The Dominant Collective Proteome of Gut Microbes Differs between Patients with Quiescent Crohn's Disease and Healthy Controls

Supplementary Methods

Supplementary method 1

Preparation of bacterial fractions

Only one stool was processed at a time in order to minimize the processing time. All materials and solutions were sterilized and equilibrated in an anaerobic chamber before use. All steps except ultracentrifugation were performed in the anaerobic chamber. Biological samples and all materials were maintained at 4° C or at similar cold temperature (on ice) during the entire extraction process in order to minimize *de novo* protein synthesis and enzymatic activities. The microbiota was separated from the fecal matrix by flotation in a preformed Nycodenz continuous gradient (online supplementary figure 1). Briefly, Ultra-Clear™ centrifuge tubes (25×89 mm, Beckman Instruments, CA, USA) were filled with 16 ml of a degassed 20% w/v Nycodenz solution in 1X PBS containing 0.03% w/v Na-deoxycholate (density 1.108 mg/ml), and kept vertically at –80°C for at least 24 h. The tubes were then thawed at room temperature before use.[1] This allowed a highly reproducible continuous gradient of Nycodenz to be self-formed without ultracentrifugation, with a density range of 1.032–1.221. For each replicate, a stool aliquot (1.25-2.00 g according to the density measured by weighing 1 ml stool aspirated into a 1-ml syringe with the tip cut off) was weighted into a 50 ml BD Falcon™ Conical Tube, and completed to 6.35 g with cold 1X PBS containing 0.03% w/v Na-deoxycholate. Cell integrity was preserved as much as possible by gentle homogenization with a glass rod in presence of a non-toxic concentration of Na-deoxycholate, a detergent naturally present in the human hindgut. This favored detachment of bacteria from the fecal matrix while minimizing lysis often seen under current dispersion procedures by mechanical treatments (blending, rotating or ultrasonication) and possible induction of gene expression by chemical detergents such as Tween.[2, 3] Then 18 ml of a cold 60% w/v Nycodenz solution in 1X PBS containing 0.03% w/v Na-deoxycholate (density 1.3112 g/ml) were added with further homogenization. Finally, this heavy stool suspension in Nycodenz (final density ~ 1.230 mg/ml) was gently aspirated into a 20-ml syringe equipped with a home-made needle (2 mm-internal diameter). The syringes containing the duplicate stool suspensions were plugged, transferred out of the anaerobic chamber, and used for layering the suspension underneath the preformed continuous Nycodenz gradient. Making the sample heavy with concentrated Nycodenz and loading it below the Nycodenz gradient instead of current sample loading on top of a Nycodenz cushion suppressed exposure to air at the ultracentrifugation stage. Then during harmless low-speed ultracentrifugation in a swinging SW-28 rotor (Beckman, 14,567 × *g*, 45 min, 4°C), the fecal matrix sedimented to the bottom of the tube while bacterial cells migrated up to their buoyant density (*d* 1.110-1.190). They stabilized in an intermediate air-free, translucent and well-visible 1-2 cm thick layer (online supplementary figure 1) that was easy to pipette out. This clearly differed from the discontinuous gradients of the prior art methods where bacterial cells are concentrated in a thin and dense layer sandwiched between a lower and a greater density, and are often contaminated by underlying solid particles.[4] After centrifugation, the tubes were transferred back to the anaerobic chamber within their swinging bucket. The upper cell-free fraction was discarded, and the intermediate bacterial cell-containing fraction (5-7 ml) was collected, divided into two 50-ml BD Falcon™ Conical Tubes each made up to 50 ml with cold Tris saline (20 mM Tris, 138 mM NaCl, 2.7 mM KCl, 0.03% w/v Na-deoxycholate, pH 7.4) for washing. Cells were spun down at 3,500 × *g* for 7 min in a swing-out rotor(Sigma 3K15 centrifuge, rotor 11133), the supernatant was discarded by aspiration and the washing was repeated once. The two bacterial pellets were finally pooled by resuspension in Tris saline, making up a final volume of 1.5 ml, which was transferred into 2-ml Sarstedt tubes. A 150-µl aliquot of bacterial suspension was kept to assess conservation of the microbial diversity while the remainder was gently centrifuged and drained of excess fluid by inverting and resting on sterile surgical gauze compress. The bacterial pellets were kept at –80°C until further analyses.

