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

DOE award number:
DE-FG02-04ER63940

Recipient:
Dr. Monica Riley, Marine Biological Laboratory, Woods Hole, MA 02543
Transferred to Dr. Margrethe H. Serres, Marine Biological Laboratory, Woods Hole, MA 02543 in January 2006.

Project title:
Functional Analysis of *Shewanella*, a cross genome comparison.

Director/principal investigator: Dr. Margrethe H. Serres

Consortium: Affilated with the Shewanella Federation research group lead by Dr. Jim K. Fredrickson.

2.
The report does not patentable material or protected data.

3. Executive summary

The bacterial genus *Shewanella* includes a group of highly versatile organisms that have successfully adapted to life in many environments ranging from aquatic (fresh and marine) to sedimentary (lake and marine sediments, subsurface sediments, sea vent). A unique respiratory capability of the *Shewanellas*, initially observed for *Shewanella oneidensis* MR-1 (17), is the ability to use metals and metalloids, including radioactive compounds, as electron acceptors. Members of the *Shewanella* genus have also been shown to degrade environmental pollutants i.e. halogenated compounds, making this group highly applicable for the DOE mission. *S. oneidensis* MR-1 has in addition been found to utilize a diverse set of nutrients and to have a large set of genes dedicated to regulation and to sensing of the environment.

The sequencing of the *S. oneidensis* MR-1 genome facilitated experimental and bioinformatics analyses by a group of collaborating researchers, the Shewanella Federation. Through the joint effort and with support from Department of Energy *S. oneidensis* MR-1 has become a model organism of study (6). Our work has been a functional analysis of *S. oneidensis* MR-1, both by itself and as part of a comparative study. We have improved the annotation of gene products, assigned metabolic functions, and analyzed protein families present in *S. oneidensis* MR-1. The data has been applied to analysis of experimental data (i.e. gene expression, proteome) generated for *S. oneidensis* MR-1. Further, this work has formed the basis for a comparative study of over 20 members of the *Shewanella* genus. The species and strains selected for genome sequencing represented an evolutionary gradient of DNA relatedness, ranging from close to intermediate, and to distant. The organisms selected have also adapted to a variety of ecological niches. Through our work we have been able to detect and interpret genome similarities and differences between members of the genus. We have in this way contributed to the understanding of speciation and adaptation in Shewanella in relation to its metabolic and respiratory capabilities and its environmental interactions.

Through the collaboration with several research groups focused on the biology of *S. oneidensis* MR-1 and other members of the *Shewanella* genus, we have been able to link bioinformatically based predictions and analyses to experimental data in a productive manner. In additon to improving our understanding of *S. oneidensis* MR-1, our work is also applicable to the understanding of other environmentally versatile microbes.

4. Compare of the actual accomplishments with the goals and objectives of the project.

We carried out our proposed work of continuing the functional analysis of the *S. oneidensis* MR-1 genome. Up to date information on gene product functions, metabolic pathways, and relevant literature references has been disseminated to members of the Shewanella Federation and to the public through our web accessible pathway genome database ShewCyc as well as the Shewanella Knowledgebase developed at ORNL. The annotations have been and continue to be applied to the analysis of experimentally derived datasets. Our protein family studies have contributed to the improved gene product annotations as well as highlighted functions that have undergone diversification and specialization in their host organism. We hosted a colloquium at the 2008 ASM general meeting on the topic presenting some of our data. The comparative work was initially proposed against the genomes of other model organisms. However, we changed our focus to comparative analyses against other *Shewanella*s when these genome sequences becaume available. This was done in order to contribute better to the overall goal of the Shewanella Fedration. Our work has been presented in several posters, presentations and publications (see section on products developed).

5. Summary of project activities for the entire period of funding.

5.1 Gene product annotations.

The effort of improving the functional description for the *S. oneidensis* MR-1 (hereafter referred to as MR1) gene products has been an ongoing effort through the life of the project. It has included the use of a controlled vocabulary for consistency in description of gene products to improve function assignments and to make the comparative analyses more intuitive. We have monitored the primary literature for new functional characterization of homologous gene products in *Shewanella* and in other microbes. Information has also been mined from function specific databases i.e. Transport Classification Database (www.tcdb.org), TransportDB (www.membranetransport.org), MiST (http://genomics.ornl.gov/mist), DBD Transcription Factor Prediction Database (www.transcriptionfactor.org), MEROPS Peptidase Database (merops.sanger.ac.uk), and Insertion Sequence (IS) Finder Database (www-is.biotoul.fr/is.html). The annotation effort has been done in collaboration with the other members of the Shewanella Federation research team, mainly Dr. M.F. Romine (Pacific Northeast National Lab). Our work is done in coordination with the new and revised gene calls published by Romine et al. (20). In addition we also captured information from proteome analyses (5) (14), gene expression studies (4) (2) (3), and from other genome wide studies such are the regulatory analysis by Liu et al. (16) and Kazakov et al (13).

The current annotations of the MR1 gene products can be accessed on line at our pathway genome database ShewCyc and at the Shewanella Knowledgebase (see below for details). Of the 4561 gene products, 4413 are protein coding genes (CDSs) and 148 are RNAs. We have also identified 153 gene products that are encoded by fused genes in MR1. The proteins encoded by fused genes have two or more activities in separate parts of the molecule. These proteins are often misannotated, mainly due to including only one of the activities in the gene product description. For our analyses we separated the fused genes into modules representing each of the independent activities. We then assigned module specific functions. The separated gene products were labeled with their locus tag followed by an extension indicating their order in the protein (i.e. "_1" is the most N-terminal module). We found that metabolic enzymes, transporters, and regulatory proteins were most frequently fused in MR1. The relatively high number of fused genes in MR1 is likely due to the many regulatory proteins that are found in this organism. A comparison of the current annotation with the original one published by Heidelberg et al. (10), showed that 66% of the gene product descriptions have been improved. These improvements include differences to 562 Enzyme Commision (EC) number assignments, either representing new assignments (79%) or reassignments.

We classified the genes according to their gene product type (see a complete list of categories in the legend for Figure 1). The most abundant gene product types were found to be Enzymes, Unknown functions, Regulators and Transporters (Fig. 1).

