SUPPORTING INFORMATION

Genomic, proteomic and metabolite characterization of gemfibrozil-degrading organism *Bacillus* sp. GeD10

Henrik Kjeldal[1], Nicolette A. Zhou[1,2], Dirk K. Wissenbach[3], Martin von Bergen[1,3,4], Heidi L. Gough[2], and Jeppe L. Nielsen[1]*

[1]Aalborg University, Department of Chemistry and Bioscience; Fredrik Bajers Vej 7H, DK-9220 Aalborg, Denmark

[2]University of Washington, Department of Civil and Environmental Engineering; More Hall 201 Box 352700, Seattle, Washington 98195-2700, USA.

Helmholtz – Centre for Environmental Research – UFZ, [3]Department of Molecular Systems Biology; Permoserstr. 15, D-04318 Leipzig, Germany.

[4]Institute of Biochemistry, Faculty of Biosciencs, Pharmacy and Psychology, University of Leipzig, Germany.

*corresponding author, (+45) 99 40 85 06; fax (+45) 98 14 18 08; jln@bio.aau.dk

Submitted to Environmental Science & Technology

Table of contents

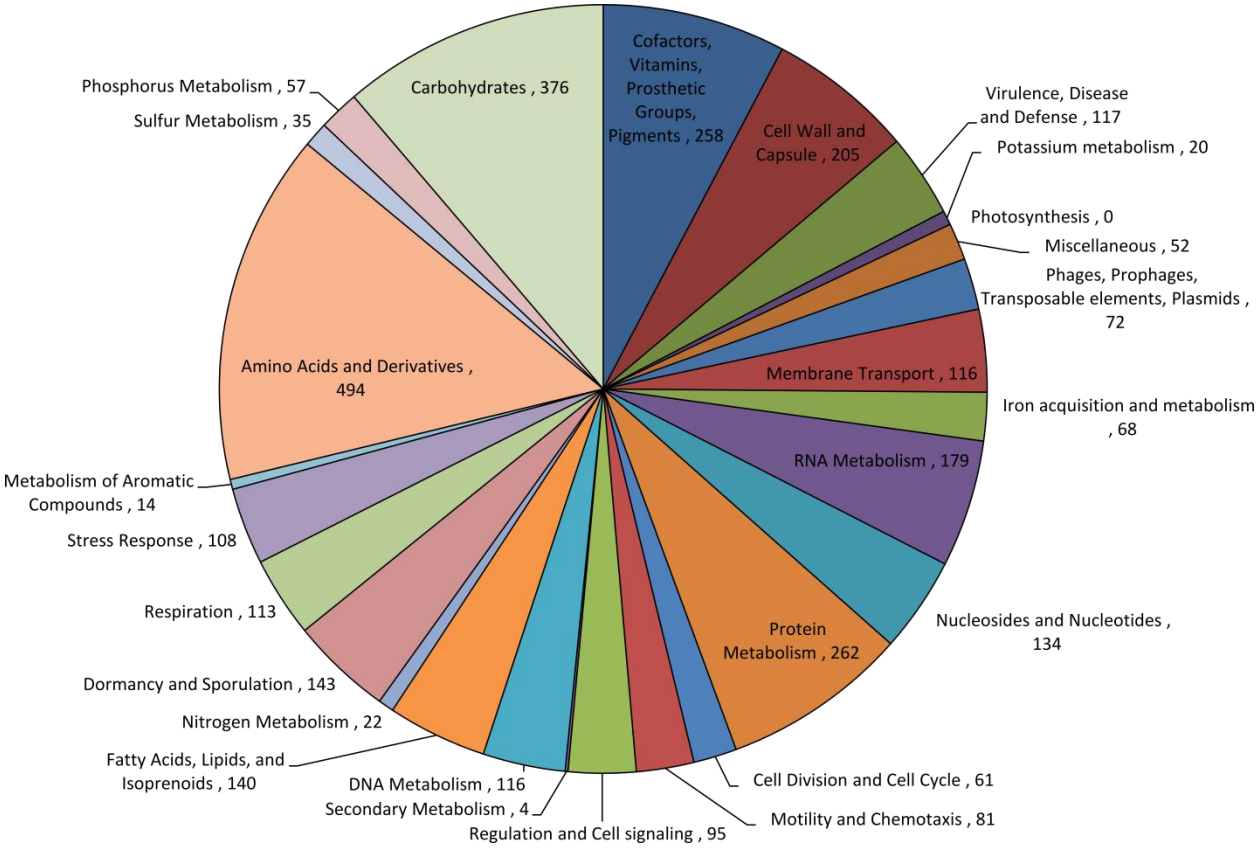


Figure S1. Distribution of functionally annotated genes obtained from RAST. The functional category is indicated followed by the subsystem feature count (number of genes identified in the corresponding category).

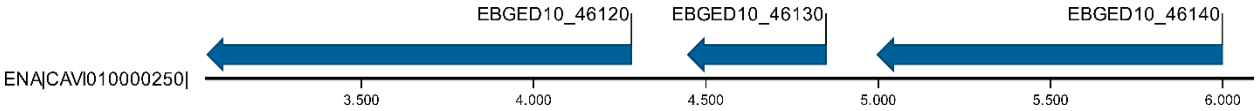


Figure S2. Gene cluster on contig CAVI010000250 containing the putative P450 hydroxylase (EBGED_46120) and two uncharacterized proteins (EBGED10_46130 and EBGED10_46140).

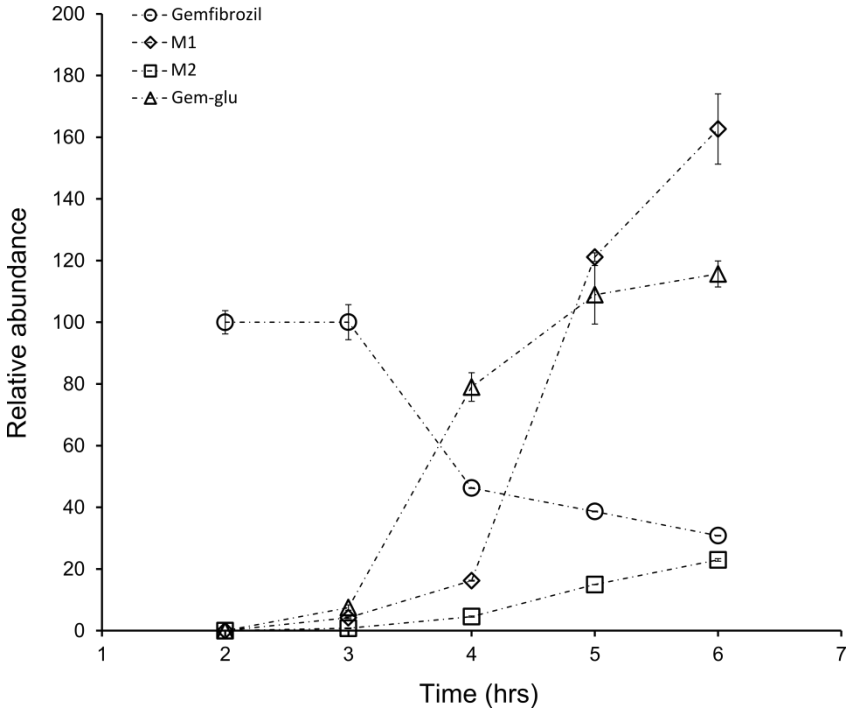
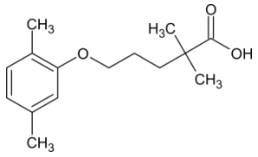
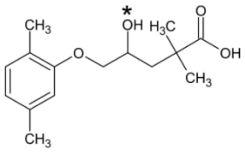
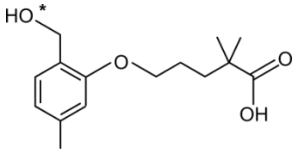
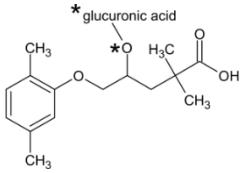


