Final Scientific/Technical Report prepared for The U.S. Department of Energy

DOE award number:
FG02-08ER64511

Recipient:
Dr. Margrethe H. Serres, Marine Biological Laboratory, Woods Hole, MA 02332

Project title:
Integrated Genome-Based Studies of *Shewanella* Echophysiology

Director/principal investigator:
Margrethe H. Serres

Consortium:
The *Shewanella* Federation research group lead by James K. Fredrickson.

The report does not contain patentable material or protected data.
The report summarizes the work done by Dr. Margrethe H. Serres contributing to the joint effort by the *Shewanella* Federation to do an integrated genome-based study of *Shewanella* ecophysiology.

Executive summary (from *Shewanella* Federation proposal):

Shewanella oneidensis MR-1 is a motile, facultative γ-Proteobacterium with remarkable respiratory versatility; it can utilize a range of organic and inorganic compounds as terminal electron acceptors for anaerobic metabolism. The ability to effectively reduce nitrate, S0, polyvalent metals and radionuclides has established MR-1 as an important model dissimilatory metal-reducing microorganism for genome-based investigations of biogeochemical transformation of metals and radionuclides that are of concern to the U.S. Department of Energy (DOE) sites nationwide. Metal-reducing bacteria such as *Shewanella* also have a highly developed capacity for extracellular transfer of respiratory electrons to solid phase Fe and Mn oxides as well as directly to anode surfaces in microbial fuel cells. More broadly, *Shewanellae* are recognized free-living microorganisms and members of microbial communities involved in the decomposition of organic matter and the cycling of elements in aquatic and sedimentary systems.

To function and compete in environments that are subject to spatial and temporal environmental change, *Shewanella* must be able to sense and respond to such changes and therefore require relatively robust sensing and regulation systems. The overall goal of this project is to apply the tools of genomics, leveraging the availability of genome sequence for 18 additional strains of *Shewanella*, to better understand the ecophysiology and speciation of respiratory-versatile members of this important genus. To understand these systems we propose to use genome-based approaches to investigate *Shewanella* as a *system of integrated networks*; first describing key cellular subsystems – those

involved in signal transduction, regulation, and metabolism - then building towards understanding the function of whole cells and, eventually, cells within populations. As a general approach, this project will employ complimentary "top-down" – bioinformatics-based genome functional predictions, high-throughput expression analyses, and functional genomics approaches to uncover key genes as well as metabolic and regulatory networks. The "bottom-up" component employs more traditional approaches including genetics, physiology and biochemistry to test or verify predictions. This information will ultimately be linked to analyses of signal transduction and transcriptional regulatory systems and used to develop a linked model that will contribute to understanding the ecophysiology of *Shewanella* in redox stratified environments. A central component of this effort is the development of a data and knowledge integration environment that will allow investigators to query across the individual research domains, link to analysis applications, visualize data in a cell systems context, and produce new knowledge, while minimizing the effort, time and complexity to participating institutions.

Compare Goals with Accomplishments

We have successfully contributed to the overall objectives of the *Shewanella* Federation proposal throught genome annotation efforts, cross genome comparisons and the building and curation of pathway genome databases. Our work has contributed to four posters and seven publications (see below) and has laid the foundation for further analysis of interactions between co-exisiting microbes, including a member of the *Shewanella* genus.

Project acitivities

The overall role of our group was to integrate with projects undertaken by members of the *Shewanella* Federation and to provide bioinformatics support for these projects. The main objectives that were to be addressed by the *Shewanella* Federation as a group included:

1. Elucidate the Environmentally-Controlled Signal Transduction and Transcriptional Regulatory Systems Crucial for the Ecological Success of *Shewanella*.

2. Characterize the Electron Transfer Networks and Central Metabolic Pathways Involved in Energy Conversion in *Shewanella.*

3 Develop an Integrated Model of *Shewanella* Metabolism, Sensing & Regulation

During the project period with have contributed to these objectives mainly through three major avenues: A) Genome Annotation, B) Comparative Genomics, and C) Metabolic Pathway Reconstruction. We have coordinated our efforts chiefly through the Pacific Northwest National Lab, but have also been collaborating with groups at the Boston University and at University of Texas. Our results are described in detail in the sections below.

A. ANNOTATION

A. 1. The dataset

Our project included the analysis of twenty strains belonging to the genus *Shewanella*. This is a group of phylogenetically related organisms that has evolved to exist in a wide range of ecological environments ranging from a wood pier in brackish water (*Shewanella* sp. ANA-3) to an Amazon river sediment (*S. amazonensis* SB2B) and a hydrothermal vent (*S. loihica* PV-4), see Table 1 for a list of the strains included in the analysis. The environments these *Shewanellas* are found in vary in their type or amount of nutrients available, their temperature, and in their pressure (from land to deep ocean). The genome sequences for all of the selected organisms have been determined. As seen in Table 1 the majority of the strains are associated with marine environments, although these environments have different salinity, temperature and depth.

It is our goal to search these sequences and the functions they encode for traits that are common to one or a few of the organisms and that can be related to the survival in their respective environments. Such traits would be evidence of adaptation, and we are interested in identifying whether these adaptive traits are derived from early ancestors via gene duplication and functional divergence (vertical transfer) or whether they are obtained via uptake of genes from other organisms co-existing in their environment (lateral gene transfer).

Table 1. Overview of the *Shewanella* strains analyzed.

ID	Strain	Origin	Isolation site	Pressure	Temp.
CN32	*S. putrefaciens* CN-32	Albuquerque, NM	Subsurface, shale-sandstone, 250 m	peizsosensitive	mesophilic
W3181	*Shewanella* sp. W3-18-1	Washington Coast, Pacific Ocean	Marine sediment, under 997 m of oxic water	peizsosensitive	mesophilic
OS155	*S. baltica* OS155	Baltic Sea	Sea water, oxic zone, 2 ml/l O2, 90 m	peizsosensitive	mesophilic
OS185	*S. baltica* OS185	Baltic Sea	Sea water, oxic-anoxic interface, 120 m	peizsosensitive	mesophilic
OS195	*S. baltica* OS195	Baltic Sea	Sea water, anoxic zone, 140 m	peizsosensitive	mesophilic
OS223	*S. baltica* OS223	Baltic Sea	Sea water, oxic-anoxic interface, 120 m	peizsosensitive	mesophilic
ANA3	*Shewanella* sp. ANA-3	Woods Hole, MA, USA	Brackish water, As-treated wooden pier	peizsosensitive	mesophilic
MR4	*Shewanella* sp. MR-4	Black Sea	Sea water, oxic zone, 16oC, 5 m	peizsosensitive	mesophilic
MR7	*Shewanella* sp. MR-7	Black Sea	Sea water, anoxic zone, high NO3, 60 m	peizsosensitive	mesophilic
MR1	*S. oneidensis* MR-1	Lake Oneida, NY, USA	Sediment, anaerobic, Mn(IV) reduction	peizsosensitive	mesophilic
Sden	*S. denitrificans* OS217	Baltic Sea	Sea water, oxic-anoxic interface, 120 m	peizsosensitive	mesophilic
Sama	*S. amazonesis* SB2B	Amazon River, Amapa, Brazil	Sediment, suboxic redox conditions, 1 m	piezosensitive	growth over 35 C
PV4	*S. loihica* PV-4	Hawaiian Sea mount	Iron-rich mat, hydrothermal vent, 1,325 m	piezosensitive	growth over 35 C
Sfri	*S. frigidimarina* NCIMB 400	Coast of Aberdeen, UK	Sea water, North Sea	peizsosensitive	cold-adapted
Spie	S. *piezotolerans* WP3	West Pacific site	Sediment, under 1914 m of water	piezophilic	cold-adapted
Shal	S. *halifaxens* HAW-EB4	Halifax Harbor, Nova Scotia, Canada	Sediment, munitions-dumping area, 215 m	piezotolerant	mesophilic
Spea	S. *pealeana* ANG-SQ1	Woods Hole Harbor, MA	Squid nidamental gland	piezotolerant	mesophilic
Swoo	*S. woodyi* MS-32	Alboran Sea 2-300m	Squid ink, sea water	piezotolerant	mesophilic
Sben	*S. benthica* KT99	Pacific Ocean, Tonga-Kermadec Trench	Deep sea high pressure zone, 9000m	piezotolerant	mesophilic
Ssed	*S. sediminis* HAW-EB4	Halifax Harbor, Nova Scotia, Canada	Sediment, munitions-dumping area, 215 m	piezotolerant	mesophilic