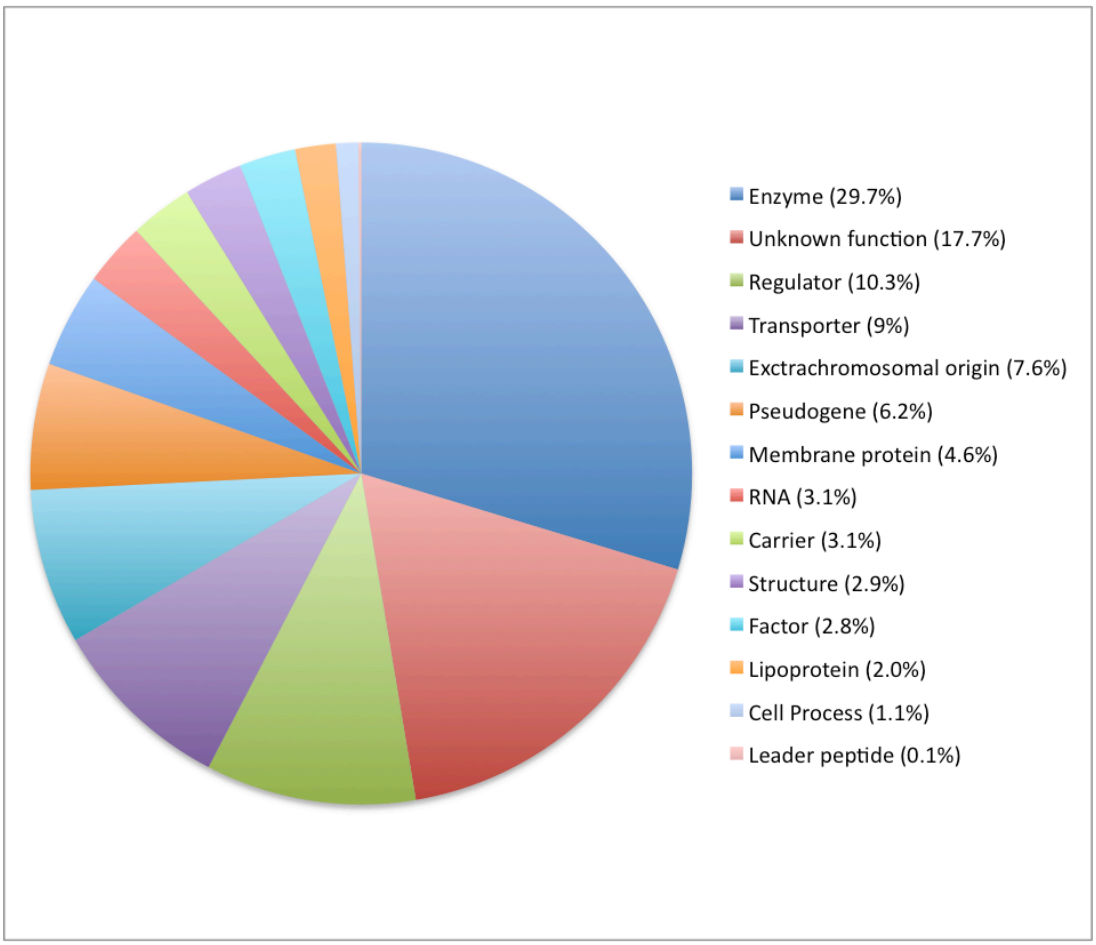

Figure 1. Distribution of gene product types encoded in ***S. oneidensis*** MR-1.

The distribution of gene product types in MR1 is comparable to that seen for *Escherichia coli* K-12 (19). However, one exception is that MR1 contains a large number of pseudogenes. This can likely be attributed to the many insertion sequences and miniature inverted-repeat transposable elements that are present in the MR1 genome sequence (20).

MR1 gene products were also characterized according to the metabolic or cellular role they play in the cell. We used the TIGR cellular role classification system consisting of a "Main role" and a "Sub role" (18). Among our MR1 cellular role assignments, 70% were to functional categories while the remaining 30% were to roles of unknown function (i.e. Unknown role, Hypothetical proteins). A complete list of the cellular role assignments can be accessed at the Shewanella Knowledgebase (see below), and a list of the main cellular roles that were assigned to the MR1 geneproducts is shown in Table 1.

Table 1. Main cellular role assignment for *S. oneidensis* MR-1 gene products according to the TIGR classification system.

No. Assignments	Main Cellular Role
432	Mobile and extrachromosomal element functions
404	Transport and binding proteins
384	Energy metabolism
372	Regulatory functions
313	Cell envelope
288	Cellular processes
271	Proein fate
209	Disrupted reading frame
180	Protein synthesis
155	Biosynthesis of cofactors, prosthetic groups, and carriers
155	DNA metabolism
103	Signal transduction
96	Amino acid biosynthesis
64	Fatty acid and phospholipid metabolism
64	Purines, pyrimidines, nucleosides, and nucleotides
62	Transcription
57	Central intermediary metabolism

We made an effort to annotate predicted proteins of unknown function. Of the 1538 gene products originally described as 'conserved hypothetical proteins' or 'hypothetical proteins', there are now assignments for 840 (55%) of them. These new function assignments include 205 enzymes. Some of the new enzyme activities were based on new experimental evidence in MR1, i.e. PrpF (7, 9) and NagK (26) or in other microrganisms, i.e. YhbJ (11). Other new assignments were based on the protein belonging to a protein family. In the latter case a more general function assignment such as thioesterase family protein was often used. We found that some of the unknown function proteins were encoded by the MR1 prophages MuSO1, MuSO2, or LambdaSo. Other new function assignments included "membrane protein" for proteins predicted to be located in the inner or outer membrane and "expressed protein" for proteins that were detected by proteome analyses.

Finally, we made use of the structural conservation of protein sequences to assign proteins to protein families. As protein sequences diverge, they loose similarity at the sequence level while maintaining a similarity at the structural level. Hence we used predicted structural domains for the MR1 proteins from the Superfamily database (25) to assign general functions to the proteins of unknown function.

We have deposited the newly assigned gene product annotations for MR1 in the *S. oneidensis* MR-1 genome database hosted by the Oakridge National Laboratory (ORNL) computational group (https://compbio.ornl.gov/microbial/sone/). The aim of this database is to generate an updated Genbank version for the genome. Work is in the process for generating such a deposit. The updated annotations is also entered in our web accessible MR1 pathway genome database ShewCyc (http://pathways.mbl.edu/).

5.2. Database for metabolic pathways and cell roles

In order to capture the annotated MR1 genome in a metabolic framework, we built a pathway genome database (PGDB) for MR1 using the Pathway Tools software (1, 12). Our MR1 database, ShewCyc, is installed on a local web server that is accessible through the internet: http://pathways.mbl.edu/. The database is continually updated with new annotations and pathway predictions.

