Supporting information for:

Identification of Bacterial Species by Untargeted NMR Spectroscopy of the Exo-metabolome

T. L. Palama[a†], I. Canard[b], G. J. P. Rautureau[a], C. Mirande[b], S. Chatellier[b], and B. Elena-Herrmann[a]

[a.] *Université de Lyon, Institut des Sciences Analytiques (CNRS/ENS Lyon/UCB Lyon1), Centre de RMN à Très Hauts Champs, 69100 Villeurbanne, France. Email : benedicte.elena@ens-lyon.fr*
[b.] *bioMérieux, Innovation Unit – Microbiology Research, 38390 La Balme-les-Grottes, France.*
† Present address: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés – INSA. 135 Avenue de Rangueil. 31077 Toulouse Cedex 4

Table of Contents:

Detailed experimental procedures:

Bacterial samples:

We studied a total of 576 samples corresponding to 6 bacterial species x 8 strains x 3 different cultures (biological replicates) x 4 collecting times (T0, Te, Te+1h and Te+2h).

Te corresponds to the middle of the growth exponential phase and was determinate previously for each species in the bacterial growth conditions used in this study. The collection time point was standardized in order to analyze the same metabolic state for each species. The bacterial inoculum was prepared and calibrated to have the same amount of bacteria at T0 and Te for each species.

Bacteria culture and sample preparation:

- Clinical isolates of bacterial species, stored at -80°C, were thawed, and pre-cultured in Columbia agar + Sheep blood 5% (2 consecutive sub-cultures)
- Culture in Mueller-Hinton (MH) liquid medium until exponential phase
- Suitable inoculum is taken to have a concentration at T0 comprised between 2 and 2.5 MacFarland units (MFc)
- Incubation of the inoculum at 37°C in MH liquid medium (5mL)
- From the same inoculum: 1mL of the culture medium is collected at time T0 (beginning of incubation), Te (mid exponential phase), Te+1h and Te+2h, in eppendorf tubes.
- Each sample is then centrifugated for 10 min at 4000g
- Supernatants are collected in eppendorf tubes and stored at -80°C until NMR analysis

NMR analysis:

Sample preparation was as follows: 60µL of a mixture containing 1.25M KH_2PO_4 phosphate buffer (pH=7,4) in D2O with 2mM NaN_3 and 0.1% trimethylsilyl propionate (TMSP) was added to 540 µL supernatant samples. Both solutions were mixed thoroughly and 550 µL were then transferred to 5mm NMR tubes and sorted in 96-tubes racks.

All NMR experiments were carried out on a Bruker 600 MHz NMR spectrometer equipped with a 5mm TCI cryoprobe and a SampleJet autosampler, enabling high throughput data acquisition for large collections of samples. The temperature was controlled at 27°C throughout the experiments, and the samples were kept refrigerated at 4°C during a waiting time of less than 24h in the autosampler, before NMR analysis. Standard ^{1}H 1D NMR pulse pulse sequence nuclear Overhauser effect spectroscopy (NOESY) with z-gradient and water presaturation (Bruker pulse program *noesygppr1d*) were applied on each sample to obtain corresponding metabolic profiles. A total of 128 transient free induction decays (FID) were collected for each experiment with a spectral width of 20 ppm. The relaxation delay was set to 4 s. The NOESY mixing time was set to 10 ms and the 90° pulse length was automatically calibrated for each sample at around 14.47 µs. The total acquisition time of each sample was 13 min 48 sec.

Data processing:

All FIDs were multiplied by an exponential function corresponding to a 0.3 Hz line-broadening factor, prior Fourier transformation. ^{1}H-NMR spectra were manually phased and referenced to the TSP signal (δ = - 0.016 ppm at pH 7.4) using Topspin 3.1 (Bruker GmbH, Rheinstetten, Germany). Extraction of a data matrix for multivariate statistical analysis from the ^{1}H NMR profiles was done using the Statistics toolbox of AMIX (Bruker Biospin). Spectra were integrated from 0.3 to 10 ppm at a step of 0.01 ppm but excluding the regions of residual water and polyethylene glycol plastic contaminant signal, at 4.68-4.88ppm and 3.7-3.73ppm, respectively. Normalization on total intensity was performed. The resulting data matrix contains 947 NMR variables.

