SUPPLEMENTARY MATERIAL

Supplementary material includes:

Materials and Methods

1. Organisms used in this study: The organisms used in this study, their genomic features, gene-content, and accession numbers of the versions of the genomic sequences used in the study are provided in Table S1.

2. Identification of orthologs: Protein coding sequences belonging to mobile genetic elements including plasmids, prophages, transposons and IS elements were identified by their annotations and by using the databases IS finder (1) and ACLAME (2). These proteins were omitted for ortholog comparisons. Pseudogenes were identified by visual inspection of aligned sequences as described previously for *S. oneidensis* MR-1 (3).

Orthologs were identified for the ten *Shewanella* genomes by combination of three methods: i) protein-protein pair-wise reciprocal BLAST (blastp) (4), ii) reciprocal protein-genomic sequence best match (tblastn), and iii) Darwin pair-wise best hit. For the protein-protein best match, INPARANOID v.1.35 (5) with a BLOSUM62 matrix, a minimum bit score of 30, and an alignment of 50% or more was used. Pairwise ortholog tables were then analyzed by Perl scripts to identify complete ortholog graphs (http://mathworld.wolfram.com/CompleteGraph.html), where the nodes of the graphs are the proteins and the edges are the INPARANOID ortholog connections. Complete ortholog graphs have n nodes and $\binom{n}{2} = n(n-1)/2$ edges, where n is the number of input genomes. Protein-genomic sequence best match was performed as follows: all protein-coding sequences from one genome (i.e., the reference) were searched against the genomic sequence of another genome (i.e., the database) using tblastn (protein sequence against translated nucleotide database). The best match, when better than at least 50% overall amino acid identity (recalculated to an identity along the entire sequence) and an alignable region >70% of the length of the query sequence, was extracted from the database genome using a custom PERL script and searched

against all proteins in the reference genome for reciprocal best matches. Darwin v. 2.0 (6) was used to generate pair-wise alignments with a Pam distance of 100 or less over at least 70% of the protein lengths. The closest sequence match was extracted for each protein. The results of the three methods were combined and manually adjusted taking into account genome neighborhood information.

3. Identification of gene duplications: Alignments of the *Shewanella* proteins were generated using Darwin 2.0 (6). Proteins were aligned over at least 83 amino acids or ≥70% of the sequence lengths. Strain specific duplications were identified from the data set as the proteins that had a better match within the genome they reside (excluding the self match) when compared against all other genomes used in the study.

4. Proteome analysis: Cultures were grown aerobically in Tryptic Soy Broth to final Optical Density, OD=0.5. Cells were lysed, proteins extracted and digested with trypsin, and the resulting peptides analyzed by mass spectrometry as previously described (7, 8). Each observed protein was identified by at least one unique, fully tryptic, peptide detected and at least two peptides overall in the tandem mass spectrometry data at a Peptide Expectation score of ≤ -8.3 and assigned an amino acid sequence using the X!Tandem algorithm (9). To calculate the degree of similarity in total proteins expressed by the organisms used in this study, the following analysis was performed. Each protein-coding gene in the genome was scored based on the detection of the corresponding protein in the proteomic profile using an arbitrary scoring scheme: proteins with no detectable unique peptides were scored as absent (i.e., "0" value), proteins with two unique peptides were scored present (i.e., "1" value). Detecting two or more unique peptides represents a high-confidence threshold for accurate protein identification in the expressed proteome (7). The organisms were subsequently clustered based on the complete scoring matrix for all genes found in

the MR-1 main chromosome (n = 4,318) or the core genes only (n = 2,128) using the Cluster 3.0 software with single-linkage and Euclidean distance (10) (genes not encoded in the corresponding genome were scored as absent in the proteomic profiles). Therefore, the branch lengths of the derived cladogram (i.e., figure 3) reflected the degree of similarity among organisms in expressed orthologous proteins under the conditions tested. Scoring the proteins according to an alternative scheme that reflected the fact that the more peptides per protein detected the more certain it is that the corresponding gene was indeed expressed, i.e., "0" value for absence of unique peptides, "0.3" for one unique peptide detected, "0.6" for two peptides and "1" for more than two peptides, provided very similar results with the 0/1 scoring scheme (data not shown).