Supplementary method 2

Bacterial diversity and composition profiling by pyrosequencing

Total DNA was extracted as previously described from 150 mg stool aliquots or one tenth of the bacterial pellets.[5] Microbiota composition of all crude samples and half of the corresponding extracted microbial pellets was analyzed using 454 pyrosequencing of the 16S rDNA V3-V4 region (V3F: TACGGRAGGCAGCAG,V4R:GGACTACCAGGG TATCTAAT). A total of 16 DNA samples were pyrosequenced at Genoscreen (Genoscreen, Lille, France) using GS-FLX-Titanium technology following manufacturer's instructions (Roche). Sequences were trimmed for adaptors and PCR primers removal and binned for a minimal sequence length of 300 bases, a minimal base quality threshold set at 27 and 15% of tolerated N. Resulting sequences (28,466) were assigned to different

taxonomic levels (from phylum to genus) using the RDP database (release 10, update 26).[6] Assignment was not further extended to operational taxonomic units since most taxonomic assignments of predicted proteins were not bellow the genus level.
Estimates of phylotypes richness were calculated according to the bias-corrected Chao1 estimator, while diversity was assessed by the Simpson index.[7, 8]

Supplementary method 3

Two-dimensional fluorescence difference gel electrophoresis (2D-DIGE)

Freshly prepared lysis buffer (1.5 ml) containing 8.75 M urea (Pharmacia), 2.5 M thiourea (Sigma), 5% w/v CHAPS (Sigma), 75 mM DTT (Sigma), and 31.25 mM dihydrate spermine base (Fluka), was directly added to each frozen bacterial pellet. The pellets were dispersed by vigorous vortexing and incubated at room temperature for 1 h with periodic vortexing. The lysates were then transferred into 6-ml polyallomer ultracentrifuge tubes (Beckman) and centrifuged in a Beckman SW55 Ti rotor at 24,5419 × *g* for 1 h at 18°C. The supernatants were neutralized with concentrated HCl, protein concentrations determined using the GE Healthcare 2-D Quant Kit, and 150 µg protein aliquots stored at −80°C. The aliquots were stripped of non-protein contaminants using the GE Healthcare 2-D Clean-Up Kit, and dissolved in 50 µl of a Tris-buffered solution (7 M urea, 2 M thiourea, 4% w/v CHAPS, 15 mM Tris pH 8.5). The pH was checked on pH strips (Sigma) and if necessary, adjusted to 8.5 with a 100 mM Tris solution, pH 8.5. The protein concentration was determined using the 2-D Quant Kit and aliquots of 50 µg proteins were used for DIGE experiments. Importantly, our extraction and solubilization protocols allowed the breakage of even the most robust Gram-positive micro-organisms (i.e. *Rumincoccus* spp.) and the efficient recovery of their cytosolic proteins, as inferred from preliminary yield experiments where Gram-positive and Gram-negative organisms where independently cultured and then pooled in different proportions before lysis and solubilization.

Each protein sample (50 µg) was labeled with 200 pmoles of Cy3 or Cy5 CyDye DIGE Fluor minimal dye (GE Healthcare) following the manufacturer's instructions with a dye swap design. Typically, the two protein samples from a given patient-control pair were differentially labeled with Cy3 or Cy5, respectively, and their replicates were reversely labeled with Cy5 or Cy3 (table 2). The labeling reaction (30 min on ice in the dark) was quenched by addition of 1 µl of 10 mM lysine. Samples were vortexed and incubated on ice for 10 min in the dark. They were combined in pairs, and each pair was separated on the same first and second dimension gel (12 gels in total) together with 50 µg of a Cy2-labeled internal standard made of a pool of equal protein amounts from each of the 24 samples included in the study. The internal standard helped to overcome gel-to-gel variability inherent to 2D-gel experiments.