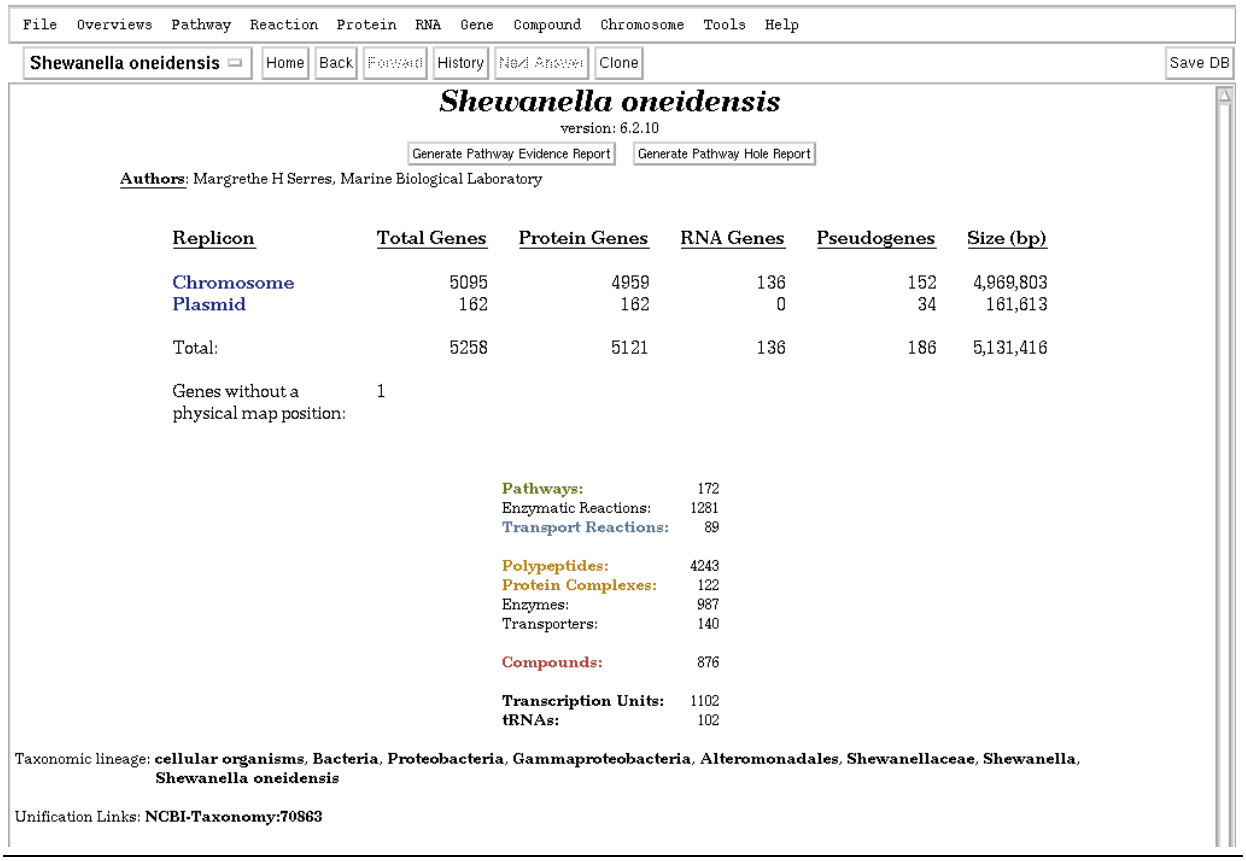

Figure 2. The *S. oneidensis* MR-1 pathway genome database .

A summary of the current status of the MR1 PGDB is shown in Fig. 2. The database may be queried for information on genes, proteins, pathways, and compounds. Each gene product is represented by a gene page and a protein page. The gene page contains information on gene nomenclature, genome location, unit of transcription, regulator(s), membership in paralogous protein family, and on cellular role assignments according to the MultiFun classification system (22). The protein page contains information on the gene product name and synonyms, location, size, cell role, enzymatic reactions, and pathway memberships. Curated notes relating to the gene product with references to the literature is also eneterd to the gene product page. There are distict pages depicting the pathways (Fig. 3). Hyperlinks are used to migrate between the gene, protein and pathway pages. Currently the MR1 geneproducts have been assigned to 1281 enzymatic reactions, 89 transport reactions, and 172 metabolic pathways.

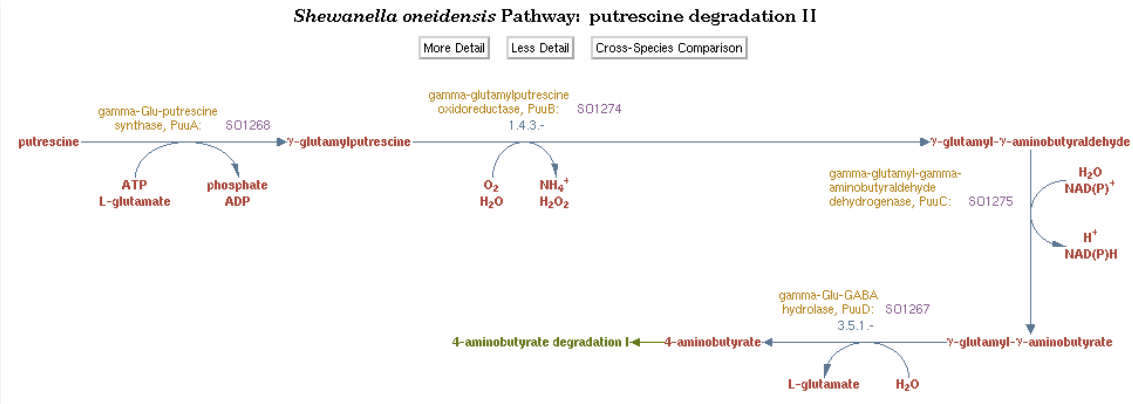

Figure 3. Putrescine degradation pathway in *S. oneidensis* MR-1.

We were also able to capture some of the MR1 regulatory interactions in the PGDB. Transcriptional regulation predictions for MR1 has been generated by Dr. Dmitry Rodionov (http://regprecise.lbl.gov). Regulatory information from RegPrecise, the literature, and from gene expression studies carried out by members of the Shewanella Federation has been incorporated in to the MR1 database. An overview of the transcriptional regulation of MR1 was generated using the Pathway Tools software (Fig. 4).