Multivariate data analysis:

Principal component analysis (PCA) and hierarchical clustering analysis (HCA) were performed using SIMCA-P 13 (Umetrics, Umea, Sweden) with scaling based on the Pareto method.

Metabolites identification:

Metabolite identification was achieved by comparing spectra with databases such as HMDB, Chenomx NMR Suite (Chenomx Inc., Edmonton, Canada) and B-BIOREFCODE (Bruker BioSpin), or by spiking a reference compound (PEG) in the media. Identification of the metabolites was further verified with homonuclear and heteronuclear 2D NMR experiments such as ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, ^{1}H-^{1}H TOCSY and J-resolved experiments.

Table S1: List of the 48 bacterial strains cultured and analyzed by NMR spectroscopy:

Bacterial species	Te (hours)	Strain number (API)
Enterococcus faecalis	1.5	1207053 8311066 9103048 9103050 9103047 0206137 1006023 0911008
Escherichia coli (lactose fermentating isolates)	1.5	1110022 1110027 1110028 1110032 1009235 1012283 0003001 7705035
Proteus mirabilis	1.5	0405048 9809009 1304018 0606273 9203148 9203147 9203139 9203138
Pseudomonas aeruginosa	6	9003025 9207016 9504024 9208009 7509005 1010250 1012279 1012280
Staphylococcus aureus	5	1208029 8803116 8311065 1207030 0804211 7504036 8803069 1206004
Staphylococcus saprophyticus	5	9906026 9906015 9809082 8411183 0308057 9803083 0305127 1206030

Te: Growth time corresponding to the middle of the exponential phase in the studied growth conditions.

Figure S2: Representative ^{1}H NMR profiles (600 MHz) for the different media after bacterial culture and for reference Mueller Hinton medium

