

Supplementary Figures

Comparative genomics reveals a novel genetic organization of the *sad* cluster in the sulfonamide-degrader '*Candidatus* *Leucobacter sulfamidivorax*' strain GP

Ana C. Reis^{a,b}, Boris A. Kolvenbach^b, Mohamed Chami^c, Luís Gales^{d,e,f}, Conceição Egas^{g,h},

Philippe F.-X. Corvini^b, Olga C. Nunes^{a*}

^aLEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, Department of Chemical Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal

^bInstitute for Ecopreneurship, School of Life Sciences, University of Applied Sciences Northwestern Switzerland, Gruendenstrasse 40, 4132 Muttenz, Switzerland

^cBioEM lab, Biozentrum, University of Basel, CH-4058 Basel, Switzerland

^di3S – Instituto de Investigação e Inovação em Saúde, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

^eIBMC – Instituto de Biologia Molecular e Celular, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal

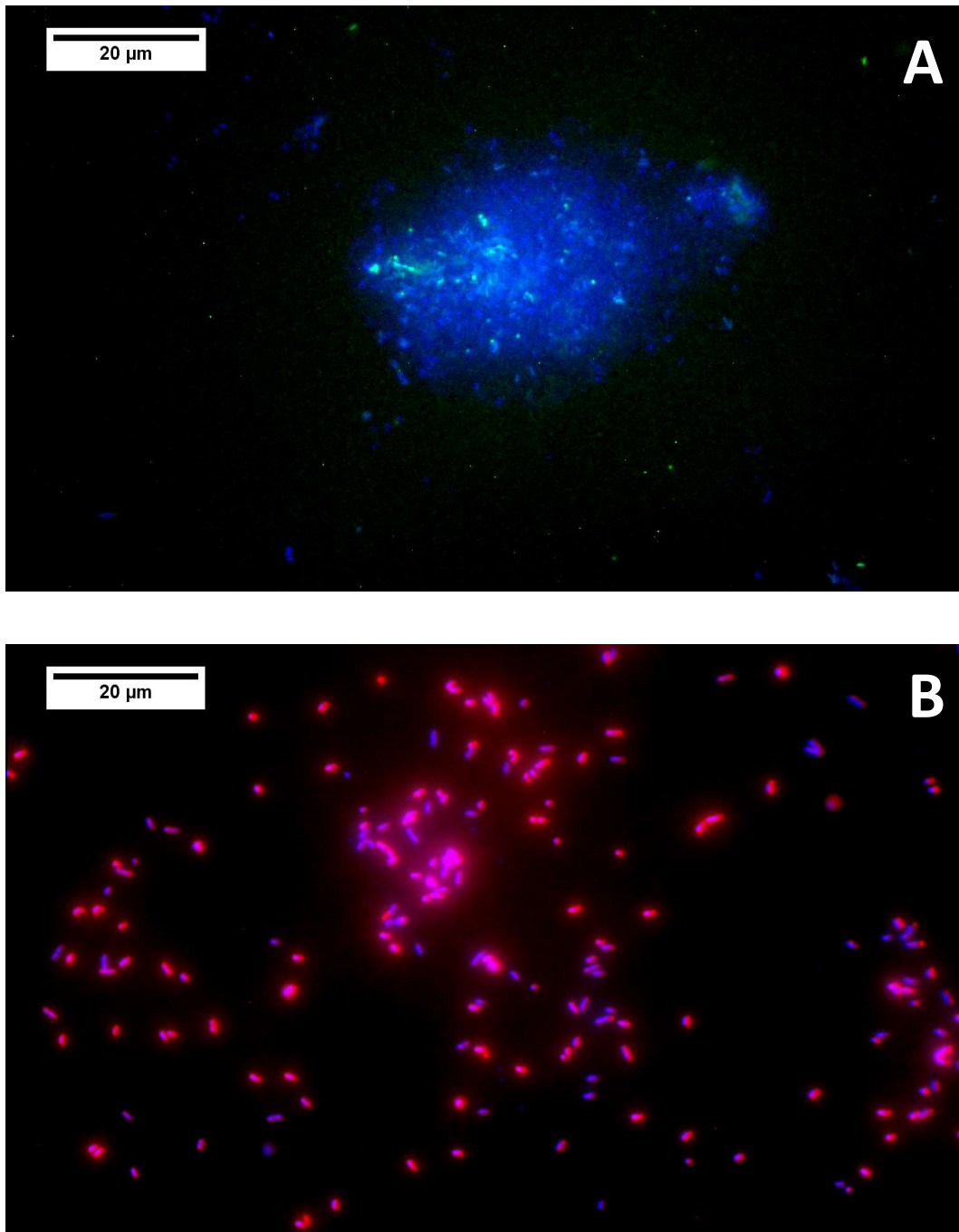
^fICBAS – Instituto de Ciências Biomédicas Abel Salazar, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal

^gNext Generation Sequencing Unit, Biocant, BiocantPark, Núcleo 04, Lote 8, 3060-197 Cantanhede, Portugal

^hCenter for Neuroscience and Cell Biology, University of Coimbra, 3004-504 Coimbra, Portugal

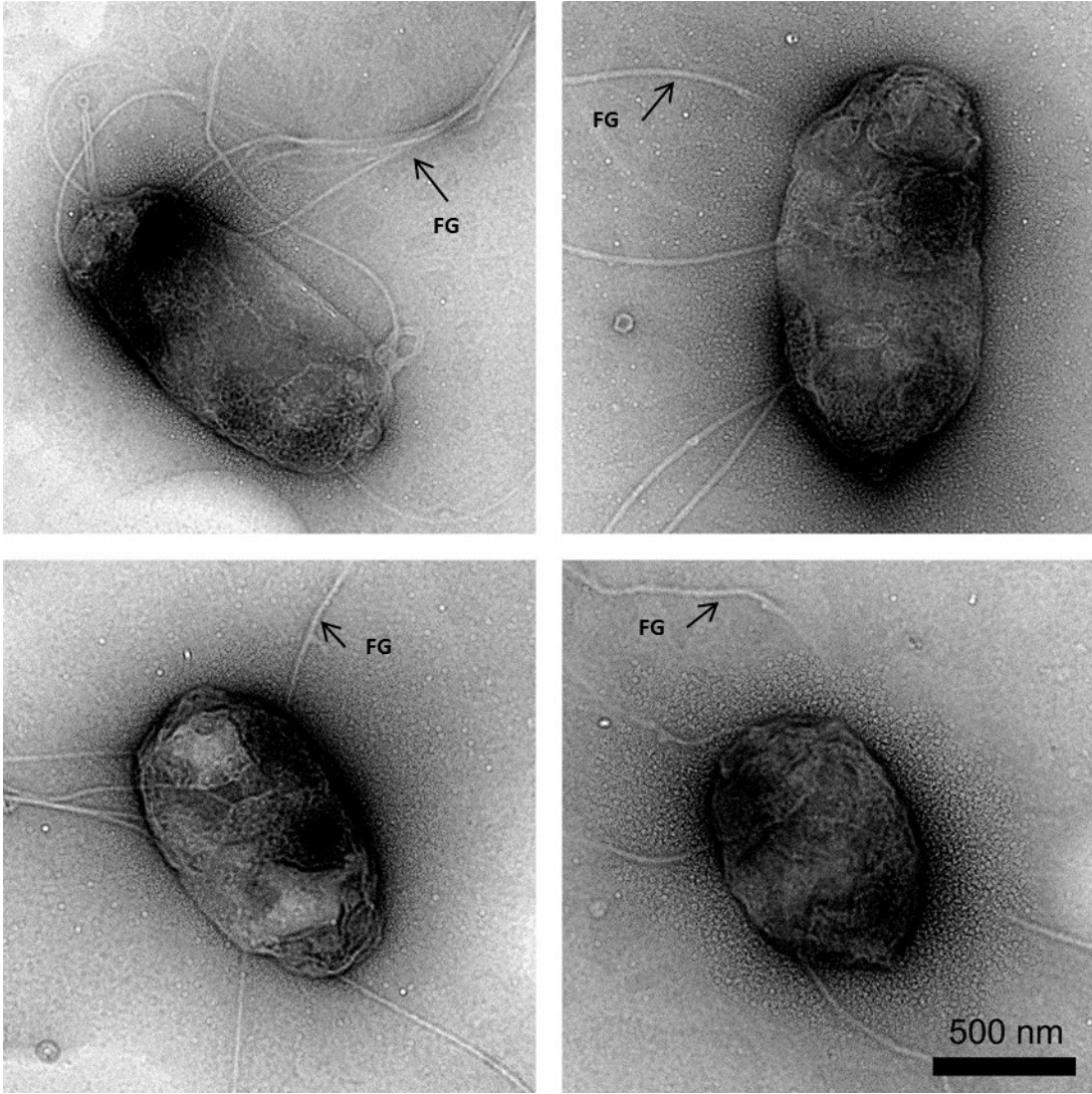
* Corresponding author: Olga C. Nunes: tel.: +351 225081917; e-mail address:

opnunes@fe.up.pt

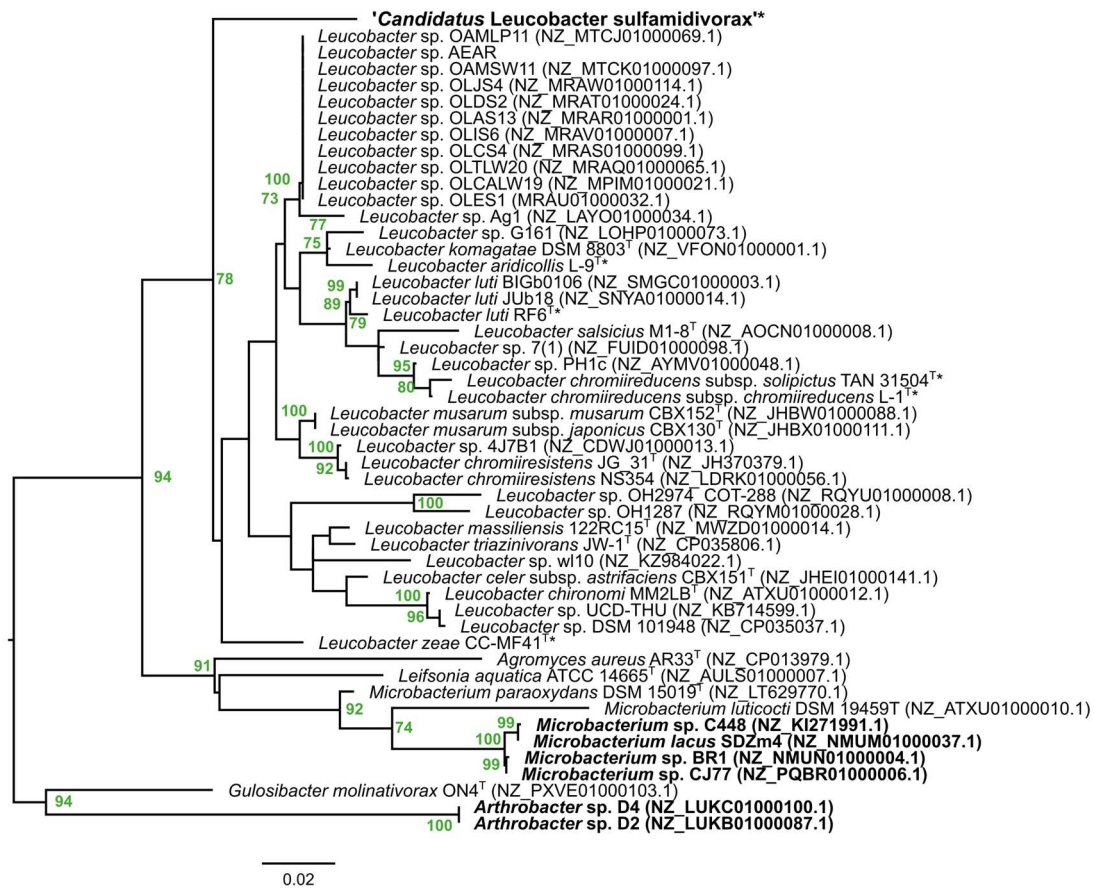


Supplementary Figure 1 – Fluorescence microscopy composite images of DAPI-stained cells of the microbial consortium (blue) and (A) cells hybridized with the modified ActORD1 FISH probe (stains strain GP, 5' fluorophore: FAM, green, sequence: 5'- CACCAGGAATTCCAATCTCC-3', original probe accession number: pB-1931, reference: [1]) or with (B) cells hybridized with Alca2 FISH probe (stains strain PR1; 5' fluorophore: Cy3, orange/red; sequence: 5'- CATCTTTCTTTCCGAACCGC-3'; probe accession number: pB-2127, reference: [2]).