The numbers of expressed proteins (based on the two unique peptides detected cut-off) were 1140 (S. oneidensis MR-1), 1294 (Shewanella sp. MR-4), 1229 (Shewanella sp. MR-7), 1144 (Shewanella sp. ANA-3), 1402 (Shewanella sp. W3-18-1), 1483 (S. putrefaciens CN-32), 1295 (S. loihica PV-4), 1050 (S. amazonensis SB2B), 1442 (S. frigidimarina NCIMB 400), and 1356 (S. denitrificans OS217). Expressed proteins for each strain are provided in Table S6.

5. Two-dimensional proteomic analysis: Frozen cell aliquots were mixed with 2 volumes of a solution containing 9M urea, 2% 2-mercaptoethanol, 2% ampholytes (pH 8-10) (Bio-Rad, Hercules, CA), and 2% Nonidet P40. The mixture was centrifuged at 435,000 *g* for 10 min and soluble denatured proteins were recovered and their concentration measured using a modification of the Bradford protein assay (11). Protein separation was carried out as described previously (12). Silver nitrate was used for protein detection and images were digitized using an Eikonix 1412 scanner interfaced with a VAX 4000-90 workstation. Regions of the patterns compared across the ten different *Shewanella* strains ware shown in supplementary figure S5. Pattern regions were selected on the basis of consistent protein detection in the 2DE gels from more than one

Shewanella species. Clustering analysis was performed based on presence or absence of a protein spot.

6. Growth studies. Cultures were grown at 15^{o}C (*S. frigidimarina*) or at room temperature (all other strains) in Minimal Media (13) supplemented with electron acceptors. Anaerobic cultures were grown with 20 mM lactate as a carbon source and one of the following electron acceptors; FeOx (30 mM), Mn-x (3 mM), Co(III) (5 mM), Cr(VI) (0.1 mM), As(V) (2mM), or Se(IV) (0.1 mM). Growth was measured by the ability to reduce the metals or metalloids.

SUPPLEMENTARY TABLES

Table S1. Genome features of the ten sequenced *Shewanella* organisms used in the study

Species or strain	Genome Size[a]	Number of genes	Protein coding genes (%coding)	tRNAs & rRNAs	Partially shared genes[b] (%)	Strain specific genes[b] (%)	Accession Number
S. oneidensis MR1	5,131,416	4,561; 184	4,318 (83); 149 (69)	130	1,401 (33.4)	628 (14.9)	AE014299; AE014300
S. sp. W3-18-1	4,708,380	4,217	4,044 (85)	129	1,624 (40.1)	258 (6.4)	CP000503
S. putrefaciens CN-32	4,659,620	4,134	3,972 (85)	129	1,623 (40.9)	180 (4.5)	CP000681
S. sp. MR-7	4,799,109	4,178; 8	4,006 (85); 8 (56)	144	1,691 (42.4)	162 (4.0)	CP000444; CP000445
S. sp. MR-4	4,706,287	4,084	3,924 (85)	141	1,661 (42.4)	91 (2.3)	CP000446
S. sp. ANA-3	5,251,146	4,268; 251	4111 (85); 249 (83)	135	1,762 (40.3)	442 (10.1)	CP000469; CP000470
S. denitrificans OS217	4,545,906	3,905	3,754 (84)	127	737 (19.7)	827 (22.2)	CP000302
S. frigidimarina NCIMB400	4,845,257	4,199	4,029 (84)	133	1,097 (27.5)	726 (18.2)	CP000447
S. loihica PV-4	4,602,594	3,993	3,859 (85)	124	1,196 (30.9)	494 (12.8)	CP000606
S. amazonensis SB2B	4,306,142	3,785	3,645 (88)	130	1,250 (32.4)	440 (11.4)	CP000507

