotechnol. 2015;

26. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009; 75: 7537–41. doi: 10.1128/AEM.01541-09 PMID: 19801464

27. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011; 27: 2194–200. doi: 10.1093/bioinformatics/btr381 PMID: 21700674

28. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007; 73: 5261–7. doi: 10.1128/AEM.00062-07 PMID: 17586664

29. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics. 2012; 28: 2106–13. doi: 10.1093/bioinformatics/bts342 PMID: 22711789

30. Soergel DAW, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 2012; 6: 1440–4. doi: 10.1038/ismej.2011.208 PMID: 22237546

31. Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J Microbiol Methods. 2011; 84: 81–7. doi: 10.1016/j.mimet.2010.10.020 PMID: 21047533

32. Jumpstart Consortium Human Microbiome Project Data Generation Working. Evaluation of 16S rDNA-based community profiling for human microbiome research. PLoS One. 2012; 7. doi: 10.1371/journal.pone.0039315 PMID: 23349657

33. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014; 12: 87. doi: 10.1186/s12915-014-0087-z PMID: 25387460

34. Spangler R, Goddard NL, Thaler DS. Optimizing Taq polymerase concentration for improved signal-to-noise in the broad range detection of low abundance bacteria. PLoS One. Public Library of Science; 2009; 4: e7010. doi: 10.1371/journal.pone.0007010 PMID: 19753123

35. Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLoS One. 2014; 9: e97876. doi: 10.1371/journal.pone.0097876 PMID: 24837716

36. Makarova KS, Aravind L, Wolf YI, Tatusov RL, Minton KW, Koonin E V, et al. Genome of the extremely radiation-resistant bacterium Deinococcus radiodurans viewed from the perspective of comparative genomics. Microbiol Mol Biol Rev. 2001; 65: 44–79. doi: 10.1128/MMBR.65.1.44–79.2001 PMID: 11238985

37. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, et al. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. 2010; 4: 642–7. doi: 10.1038/ismej.2009.153 PMID: 20090784

38. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One. 2012; 7: e43093. doi: 10.1371/journal.pone.0043093 PMID: 22905208

39. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One. 2011; 6: e27310. doi: 10.1371/journal.pone.0027310 PMID: 22194782

40. Miyazaki M, Koide O, Kobayashi T, Mori K, Shimamura S, Nunoura T, et al. Geofilum rubicundum gen. nov., sp. nov., isolated from deep subseafloor sediment. Int J Syst Evol Microbiol. 2012; 62: 1075–80. doi: 10.1099/ijs.0.032326–0 PMID: 21705444

41. Nishida H. Genome DNA Sequence Variation, Evolution, and Function in Bacteria and Archaea. Curr Issues Mol Biol. 2012; 15: 19–24. PMID: 22772895

42. Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmento H, et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environ Microbiol. 2013; doi: 10.1111/1462-2920.12250 PMID: 24102695

43. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. Nature Publishing Group; 2012; 6: 1621–4. doi: 10.1038/ismej.2012.8 PMID: 22402401

44. Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, et al. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol. 2013; 531: 371–444. doi: 10.1016/B978-0-12-407863-5.00019–8 PMID: 24060131

45. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. Appl Environ Microbiol. 1998; 64: 3724–30. PMID: 9758791

46. Kim M, Yu Z. Variations in 16S rRNA-based microbiome profiling between pyrosequencing runs and between pyrosequencing facilities. J Microbiol. 2014; 52: 355–65. doi: 10.1007/s12275-014-3443-3 PMID: 24723104

47. Ge Y, Schimel JP, Holden PA. Analysis of run-to-run variation of bar-coded pyrosequencing for evaluating bacterial community shifts and individual taxa dynamics. PLoS One. 2014; 9: e99414. doi: 10.1371/journal.pone.0099414 PMID: 24911191

48. Zhan A, He S, Brown EA, Chain FJJ, Therriault TW, Abbott CL, et al. Reproducibility of pyrosequencing data for biodiversity assessment in complex communities. Faith D, editor. Methods Ecol Evol. 2014; 5: 881–890. doi: 10.1111/2041-210X.12230

49. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PLoS One. 2014; 9: e94249. doi: 10.1371/journal.pone.0094249 PMID: 24722003

50. Frey KG, Herrera-Galeano JE, Redden CL, Luu T V, Servetas SL, Mateczun AJ, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. BMC Genomics. 2014; 15: 96. doi: 10.1186/1471-2164-15-96 PMID: 24495417

51. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using illumina paired-end technology. PLoS One. 2013; 8: e66643. doi: 10.1371/journal.pone.0066643 PMID: 23799126

52. Ward DM, Cohan FM, Bhaya D, Heidelberg JF, Kühl M, Grossman a. Genomics, environmental genomics and the issue of microbial species. Heredity (Edinb). 2008; 100: 207–19. doi: 10.1038/sj.hdy.6801011

53. Cohan FM. What are bacterial species? Annu Rev Microbiol. 2002; 56: 457–487. doi: 10.1146/annurev.micro.56.012302.160634 PMID: 12142474

54. Janda JMJ, Abbott SSL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol. 2007; 45: 2761–4. doi: 10.1128/JCM.01228-07 PMID: 17626177

55. Epstein SS. The phenomenon of microbial uncultivability. Curr Opin Microbiol. Elsevier Ltd; 2013; 16: 636–42. doi: 10.1016/j.mib.2013.08.003 PMID: 24011825

56. Hugenholtz P. Exploring prokaryotic diversity in the genomic era. Genome Biol. 2002; 3.

57. Sanmiguel C, Gupta A, Mayer EA. Gut Microbiome and Obesity: A Plausible Explanation for Obesity. Curr Obes Rep. 2015; 4: 250–261. doi: 10.1007/s13679-015-0152-0 PMID: 26029487

58. Bleau C, Karelis AD, St-Pierre DH, Lamontagne L. Crosstalk between intestinal microbiota, adipose tissue and skeletal muscle as an early event in systemic low-grade inflammation and the development of obesity and diabetes. Diabetes Metab Res Rev. 2014; doi: 10.1002/dmrr.2617 PMID: 25352002

59. Guida S, Venema K. Gut microbiota and obesity: Involvement of the adipose tissue. J Funct Foods. 2015; 14: 407–423. doi: 10.1016/j.jff.2015.02.014

60. Branton WG, Ellestad KK, Maingat F, Wheatley BM, Rud E, Warren RL, et al. Brain microbial populations in HIV/AIDS: α-proteobacteria predominate independent of host immune status. PLoS One. Public Library of Science; 2013; 8: e54673. doi: 10.1371/journal.pone.0054673

61. Erny D, Hrabě de Angelis AL, Jaitin D, Wieghofer P, Staszewski O, David E, et al. Host microbiota constantly control maturation and function of microglia in the CNS. Nat Neurosci. 2015; doi: 10.1038/nn.4030 PMID: 26030851

62. Wang Y, Kasper LH. The role of microbiome in central nervous system disorders. Brain Behav Immun. 2014; 38: 1–12. doi: 10.1016/j.bbi.2013.12.015 PMID: 24370461

63. Itzhaki RF, Wozniak MA, Appelt DM, Balin BJ. Infiltration of the brain by pathogens causes Alzheimer's disease. Neurobiol Aging. 25: 619–27. doi: 10.1016/j.neurobiolaging.2003.12.021 PMID: 15172740

64. Friedland RP. Mechanisms of molecular mimicry involving the microbiota in neurodegeneration. J Alzheimers Dis. 2015; 45: 349–62. doi: 10.3233/JAD-142841 PMID: 25589730

65. Catanzaro R, Anzalone M, Calabrese F, Milazzo M, Capuana M, Italia A, et al. The gut microbiota and its correlations with the central nervous system disorders. Panminerva Med. 2015; 57: 127–43. PMID: 25390799