In addition, the *Shewanella* strains selected for this study represent an "evolutionary gradient", where some strains are phylogenetically more closely related than others (See Fig. 1). The average nucleotide identity (ANI), a measure of sequence identity between genome sequences [10], range from 99% (CN32 to W3181) to 74.7% (Sama vs Spie). By selecting this set of organisms we were hoping to shed light on functional divergences between closely as well as more distantly related members of the genus. A related study of ten *Shewanella* species has already been published [9]. In the current study additional and more divergent strains are included.

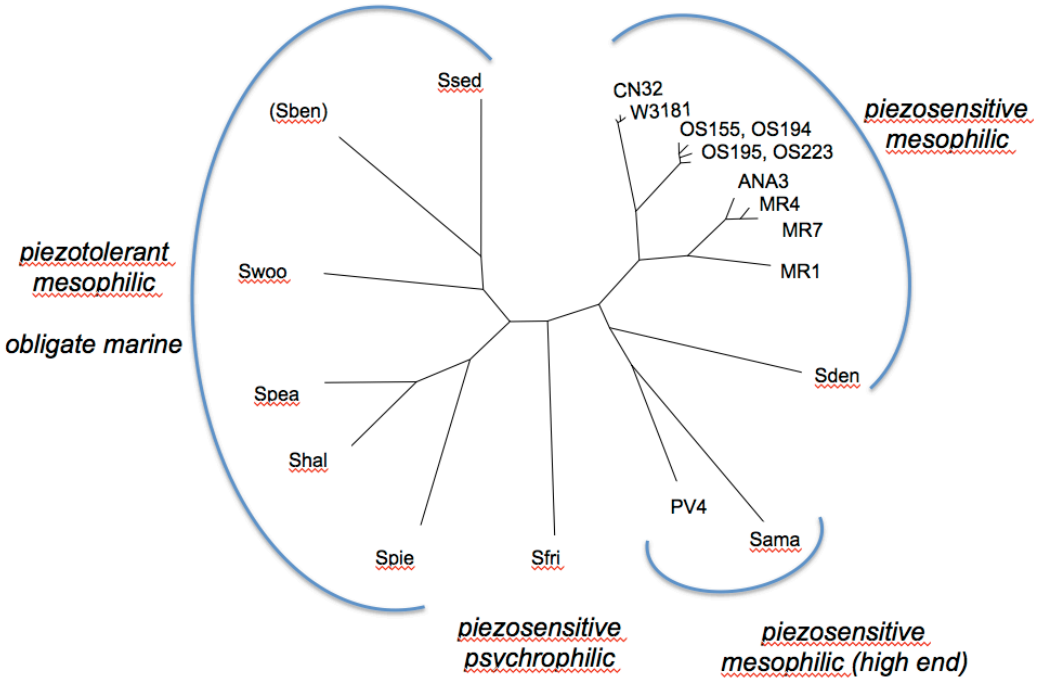


Fig. 1. Phylogenetic relationship of the *Shewanella* strains used in this study. The phylogenetic tree is based on 16S rRNA sequence similarity. See Table 1 for explanation of strain abbreviations.

A. 2. Genome annotations

A functional comparison of organisms is greatly enhanced when the gene product annotations are well curated. To efficiently improve the automated annotations of the *Shewanella* genomes, we made use of the orthologous relationships between the genomes. Orthologs were identified from pairwise best bi-directional hits followed by evaluation of gene order. We expanded the *Shewanella* ortholog table used in previously published work [9] to include all available genome sequences. We identified approximately 18380 *Shewanella* orthologs, where 1900 genes (10%) were common to all strains (the core), 10440 genes (57%) were unique to one of the strains, and 6040 genes (33%) were present in two or more of the strains. Approximately half of the strain-specific genes were predicted to be mobile, based on genome context and lack of sequence similarity to other *Shewanella* strains.

We made use of our up-to-date manual curation of the *S. oneidensis* MR-1 gene products and propagated these to the ortholog table. This included 4430 genes. In addition we inferred function assignments from published work on orthologs or homologs. Several databases i.e. TCDB (transport proteins) [17], RegPrecise (transcriptional regulators) [13], Merops (peptidases) [15], ACLAME (mobile genetic elements) [11] were used to improve the gene product descriptions. We also predicted pathways or cellular roles for the encoded gene products. Currently

10130 of the orthologs have been assigned to 300 functional roles such as metabolic pathways, transport, and sensory/regulatory functions. The most common roles assigned to the dataset were DNA mobilization – IS elements (2040; 20.1 %), transport (1054; 10.4 %), one component regulator (540; 5.3%), LPS synthesis (330; 3.3 %), and respiration (320; 3.2%). The large fraction of DNA mobilization – IS element orthologs can be explained by their strain specificity; this group of proteins are mostly either present in one or in a few strains. The high number of transport and transcriptional regulators could be reflective of the level of curation of these gene products in the dataset or of the importance of these functions for the Shewanella genus. The high number of LPS synthesis genes is interesting taking into account the unique and diverse anaerobic respiratory pathways of this genus; paths that are located in the outer and inner membranes and that may affect the membrane composition.

Similarly, around 2000 *Shewanella* orthologs were assigned to 210 SEED subsystems via the automatic RAST annotation pipeline. Although we intended to, we did not curate or develop new SEED subsystems for the *Shewanella* genes such as the N-acetyl glucosamine degradation or N-acetylgalactosamine subsystems *[12, 19]*.

In order to better describe proteins that did not have function assignments, we made use of domain descriptions or domain identities. Domains represent conserved regions in a protein that has a specific role, i.e. substrate binding, cofactor binding, protein-protein interaction site, and that may be used to decipher some functional information. To identify protein domains in the *Shewanella* genomes, we compared the encoded protein sequences against the Pfam database [14]. 75% of the *Shewanell*a proteins contained Pfam domains (Fig. 2). Some of these domains were common to all of the strains while others were either unique to one strain or present in 2-19 strains. We used the presence of Pfam domains and information about these domains to assign a broader family function where possible. Alternatively, we listed the domain identity in the gene product description where the domains are of unknown functions (i.e. DUF domains) [1].
There are currently 756 orthologs of unknown function with a DUF domain assignment.