Isolate	Place of isolation
S. oneidensis MR-1	Freshwater sediments, Lake Oneida, NY, USA
Shewanella sp. W3-18-1	Marine sediments, Pacific Ocean (630 m, 5-6 cm core), WA, USA
S. putrefaciens CN-32	Subsurface sandstone, New Mexico, NM, USA
Shewanella sp. MR-7	60 m depth, anoxic, Black Sea
Shewanella sp. MR-4	5 m depth, oxic, Black Sea
Shewanella sp. ANA-3	Arsenic treated wood in brackish estuary, Woods Hole, MA, USA
S. denitrificans OS217	Marine oxic/anoxic zone, Gotland Deep, Baltic Sea
S. frigidimarina NCIMB 400	Marine, North Sea, UK
S. loihica PV-4	Marine, Naha Vents, Hawaii
S. amazonensis SB2B	Marine sediments, Amazon River delta, Brasil

[a] Genome size was computed as the sum of chromosome and plasmid.

[b] Only orthologous genes outside of the core were considered in this analysis.

Table S2. Complete gene ortholog table among the ten *Shewanella* genomes. Orthologous relationships were determined for the *Shewanella* proteins by their sequence similarity, 50% amino acid identity over ≥70% of the protein length, and by gene neighborhood analysis. The proteins are listed in the table by their locus tags. Gene product descriptions and the genome fraction they belong to (core, dispensable or unique) are also included. Proteins encoded by mobile elements including plasmids, prophages, transposons, and insertion sequences are not included in the table. Proteins encoded by pseudogenes are marked by "*".
Table S2 is provided as a separate Microsoft Excel file.

Table S3. List of genes encoded in mobile elements and pseudogenes

Species or strain	CDSs encoded by mobile elements:					Pseudogene
	Plasmid	***IS***	***Tn***	***Prophag***	***Total***	
Shewanella oneidensis MR1	108	237		178	523	230
Shewanella sp. W3-18-1		81	20	137	238	70
Shewanella putrefaciens CN-32		84	25		109	67
Shewanella sp. MR-7	10	23		47	80	40
Shewanella sp. MR-4		29			29	30
Shewanella sp. ANA-3	231	44	35		310	62
Shewanella denitrificans OS217		70	35	49	154	79
Shewanella frigidimarina NCIMB400		68		51	119	63
Shewanella loihica PV-4		14	15		29	9
Shewanella amazonensis SB2B		13			13	13
Total	349	663	130	462	1604	663

* Individual CDSs were counted only in one category.

Table S4. Metabolic pathways present in the Shewanella core genome according to the BioCyc pathway schema (http://biocyc.org/)

Core Metabolic Pathways
(deoxy)ribose phosphate degradation
2'-deoxyribonucleotide/ribonucleoside metabolism (aka de novo biosynthesis of pyrimidine deoxyribonucleotides)
5-phosphoribosyl 1-pyrophosphate biosynthesis
acetate metabolism
acetoacetate degradation
alanine biosynthesis III
alanine utilization (NAD-dependent) (alanine deg. IV)
aminobutyrate utilization
ammonia assimilation cycle
arginine biosynthesis
arginine utilization
asparagine - aspartate interconversion
asparagine biosynthesis I
aspartate biosynthesis I
aspartate degradation II
ATP proton motive force interconversion
biotin biosynthesis
chorismate biosynthesis
cyclic regeneration of tetrahydrobiopterin
cyclopropane fatty acid (CFA) biosythesis
cysteine biosynthesis
cysteine degradation
de novo biosynthesis of pyrimidine ribonucleotides
degradation of purine deoxyribonucleosides
degradation of pyrimidine deoxyribonucleosides
degradation of pyrimidine ribonucleosides
DNA degradation
dTDP-rhamnose biosynthesis
eicosapentenoic acid biosynthesis
Entner-Doudoroff pathway
ethanol degradation
fatty acid oxidation pathway I
folic acid biosynthesis
formaldehyde degradation
formaldehyde degradation, plasmid copy
gluconeogenesis
glutamate - aspartate interconversion
glutamate biosynthesis
glutamate utilization
glutamine - glutamate interconversion
glutathione biosynthesis
glycine biosynthesis
glycine biosynthesis from L-threonine
glycine cleavage
glyoxylate bypass
heme biosynthesis I
histidine biosynthesis
histidine utilization
homoserine biosynthesis
interconversion of aspartate and asparagine