Isoelectric focusing (IEF) was performed on 24 cm pH 4–7 linear IPG DryStrips (GE Healthcare) using a Protean IEF Cell (Bio-Rad). We chose a pI range of 4-7 based on preliminary assays showing that most proteins that we could extract and solubilize from intestinal microbiota migrated within this range in 2D experiments, and that using higher-coverage pI ranges, even if non-linear (e.g., non-linear wide range 3-10), dramatically increased spot overlap in areas of interest while revealing only few additional spots with extreme pH values. After active rehydration (50 V for 12 h), IEF was programmed as follows: linear ramp to 300 V over 1.5 h, linear ramp to 3000 V over 2 h, 3000 V constant for 2.5 h, linear ramping to 10000 V over 2 h, and finally 10000 V constant for 5 h. After IEF, strips were equilibrated and proteins were reduced and alkylated. Then, second dimension SDS-PAGE was carried out on 25×20 cm 11% polyacrylamide gels which were home-cast in an Ettan DALT six Gel Caster (GE Healthcare) between low fluorescence glass plates (GE Healthcare). Gel electrophoresis was run at 18°C in an Ettan DALT Six Electrophoresis unit (GE Healthcare) at constant power of 5 watts per gel for 30 min and then 17 watts per gel until bromophenol blue had reached the bottom of the gels (about 5 h). Only three gels were run simultaneously to allow scans on the same day, right after the second dimension. Experimental molecular weights were calculated from digitized images using molecular weight marker proteins provided by Biorad (Precision Plus Protein Standards, all blue, 10–250 kDa).

The gels were immediately scanned directly between the glass plates with a Typhoon 9410 Variable Mode Imager (GE Healthcare) using the following settings: 488-nm (blue) laser with a 520BP40 emission filter for Cy2; 532-nm (green) laser with a 580BP30 emission filter for Cy3; and 633-nm (red) laser with a 670BP30 emission filter for Cy5. All gels were pre-scanned at low resolution (1000×1000 μm^2/pixel) to set the appropriate photomultiplier tube voltage (PMT) for each channel so that spot intensities were within the linear range 40-80000 U with no saturation of spots except for the most intense ones on each image. Those optimal PMT values were then used for the final high-resolution scans (100×100 μm^2/pixel). After scanning, the gels were removed from their glass plates, and fixed in ethanol/acetic acid/water (50:5:45, v/v/v) for 1 h. Finally, they were washed in ethanol/water (50:50, v/v) for 20 min and kept at 4°C in acetic acid/glycerol/ethanol/water (1:5:10:84, v/v/v/v) for later picking.

Images were cropped on the ImageQuant software (Molecular Devices) to remove areas extraneous to the gel image prior to analysis. Analysis of the multiplexed DIGE images was performed using Progenesis SameSpots software (Nonlinear Dynamics, version 2.0) where all images are aligned before performing spot detection jointly in all images, thus avoiding difficulties due to missing values in multi-gel analyses. Spot volumes were exported to the statistical environment 'R' for further analysis.

Supplementary method 4

Normalization and statistical analysis of 2D-DIGE data

A total of 6 Crohn patients were compared to controls in a dye-swap design on 12 gels. An internal standard was measured in the third channel. We followed the recommendation of Nonlinear Dynamics to subtract the estimates of local background signal provided by the SameSpots image analysis software. In an initial conservative normalization step, for each of the 36 gel images (12 gels x 3 channels), differences in labeling yields / dye fluorescence efficiencies were compensated by a constant scaling factor, and remaining global differences in background fluorescence were balanced by a constant offset. Appropriately for a matching additive–multiplicative error model, a variance stabilizing asinh-transform was applied; for ease of interpretability of fold-changes a gauge for asymptotic equivalence to a $\log_2$-scale was chosen ('asinh2'). For sufficiently strongly expressed proteins, particularly for all the picked spots, differences in the transformed data can be interpreted as 'log-ratios'. Normalization parameters were obtained robustly for the most invariant subset of 55% of spots in all 12 x 3 images using the 'vsn' package.[9]