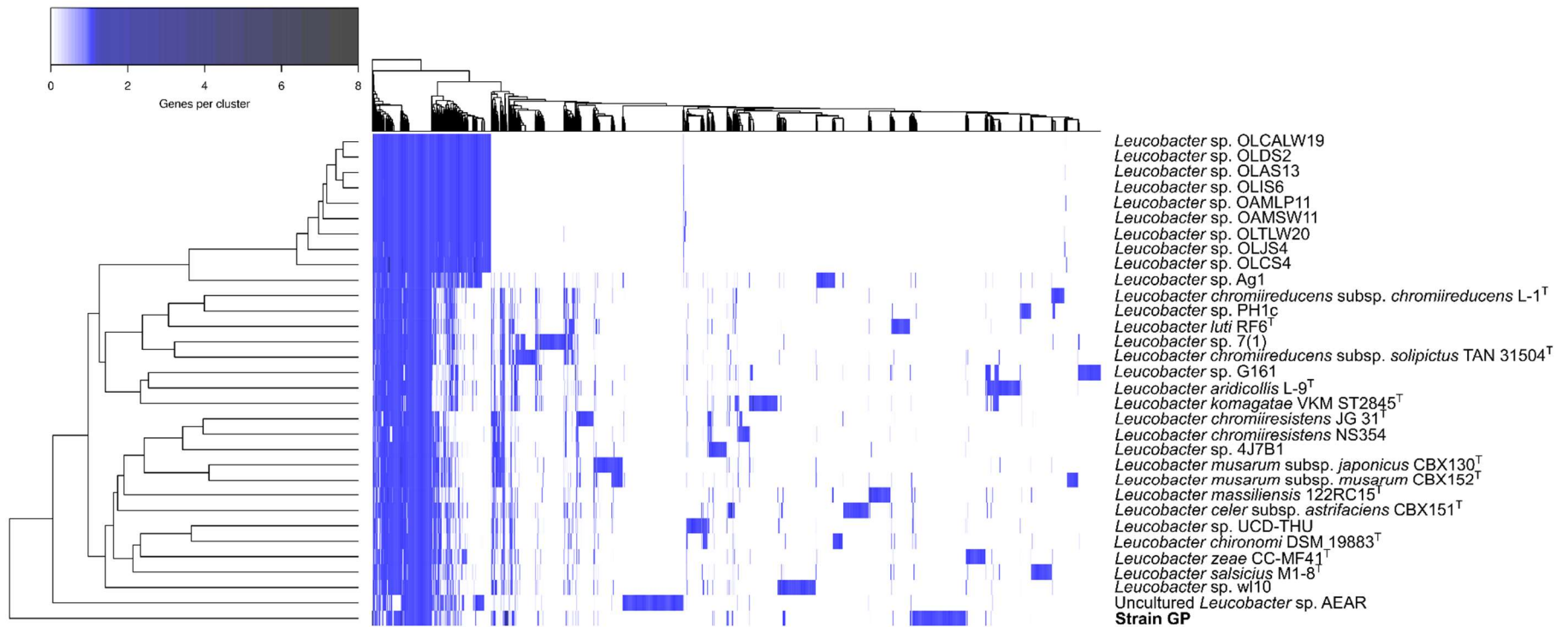
Fluorescence in situ hybridization (FISH) was performed as previously described [3] with a few modifications. To analyze strain GP (A) the cells of the microbial consortium were fixed with ethanol for 16 h at 4 °C. Prior to hybridization the cells were pre-treated with achromopeptidase (Sigma) at 60 U/ml for 60 seconds at 37 °C. While, to analyze *Achromobacter denitrificans* PR1 cells (B) the microbial consortium was fixed with 4 % (v/v) formaldehyde (Sigma) for 3 h at 4 °C and no pre-treatment was performed. Probe concentration was 3 or 10 ng/µl for strain PR1 and GP, respectively. Formamide (Sigma) concentration was 25 or 50% (v/v) for PR1 and GP, respectively.



Supplementary Figure 2 -Electron micrographs of negatively stained *Achromobacter denitrificans* PR1 showing the presence of peritrichous flagella (FG).



Supplementary Figure 3 – Cladogram of the 16S rRNA gene inferred from maximum likelihood estimation with MEGA6 with the best-fitting model: TN93+G+I [4]. *Leucobacter* spp. strains sequenced in this study are marked with an asterisk, and sulfonamide degraders are shown in bold. The tree was rooted at the outgroup and visualized with FigTree [5]. The scale bar represents the number of expected substitutions per site. Bootstrap values were inferred from 1000 replicates, values above 70% are shown at the corresponding nodes.