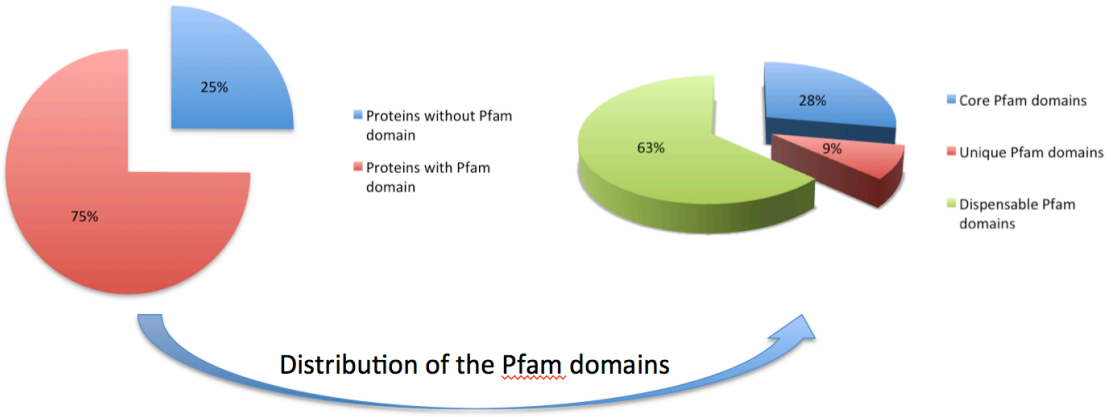


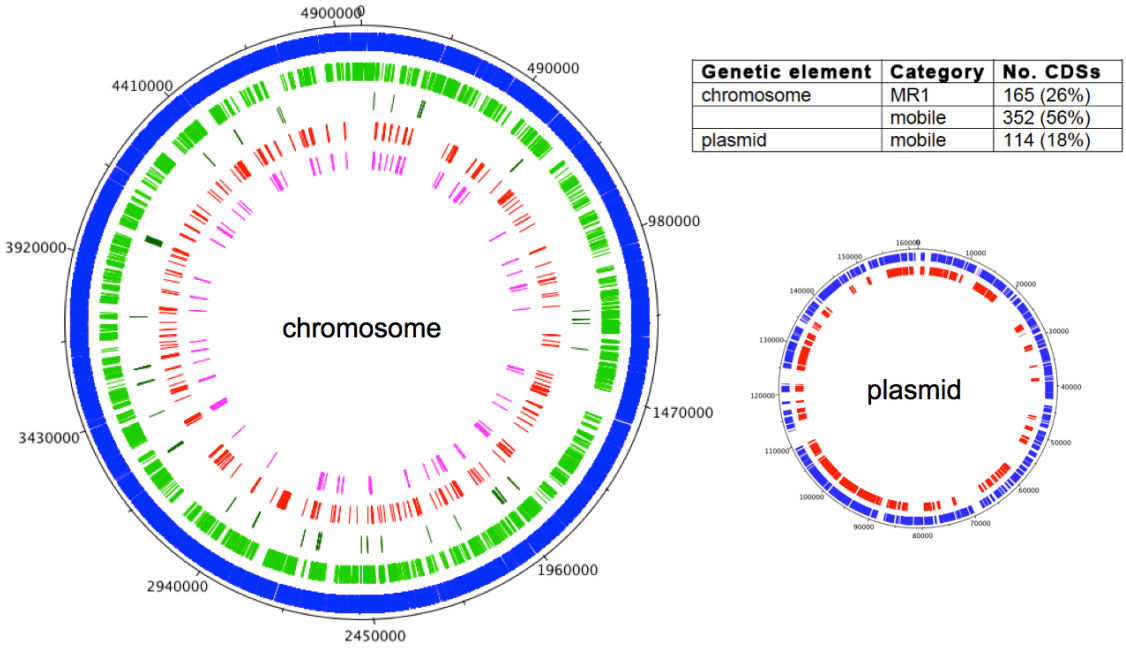
Figure 2. Pfam domains identified in *Shewanella* genome sequeces.
Core Pfam domains were present in all twenty strains, dispensable Pfam domains were present in 2-19 strains, while unique Pfam domains were found in only one of the strains.

A. 3. Evidence for adaptations

The best studied of the *Shewanella* strains is *S. oneidensis* MR-1 [6]. Although this organism is most famous for its ability to reduce metals for anaerobic respiration [4, 8, 18], it is becoming a model organism for bioenergy and remediation [5].

We made use of the *Shewanella* ortholog table to find the genes that were either unique to *S. oneidensis* MR-1 or lost by this strain compared to the rest of the *Shewanellas* with the hopes of shedding light on function important for adaptation. Figure 3 shows genes encoded by the chromosome and plasmid of *S. oneidensis* MR-1. We identified 631 unique CDSs that were encoded either on the chromosome or the plasmid. We also found 74% of these genes to be flanked by mobile elements while 26% appeared to be integrated into the genome. This suggests that lateral gene transfer has played a significant role in obtaining new genetic information by *S. oneidensis* MR-1.

Some of the functions encoded by the *S.oneidensis* MR-1 specific genes are listed in Table 2. We observed that several of these loci consist of single genes. However, we also detected loci consisting of 3-13 consecutive genes. Functions encoded by these gene clusters include capsule biosynthesis, virulence-related functions, chemotaxis and regulation.



Genetic element	Category	No. CDSs
chromosome	MR1	165 (26%)
	mobile	352 (56%)
plasmid	mobile	114 (18%)

Figure 3. *S. oneidensis* MR-1 gene overview.
The figure shows all genes (blue), core genes shared by all *Shewanella* strains (light green), tRNAs (dark green), *S. oneidensis* MR-1 specific (unique) genes (red), and *S. oneidensis* MR-1 specific (unique) genes not flanked by mobile elements (pink).

Table 2. *S. oneidensis* MR-1 specific genes

Locus Tag	No. CDSs	Gene Product Functions	Blast vs. NR
SO3174-85	13	capsule-EPS-LPS biosynthesis	mix
SO2539-47	9	two-component regulators	*Alteromonas*
SO4355-62	8	respiratory complex, Mtr-like	*Shewanella*
SO0181-88	8	virulence/interaction proteins	*Saccharophagus*
SO4639-43	4	toxin/anti-toxinsystem	mix
SO4539-44	4	carbohydrate utilization	*Shewanella*
SO3302-06	4	virulence/adhesion proteins	mix
SO2323-26	4	chemotaxis	*Reinekea*
SO0911-13	3	virulence, regulation	mix
SO0091-93	3	pyrimidine utilization	*Vibrio*
SO2317-20	3	chemotaxis	*Reinekea*
SO2143-45	3	two-component regulators	*Alteromonadales*

To search for the potential origins of the unique genes, we performed a blastp search against NCBI's non-redundant (NR) database. The genera of the most significant hits were extracted, are these shown in the last column of Table 2. Based on the result we see that some strain-specific genes appear to have arisen by gene duplication (closest hit is *Shewanella*) while others appear to have originated outside of the *Shewanella* genus.

Table 3. Functions and cellular roles lost by *S. oneidensis* MR-1.

Missing in	Function	Role
MR-1	periplasmic nitrate reductase, Nap	respiration
MR-1, OS155, Sden, Sama, Ssed, Sben, Spea, Shal	Cytochrome bo3 ubiquinol oxidase, Cyo	respiration
MR-1, Sden	Na translocating OAA decarboxylase, Oad	C-metabolism
MR-1, OS155, OS185, OS195	arabinose utilization	C-metabolism
MR-1, Sden	RND efflux transporter	transport
MR-1, Sden, Sfri, Sama	Na:divalent anion symporter	transport
MR-1, ANA3, MR-4, MR-7	Na:alanine symporter, AgcS	transport
MR-1, CN32, W3181, S200, PV4, Sama	Na:dicarboxylate symporter	transport
MR-1, Ssed, Sben, Swoo, Spea, Shal	Na:proline symporter	transport

We were also interested in the genes that were lost in *S. oneidensis* MR-1 and maintained in several of the other *Shewanella* strains. Most of these genes encoded proteins of unknown function, but examples of the functions that could be identified are shown in Table 3. These include respiration (cytochrome oxidase, nitrate reductase), metabolism (arabinose utilization), and transport (amino acid, carboxylate, ions). Several of the transporters identified are sodium symporters. The loss of these proteins may be a result of this strain residing in fresh water environments.