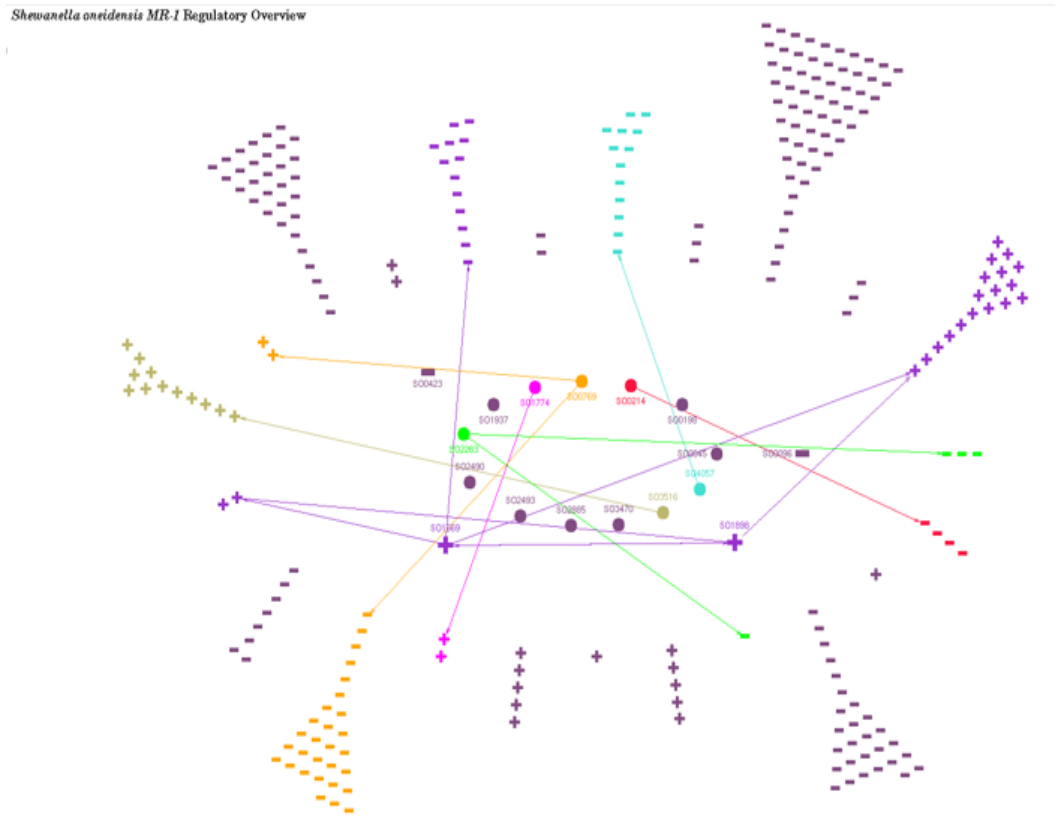

Figure 4. Overview of regulatory interactions in *S. oneidensis* MR-1.

ShewCyc has also been used to generate genome overview pictures that summarizes the MR1 gene products and their pathway assignment (Fig. 5) Classes of pathways (nucleoside/nucleotide biosynthesis, cell structure biosynthesis, amino acid degradation etc.) are clustered in the figure as indicated. The symbols used in Fig. 5 includes Δ (amino acids), □ (carbohydrates), ◊ (proteins), {} (purines/pyrimidines), ∇ (cofactors), T (tRNAs), and O (other). The overview feature has proven to be useful for interpreting genomic and experimental data. Figure 5 shows a cellular overview of MR1 where the enzymatic reactions have been color coded according to whether the respective enzymes have paralogs in MR1 (red) or not (yellow). The presence of paralogous metabolic enzymes can give us further insight into metabolic pathways that have evolved by gene duplications followed by divergence. We have also used gene expression data and proteome data as overlays to the cellular overview schema. Such MR1 datasets have been generated to capture the induction or repression of metabolic pathways in an easily interpretable framework. We have also used this method to aid in our gene product function preditions.

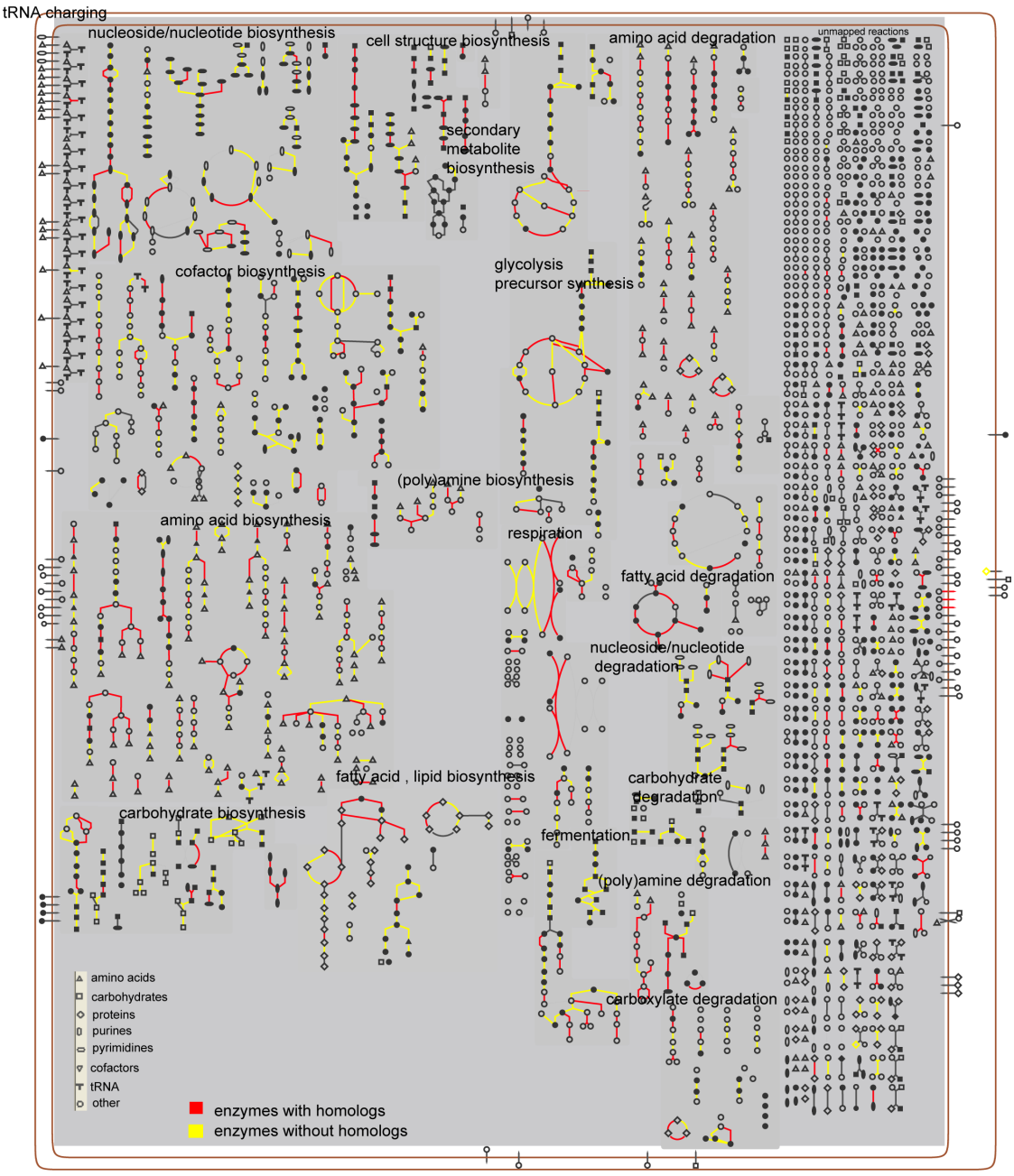

Figure 5. Cellular overview of *S. oneidensis* MR-1 functional assignments.