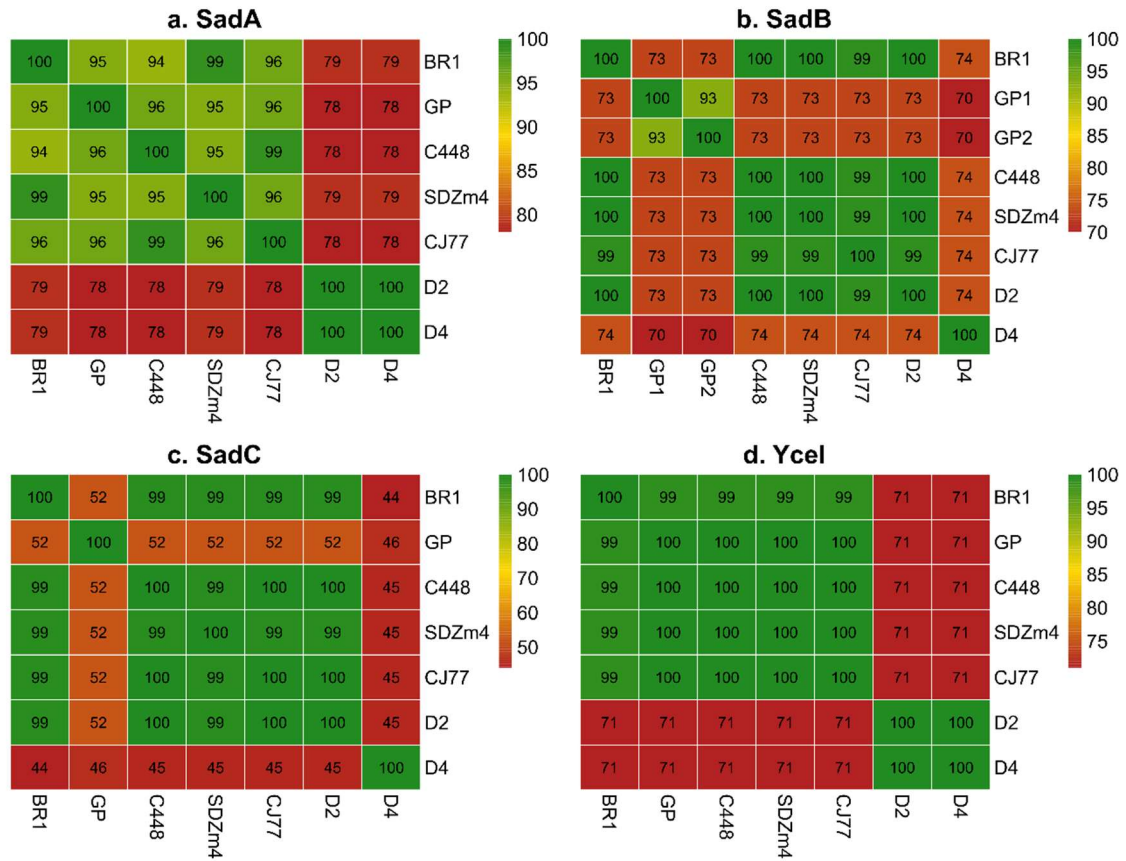
Supplementary Figure 4 – Presence/absence heatmap representation and dendrograms of the 12,998 orthologs gene clusters found in the pangenome of *Leucobacter* spp. and strain GP obtained with the GET_HOMOLOGUES package [6]. Each column represents a different gene cluster which can be absent (white) or present (blue) in each strain. As paralogs were included in the analysis, some clusters have more than one homolog per genome, and these are shown in darker blue.



Supplementary Figure 5 – Visualization of the reads of the strain GP’s MAG on the Interactive Genomics Viewer (IGV) [7] mapping to the reference genome and annotations of *Leucobacter chironomi* strain DSM 19883^T (assembly accession number GCA_000421845.1). This region from strain DSM19883^T (ATXU0100005.1:1..268438) contains the genes from the purine *de novo* biosynthetic pathway and the porphyrin and chlorophyll metabolism pathway (left to right): phosphoribosylformylglycinamide synthase subunit PurQ (accession no. WP_017883592.1, locus tag H629_RS0106495); porphobilinogen synthase HemB (accession no. WP_024356487.1, locus tag H629_RS0106505); porphobilinogen deaminase HemC (accession no. WP_084705356.1, locus tag H629_RS14980); uroporphyrinogen decarboxylase HemE (accession no. WP_024356489.1, locus tag H629_RS0106525); glutamyl-tRNA reductase HemA (accession no. WP_024356490.1, locus tag H629_RS0106530).



Supplementary Figure 6 – Visualization of the reads of the strain GP's MAG on the IGV [7] mapping to the reference genome and annotations of *Leucobacter chironomi* strain DSM 19883^T (assembly accession number GCA_000421845.1). This region from strain DSM19883^T (ATXU01000008.1:1..186096) contains the genes related to amino acid metabolism and from the glutathione and *L*-cysteine ABC transporter pathway (left to right): leucine--tRNA ligase (accession no. WP_017793981.1, locus tag H629_RS0110150); alpha/beta hydrolase (accession no. WP_010837840.1, locus tag H629_RS0110155); thiol reductant ABC exporter subunit CydC (accession no. WP_024357158.1, locus tag H629_RS0110165); thiol reductant ABC exporter subunit CydD (accession no. WP_024357159.1, locus tag H629_RS0110170).



Supplementary Figure 7 – Heatmaps representing amino acid identity (BLASTp) of the SadABC (a, b and c) complex and YceI transporter (d) among isolates from the *Microbacterium* genus (strains BR1, C448, SDZm4 and CJ77), *Arthrobacter* genus (strains D2 and D4) and strain GP.

(a) Acyl-CoA dehydrogenase, N-terminal (Pfam Acyl-CoA_dh_N)