B. COMPARATIVE GENOMICS

B. 1. Comparison of protein families in *Shewanella* strains

Protein families are made up of sequence similar proteins that encode for same or similar functions. Some protein families are enzymes that share the same substrate, product, or cofactor. Other families consist of transport proteins that transport the same or similar substrates or of transcriptional regulators that bind related effector

molecules. The abundance and size of specific protein families give insight into functions that are of importance to the organism. I.e. enrichment of amino acid uptake protein families in an organism suggests that it uses amino acids as its carbon or nitrogen source.

We determined the protein families present in the *Shewanella* genomes based on their Pfam domain content. Although some proteins contain more than one Pfam domain, we chose to use these domains to cluster proteins into sequence similar groups. We found that the set of largest protein families in each of the *Shewanella* strains were approximately the same for all of the organisms analyzed (Fig. 4). Based on the functions that are encoded by these protein families, it is apparent that *Shewanellas* dedicate a significant portion of their genetic resource to interactions with the environment.

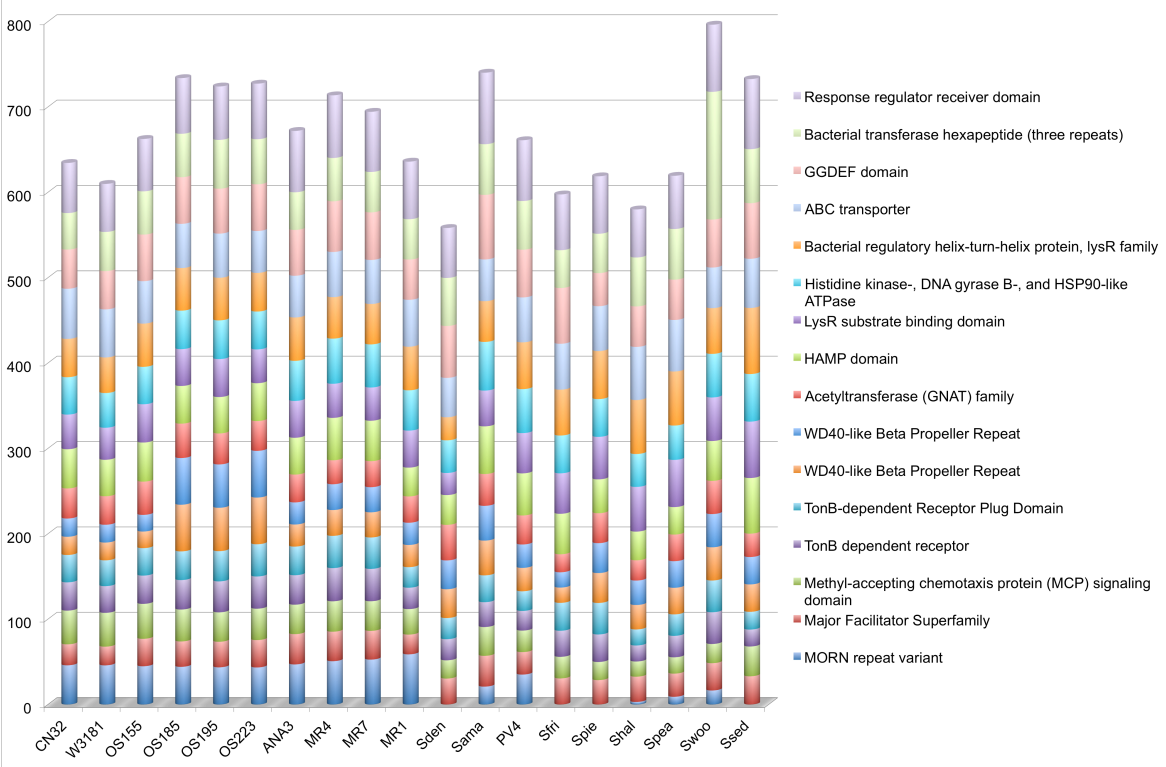


Figure 4. Most abundant protein families in *Shewanella.*
The number of proteins belonging to the families is shown on the y-axis.

Functions encoded by the most abundant protein families include chemotaxis proteins, two-component regulators, and signaling protein - all functions involved in sensing and responding to environmental conditions.

Some of the *Shewanella* strains are known to be more tolerant to salinity than others. We therefore compared protein families involved in transport of sodium between the strains. Pfam domains representing sodium transporters were

identified and their prevalence in the genomes determined (Fig. 5).

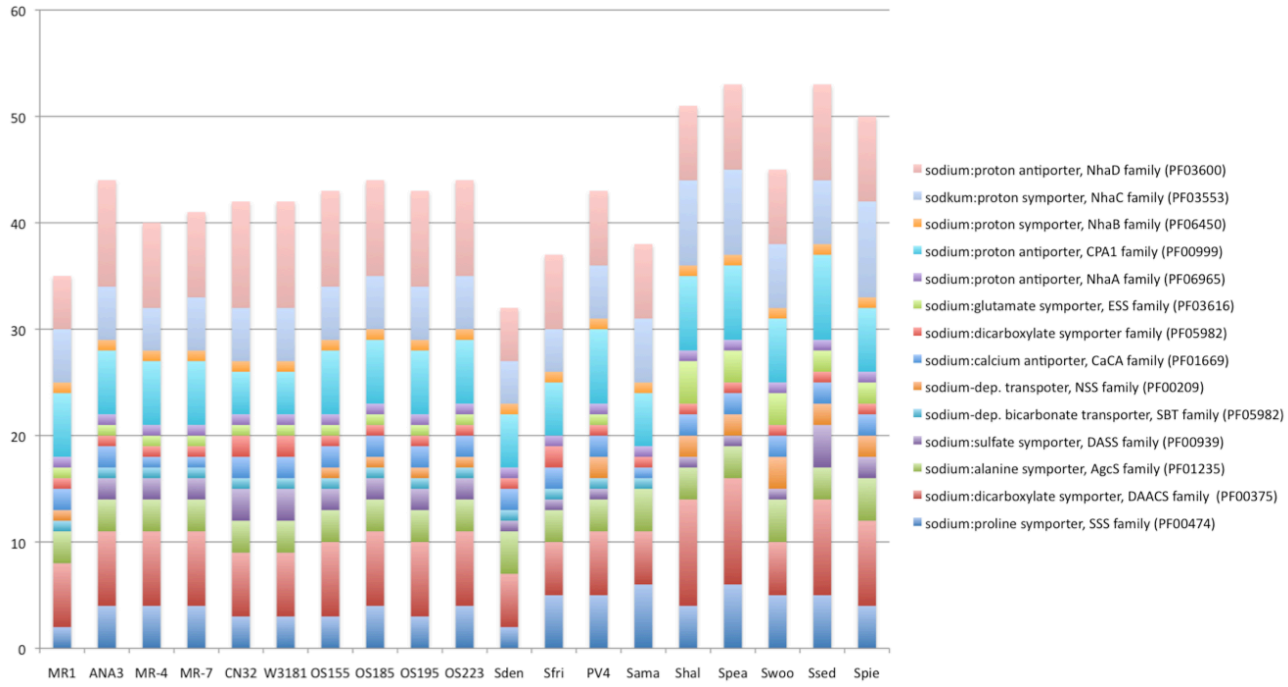


Figure 5. Abundance of *Shewanella* sodium transporter families.
The number of proteins belonging to the families is shown on the y-axis.

Overal. the obligate marine *Shewanella* strains, including *S. halifaxens, S. pealeana, S. sediminis*, and *S. piezotolerans*, contained a higher number of sodium transport proteins. The lowest number of sodium transporters is shown for *S. oneidensis* MR-1 and *S. denitrificans*. *S. oneidensis* MR-1 was isolated from fresh water sediments, and might have lost their sodium transporters while adapting to the fresh water environment. *S. denitrificans* has in general undergone a lot of gene rearrangements, involving both loss and gain of functions.