For differential expression analysis, we applied two complementary methods, both established and commonly used in microarray gene expression analysis. They represent different approaches to the challenge of comparing small groups for thousands of variables. The classical statistic for such comparisons, the *t*-test, is particularly sensitive to the typically noisy estimates of the within-group variances from small numbers of replicates. In a 'hierarchical' approach, this variance is uniformly estimated for all proteins in a hierarchical ANOVA model, as implemented by the 'fspma' package.[10] In a 'per gene' approach, the variance is estimated separately for each protein but then regularized towards an overall average variance using an empirical Bayesian approach, as implemented in the 'limma' package.[11] This moderates unusually extreme variance estimates. As variances observed across biological replicates are expected to differ amongst proteins, both approaches form different approximations and are expected to give different and complementary results. Finally, significance estimates were corrected for multiple testing for strong control of the false discovery rate (FDR) under arbitrary dependency structures following Benjamini-Yekutieli.[12]

Supplementary method 5

Spot exision and nanoLC-MS/MS analysis

Nine gels (Table 2) were post-stained with SYPRO® Ruby (BioRad) following the manufacturer's instructions. Spots of interest were robotically excised under computer-assisted visual control into polypropylene 96-well plates by a BioRad Exquest Spot Cutter equipped with a 1.5 mm diameter head. In-gel digestion was performed with the Progest system (Genomic Solution) according to a standard trypsinolysis.[13] LC-MS/MS analyses were performed on Ultimate 3000 LC system (Dionex, Voisins-le-Bretonneux, France) connected to a LTQ-Orbitrap Discovery mass spectrometer (Thermo Fisher, USA) via a nanoelectrospray ion source. Briefly, tryptic peptide mixtures (4 µl) were loaded at 20 µl/min flow rate onto a desalting precolumn Pepmap C18 (0.3 x 5 mm, 100 Å, 5 µm, Dionex). After 4 min, the precolumn was connected to the separating nanocolumn Pepmap C18 (0.075 x 15cm, 100Å, 3 µm) and the peptides were eluted with a 2-36 % linear gradient of buffer B (80 % acetonitrile, 0.1 % formic acid) in buffer A (2 % acetonitrile, 0.1 % formic acid) at 300 nl/min over 30 min. Ionization was performed on liquid junction with a spray voltage of 1.3 kV applied to non-coated capillary probe (PicoTip EMITER 10-µm tip inner diameter, New Objective, USA). Peptide ions were automatically analyzed by the data-dependent method in Xcalibur software (version 2.0.7) as follows: full scan MS (m/z 300-1600) in the Orbitrap analyzer, and MS/MS on the four most abundant precursors in the LTQ linear ion trap. In this study, only the doubly charged precursor ions were subjected to MS/MS fragmentation with a 1.5 min exclusion window, and with classical peptide fragmentation parameters (Qz = 0.22, activation time = 50 ms, collision energy = 35%).

Supplementary method 6

Database searching

We used the X!Tandem software to match experimental spectral data from each spot against theoretical fragmentation predicted from the MetaHit database (3,299,822 complete or incomplete ORFs annoted for functions and taxonomy),[14] associated with the human Swiss-Prot database and an in-house contaminant database. The X!Tandem search parameters were the following: one trypsin missed cleavage allowed, alkylation of cysteine and conditional oxidation of methionine. Precursor and fragment ion mass tolerances were set at 10 ppm and 0.5 Da, respectively. A refined search was added with similar parameters, except that the semi-tryptic peptides and the possibly N-terminal acetylated proteins were included. All peptides matched with an E-value lower than 0.05 were parsed with an in-house program written in Java (http://pappso.inra.fr/bioinfo/xtandempipeline/). Proteins identified with at least two peptides and a log(E-value) lower than -2.6, were validated. Proteins identified with a same set of peptides were assembled into a same sub-group. Finally, sub-groups of proteins identified with at least one common peptide were assembled into a same group, and the specific peptides for each sub-group within a same group were highlighted in the final protein list (supplementary excel file). Functional and taxonomic assignments of the predicted proteins were described previously.[14] If a protein was conserved in many species, it was assigned to the lowest common ancestor. When functional and/or phylogenetic description were incomplete, manual blasts relying on KEGG, UniProtKB and InterProScan databases enriched with new sequences or drafts of individual genomes were carried out to gain knowledge on the protein sequences. Importantly, using the grouping algorithm included in the X!TandemPipeline resulted in subgroups and groups of proteins with a common functional assignment.[15] Phylogenetic lineage of proteins within a same subgroup or even a same group,