Figure S3. Relative abundance of metabolites of *Bacillus* sp. GeD10 detected during gemfibrozil degradation as a function of time. Relative abundance is the abundance of metabolites relative to the abundance of gemfibrozil at the start of the experiment. Degradation of gemfibrozil (circles) was observed following 3 hrs, at which time an increase in abundance was also observed for metabolites M1 (diamonds), M2 (squares) and Gem-glu (triangles). Error bars indicate standard deviation (n = 2).

Table S1. Metabolites identified during gemfibrozil degradation by *Bacillus* sp. GeD10. An asterisk indicates that the identified metabolite is putative (the exact position of the modification on the side chain or the aromatic ring could not be determined). The Gem-glu metabolite was detected as a sodium adduct, hence the observed m/z corresponds to $[M+Na]^-$. m/z of fragments as well theoretical m/z of precursors can be found in the supplementary data (Supporting Information).

Gemfibrozil metabolites	**Identifier**	**Structure**	**Retention time (min)**	**m/z**
Gemfibrozil	-		21.7	121.0645
HO-alkyl	M1		20.6	265.14453
HO-ring	M2		20.8	265.14453
HO-glucuronic acid conjugate	Gem-glu		18.4	463.1735

*For clarity only one of the possible isomers has been depicted, i.e. other hydroxylation /glucuronidation sites exist.

DNA extraction

DNA was extracted using the FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA, USA). Following extraction, a cleanup of the DNA was performed by mixing 100 µL of DNA 1:1 with phenol:chloroform:isoamyl alcohol (25:24:1, pH ~7.5), vortexing, centrifuging (10,000 x *g*, 10 min), and subsequently transferring the upper layer to a fresh tube. Next, 1 µL of 20 mg/mL glycogen, 10 µL of 7.5 M NH_4Ac, and 250 µL 99% (v/v) ethanol was added. The mixture was incubated at room temperature for 5 min, subsequently centrifuged (10,000 x *g*, 15 min), and the supernatant was then discarded. The pellet was dried for 15 min then dissolved in 100 µL water. A library for paired-end sequencing was prepared using the Nextera DNA Sample Preparation Kit (Illumina, San Diego, CA, USA). Everything was prepared following the manufacturer's instructions with one exception: the clean-up of the tagmented DNA was carried out using the MinElute Reaction Cleanup Kit (Qiagen Inc., Valencia, CA, USA).

Protein extraction

Cells were washed twice with phosphate-buffered saline (PBS) followed by centrifugation (10,000 x *g*, 4 ºC, 15 min). Cell pellets were resuspended in 10 mL PBS supplemented with protease inhibitors (Complete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets; 1 Tablet in 50 mL solution; Boehringer Mannheim, Mannheim, Germany). The proteins were extracted as follows: cells were disrupted by sonication (cycle 0.5, 10 min, 20W power output; Sonopuls HD2200, SH213G Booster Horn, TT13 sonotrode, Bandelin Electronic, Berlin, Germany) and the cell debris was removed by centrifugation (6000 x *g*, 4 ºC, 4 min). The pellet was resuspended in 10 mL PBS supplemented with protease inhibitor cocktail tablets. The proteins were concentrated by acetone precipitation as follows: four times the sample volume of -20ºC acetone was added to the sample. Samples were vortexed and incubated for 60 min at -20°C and were subsequently centrifuged (15,000 x *g*, 10 min). The supernatant was removed and the pellet was resuspended in 50 mM triethylammonium bicarbonate. The BCA Protein Assay (Thermo Fisher Scientific) was used to quantify the protein content prior to in-gel digestion according to the manufacturer's instructions.