5.3. Shewanella Cross Genome Comparisons, commonalities and differences in protein functions.

A comparison of MR1 against other model organisms including *E. coli*, *Vibrio cholerae*, *Geobacter sulfurreducens*, *Pseudomonas aeruginosa*, and *Bacillus subtilis* was proposed. The comparisons were to be done at the protein level and at the protein family level. Sequence similarity searches were done using Darwin (8). We selected pair-wise alignments between CDSs of MR1 and the four organisms where 83 or more amino acids aligned over at least 45% of the sequences at a PAM distance of 100 or less. These values had been previously determined based on work with *E. coli*. Sequence matches were found for 1561, 1359, 593, and 480 MR1 proteins in *V. cholerae*, *P. aeruginosa*, *G. sulfurreducens*, and *B.subtilis*, respectively. In fact, when comparing MR1 CDSs to those of 107 bacterial genomes, *V. cholerae* and *P. aeruginosa* were among the top three organisms with best sequence matches to MR1. We did not complete our analyses against the selected genomes due to a change in the focus of our research. But we hope to include this work for a future publication. We believe that the additional knowledge we have gained while working with MR1 will be helpful for comparing the MR1 gene products and protein families with those encoded in the genomes of the four reference organisms.

Our genome comparison targets were changed in conjunction with work done by the Shewanella Federation, a collaborative group of Shewanella resesarchers. In order to benefit the work done by this group we joined the effort to do comparative studies of *Shewanella* genome sequences. Our initial targets were the first ten genome sequences to be completed (Table 2).

Table 2. Members of the *Shewanella* genus whose genome sequences were compared.

Strains	Environment isolated from	Abbrev.
S. oneidensis MR-1	freshwater sediments, Lake Oneida, NY, USA	MR1
Shewanella sp. ANA-3	arsenic treated wood in estuary, Woods Hole, MA	ANA3
Shewanella sp. MR-4	oxic zone of water column, Black Sea	MR4
Shewanella sp. MR-7	anoxic zone of water column, Black Sea	MR7
S. putrefaciens CN-32	subsurface sandstone, New Mexico, NM	CN32
Shewanella sp. W3-18-1	marine sediment, Pacific Ocean	W3181
S. frigidimarina NCIMB400	marine North Sea, UK	Sfri
S. denitrificans OS217	marine oxic/anoxic zone, Baltic Sea	Sden
S. loichica PV-4	marine Naha Vents, Hawaii	PV4
S. amazonensis SB2B	marine sediments, Amazon River delta, Brasil	Sama

The *Shewanella* strains were selected for sequencing based on their similarity at the 16S rRNA level. The goal was to obtain genome sequences belonging to one genus but whose relatedness formed a gradient from closely to intermediately to distantly related strains or species. The Average Nucleotide Identity (ANI) values (15) were calculated for the ten *Shewanella* genomes. ANI values represent a sensitive method to calculate evolutionary relatedness among DNA sequences. The *Shewanella* ANI values ranged from 96.5-98.4% (close) to 79-92% (intermediate) to 69.7-72% (distant). An ANI values of 95% or above correspond to the species level.

A significant effort went into identifying protein orthologs in the ten *Shewanella* genomes. The orthologs are important for genome comparisons as well as populating the annotations of the non-MR1 genomes with the highly curated MR1 annotations. Orthologs were identified for the ten *Shewanella* genomes by a combination of three methods; protein-protein pair-wise reciprocal BLAST (blastp), reciprocal protein-genomic sequence best match (tblastn), and Darwin pair-wise best hit. The output of the three methods was combined and manually curated. We further supplemented the evaluation with neighborhood analyses to decipher paralogous from orthologous relationships. An ortholog editor (database) was built for the *Shewanella* dataset by Shewanella Federation Collaborators at Oak Ridge National Laboratory under the leadership Dr. Ed Uberbacher. The ortholog editor is a component of the Shewanella Knowledgebase (http://shewanella-knowledgebase.org). In addition to curating the groups of orthologous proteins, we have been involved in assigning functions and cellular roles to the gene products. The *Shewanella* ortholog editor has already been valuable in cross genome comparisons (26), and is currently used in several ongoing research projects.

In our work we identified a total of 9,782 non-redundant *Shewanella* proteins that make up the pan genome. To remove the redundant proteins, we chose one representative from each of the ortholog sets. Of the genes making up the pan genome, 22% were found to be present in all of the *Shewanella* genomes encoding the core functions shared by the Shewanella strains. Figure 6 depicts the core Shewanella functions and metabolic pathways. A study of the functions encoded by the core genome suggests that all members of the *Shewanella* genus are able to degrade proteinaceous material. We found the core sequences to encode for degradation pathways for all but two of the amino acids. Further, a number of amino acid and peptide transporters are also shared among the *Shewanellas*. As was observed for MR1 (21), none of the other *Shewanella* genomes contain the phosphofructokinase (Pfk) gene. This suggests that instead of using the glycolytic pathway, these organisms instead degrade sugars via the pentosephosphate pathway (4).