SadA [Strain GP]	50	HEHAPDSDRDRRVSEVVIDALEELDLELDFQVCTPRRYGGFQSNFRTLFLFELTA	99
SadA [<i>Microbacterium</i> sp. BR1]	50	HEHAPDSDRDRRVSEVVIDGLEELDLELDFQVCTPRRYGGFQSNFRTLFLFELTA	99
SadA [<i>Arthrobacter</i> sp. D2]	41	REHALDSDRDRRVSVQVVIDKMEELDLELDFQVCTPRRYGGFQANFRTLFLDLTA	90
SadA [<i>Arthrobacter</i> sp. D4]	41	REHALDSDRDRRVSVQVVIDKMEELDLELDFQVCTPRRYGGFQANFRTLFLDLTA	90
SadB1 [Strain GP]	20	EANGAQGETDRRVLQDSIDALEAIGAFRVTPPKYGGFQGLSQDHVDVAR	69
SadB2 [Strain GP]	3	EANGAKGEADRRVAQESIDALEAIGAFRVTPPKYGGFQGLSQDHVDVAR	52
SadB [<i>Microbacterium</i> sp. BR1]	25	EKNAAQGEAERRVSVQESIDALEAIGAFRVTPPAKYGGYEGDSRAQVDVGA	74
SadB [<i>Arthrobacter</i> sp. D2]	25	EKNAAQGEAERRVSVQESIDALEAIGAFRVTPPAKYGGYEGDSRAQVDVGA	74
SadB [<i>Arthrobacter</i> sp. D4]	21	RANGEQGEQNRRAQESI EALEAVGAFRVTPPSRFGGFQGLSDRAQVDVSR	70
SadA [Strain GP]	100	E I A R G D G G T A W A F A L L N S N A W G V G T Y S R E A Q	130
SadA [<i>Microbacterium</i> sp. BR1]	100	E I A R G D G G T A W A F A L L N S N A W D V G T Y S R E A Q	130
SadA [<i>Arthrobacter</i> sp. D2]	91	E I A R G D G G T A W A F G L L N S N A W D V G T Y S R Q A Q	121
SadA [<i>Arthrobacter</i> sp. D4]	91	E I A R G D G G T A W A F G L L N S N A W D V G T Y S R Q A Q	121
SadB1 [Strain GP]	70	A V G R G D G G T G W I T A L I N M A G W L T A L L P D Q A Q	100
SadB2 [Strain GP]	53	A V G R G D G G T G W I T A L I N M A G W L T A L L P E Q A Q	83
SadB [<i>Microbacterium</i> sp. BR1]	75	A V G K G D G G T A W V V A L T N I A N W L T A L Y P E K A Q	105
SadB [<i>Arthrobacter</i> sp. D2]	75	A V G K G D G G T A W V V A L T N I A N W L T A L Y P E K A Q	105
SadB [<i>Arthrobacter</i> sp. D4]	71	A I G K A D G G T A W V V A L I N I S N W L T S L Y P R Q A Q	101

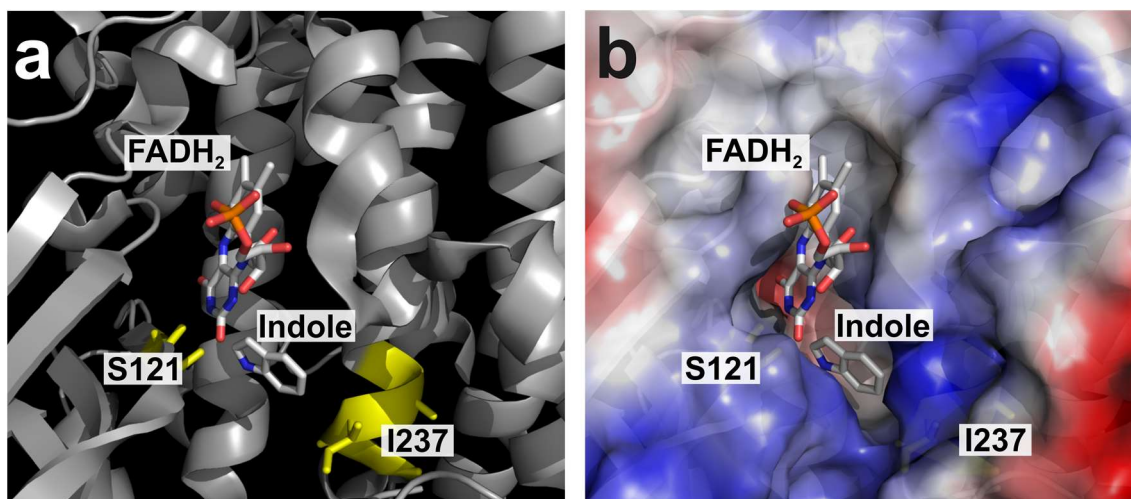
(b) Acyl-CoA dehydrogenase, middle domain (Pfam Acyl-CoA_dh_M)

