High-throughput sequencing of bacterial communities

One aliquot (100 mg) of each fecal sample was mixed with 500 µl phenol:chloroform:isoamyl, 900 µl of ASL buffer (Qiagen, QIAamp® stool mini kit) and 200 mg zirconia beads. Cells were lyzed using a bead beater (FastPrep-24; MP Biomedicals LLC) at 6.5 M/s, 0.5 min (4 cycles interrupted by 3 minutes on ice) and incubated at 95 °C (5 min, 800 rpm). After centrifugation (1 min, 13,000 g) the aqueous phase was collected into a fresh Eppendorf tube and filled up to 1.2 ml with ASL buffer. DNA was purified using the QIAamp® stool mini kit following the manufacturer's instructions (Qiagen, QIAamp® stool mini kit), with addition of RNAse treatment (37 °C, 30 min). DNA concentrations and purity (A260/A280 ratio) was determined by spectrophotometric analysis (ND-1000 spectrophotometer, NanoDrop Technologies, Willigton, USA). The V4 region (253 bp) of 16S rRNA genes was amplified (30 cycles) as described previously [1, 2]. Amplicons were purified using the AMPure XP system (Beckmann), pooled in equimolar amount and sequenced in paired-end modus (PE200) using the MiSeq system (Illumina Inc) following the manufacturer's instructions and a final DNA concentration of 16 pM and 5 % (v/v) PhiX standard library.

Sequence analysis

Raw read files were demultiplexed (allowing a maximum of 2 errors in barcodes) and each sample was processed using usearch [3] following the UPARSE approach [4]. First, all reads were trimmed to the position of the first base with quality score <3 and then paired. The resulted sequences were size filtered excluding those with assembled size <220 and >280 nucleotides. Paired reads with expected error >3 were further filtered out and the remaining sequences were trimmed by 10 nucleotides on each side to avoid GC bias and non-random base composition. For each sample, sequences were de-replicated and checked for chimeras with UCHIME [5]. Sequences from all samples were merged, sorted by abundance, and operational taxonomic units (OTUs) were picked at a threshold of 97 % similarity. Finally, all sequences were mapped back to the representative sequences resulting in one

OTU table for all samples. Only those OTUs with a relative abundance above 0.5 % total sequences in at least one sample were kept. RDP classifier [6] was used to assign taxonomic classification to the OTUs representative sequences (80 % confidence). A phylogenetic tree was constructed using fasttree [7]. OTU counts were expressed as proportions of total sequences per sample and normalized to the minimum count of total sequences observed across samples. For estimation of diversity within samples (*alpha*-diversity), the Shannon index was calculated and transformed to the corresponding effective number of species as described by Jost [8]. Sequences of candidate OTUs were assigned to taxa with valid names in nomenclature using the EzTaxon database [9].

High-resolution mass spectrometry

Sample preparation was performed on dry ice. Fecal samples (5 mg each) were placed into tubes containing ceramic beads (NucleoSpin® Bead Tubes, MACHEREY-NAGEL GmbH & Co. KG) and a stainless metal bead (QIAGEN®, 5 mm). A volume of 1 mL cold methanol (-20 °C) (Fluka® Analytical, LC-MS CHROMASOLV®) was added in each tube followed by homogenization (5 min, 30 1/s) using a TissueLyser II (QIAGEN®). After centrifugation (14,000 rpm, 10 min, 4 °C), supernatants were collected into new Eppendorf tubes and stored at -80 °C until further processed. Prior to MS measurement, a volume of 100 µL of each sample was diluted in 300 µL of methanol and the total volume was placed into maximum recovery vial (Waters®, 12x32 mm). Samples were measured using a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) (BRUKER® solariX, Daltonic GmbH, Bremen, Germany) equipped with a 12 Tesla superconducting magnet and an Apollo II ESI source. The instrument was externally calibrated on clusters of arginine (m/z of 173.10440, 347.21607, 521.32775 and 695.43943) dissolved in methanol at a concentration of 5 mg/mL. Samples were placed in an autosampler (GILSON®) at -4 °C in a randomized order and were injected into the mass spectrometer at a flow rate of 120 µL/h. Data were acquired in the negative ion acquisition mode with the following parameters: capillary voltage -3,600 V, ion accumulation time -0.3 s, flight

time to acquisition cell -1 ms, drying gas flow rate -4 L/min, drying gas temperature -180 °C, nebulizer gas pressure -2 bar. A total of 500 scans were accumulated per sample in the mass range of 120-1000 m/z.