Sodium transport family comparisons can also be done family by family. In Fig. 6 we show how the 65 sodium:proline symporter, SSS family members can be clustered at different similarity levels. By changing the threshold used in the clustering from low similarity (< pam 150) to high similarity (< pam 50), similarity between distantly related members of a family is not detected resulting in the formation of more clusters. At pam 50, the clusters correspond to ortholog groups. The colors of the nodes in Fig. 6 represent different strains.

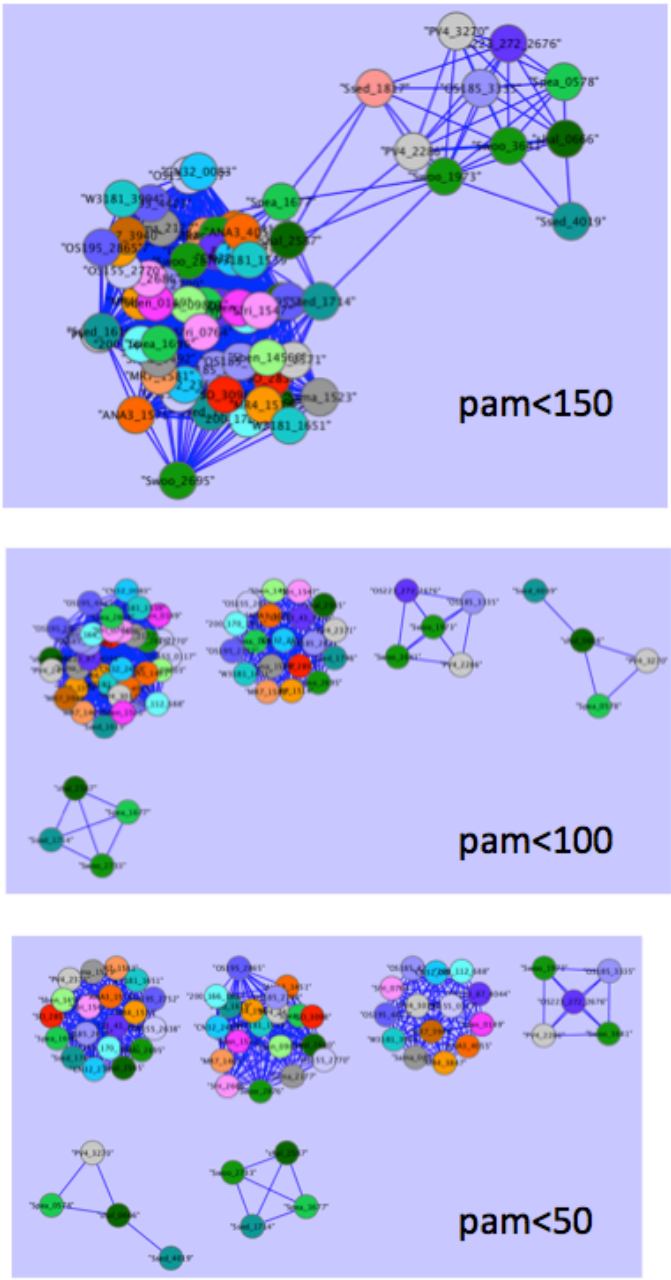


Figure 6. Clustering of *Shewanella* sodium:proline symport family members. The colors of the nodes represent different strains.

B. 2. Protein family comparisons and adaptation

Protein families may give insight into functions that are in abundance or absent from organisms. As shown in Fig. 5, comparing the sodium transport proteins revealed that *S. oneidensis* MR-1 and *S. denitrificans* have fewer sodium transport proteins, indicating that they have adjusted their metabolisms to a less saline

environment. A comparison of the protein families among the *Shewanell*a strains using Multi Experiment Viewer (MeV) software indicated a set of protein families that differed in abundance for the piezotolerant obligate marine *Shewanella* strains (Fig. 7).

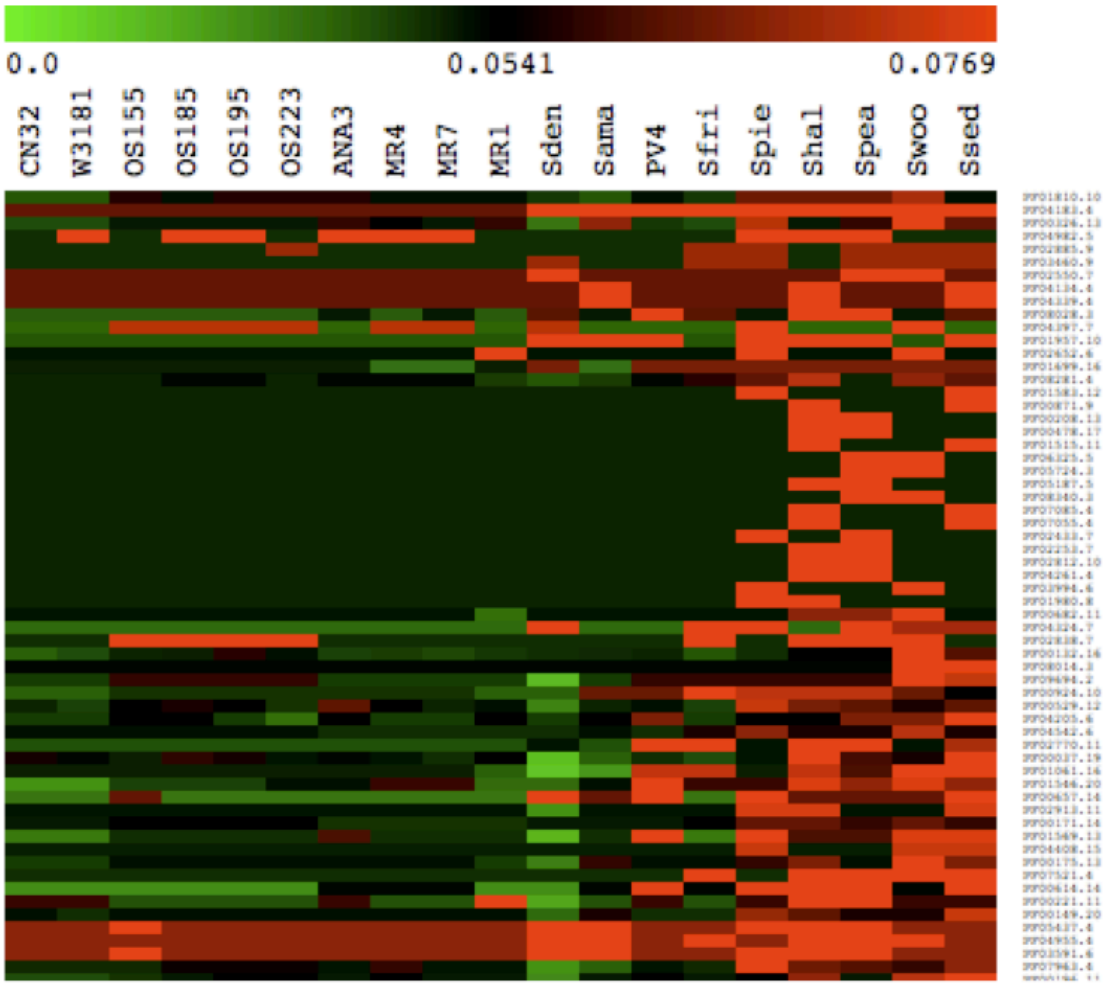


Figure 7. *Shewanella* protein family abundance.
Multi Experiment Viewer (MeV) was used to identify protein families that show a positive trend (red = expansion of protein families) or negative trend (green = reduction of protein families).

A list of the protein families which were found to differ in size for *S. piezotolerans, S. halifaxens, S. pealeana, S. woodyi*, and *S. sediminis* compared to the remaining *Shewanella* strains are shown in Table 4. Protein families with a positive trend are larger while protein families with a negative trend are smaller.

Table 4. *Shewanella* protein family expansion/reduction.