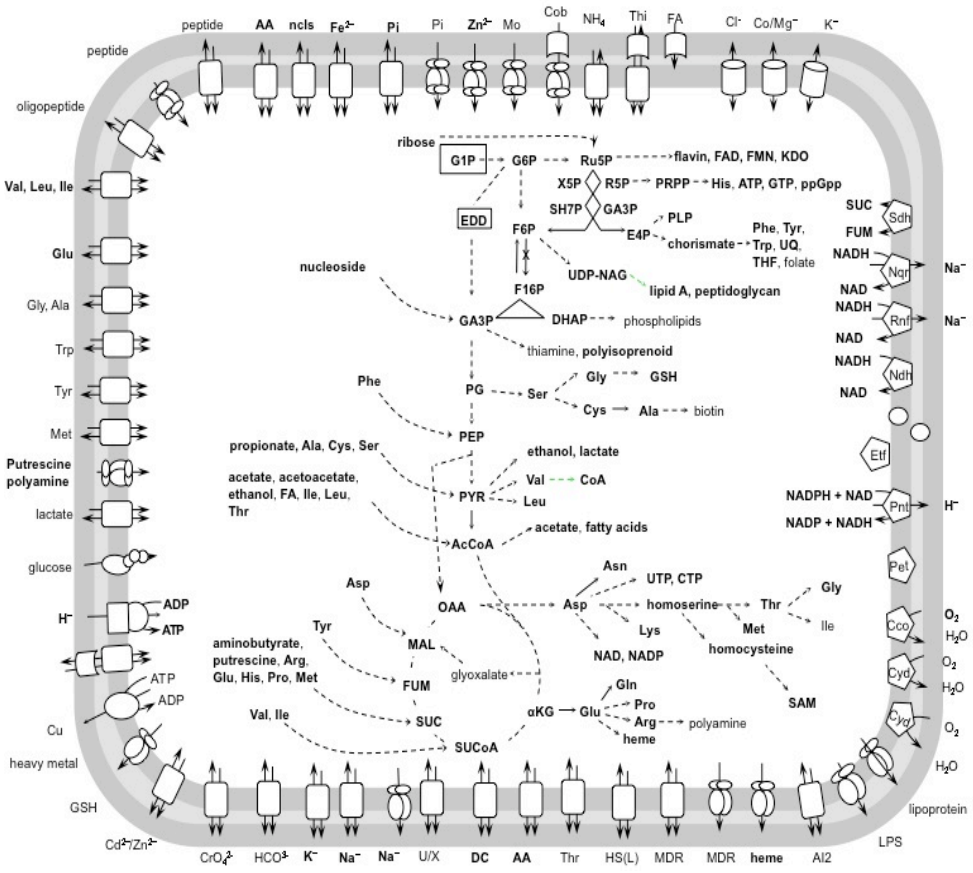

Figure 6. Cellular overview of metabolic pathways that are conserved in *Shewanella*.

Beyond the core gene set, we found that 30% of the *Shewanella* proteins were present in two or more of the strains and that 48% of the proteins were unique to one of the *Shewanella* strains. These two latter groups, representing 78% of the non-redundant proteins, are considered the variable *Shewanella* genome fraction. The variable genome is believed to encode the functions that reflect environmental (niche) adaptation. The *Shewanella*'s have successfully adaptated to a range of aquatic and sedimentary niches. From the variable genome we identified genes specific to closely related *Shewanellas*. For example MR1, ANA-3, MR4 and MR7 (the MR cluster) all encode a selenium-containing formate dehydrogenase as well as a transporter for the import of the molybdenum cofactor required by the dehydrogenase. Proteins found to be unique to each of the genomes included insertion sequences, prophages, outer membrane proteins, TonB-dependent receptors, proteases, restriction-modification enzymes, glycosylases, and polysaccharide biosynthesis enzymes. Some of these strain specific proteins could be linked to metabolic fitness or to interaction with the environment. For example, the strain-specific polysaccharide biosynthesis genes may encode unique surface properties that influence the ability of the particular strain to adapt and survive in its environment. Overall we found it difficult to relate functions in the variable genome to functions

needed in their environment. This can be attributed to lack of knowledge about he environment the strain was isolated from as well as the many strain functions that not yet are annotated (known).

A cross genome study of the ten Shewanella strains has been submitted for publication. As part of the collaborative effort, we analyzed the conservation of various genotypic and phenotypic *Shewanella* traits (features) relative to their sequence divergence (Fig. 7). The aim of the work was to determine whether the different features evolve more or less rapidly over evolutionary time in *Shewanella*. Features compared included the presence or absence of orthologous genes, metabolic pathways, respiratory complexes, gene product types, and aerobic growth phenotypes. According to Figure 7, metabolic pathways and growth phenotypes are less prone to change compared to the presence of orthologs. The least amount of conservation, or largest change, was seen for the cytochromes, reflecting electron transport diversity, and for TonB and MCP, two outer membrane proteins that interact directly with the environment.

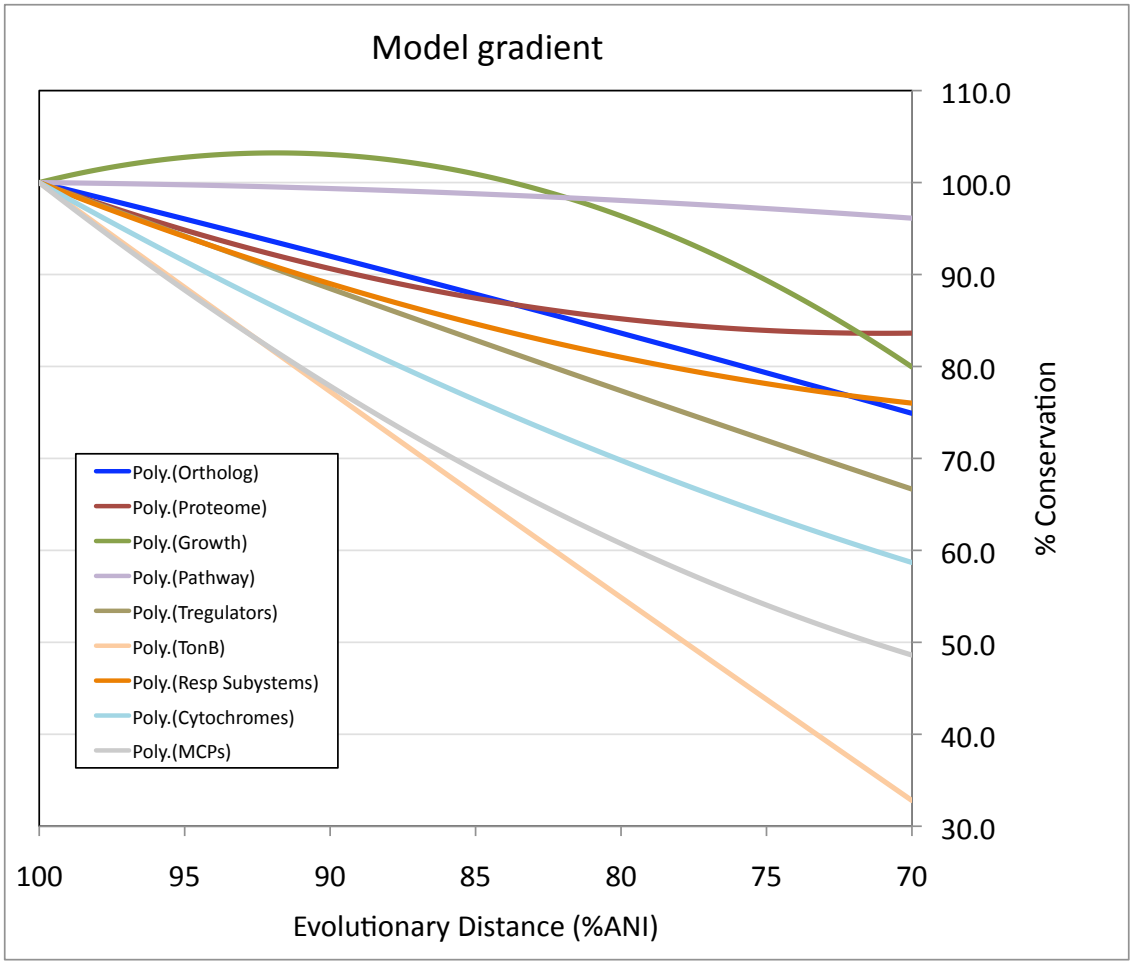

Figure 7. Conservation of genotypic and phenotypic traits in *Shewanella.* Ten *Shewanella* genome sequences were pair-wise compared for selected genotypes and phentoypes. The conservation of the traits were plotted against the evolutionary distance of the genomes.