isoleucine biosynthesis
isoleucine degradation
isoleucine valine biosynthesis
isopentenyl diphosphate biosynthesis (methylerythritol phosphate pathway)
KDO biosynthesis -- including transfer to lipid IVA
leucine biosynthesis
leucine utilization
lipid-A-precursor biosynthesis
lipoate biosynthesis
L-serine/L-threonine utilization
lysine and diaminopimelate biosynthesis
mannose-sensitive hemagglutinin pili biogenesis
methionine biosynthesis
methionine utilization II
molybdenum (molybdopterin) biosynthesis
N-acetylglucosamine metabolism, UDP- N-acetylglucosamine biosynthesis
O-antigen biosynthesis
pantothenate and coenzyme A biosynthesis
pentose phosphate shunt, non-oxidative
pentose phosphate shunt, oxidative branch
peptidoglycan biosynthesis
phenylalanine biosynthesis
phenylalanine degradation I
ppGpp metabolism
proline biosynthesis
proline utilization
purine nucleotides de novo biosynthesis I
putrescine catabolism via gamma-glutamate-putrescine
pyridine nucleotide synthesis
pyridine nucleotide synthesis (NAD biosynthesis I)
pyridoxal 5'-phosphate biosynthesis
riboflavin and FMN and FAD biosynthesis
S-adenosyl methionine biosynthesis
salvage pathways of purine and pyrimidine nucleotides
salvage pathways of pyrimidine deoxyribonucleotides
serine biosynthesis
sulfate assimilation pathway
sulfur metabolism (sulfate assimilation to cysteine via L-serine)
sulfur metabolism (sulfate assimilation to methionine via cysteine)
superpathway of fatty acid biosynthesis
thioredoxin redox
threonine biosynthesis
threonine degradation I
threonine degradation II
threonine degradation IV
tricarboxylic acid cycle
tRNA charging pathway
tryptophan biosynthesis
type IV pili biogenesis
tyrosine biosynthesis
tyrosine degradation
ubiquinone biosynthesis
valine biosynthesis
valine utilization

Proteins encoded in the *Shewanella* core genome were assigned to pathways according to the BioCyc pathway schema (http://biocyc.org/) based on sequence similarity to known metabolic enzymes. Metabolic pathways shared by all ten strains are listed in the table and depicted in Figure S3. Abbreviations for intermediates and proteins used in Figure S3 include: A, amino acids; AcCoA, acetyl-CoenzymeA; AI2, autoinducer 2; Cco, Cbb3-type cytochrome c oxidase; CoA, Coenzyme A; Cob, cobalamin; Cyo, cytochrome c oxidase; DC, dicarboxylate; DHAP, dihydroxyacetone phosphate; EDD, Entner-Doudoroff Pathway; Etf, electron transfer flavoprotein; E4P, erythrose-4-phosphate; FA, fatty acids; FUM, fumarate; F16P, fructose-1,6,-diphosphate; F6P, fructose-6-phosphate; GA3P, glyceraldehydes-3-phosphate; GSH, glutathione; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; HS(L), homoserine(lactone); KDO, 3-deoxy-D-manno-octulosonate; LP, lipoprotein; MAL, malate; MDR, multidrug; Nap, periplasmic nitrate reductase; Ndh, NADH dehydrogenase II; Nqr, NADH-quinone reductase; OAA, oxaloacetate; PEP, phosphoenolpyruvate; Pet, ubiquinol-cytochrome c reductase; PG, phosphoglycerate; PLP, pyridoxal-5-phosphate; Pnt, NAD(P) transhydrogenase; PYR, pyruvate; Rnf, NADH:ubiquinone oxidoreductase; Ru5P, ribulose-5-phosphate; R5P, ribose-5-phosphate; SAM, S-adenosyl methionine; Sdh, succinate dehydrogenase; SH7P, sedoheptulose-7-phosphate; SUC, succinate; SUCoA, succinyl-CoenzymeA; THF, tetrahydrofolate; Thi, thiamine; UQ, ubiquinone; U/X, uracil/xanthine; X5P, xylulose-5-phosphate; αKG, alpha-ketoglutarate.