was also similar or close, except within groups of highly conserved proteins.

Supplementary method 7

Validation of CD-associated candidate proteins using SRM-based targeted proteomics

Chemical. Modified porcine trypsin was obtained from Promega (Madison, WI), crude isotopically-labeled standard peptides (PEPotec SRM Peptides) for LC-SRM assays were synthetized by Thermo Fisher Scientific (Ulm, Germany). Methanol and phosphoric acid were purchased from Fischer scientific (Loughborough, United Kingdom), and all other reagents and chemicals were purchased from Sigma Aldrich (St. Louis, MO). All buffers were prepared with Milli-Q water.

Experimental workflow. Protein samples were purified using a monodimensional stacking SDS-PAGE protocol, followed by protein reduction, alkylation and in-gel trypsin digestion. The tryptic peptides obtained were extracted and analyzed by nanoLC-SRM. All twelve individual samples were analyzed (6 HC controls and 6 DC patients, Table 1), each of them with two SRM methods monitoring different subsets of transitions and injected in triplicate for each method.

Monodimensional Stacking SDS-PAGE Purification. Bacterial fractions were prepared as described in Supplementary method 1 and pellets were solubilized and quantified using the GE Healthcare 2-D Quant kit as described in Supplementary method 3. The aliquots were stripped of non-protein contaminants using the GE Healthcare PlusOne SDS-PAGE Clean-Up Kit. Each sample was diluted with denaturating buffer in order to obtain a protein concentration of 4µg/µl. A quality control pool made of equal fractions from each sample, was created and prepared as the individual samples.

Stacking SDS-PAGE purification was carried out using a 5% polyacrylamide stacking gel and a 10% running gel in a Mini PROTEAN cell (Bio-Rad). Running gels are only used to hold stacking gels in place. The samples were migrated over 2 cm in the stacking gel. Twenty-five µl of each sample, i.e. 100 µg of proteins, in freshly prepared denaturating buffer (GE Healthcare PlusOne SDS-PAGE Clean-Up Kit) were loaded per lane and electrophoresed at 50 V for 45min. After migration, the gels were washed with water and fixed using 50:50 methanol:water (v:v)/3% Phosphoric acid. Gels were stained by a colloidal blue method (G250, Fluka, Buchs, Switzerland). Molecular mass markers (10−250 kDa, Precision Plus Protein™ Unstained Standards, Biorad) were used as migration control. Migration front bands were excised manually using a ruler and a bistoury.

Target Peptide Selection. Among proteins discovered in the 2D-DIGE non-targeted approach, 13 proteins were selected for validation using LC-SRM. Using the METAHIT Database[14] as a reference database, peptides that were found to be unique and specific for a protein or a group of proteins (same function from phylogenetically close bacterial strains) were chosen. Priority was given to peptides that had already been identified in previous shotgun experiments acquired on equivalent samples, preferentially without fractionation, and showing high-quality MS/MS spectra. Chosen peptides were 7 to 25 amino acids long, contained no miscleavage and no methionine in their sequences.