66. Everard A, Lazarevic V, Gaïa N, Johansson M, Ståhlman M, Backhed F, et al. Microbiome of prebiotic-treated mice reveals novel targets involved in host response during obesity. ISME J. 2014; 8: 2116–30. doi: 10.1038/ismej.2014.45 PMID: 24694712

67. Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, et al. A catalog of the mouse gut metagenome. Nat Biotechnol. 2015; 33: 1103–8. doi: 10.1038/nbt.3353 PMID: 26414350

68. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006; 444: 1027–1031. doi: 10.1038/nature05414 PMID: 17183312

69. Ley RE, Bäckhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci. 2005; 102: 11070–5. doi: 10.1073/pnas.0504978102 PMID: 16033867

70. Wang L, Llorente C, Hartmann P, Yang A-M, Chen P, Schnabl B. Methods to determine intestinal permeability and bacterial translocation during liver disease. J Immunol Methods. 2015; 421: 44–53. doi: 10.1016/j.jim.2014.12.015 PMID: 25595554

71. De Punder K, Pruimboom L. Stress induces endotoxemia and low-grade inflammation by increasing barrier permeability. Front Immunol. 2015; 6: 223. doi: 10.3389/fimmu.2015.00223 PMID: 26029209

72. Kruis T, Batra A, Siegmund B. Bacterial translocation—impact on the adipocyte compartment. Front Immunol. 2014; 4: 510. doi: 10.3389/fimmu.2013.00510 PMID: 24432024

73. Ono S, Tsujimoto H, Yamauchi A, Hiraki S, Takayama E, Mochizuki H. Detection of microbial DNA in the blood of surgical patients for diagnosing bacterial translocation. World J Surg. 2005; 29: 535–9. doi: 10.1007/s00268-004-7618-7 PMID: 15776295

74. Wang Z, Zhang L, Guo Z, Liu L, Ji J, Zhang J, et al. A unique feature of iron loss via close adhesion of Helicobacter pylori to host erythrocytes. PLoS One. 2012; 7: e50314. doi: 10.1371/journal.pone.0050314 PMID: 23185604

75. Horzempa J, O'Dee DM, Stolz DB, Franks JM, Clay D, Nau GJ. Invasion of erythrocytes by Francisella tularensis. J Infect Dis. 2011; 204: 51–9. doi: 10.1093/infdis/jir221 PMID: 21628658

76. Brekke O-L, Hellerud BC, Christiansen D, Fure H, Castellheim A, Nielsen EW, et al. Neisseria meningitidis and Escherichia coli are protected from leukocyte phagocytosis by binding to erythrocyte

complement receptor 1 in human blood. Mol Immunol. 2011; 48: 2159–69. doi: 10.1016/j.molimm.2011.07.011 PMID: 21839519

77.   Tedeschi GG, Bondi A, Paparelli M, Sprovieri G. Electron microscopical evidence of the evolution of corynebacteria-like microorganisms within human erythrocytes. Experientia. 1978; 34: 458–60. PMID: 639937

78.   Yamaguchi M, Terao Y, Mori-Yamaguchi Y, Domon H, Sakaue Y, Yagi T, et al. Streptococcus pneumoniae invades erythrocytes and utilizes them to evade human innate immunity. PLoS One. 2013; 8: e77282. doi: 10.1371/journal.pone.0077282 PMID: 24194877

79.   Damgaard C, Magnussen K, Enevold C, Nilsson M, Tolker-Nielsen T, Holmstrup P, et al. Viable Bacteria Associated with Red Blood Cells and Plasma in Freshly Drawn Blood Donations. PLoS One. 2015; 10: e0120826. doi: 10.1371/journal.pone.0120826 PMID: 25751254

80.   Lipford GB, Heeg K, Wagner H. Bacterial DNA as immune cell activator. Trends Microbiol. 1998; 6: 496–500. PMID: 10036729

81.   Häcker G, Redecke V, Häcker H. Activation of the immune system by bacterial CpG-DNA. Immunology. 2002; 105: 245–51. PMID: 11918685