Trend	Pfam	Function	Family Size
positive	PF00132.16	Bacterial tansferase hexapeptide	43 - 150
positive	PF00005.19	ABC transporter	46 - 62
positive	PF03466.12	transcriptional regulator, LysR	26 - 66
postiive	PF00037.19	4Fe-4S binding domain	8 - 48
positive	PF01266.16	FAD dependent oxidoreductase	11 - 26
negative	PF07661.5	Morn repeat variant	0 - 58
negative	PF08447.3	PAS fold	8 - 25
negative	PF01590.18	GAF domain	6 - 17
negative	PF07012.4	Curlin associated repeat	0 - 30
negative	PF01380.14	SIS domain	4 - 9

A link between the functions of the protein families that are increased or reduced in the piezotolerant obligate marine strains and the environment they live in is not clear. The bacterial transferase hexapeptide domain is present in acyltransferases, and the ABC transport domain is that of the ATP-binding component of the transporter. The 4Fe-4S and FAD dependent oxidoreductases may be used for respiratory paths that are present in this group of *Shewanellas*. SIS (Sugar ISomerase) domains are less prevalent among the piezotolerant obligate marine strains possibly suggesting that sugar degradation is less common in these environments.

Differences in protein family sizes can further lead one to genomic loci where there are differences in the gene content. As shown in Fig. 8, a set of 4 *Shewanella* orthologs unique to the piezotolerant obligate marine strains was identified from protein family abundance data.

ortholog	Pfam	TIGRfam	
sog12634	PF00165.15	ND	transcriptional regulator, AraC family
sog10696	ND	ND	permease, MFS family
sog12633	PF00529.12	ND	secretion protein, HlyD family
sog12631	PF00990.13	TIGR00254	diguanylate cyclase

Figure 8. Protein family differences reveal genomic loci of interest.
A loci containing a transcriptional regulator, transport proteins, and a protein for the synthesis of the regulatory compound cyclic-di-GMP is present in piezotolerant obligate marine strains and absent in the non-piezotolerant *Shewanella* strains. It is possible that cyclic-di-GMP plays a role in adaptation to the depth or to the marine environment.

B. 3. Comparison of sequence signatures in *Shewanella*

To find signs of adaptive changes from selective pressures on protein-coding genes we calculated the ratio of non-synonymous substitutions per non-synonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks); the Ka/Ks ratio. Core ortholog groups present in all *Shewanella* genomes were used to perform a Ka/Ks analysis with SNAP (Synonymous Nonsynonymous Analysis Program). A pairwise comparison of the Ka/Ks values was done between the proteins encoded in the piezotolerant strains (*S. piezotolerans, S. halifaxens, S. pealeana, S. woodyi*, and *S. sediminis*) and those encoded in the remaining *Shewanella* strains was done. The significance of the differences in the Ka/Ks values was calculated using SAM or the Student's T-test in the Multi Experiment Viewer (MeV).

Using SAM we identified 378 proteins that were significantly different in their Ka/Ks ratio between piezotolerant and non-piezotolerant strains (Fig. 9). We analyzed these proteins in the SheonCyc, a BioCyc pathway genome database developed for *S. oneidensis* MR-1, and found that the altered proteins are involved in several metabolic pathways (Fig. 10). Pathways potentially under selective pressure include outer and inner membrane proteins and cofactor biosynthesis proteins.

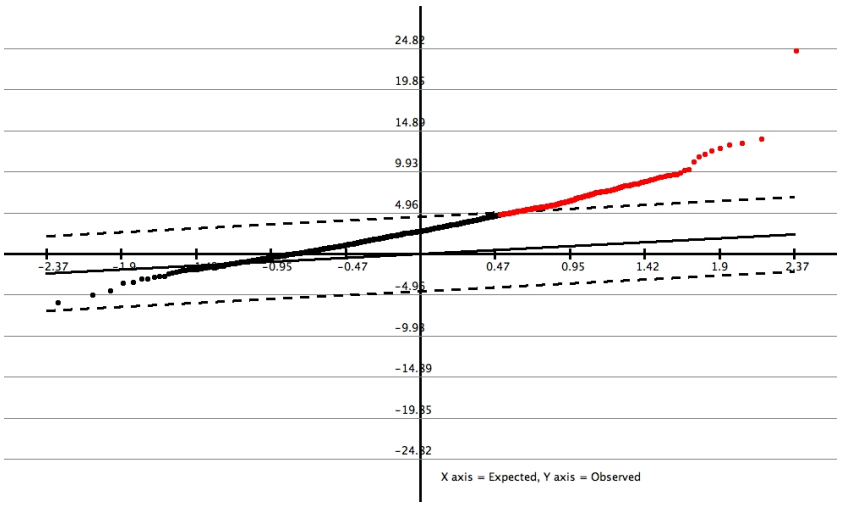


Figure 9. Ka/Ks substitution differences beween *Shewanella* proteins.
Ka/Ks ratio of the core *Shewanella* proteins in piezotolerant vs non-pieoztolerant *Shewanella* strains were determined using SAM. Significantly different ratios are shown in red.

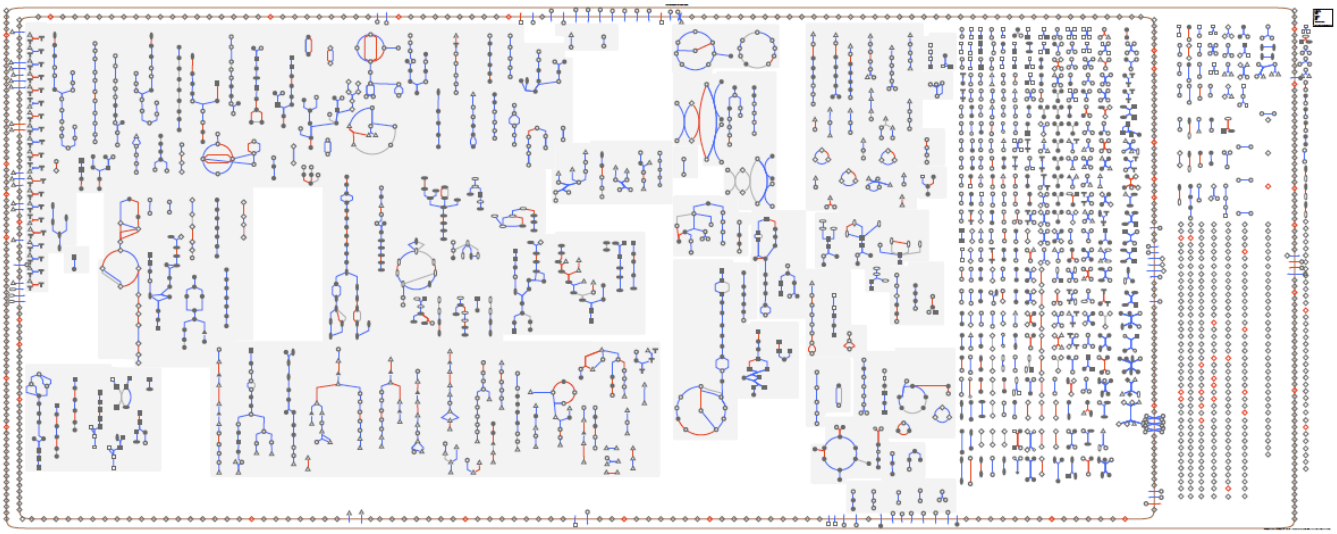


Figure 10. Metabolic overview of *Shewanella*.
Proteins found to have significant sequence differences in the piezotolerant vs non-piezotolerant strains are indicated in red in the overview figure.

Using the Student T-test we identified 675 proteins that were significantly altered between the piezotolerant and non-piezotolerant strains. Proteins with the highest T-values are shown in Table 5.