SadA [Strain GP]	149	A A G S T A S A R K V D G G Y V I S G R W P Y A S G S L H A Q W V N L G F D . V E I D G A P V R M M	197
SadA [<i>Microbacterium</i> sp. BR1]	149	P A G P T A S A R K V D G G Y V I S G R W P Y A S G S L H A Q W A E L G F V . V E I D G A P V R M M	197
SadA [<i>Arthrobacter</i> sp. D2]	140	A A G P T A S A R R V D G G Y L I S G R W P Y A S G S L H A Q W A E L G F D . V E I D G E T V W M M	188
SadA [<i>Arthrobacter</i> sp. D4]	140	A A G P T A S A R R V D G G Y L I S G R W P Y A S G S L H A Q W A E L G F D . V E I D G E T V W M M	188
SadB1 [Strain GP]	116	. . A T N G K T R R V P G G Y I V S G E W G Y A S G S W H A E W S F L G A E L V D E N G D F D D A A	163
SadB2 [Strain GP]	99	. . A T N G K T K R V P G G Y I V S G E W G Y A S G S W H A E W S F L G A E L V D E D G N F D D A A	146
SadB [<i>Microbacterium</i> sp. BR1]	121	. . A T N G K T Q R V D G G Y L V S G E W S Y N S A S W H T Q W A I L G A E L V D E N G D F V D T A	168
SadB [<i>Arthrobacter</i> sp. D2]	121	. . A T N G K T Q R V D G G Y L V S G E W S Y N S A S W H T Q W A I L G A E L V D E N G D F V D T A	168
SadB [<i>Arthrobacter</i> sp. D4]	117	. . A T S G T T R R V D G G Y V V S G E W P Y A S G S L H S D W A I V G A N L V D E D G N F D D A A	164
SadA [Strain GP]	198	S . L V P M D E V T L E D T W Y V A G M R G S G S N T V V G T	227
SadA [<i>Microbacterium</i> sp. BR1]	198	T . L V P M D E V T L E D T W Y V A G M R G S G S N T V V G T	227
SadA [<i>Arthrobacter</i> sp. D2]	189	T . L L P I N E V T V E D T W Y V A G M R G S G S N T I V G R	218
SadA [<i>Arthrobacter</i> sp. D4]	189	T . L L P I N E V T V E D T W Y V A G M R G S G S N T I V G R	218
SadB1 [Strain GP]	164	Q L L I P K S D L G Y Q D T W Y V A G M R S S G S N T W T A D	194
SadB2 [Strain GP]	147	Q L L I P K T D L G Y K D T W Y V A G M R S S G S N T W V A E	177
SadB [<i>Microbacterium</i> sp. BR1]	169	Q L L I P R S D L G F K D I W H V A G M R S S G S N A L S A T	199
SadB [<i>Arthrobacter</i> sp. D2]	169	Q L L I P R S D L G F K D I W H V A G M R S S G S N A L S A T	199
SadB [<i>Arthrobacter</i> sp. D4]	165	Q L L I P R S E F A Y K D T W Y V A G M R S S G S N T L I A N	195

(c) Acyl-CoA dehydrogenase, C-terminal (Pfam Acyl-CoA_dh_Z)