Table S5. Metal and metalloid reduction for the ten *Shewanella* organisms

Strain	FeOx (30 mM)	M-x (3 mM)	Co(III) (5 mM)	Cr(VI) (0.1 mM)	As(V) (2 mM)	Se (IV) (1 mM)
MR-4	+	+	-	-	-	+
MR-7	+	+	-	-	-	+
ANA -3	+	-	-	+	+	+
MR-1	+	+	+	+	-	+
W3-18-1	+	+	-	-	-	+
CN-32	+	+	+	+	+	+
OS217	-	-	ND	-	ND	ND
NCIMB400	-	+	-	+	+	-
PV-4	-	+	+	+	+	+
SB2B	+	+	+	+	ND	+

ND: Not Determined

Shewanella cultures were grown anaerobically in Minimal Media (13) in the presence of 20 mM lactate as an electron donor and several metals or metalloids (see table for details) as an electron acceptor. Anaerobic respiration was considered as positive when cell growth and metal or metalloid reduction occurred.

Table S6. Proteins detected in the high-precision mass spectrometry protein profiles for each strain. The table lists proteins detected in the proteome analysis (see Materials and Methods) by 2 or more peptides. Proteins are listed by their locus tags. The number of peptides detected per protein and the gene product description is included.
Table S6 is provided as a separate Microsoft Excel file.

SUPPLEMENTARY FIGURES

Figure S1.

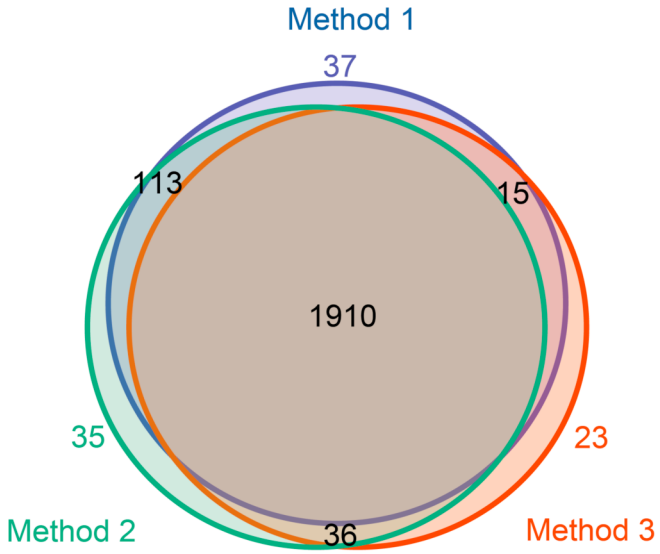

Fig. S1. Comparisons between the three different methods employed to identify the core gene set for the ten strains. Method 1: protein-protein pair-wise reciprocal Blastp (INPARANOID); Method 2: reciprocal protein-genomic sequence best match (tblastn); Method 3: Darwin pair-wise best hit. The final orthologous gene core (available in Table S2) comprised of the core genes identified by all three methods, supplemented by the manually inspected and verified orthologous genes predicted by one or two of the methods.

Figure S2.