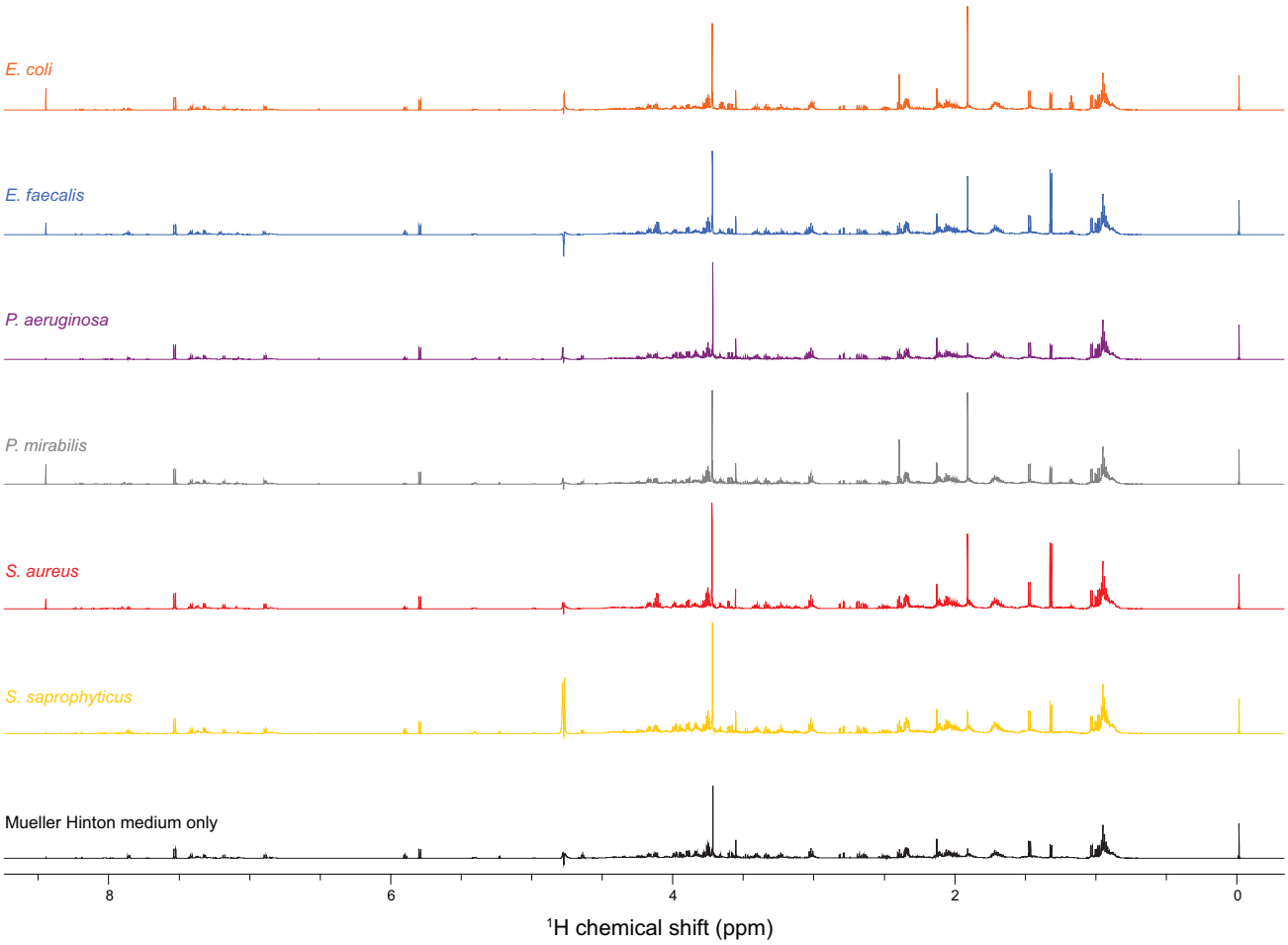

Table S3: List of the 43 metabolites identified in the cultured media by NMR spectroscopy:

Compounds	^{1}H Chemical shift (ppm) and mutiplicity
1-methylhistidine	7.02 (s), 7.84 (s)
2-hydroxybutyrate	0.88 (t)
2-methylglutarate	1.06 (d)
2-oxoisocaproate	2.60 (d)
Acetate	1.91 (s)
Adenine	8.20 (s), 8.24 (s)
Alanine	1.47 (d), 3.78 (q)
Arginine	3.24 (t), 3.77 (t)
Aspartate	2.67 (dd), 2.80 (dd), 3.90 (dd)
Citrate	2.65 (d), 2.52 (d)
Ethanol	1.17 (t)
Formate	8.44 (s)
Fumarate	6.51 (s)
Glucose	3.23 (dd), 3.39 (d), 3.45 (m), 3.47 (t), 3.52 (dd), 3.88 (m), 4.65 (d),
Glutamate	2.04 (m), 2.34 (dd), 3.75 (dd)
Glutamine	2.44 (dd)
Glycine	3.55 (s)
Glycogen	5.40 (d)
Histidine	7.08 (d), 7.88 (d)
Histamine	7.07 (d), 7.83 (d)
Hypoxanthine	8.20 (s), 8.18 (s)
Isoleucine	0.93 (t), 1.00 (d), 1.25 (m), 1.46 (m), 1.97 (m), 3.66 (d)
Ketoleucine	2.6 (d), 2.08 (m), 0.92 (d)
Lactate	1.32 (d), 4.11 (q)
Lactose	4.67 (d), 5.24 (d)
Leucine	0.95 (d), 0.96 (d), 1.68 (m), 3.72 (m)
Lysine	1.44 (m), 1.50 (m), 1.72 (m), 1.89 (m), 3.02 (t), 3.73 (d)
Methionine	2.11 (m), 2.13 (s), 2.19 (m), 2.63 (t), 3.85 (dd)
Phenylalanine	3.12 (dd), 3.27 (dd), 3.98 (dd), 7.32 (dd), 7.37 (t), 7.42 (d)
Proline	1.99 (m), 2.06 (m), 2.34 (d), 3.33 (m), 3.41 (m), 4.12 (dd)
Propionate	1.04 (t)
Propylene glycol	1.13 (d)
Pyroglutamate	2.02 (m), 2.39 (m), 2.40 (m), 2.50 (m), 2.50 (m), 4.17 (dd)
Serine	3.98 (dd), 3.93 (dd), 3.83 (dd)
Succinate	2.39 (s)
Sucrose	5.40 (d)
Threhalose	5.18 (d)
Threonine	3.58 (d), 4.25 (m)
Tyrosine	3.04 (dd), 3.19 (dd), 3.93 (dd), 6.89 (d), 7.18 (d)
Tyramine	2.93 (t), 3.25 (t), 6.91 (d), 7.22 (d)
Uracil	5.79 (d), 7.53 (d)
Uridine	4.12 (dd), 4.22 (t), 4.35 (dd), 5.89 (d), 5.91 (d), 7.86 (d)
Valine	0.98 (d), 1.03 (d), 2.26 (m), 3.60 (d)