Transition Selection. In order to select the best transitions for each peptide, four randomly chosen protein samples were prepared as described above, pooled together, and injected in the same microLC-SRM system used for sample analysis. At least 6 transitions (including y- and b-type ions) were monitored for each peptide in an unscheduled method. This allowed determining the retention times of all targeted peptides, verifying endogenous and isotopically-labelled peptides co-elution, eliminating interfered transitions and adjusting the isotopically-labelled peptides concentrations. A concentration-balanced mixture of the crude peptides was prepared in order to obtain comparable signal intensities between light and heavy transitions (the peptides were split into 4 groups defined by signal intensities and diluted 2400, 1200, 600 or 300 times from the purchased stock solutions). For each peptide, at least three transitions were eventually monitored in order to identify the peptide and the quantification was done only on non-interfered transitions showing coefficients of variation lower than 20%. The complete list of measured transitions is presented in supplementary Table 2.

MicroLC-SRM Analyses. After in-gel reduction and alkylation using a MassPrep Station (Waters, Milford, MA), the protein bands excised from the stacking gel were in-gel digested using a 1:100 trypsin:protein ratio (Promega , Madison, WI) overnight at 37 °C. The resulting tryptic peptides were extracted using 60% acetonitrile in 0.1% formic acid for 1h at room temperature. Equal amounts of the concentration-balanced mixture of stable isotope-labeled crude peptides were spiked in each peptide extract. The total volume was reduced in a vacuum centrifuge and adjusted to 15µl using 0.1% formic acid in water before microLC-SRM analysis. Peptides were analyzed on a Dionex Ultimate 3000 system coupled to a TSQ Vantage Triple Quadrupole instrument (Thermo Fischer Scientific, San Jose, CA, USA). For each analysis, a volume of 1.5µL of sample, i.e. 10µg of protein, was injected and trapped on a precolumn (Zorbax C18 stable bond, 5 µm, 1.0 × 17 mm, Agilent Technologies) then separated on a C18 column (Zorbax 300 SB C18, 3.5 µm, 150 × 0.3 mm, Agilent Technologies). The peptides were eluted with a linear gradient of 2% acetonitrile/98% water/0.1% formic acid (solvent A) and 98% acetonitrile/2% water/0.1% formic acid (solvent B). Trapping was performed for 3 min at a flow rate of 50 $\mu L \cdot min^{-1}$ with solvent A. Elution was performed at a flow rate of 5 $\mu L.min^{-1}$ using a two-step optimized gradient : Step One (Elution gradient): 3min at 5% B; from 5% to 35% B in 40 min; 5min at 80% B; 2min at 5% B; Step Two (Column washing and regeneration gradient): from 5% B to 50% B in 5min; 2min at 80% B; 15min at 5% B. For optimal microLC-SRM, the TSQ vantage mass spectrometer was operated with the following parameters. Triplicate injections of each sample were performed with two distinct methods, each monitoring a subset of all transitions (supplementary Table 4). The system was operated in positive mode, the ion spray voltage was set at 3000V, the capillary temperature at 300°C, the

nitrogen collision gas pressure was set to 1.5 mTorr, Q1 and Q3 resolution set to 0.7 Da and the collision energy was calculated using the following equation: CE=0.03 x (Precursor ion m/z) + 2.905. Scheduled SRM was used for data acquisition, each transition was monitored during a 7 minutes time window centered at previously determined peptide retention times, with a cycle time of 2.6 s and minimal dwell times of 22ms and 25ms for the SRM method 1 and 2, respectively. The system was controlled by Chromeleon Xpress software (v. 6.8) for the liquid chromatography system and Xcalibur (v. 2.1.0) software for the mass spectrometry system.

MicroLC-SRM Data Analysis and Quality Controls. The Skyline open-source software package[16] was used to merge the results obtained with the two SRM methods for each sample, visualize the SRM data, perform peak picking and integration of transition peak areas. The overall reproducibility of the experiment was verified by calculating light/heavy area ratios for each transition, and verifying that coefficients of variation were lower than 20% for triplicate injections.