5.4. Analyzing groups of sequence similar proteins in MR1.

Groups of sequence similar proteins, paralogous proteins, are believed to have arisen through gene duplication and divergence, and hence they are reflective of the evolutionary history of an organism. Although some group members may have arisen through horizontal transfer, we did not attempt to distinguish these in our studies. The largest protein groups found in MR-1 are shown in Table 1 and the complete list of 418 groups can be accessed through ShewCyc.

Table 3. The largest groups of sequence similar proteins in *S. oneidensis* MR-1.

Group Size	Group Function
64	response regulator
52	ABC superfamily transporter, ATPase subunit
51	ISSod4 transposase
49	transcriptional regulator, LysR family
39	ISSod1 transposase
31	sensor histidine kinase
31	diguanylate cyclase
28	methyl-accepting chemotaxis protein
26	secretion protein (HlyD family)
22	cyclic diguanylate phosphodiesterase
19	unknown function
16	ISSod3 transposase
14	oxidoreductase, short chain dehydrogenase/reductase family
14	ABC superfamily transporter, inner membrane subunit
13	ISSod11 transposase
12	ATP-dependent RNA helicase
12	cation efflux protein, AcrB/AcrD/AcrF family
11	NDP-epimerase/isomerase/reductase/dehydratase
11	Sigma54 specific transcriptional regulator
10	flavoprotein

One may propose that an increased abundance of a protein-type in an organism indicates that this protein (or its activity) is of importance to this organism and possibly reflect a functional specialization related to environmental conditions. Based on the list of the largest MR1 protein families, we suggest that regulation and environmental sensing are functions that are important to MR1. One and two component regulators make up some of the largest groups (i.e. LysR family transcriptional regulators, sensor histidine kinases and response regulators). The significance of these types of regulators was also suggested from other MR1 analyses (6). Signaling by cyclic diGMP (c-di-GMP) appears also to be important and highly regulated in MR1. The genome encodes 31 enzymes for synthesis

of c-di-GMP (diguanylate cyclases) and 22 enzymes for its degradation (cyclic diguanylate phosphodiesterase). Thormann et al. showed that c-di-GMP is involved in biofilm stability in MR1 (24). Also, the number of large transposase families reflect the many IS elements in MR1 and the high number of pseudogenes that are found in the genome.

We proposed to do a comparison of sequence similar protein groups in selected genomes. The goal was to detect whether differences in the sizes of comparable groups could be linked to 1) increased diversity of related functions or metabolisms and 2) adaptation to specific environmental conditions. Previous work of ours has suggested such connections (23). A comparison of the largest MR1 protein groups with those present in the other *Shewanella* genomes showed significant variation of group sizes (Table 4). For example while there are 64 response regulators in MR1, Sden only contains half as many. This could reflect that Sden has specialized as a denitrifier living in a narrow zone in the water column, while MR1 has adapted to the lake sediment where it needs to sense a larger range of environmental factors (i.e. various nutrients, electron acceptors). MR1 also contains a higher number of ATPase subunits of ABC transporters relative to the other strains. Some of these are not located adjacent to other transport subunits, suggesting they may have a novel role in MR1. Not surprisingly large variations are also seen for the different transposase families. These families tend to be specific to more closely related strains. Another interesting observation is the large number of cyclic-di-GMP synthesizing diguanylate cyclases seen in all ten genomes. However, the number of diguanylate phosphodiesterases, enzymes that cleave the cyclic-di-GMP signaling molecule, are significantly higher in MR1 (22 proteins) compared to the other genomes (3-8 proteins). This signaling pathway therefore appears to be quite different in MR1 relative to the other strains analyzed.

Table 4. The largest paralogous group in *S. oneidensis* MR-1 and their sequence similar matches in 9 *Shewanella* genomes.

Group function	MR1	ANA3	MR4	MR7	CN32	W3181	PV4	Sfri	Sden	Sama
response regulator	64	43	44	44	38	39	46	39	32	45
ABC transporter, ATPase	52	28	34	35	38	37	38	39	32	33
transcriptional regulator, LysR family	48	49	39	38	39	40	46	49	23	40
transposase, ISSod4/ISSod5	44	0	0	0	0	1	0	0	4	0
transposase orfA, ISSod1	35	8	6	11	7	7	1	8	12	1
transposase orfB, ISSod1	35	0	0	0	7	3	1	0	3	1
diguanylate cyclase	**32**	**29**	**24**	**26**	**22**	**21**	**26**	**29**	**24**	**28**
sensory histidine kinase	**30**	**23**	**20**	**22**	**23**	**22**	**23**	**23**	**15**	**25**
membrane fusion protein	**27**	**32**	**27**	**26**	**29**	**25**	**32**	**32**	**15**	**23**
chemotaxis receptor	**27**	**36**	**33**	**31**	**35**	**37**	**24**	**36**	**19**	**27**
diguanylate phosphodieseterase	**22**	**6**	**5**	**5**	**7**	**8**	**3**	**6**	**5**	**8**
transposase, ISSod15	**17**	**1**	**2**	**8**	**7**	**3**	**0**	**1**	**15**	**1**
transposase, ISSod3	**15**	**0**	**1**	**0**	**0**	**3**	**0**	**0**	**7**	**0**
oxidoreductase, SDR fam.	**14**	**15**	**14**	**13**	**13**	**12**	**18**	**15**	**11**	**12**
ABC transporter, membr.	**14**	**9**	**7**	**9**	**10**	**10**	**6**	**9**	**3**	**6**
transporter	**13**	**21**	**15**	**15**	**22**	**18**	**18**	**21**	**15**	**13**
RNA helicase	**12**	**12**	**12**	**13**	**12**	**12**	**11**	**12**	**12**	**10**
transposase, ISSod11	**12**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
epimearase/dehydratase	**11**	**8**	**8**	**7**	**8**	**6**	**11**	**8**	**9**	**9**
transcriptional regulator	**11**	**17**	**15**	**14**	**15**	**15**	**15**	**17**	**14**	**14**

We also searched for protein families that were absent in only one of the ten genomes. Such cases could indicate functions that no longer are used by the organism it its niche. However, it is also possible that this function is carried out by a non-homologous gene product. Our comparative analysis shows that many of families are missing from the Sden genome (Table 5). Further, many of the functions missing in Sden are related to electron transport. This agrees with the knowledge of Sden's limited respiratory capability.