In-gel digestion

The sample buffer containing dithiothreitol (DTT) to a final concentration of 40 mM DTT and denatured at 95 °C for 5 min. Samples were loaded onto a 4-20% (w/v) precast SDS-gel (Bio-Rad, USA) and proteins were separated at 160V for 45 min and and proteins were visualized afterwards using Coomassie Brilliant Blue G250 staining. Proteins were subsequently in-gel digested using a modified version of a previously described protocol cutting each lane into four subfractions[1]. Briefly, each gel piece was further excised before being washed with ammonium bicarbonate (50 mM), reduced with 10 mM DTT, and alkylated with 55 mM iodoacetamide. Proteins were digested with 12.5 mg L^{-1} trypsin (Promega) for 8 hrs at 37ºC. Tryptic digests were extracted from the gel matrix using a buffer comprised of acetonitrile and amminoum bicarbonate (50 mM) in the ratio of 5:1.

LC-MS/MS analyses for proteomics

The LC system UltiMate 3000 RSLCnano system (Thermo Scientific) was coupled to the Q Exactive mass spectrometer (Thermo Scientific) through a Nanospray Flex ion source (Thermo Scientific). Peptides were loaded onto a trap column (Acclaim Pepmap Nanotrap column, C18, 100 Å, 100 μm i.d. × 2 cm, nanoViper fittings, Dionex) in solvent comprised of 2% (v/v) acetonitrile and 0.1% (v/v) FA , followed by separation on an analytical column (RP 75 μm i.d. Acclaim Pepmap RSLC , C18, 100 Å, 75 μm i.d. × 50 cm, nanoViper fittings, Dionex). Samples were eluted at a flow rate of 300 nL min^{-1} . Pierce LTQ VELOS ESI Positive ion calibration solution (P/N 88323) was used for calibrating (sub 5 ppm mass accuracy at R 70 000) the mass spectrometer prior to measurements.

The mass spectrometer was operated in positive mode only fragmenting precursors with assigned charge ≥ 2. An isolation windows of 2 Th was used and survey scans were acquired at 400-1200 m/z at a resolution of 70 000 at m/z 200 and fragmentation spectra at a resolution of 17 500 at m/z 200. Only precursors with assigned charge ≥ 2 were selected for fragmentation using an isolation window of 2 Th. A maximum ion injection time of 100 ms was set for MS and 75 ms for MS/MS scans.

The automatic gain control for survey scans was set to 1e6 ions and 1e5 ions for fragmentation scans. The underfill ratio was set to 3.5%, the apex trigger to 2 to 15 sec and a dynamic exclusion of 30 sec was used.

Normalized Collision Energy was set to 30 and the features “peptide match” and “exclude isotopes” were enabled.

Sample preparation for metabolite identification

Collected samples were mixed with an equal volume of acetonitrile and centrifuged for 10 min at 10,000 ×g and 4°C. Samples were decanted, concentrated and prepared as described previously by Wissenbach et al.[2] with the following modifications: the sample was following centrifugation at 10.000 x *g* evaporated under a gentle stream of nitrogen at 60 °C in order to concentrate the sample. The residue was subsequently resuspended in 50 µL of mobile phase A/B (1:1;v:v). Mobile phase (A) was comprised of 10 mM ammonium formate in acetonitrile with 0.1% (v/v) formic acid while mobile phase (B) was comprised of acetonitrile with 0.1% (v/v) formic acid.

References

(1) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **2007**, *1* (6), 2856–2860.

(2) Wissenbach, D. K.; Meyer, M. R.; Remane, D.; Weber, A. A.; Maurer, H. H. Development of the first metabolite-based LC-MS(n) urine drug screening procedure-exemplified for antidepressants. *Anal Bioanal Chem* **2011**, *400* (1), 79–88.