Metabolome analysis

A sine apodization was performed before Fourier transformation of the time-domain transient. Raw data were processed with DataAnalysis 4.1 (BRUKER®). A FTMS peak picking algorithm was applied to identify peaks in the raw data using the following parameters: relative intensity threshold (relative to the highest peak in a spectrum) -0.001 %, signal-to-noise ratio threshold -4, absolute intensity threshold -100 (default). Identified m/z signals were internally calibrated using a reference list provided by the software. Data were exported as individual .ascii files containing m/z values and corresponding intensities. Peak alignment and data matrix creation were performed by in-house algorithm written in Fortran that searches for mass signals from different spectra that are close to each other (within 1 ppm window) combining them into one group that represents a unique feature. In order to stabilize the variance, m/z values with frequency <10% across all samples were deleted. The data was additionally filtered by an algorithm written in MATLAB that creates a mass-difference network using experimental m/z values as input and, according to the constructed network, annotates the m/z values by unique molecular formulas [10, 11]. The network can be used to reveal clusters of compounds, assign them to possible metabolite classes, and hypothesize their chemical structure. Thereby, after filtering out features that could not be part of the network, every remaining feature corresponds to unique molecular formula. m/z values were also annotated using the MassTRIX webserver [12] that searches for matches between experimental and theoretical m/z values within 1 ppm error window using KEGG, HMDB, and the LipidMAPS databases as input sources. Missing values were substituted by the level of noise from original spectra.

Iron measurement

Fecal samples were freeze-dried (Christ-Heraeus), weighed and introduced into quartz vessels. Subsequently, 1 mL of suprapure HNO_3 (Merck) subjected to sub-boiling distillation (Berghoff distillation apparatus) was added. Vessels were introduced into a pressure digestion system for 10 h at 170 °C. Resulting solutions were filled up to 10 mL with Milli-Q H_2O and iron content was determined using an ICP-AES Spectro Ciros Vision system (SPECTRO Analytical Instruments). Samples were introduced at a flow rate of 0.6 mL/min using a peristaltic pump connected to a Meinhard nebulizer with a cyclone spray chamber. The measured spectral line corresponding to iron was 259.941 nm. Radiofrequency power was set to 1400 W. The flow of plasma and nebulizer gas was set to 13 L Ar/min and to approximately 0.6 L Ar/min, respectively. Three blank samples followed by one control sample of a certified standard were acquired each ten measurements. Iron levels were calculated on a computerized lab-data management system integrating sample measurements with calibration curves, output from blank and control samples, and digested sample weight.

References

1 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. The ISME journal 2012;**6**:1621-4.

2 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ*, et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America 2011;**108 Suppl 1**:4516-22.

3 Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;**26**:2460-1.

4 Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods 2013;**10**:996-8.

5 Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 2011;**27**:2194-200.

6 Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology 2007;**73**:5261-7.

7 Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS one 2010;**5**:e9490.

8 Jost L. Partitioning diversity into independent alpha and beta components. Ecology 2007;**88**:2427-39.

9 Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK*, et al.* EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. International journal of systematic and evolutionary microbiology 2007;**57**:2259-61.

10 Breitling R, Ritchie S, Goodenowe D, Stewart ML, Barrett MP. prediction of metabolic networks using Fourier transform mass spectrometry data. Metabolomics : Official journal of the Metabolomic Society 2006;**2**:155-64.

11 Tziotis D, Hertkorn N, Schmitt-Kopplin P. Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity. Eur J Mass Spectrom (Chichester, Eng) 2011;**17**:415-21.

12 Wagele B, Witting M, Schmitt-Kopplin P, Suhre K. MassTRIX reloaded: combined analysis and visualization of transcriptome and metabolome data. PloS one 2012;**7**:e39860.