Table 5. *Shewanella* proteins with significantly altered Ka/Ks ratios according to the Student T-test.

Ortholog	Piezotolerant	Others	T-value	Location	Gene Product
SOG04229	0.805 ± 0.244	0.073 ± 0.036	23	cyt	adenosine deaminase, Add
SOG00994	0.115 ± 0.018	0.0581 ± 0.018	18	IM	Na-translocatingNADH-quinone reductase subunit
SOG01068	0.141 ± 0.027	0.061 ± 0.028	17	cyt	transcription elongation factor, GreA
SOG03121	0.129 ± 0.030	0.052 ± 0.023	16	cyt	serine hydroxymethyltransferase, GlyA
SOG03942	0.121 ± 0.030	0.051 ± 0.020	15	cyt	glutamine synthetase, GlnA
SOG01787	0.330 ± 0.094	0.131 ± 0.052	15	LP	UDP-sugar hydrolase/nucleotidase, UshA
SOG03195	0.336 ± 0.139	0.058 ± 0.055	14	cyt	phosphoribosylaminoimidazole carboxylase, PurE
SOG02130	0.095 ± 0.026	0.039 ± 0.015	14	cyt	integration host factor, IhfB
SOG02925	0.079 ± 0.0181	0.039 ± 0.014	14	cyt	chemotaxis protein, CheV_3
SOG02745	0.085 ± 0.010	0.047 ± 0.020	14	cyt	amidophosphoribosyltransferase, PurF
SOG03936	0.186 ± 0.063	0.059 ± 0.026	13	IM	conserved hypothetical protein
SOG00008	0.071 ± 0.011	0.043 ± 0.012	13	cyt	chromosomal replication initiator, DnaA
SOG00549	0.121 ± 0.022	0.073 ± 0.020	13	cyt	GTP-binding protein, HflX
SOG02131	0.121 ± 0.025	0.067 ± 0.021	13	cyt	ribosomal protein S1, RpsA
SOG00015	0.078 ± 0.018	0.033 ± 0.022	13	cyt	Glycyl-t-RNA synthetase, GlyQ

The Cellular roles of the significantly altered protein sequences identified by the Student T-test are listed in Table 6. These pathways include amino acid biosynthesis, cofactor biosynthesis and cell envelope. Several of the pathways identified with to the SAM analysis (displayed in the cellular overview diagram; Fig. 10) agree with the cellular roles identified by the Student T-test.

Table 6. Cellular roles for *Shewanella* proteins with significantly altered Ka/Ks ratios according to the Student T-test.

TIGRrole	% genes changed	% genes overall
Amino acid biosynthesis	4.9	4.2
Biosynthesis of cofactors, prosthetic groups	4.5	6.0
Cell envelope	7.6	7.6
Cellular processes	8.4	1.2
Central intermediary metabolism	1.0	2.3
DNA metabolism	4.6	4.4
Energy metabolism	10.0	9.1
Fatty acid and phospholipid metabolism	3.1	2.3
Hypothetical proteins	7.0	10.0
Protein fate	7.8	7.7
Protein synthesis	10.4	7.9
Nucleoside/nucleotide metabolism	3.4	2.1
Regulatory functions	5.7	5.8
Signal transduction	0.7	1.4
Transcription	3.6	2.4
Transport and binding proteins	4.8	7.5
Unknown role	12.4	12.8

C. METABOLIC PATHWAY RECONSTRUCTION

We are working in the BioCyc pathway genome database environment [3]. Pathway genome databases (PGDBs) have been built for 17 members of the *Shewanella* genus using the Pathway-Tools software by us and by our collaborators. These databases are part of the *Shewanella* Knowledgebase and can be accessed online via Oakridge National Laboratory
[7]. We have maintaining functional updates via the *Shewanella* ortholog table. However, our group's efforts have focused on the curation of the databases for *S. oneidensis* MR-1 and *Shewanella* sp. W3-18-1. These strains have been targeted for

expermimental analyses by the *Shewanella* Federation for their bioremediation potential and their ability to be co-cultured with cyanobacteria, respectively. We are maintaining these two PGDBs and made them web accessible at http://pathways.mbl.edu/.

Currently the *S. oneidensis* MR-1 PGDB contains 5287 genes, 5138 proteins, and 128 protein complexes linked to 1033 citations, 974 enzymatic reactions, and 205 pathways. The Pathway Tools software used to build the PGDBs includes tools for omics analyses and for comparative studies between organisms based on their PGDB curations. We are also capturing regulator predictions of the transcriptional regulators from the RegPrecise database [13].

C. 1. Applications of pathway genome databases

Our metabolic pathway constructions are captured in the pathway genome databases. The Pathway-Tools software used to build the PGDBs can be used to build additional pathways using chemical compounds and reaction. We have also made use of the PGDBs to analyze experimental data obtained from our collaborators. This has resulted in several publications:

a. Large-Scale Comparative Phenotypic and Genomic Analyses Reveal Ecological Preferences of *Shewanella* Species and Identify Metabolic Pathways Conserved at the Genus Level.
A comparison of growth phenotypes and genome sequences of five *Shewanell*a strains. [16]

b. Detection of transcriptional triggers in the dynamics of microbial growth: application to the respiratorily versatile bacterium *Shewanella oneidensis*.
Identifying metabolic changes that occur during growth phase changes [2]

The pathway genome databases continue to be maintained. They are currently being used for manuscripts describing A) Analysis of *S. oneidensis* MR-1 transcriptomic and proteomic datasets and B) Interaction studies of two microbes that can be co-cultured.

References:

1. **Bateman A.**, **P. Coggill**, **and R. D. Finn**. 2010. DUFs: families in search of function. Acta Crystallogr Sect F Struct Biol Cryst Commun **66**:1148–1152.
2. **Beg Q. K.**, **M. Zampieri**, **N. Klitgord**, **S. B. Collins**, **C. Altafini**, **M. H. Serres**, **and D. Segrè**. 2012. Detection of transcriptional triggers in the dynamics of microbial growth: application to the respiratorily versatile bacterium Shewanella oneidensis. Nucleic Acids Research –.
3. **Caspi R.**, **T. Altman**, **J. M. Dale**, **K. Dreher**, **C. A. Fulcher**, **F. Gilham**, **P. Kaipa**, **A. S. Karthikeyan**, **A. Kothari**, **M. Krummenacker**, **M. Latendresse**, **L. A. Mueller**, **S. Paley**, **L. Popescu**, **A. Pujar**, **A. G. Shearer**, **P. Zhang**, **and P. D. Karp**. 2010. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Research **38**:D473–D479.
4. **Coursolle D.**, **and J. A. Gralnick**. 2010. Modularity of the Mtr respiratory pathway of Shewanella oneidensis strain MR-1. Mol Microbiol no–no.
5. **Hau H. H.**, **and J. A. Gralnick**. 2007. Ecology and biotechnology of the genus Shewanella. Annu Rev Microbiol **61**:237–258.
6. **Heidelberg J. F.**, **I. T. Paulsen**, **K. E. Nelson**, **E. J. Gaidos**, **W. C. Nelson**, **T. D. Read**, **J. A. Eisen**, **R. Seshadri**, **N. Ward**, **B. Methe**, **R. A. Clayton**, **T. Meyer**, **A. Tsapin**, **J. Scott**, **M. Beanan**, **L. Brinkac**, **S. Daugherty**, **R. T. Deboy**, **R. J. Dodson**, **A. S. Durkin**, **D. H. Haft**, **J. F. Kolonay**, **R. Madupu**, **J. D. Peterson**, **L. A. Umayam**, **O. White**, **A. M. Wolf**, **J. Vamathevan**, **J. Weidman**, **M. Impraim**, **K. Lee**, **K. Berry**, **C. Lee**, **J. Mueller**, **H. Khouri**, **J. Gill**, **T. R. Utterback**, **L. A. Mcdonald**, **T. V. Feldblyum**, **H. O. Smith**, **J. C. Venter**, **K. H. Nealson**, **and C. M. Fraser**. 2002. Genome sequence of the dissimilatory metal ion–reducing bacterium Shewanella oneidensis. Nat Biotech **20**:1118–1123.
7. **Karpinets T. V.**, **M. F. Romine**, **D. D. Schmoyer**, **G. H. Kora**, **M. H. Syed**, **M. R. Leuze**, **M. H. Serres**, **B. H. Park**, **N. F. Samatova**, **and E. C. Uberbacher**. 2010. Shewanella knowledgebase: integration of the experimental data and computational predictions suggests a biological role for transcription of intergenic regions. Database (Oxford) **2010**:baq012.
8. **Klonowska A.**, **T. Heulin**, **and A. Vermeglio**. 2005. Selenite and tellurite reduction by Shewanella oneidensis. Appl Environ Microbiol **71**:5607–5609.
9. **Konstantinidis K. T.**, **M. H. Serres**, **M. F. Romine**, **J. L. M. Rodrigues**, **J. Auchtung**, **L.-A. McCue**, **M. S. Lipton**, **A. Obraztsova**, **C. S. Giometti**, **K. H. Nealson**, **J. K. Fredrickson**, **and J. M. Tiedje**. 2009. Comparative systems biology across an evolutionary gradient within the Shewanella genus. Proceedings of the National Academy of Sciences **106**:15909–15914.
10. **Konstantinidis K. T.**, **and J. M. Tiedje**. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci USA **102**:2567–2572.
11. **Leplae R.**, **G. Lima-Mendez**, **and A. Toussaint**. 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Research **38**:D57–61.
12. **Leyn S. A.**, **F. Gao**, **C. Yang**, **and D. A. Rodionov**. 2012. N-acetylgalactosamine utilization pathway and regulon in Proteobacteria. Genomic reconstruction and