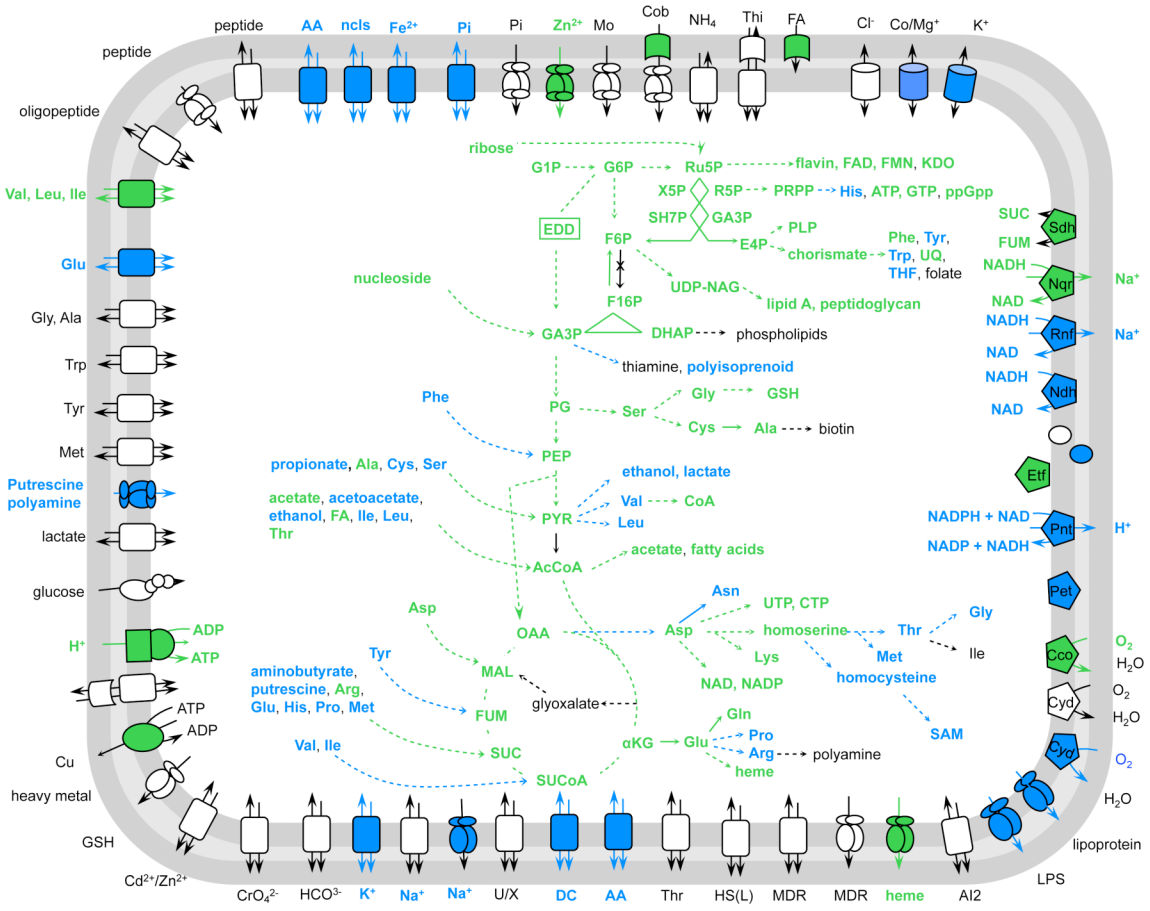

Figure S2. Metabolic overview of the *Shewanella* core. The figure includes predicted catabolic and anabolic reactions, pathway intermediates, and proteins or protein complexes encoded by the *Shewanella* core genome. Proteins were assigned to pathways according to the BioCyc pathway schema (http://biocyc.org/) based on sequence similarity to known metabolic enzymes. TransportDB was used to assign transport reactions (http://www.membranetransport.org). The figure is color coded according to protein expression during aerobic growth in Tryptic Soy Broth; expressed in all ten strains (green), expressed in one or more of the strains (blue), not expressed (black or white). A complete description of the proteins, protein complexes, and abbreviations used is available in Table S4.

Figure S3.