Quality control of the experiment was performed by examining the stability of the SRM-MS signal over time. The quality control pool was analyzed ten times over the whole course of the experiment. For each transition, coefficients of variation were calculated for light/heavy area ratios obtained during the ten repeated injections, and we set the acceptance level for coefficients of variation below 20%. All transitions met this criterion, except for those present in low amounts in the quality control pool, thus proving that the LC-SRM system was stable and performing well over the course of the experiment. Overall, all transitions used for quantification in the sample cohort showed coefficients of variation lower than 20% with a mean value of 9.4% and a median value of 8.5%. Protein relative quantification and testing for differential protein expression were performed using the R package MSstats.[17,18] The acceptance criteria for statistically different protein abundance changes between controls and CD patients were set at a p-value lower than 0.05 and a fold change higher than 2.

Supplementary method 8

Visualization and general statistical analyses

Unsupervised, hierarchical clustering of all observations was performed in the statistical environment 'R' to visualize pyrosequencing and proteomic data, using Ward's linkage and the 1-Spearman correlation coefficient as the distance metric. Correlation matrices between operational taxonomic unit (OTU) and protein datasets were calculated and displayed in heat maps using the mixOmics package. Abundance of the different phylotypes grouped into phyla and lower taxonomic groups, was compared in the bacterial communities of CD *versus* HC using a general linear model with logit (log-odds) link function ('glm' with 'quasibinomial' family). Observations were considered significant for $P \leq 0.05$. Tendencies were reported up to $P \leq 0.10$. Richness and diversity were compared between CD and HC using the Wilcoxon signed ranks test for paired data.

1. **Murayama K,** Fujimura T, Morita M, *et al*. One-step subcellular fractionation of rat liver tissue using a Nycodenz density gradient prepared by freezing-thawing and two-dimensional sodium dodecyl sulfate electrophoresis profiles of the main fraction of organelles. *Electrophoresis* 2001;**22**:2872–80.
2. **Lindahl V.** Improved soil dispersion procedures for total bacterial counts, extraction of indigenous bacteria and cell survival. *J Microbiol Methods* 1996;**25**:279–86.
3. **Frischer M,** Danforth JM, Newton Healy MA., *et al*. Whole-Cell versus Total RNA Extraction for Analysis of Microbial Community Structure with 16S rRNA-Targeted Oligonucleotide Probes in Salt Marsh Sediments. *Appl Environ Microbiol* 2000;**66**:3037–43.
4. **Muirhead RW,** Collins RP, Bremer PJ. Erosion and subsequent transport state of Escherichia coli from cowpats. *Appl Environ Microbiol* 2005;**71**:2875–9.
5. **Godon JJ,** Zumstein E, Dabert P, *et al*. Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol* 1997;**63**:2802–13.
6. **Cole JR**, Wang Q, Cardenas E, *et al*. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 2009;37:D141-5].
7. **Chao A**, Bunge J. Estimating the number of species in a stochastic abundance model. *Biometrics* 2002;**58**:531–39.
8. **Simpson EH**. Measurement of diversity. *Nature* 1949;**163**:688
9. **Huber W,** von Heydebreck A, Sültmann H, *et al*. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 2002;**18 Suppl 1**:S96-S104.
10. **Sykacek P,** Furlong RA., Micklem G. A friendly statistics package for microarray analysis. *Bioinformatics* 2005;**21**:4069–70.
11. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;**3**:Article3.
12. **Benjamini Y,** Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 2001;**29**:1165–88.
13. **Shevchenko A**, Tomas H, Havlis J, *et al*. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 2007,**1**: 2856–60.
14. **Qin J,** Ruiqiang L, Raes J, *et al*. A human gut microbial gene catalog established by deep metagenomic sequencing. *Nature* 2010;**464**:59–65.
15. **X!TandemPipeline** http://pappso.inra.fr/bioinfo/xtandempipeline/

16. **MacLean B**, Tomazela DM, Shulman N, *et al*. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010;**26**:966-8.
17. **R Development Core Team**. R: a language and environment for statistical computing RFfSCV, Austria: 2008. ISBN3-900051-07-0. URL http://www.R-project.org
18. **Chang CY**, Picotti P, Huttenhain R, *et al*. Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol Cell Proteomics* 2012;**11**:M111. 014662.