experimental characterization in Shewanella. Journal of Biological Chemistry.
13. **Novichkov P. S.**, **O. N. Laikova**, **E. S. Novichkova**, **M. S. Gelfand**, **A. P. Arkin**, **I. Dubchak**, **and D. A. Rodionov**. 2010. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. Nucleic Acids Research **38**:D111–D118.
14. **Punta M.**, **P. C. Coggill**, **R. Y. Eberhardt**, **J. Mistry**, **J. Tate**, **C. Boursnell**, **N. Pang**, **K. Forslund**, **G. Ceric**, **J. Clements**, **A. Heger**, **L. Holm**, **E. L. L. Sonnhammer**, **S. R. Eddy**, **A. Bateman**, **and R. D. Finn**. 2012. The Pfam protein families database. Nucleic Acids Research **40**:D290–301.
15. **Rawlings N. D.**, **A. J. Barrett**, **and A. Bateman**. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research **40**:D343–50.
16. **Rodrigues J. L. M.**, **M. H. Serres**, **and J. M. Tiedje**. 2011. Large-scale comparative phenotypic and genomic analyses reveal ecological preferences of shewanella species and identify metabolic pathways conserved at the genus level. Appl Environ Microbiol **77**:5352–5360.
17. **Saier M. H.** 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Research **34**:D181–D186.
18. **Wang H.**, **K. Hollywood**, **R. M. Jarvis**, **J. R. Lloyd**, **and R. Goodacre**. 2010. Phenotypic Characterization of Shewanella oneidensis MR-1 under Aerobic and Anaerobic Growth Conditions by Using Fourier Transform Infrared Spectroscopy and High-Performance Liquid Chromatography Analyses. Appl Environ Microbiol **76**:6266–6276.
19. **Yang C.** 2006. Comparative Genomics and Experimental Characterization of N-Acetylglucosamine Utilization Pathway of Shewanella oneidensis. Journal of Biological Chemistry **281**:29872–29885.

Posters:

Serres, MH, Metpally, RPR, and MF Romine (2010) Genome Signatures the Reflect Adaptation of *S. oneidensis* MR-1 to its Environment. 13th International Symposium on Microbial Ecology, Seattle, WA, August 22-27, 2010.

Fredrickson, J, Anderson, G, Baker, S, Beliaev, A, Lipton, M, Magnuson, J, Romine, M, Squier, S, Wiley, S, Bryant, D, Collart, F, Inskeep, W, Lutzoni, F, Osterman, A, **Serres, M**, Bolton, H, and D. Ward (2010) PNNL Foundational Scientific Focus Area – Biological Systems Interactions. US Department of Energy Genomic Science Awardee Meeting VIII. February 7-10, 2010.

Serres, MH, Metpally, RPR, and MF Romine (2010) Comparative Genome Analyses of Members of the Ecologically Versatile Genus Shewanella: Searching for Sequence Signatures That Reflect Environmnetal Adaptation. US Department of Energy Genomic Science Awardee Meeting VIII. February 7-10, 2010.

Serres, MH, Konstantinidis, K, Romine, MF, Rodrigues, JLM, Obraztsova, AY, Lipton, MS, McCue, LA, Fredrickson, JK, and JM Tiedje (2009) Comparative Analyses Across and Evolutionary Gradient Within the *Shewanella* Genus. US Department of Energy Genomics:GTL Awardee Workshop VII, February 8-11, 2009.

Publications:

Beg, QK, Zampieri, M, Klitgord, N, Collins, SB, Altafini, C, **Serres, MH**, Segre, D. 2012. Transcriptional and metabolic dynamics in *Shewanella oneidensis* reveals cellular integration of growth phase cues and nutrient depletion warning. Nucleic Acids Research online May 25, 2012 doi:10.1093/nar/gks467

Rodrigues, JL, **Serres, MH**, Tiedje, JM. 2011. Large scale comparative phenotypic and genomic analyses reveal ecological preferences of *Shewanella* species and identify metabolic pathways conserved at genus level. Applied and Environmental Microbiology. 77(15):5352-60.

Karpinets, TV, Romine, MF , Schmoyer, D.D, Kora, GH, Syed, MH, Leuze, MR, **Serres, MH**, Park, BH, Samatova, NF, and Uberbacher, EC. 2010. *Shewanella* knowledgebase: integration of the experimental data and computational predictions suggests a biological role for transcription of intergenic regions. Database Vol. 2010:baq012; doi:10.1093/database/baq012

Karpinets TV, Obraztsova AY, Wang Y, Schmoyer DD, Kora GH, Park BH, **Serres MH**, Romine MF, Land ML, Kothe TB, Fredrickson JK, Nealson KH, Uberbacher EC. 2010. Conserved synteny at the protein family level reveals genes underlying *Shewanella*

species cold tolerance and predicts their novel phenotypes. Funct. Integr. Genomics 10(1):97-110.

Serres, MH, Kerr, AR, McCormack TJ, Riley, M. 2009. Evolution by leaps: gene duplication in bacteria. Biol. Direct 4(1):46.

Konstantinidis KT, **Serres MH**, Romine, Rodrigues JLM, Auchtung, J, McCue LAc, Lipton MSc, Obraztsova A, Giometti CS, Nealson KH, Fredrickson JK, and Tiedje JM. 2009. Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. Proc. Natl. Acad. Sci. USA 106(37): 15909-14.

Nahum LA, Goswami,S, **Serres MH**. 2009. Protein families reflect the metabolic diversity of organisms and provide support for functional prediction. Physiol. Genomics 38: 250-260.

Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealson KH, Osterman AL, Pinchuk G, Reed JL, Rodionov DA, Rodrigues JL, Saffarini DA, **Serres MH**, Spormann AM, Zhulin IB, Tiedje JM. 2008. Towards environmental systems biology in *Shewanella*. Nat. Rev. Microbiol. 6(8): 592-603.