Figure S4: Score plot of the PCA model (PC1 vs PC2 vs PC3; N=144):

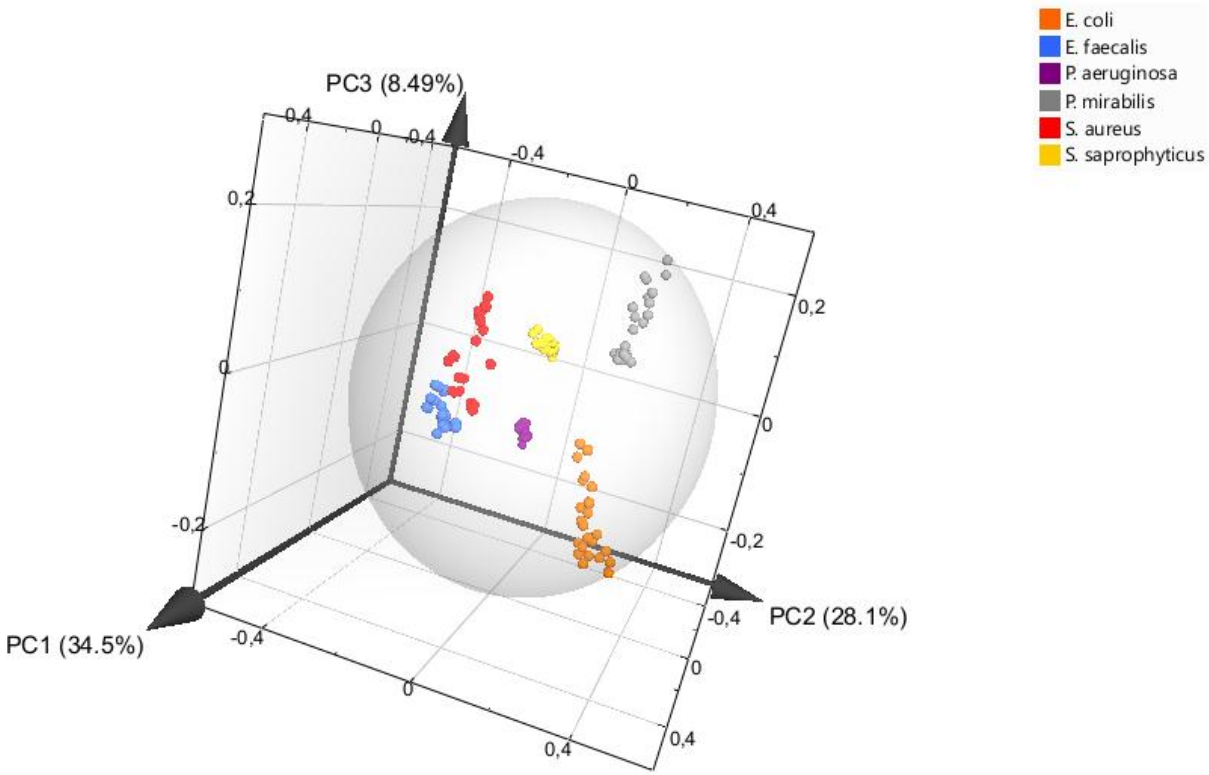

Figure S5: Score plots of the PCA model (PC1 vs PC2) of samples at T0 and Te (A), Te+1h (B) or Te+2h (C). (For each time, N=144):