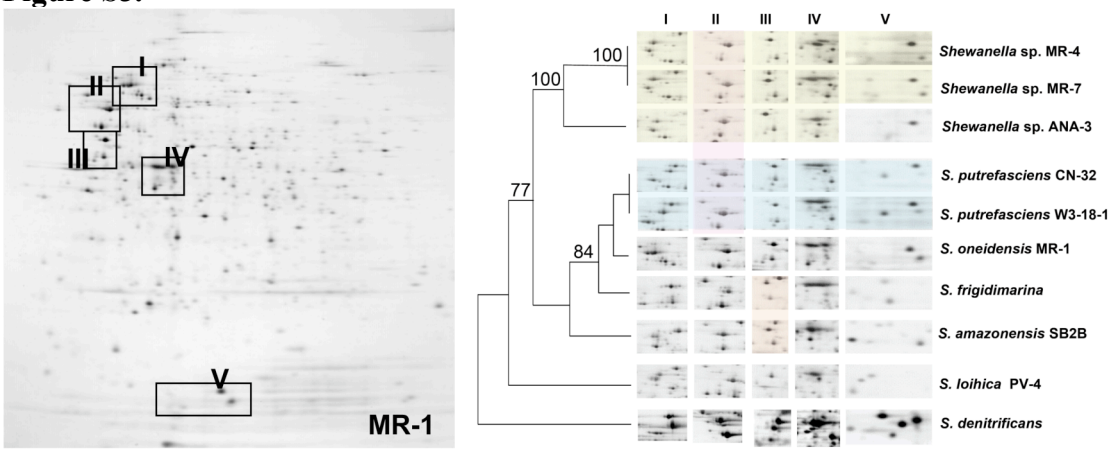

Fig. S3. Protein profile comparisons among the ten *Shewanella* organisms using 2-D protein gels. ***Left panel***: 2-D patterns of the soluble-fraction proteins isolated from *S. oneidensis* MR-1 after whole-cell lysate. Proteins were separated by two-dimensional gel electrophoresis with isoeletric focusing in the first dimension and sodium dodecyl sulfate electrophoresis in the second dimension. Proteins were detected using silver stain. Regions I through V of the patterns were selected on the basis of consistent protein detection from more than one *Shewanella* species. ***Right panel***: Groupings of similar migration patterns are indicated by color highlighting and by group number. Cladogram represents the clustering of the organisms when considering all five regions together. Note that the latter clustering is very congruent with the clustering observed based on high-precision mass spectrometry protein profiles (figure 3 of the article).

REFERENCES

1. Siguier P, Perochon J, Lestrade L, Mahillon J, & Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34(Database issue):D32-36.
2. Lima-Mendez G, Van Helden J, Toussaint A, & Leplae R (2008) Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 24(6):863-865.
3. Romine MF, Carlson TS, Norbeck AD, McCue LA, & Lipton MS (2008) Identification of mobile elements and pseudogenes in the Shewanella oneidensis MR-1 genome. *Appl Environ Microbiol* 74(10):3257-3265.
4. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
5. Remm M, Storm CE, & Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314(5):1041-1052.
6. Gonnet GH, Hallett MT, Korostensky C, & Bernardin L (2000) Darwin v. 2.0: an interpreted computer language for the biosciences. *Bioinformatics* 16(2):101-103.
7. Fang R*, et al.* (2006) Differential label-free quantitative proteomic analysis of Shewanella oneidensis cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol Cell Proteomics* 5(4):714-725.
8. Lipton MS*, et al.* (2002) Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. *Proc Natl Acad Sci U S A* 99(17):11049-11054.
9. Craig R & Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20(9):1466-1467.
10. de Hoon MJ, Imoto S, Nolan J, & Miyano S (2004) Open source clustering software. *Bioinformatics* 20(9):1453-1454.
11. Louis S. Ramagli LVR (1985) Quantitation of microgram amounts of protein in two-dimensional polyacrylamide gel electrophoresis sample buffer. *Electrophoresis* 6(11):559-563.
12. Elias DA*, et al.* (2008) The influence of cultivation methods on Shewanella oneidensis physiology and proteome expression. *Arch Microbiol* 189(4):313-324.
13. Kostka J & Nealson KH (1998) *Techniques in Microbioal Ecology* (Oxford Univ. Press, New York) 2nd Ed pp 58-78.