SadA [Strain GP]	268	V L V G A Q I G L A Q A A L D Y A L E K L P T R G V T Y T K Y A K G S D A P T N Q I A V A E A A N A	317
SadA [<i>Microbacterium</i> sp. BR1]	268	V L V G A Q V G L A Q A A L D Y A L E K L P T R G V T N T K Y A K G S D A P T N Q I A V A E A A N A	317
SadA [<i>Arthrobacter</i> sp. D2]	259	I L V A A Q I G L A Q A A L D F A L E I L P Q R G V T K T K Y R K G S E A P S N Q I A V A E A A N A	308
SadA [<i>Arthrobacter</i> sp. D4]	259	I L V A A Q I G L A Q A A L D F A L E I L P Q R G V T K T K Y R K G S E A P S N Q I A V A E A A N A	308
SadB1 [Strain GP]	237	I L A G A Q L G I G R A V L D K A I A G A . S K P I A Y T S I A H K S D S V A F Q L D I A K A A L T	285
SadB2 [Strain GP]	220	I L A G A Q L G I G R A V L D K V I A G A . A K P I A Y T S I A H K S D S V A F Q L D V A K A A L T	267
SadB [<i>Microbacterium</i> sp. BR1]	242	I L V G P Q L G M G R A V L E R V I S K A D S K A I A Y T S F E R Q S D S I A F Q L D I A K A A L L	291
SadB [<i>Arthrobacter</i> sp. D2]	242	I L V G P Q L G M G R A V L E R V I S K A D S K A I A Y T S F E R Q S D S I A F Q L D I A K A A L L	291
SadB [<i>Arthrobacter</i> sp. D4]	238	I L T G A Q L G I G R G V L E L V A E K A N K K S I A Y T S F E R Q S D S V A F Q L D I A K A A L L	287
SadA [Strain GP]	318	I D T A R M L G R R A C Y D I D A A V T N R G Q I D W A T R A R I R M D L A T I A V L C R E S I D	367
SadA [<i>Microbacterium</i> sp. BR1]	318	I D T A R M L G R R A S Y D I D A A V T N R G Q I D W A T R A R I R M D A A T I A V L C R E S I D	367
SadA [<i>Arthrobacter</i> sp. D2]	309	I D V A R M L A K R A C Y D I D A A V F N G G A I D L P T R A R I R M D T A S I A V Q C R E A I E	358
SadA [<i>Arthrobacter</i> sp. D4]	309	I D V A R M L A K R A C Y D I D A A V F N G G A I D L P T R A R I R M D T A S I A V Q C R E A I E	358
SadB1 [Strain GP]	286	L D S A D L M I E R A C R E I D E P A . A A G E Y P D Y L T R A R N R A Y V G W A A E T V S K A I E	334
SadB2 [Strain GP]	268	L D S A D L M I E R A C R E I D E P A . A A G E Y P D Y L T R A R N R A Y V G W A A E T V S R A I E	316
SadB [<i>Microbacterium</i> sp. BR1]	292	L E A A E G F A H R A T D E I D I P A . A Q G V Y P D Y L T R A R N R A Y V G W I V E H T A R A I E	340
SadB [<i>Arthrobacter</i> sp. D2]	292	L E A A E G F A H R A T D E I D I P A . A Q G V Y P D Y L T R A R N R A Y V G W I V E H T A R A I E	340
SadB [<i>Arthrobacter</i> sp. D4]	288	L D A A D M F V E R A C K E I D L P A . E A G E Y P G Y L V R A R N R A Y V G W S V E H I S R A I E	336
SadA [Strain GP]	368	K M L T A I G S A A F A S V N P L Q Q V W R D S E T A S R H A M V N	401
SadA [<i>Microbacterium</i> sp. BR1]	368	K M L T A I G S A A F A S V N P L Q Q I W R D S E T A S R H A L V N	401
SadA [<i>Arthrobacter</i> sp. D2]	359	R L L T A V G S A A F A S T S P L Q Q I W R D A G T A S R H A M V N	392
SadA [<i>Arthrobacter</i> sp. D4]	359	R L L T A V G S A A F A S T S P L Q Q I W R D A G T A S R H A M V N	392
SadB1 [Strain GP]	335	M L L T A H G S G G F A E V N A I Q R F W R D Q A V A A R H A F I L	368
SadB2 [Strain GP]	317	T L L T A H G S G G F A E V N A I Q R F W R D Q A V A A R H A F I L	351
SadB [<i>Microbacterium</i> sp. BR1]	341	M L L T A H G S G A F A E V N P L Q R L W R D Q A V A S R H A F V L	374
SadB [<i>Arthrobacter</i> sp. D2]	341	M L L T A H G S G A F A E V N P L Q R L W R D Q A V A S R H A F V L	374
SadB [<i>Arthrobacter</i> sp. D4]	337	M L L T S A G S G A F A E V N V L Q R M W R D Q A V A R H A F V L	370

Supplementary Figure 8 –Amino acid alignment with MUSCLE [8] of Acyl-CoA domains: N-terminal (a), middle (b) and C-terminal (c); between SadA and SadB homologs in *Microbacterium* sp. BR1, *Arthrobacter* sp. D2 and D4 and strain GP (SadB1: D3X82_00235; SadB2: D3X82_03160). Conserved regions within SadA and SadB and highlighted in green and conserved regions shared between all proteins are marked with an asterisk.



Supplementary Figure 9 – Close-up of the substrate-binding pocket of XiaF (PDB: 5LVW) bound to FADH₂ and indole obtained by Kugel et al. [9]. FADH₂ is the co-factor, indole the substrate and S121 and I237 are the residues that are modified in SadA of *Microbacterium* sp. BR1 and strain GP. The ribbon (a) and electrostatic surface potential (b) diagrams have been prepared with PyMol [10]. In b negative potential is shown in red and positive potential in blue.

References

1. Kyselková M, Kopecký J, Felföldi T, Čermák L, Omelka M, Grundmann GL, et al. Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol. 2008;94:439–53.
2. Sanguin H, Remenant B, Dechesne A, Thioulouse J, Vogel TM, Nesme X, et al. Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. and bacterial communities. Appl Environ Microbiol. 2006;72:4302–12.
3. Pernthaler J, Glöckner F-O, Schönhuber W, Amann R. Fluorescence in situ hybridization (FISH) with rRNA-targeted oligonucleotide probes. Methods Microbiol. 2001;30:207–26. doi:10.1016/s0580-9517(01)30046-6.
4. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013;30:2725–9. doi:10.1093/molbev/mst197.
5. Rambaut A. FigTree. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 24 Jan 2019.
6. Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79:7696–701. doi:10.1128/aem.02411-13.
7. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.
8. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. doi:10.1186/1471-2105-5-113.
9. Kugel S, Baunach M, Baer P, Ishida-Ito M, Sundaram S, Xu Z, et al. Cryptic indole hydroxylation by a non-canonical terpenoid cyclase parallels bacterial xenobiotic detoxification. Nat Commun. 2017;8:15804. doi:10.1038/ncomms15804.
10. Schrödinger L. The PyMOL molecular graphics system. 2002;Version 1.8.