Table 5. Protein groups missing in S. denittrificans OS217.

Group function	MR1	ANA3	MR4	MR7	CN32	W3181	PV4	Sfri	Sden	Sama
oxidoreductase, FeS subunit	7	9	9	8	8	8	8	5	**0**	4
tetraheme flavocytochrome	7	2	2	3	5	3	4	5	**0**	2
cytochrome b subunit	4	5	5	5	3	3	4	2	**0**	3
reductase subunit	4	3	3	4	3	2	3	2	**0**	2
decahaeme cytochrome c	4	2	3	3	2	1	2	2	**0**	2
Decahaeme cytochrome c	4	4	4	4	3	3	4	3	**0**	5
outer membrane protein	4	2	3	3	2	1	2	3	**0**	2

5.4. Interpreting MR1 expression studies.

We have used our annotation and metabolic pathway assignments to interpret experimental data generated by members of the Shewanella Federation. We analyzed the gene expression pattern during growth with different carbon sources (4). An example of the expression study is shown in Figure 8. The growth studies verified the use of the pentose phosphate pathway during growth with inosine as well as indicated that the membrane proteins SO1214 and SO1215 were involved in the import of inosine.

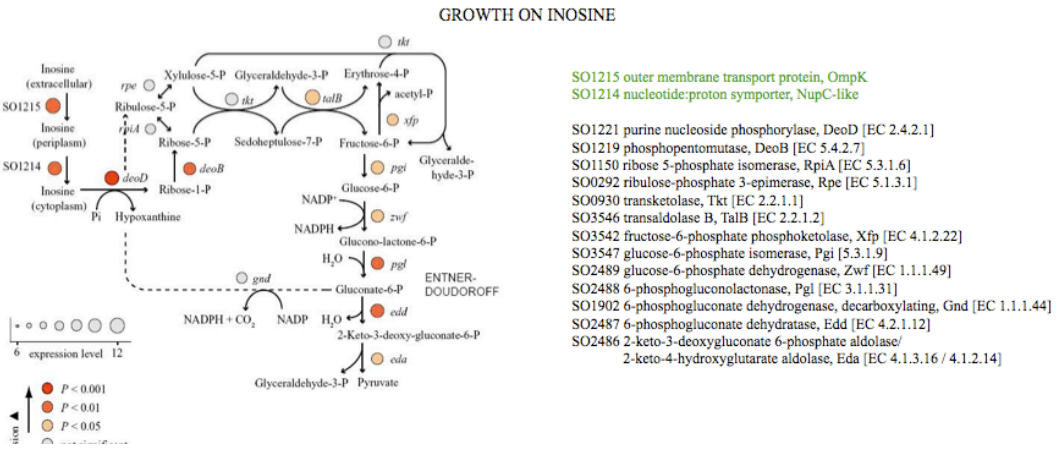

Figure 8. Gene expression pattern for an S. *oneidensis* MR-1 culture grown in the presence of inosine.

Many additional studies are in the process of being analyzed or written up for publication. Some of these are listed below:

Publications that are submitted or near to submission.

1. Inference of a Genome-wide Regulatory Network for the Metal-Reducing Microbe Shewanella oneidensis MR-1. (Driscoll , Serres, et al.)
2. Comparative systems biology across an evolutionary gradien within the *Shewanella* genus. (Konstantinidis, Serres et al.)
3. Linking phenotype and genotype of the *Shewanella* species by their quantification and correlation. (Karpinets, Obraztsova et al.)
4. A functional analysis of the reannotated *S. oneidensis* MR-1 using the ShewCyc pathway genome database. (Serres, Romine et al.)
5. Functional diversity and evolutionary divergence reflected in *Shewanella* protein families. (Metpally, Serres).

REFERENCES

1. **Caspi, R., H. Foerster, C. A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, and P. D. Karp.** 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res **36:**D623-31.
2. **Chourey, K., M. R. Thompson, J. Morrell-Falvey, N. C. Verberkmoes, S. D. Brown, M. Shah, J. Zhou, M. Doktycz, R. L. Hettich, and D. K. Thompson.** 2006. Global molecular and morphological effects of 24-hour chromium(VI) exposure on Shewanella oneidensis MR-1. Appl Environ Microbiol **72:**6331-44.
3. **Chourey, K., W. Wei, X. F. Wan, and D. K. Thompson.** 2008. Transcriptome analysis reveals response regulator SO2426-mediated gene expression in Shewanella oneidensis MR-1 under chromate challenge. BMC Genomics **9:**395.
4. **Driscoll, M. E., M. F. Romine, F. S. Juhn, M. H. Serres, L. A. McCue, A. S. Beliaev, J. K. Fredrickson, and T. S. Gardner.** 2007. Identification of diverse carbon utilization pathways in Shewanella oneidensis MR-1 via expression profiling. Genome Inform **18:**287-98.
5. **Elias, D. A., M. E. Monroe, M. J. Marshall, M. F. Romine, A. S. Belieav, J. K. Fredrickson, G. A. Anderson, R. D. Smith, and M. S. Lipton.** 2005. Global detection and characterization of hypothetical proteins in Shewanella oneidensis MR-1 using LC-MS based proteomics. Proteomics **5:**3120-30.
6. **Fredrickson, J. K., M. F. Romine, A. S. Beliaev, J. M. Auchtung, M. E. Driscoll, T. S. Gardner, K. H. Nealson, A. L. Osterman, G. Pinchuk, J. L. Reed, D. A. Rodionov, J. L. Rodrigues, D. A. Saffarini, M. H. Serres, A. M. Spormann, I. B. Zhulin, and J. M. Tiedje.** 2008. Towards environmental systems biology of Shewanella. Nat Rev Microbiol **6:**592-603.
7. **Garvey, G. S., C. J. Rocco, J. C. Escalante-Semerena, and I. Rayment.** 2007. The three-dimensional crystal structure of the PrpF protein of Shewanella oneidensis complexed with trans-aconitate: insights into its biological function. Protein Sci **16:**1274-84.
8. **Gonnet, G. H., M. T. Hallett, C. Korostensky, and L. Bernardin.** 2000. Darwin v. 2.0: an interpreted computer language for the biosciences. Bioinformatics. **16:**101-103.
9. **Grimek, T. L., and J. C. Escalante-Semerena.** 2004. The acnD genes of Shewenella oneidensis and Vibrio cholerae encode a new Fe/S-dependent 2-methylcitrate dehydratase enzyme that requires prpF function in vivo. J.Bacteriol. **186:**454-462.
10. **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.Biotechnol. **20:**1118-1123.

11. **Kalamorz, F., B. Reichenbach, W. Marz, B. Rak, and B. Gorke.** 2007. Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol.Microbiol. **65:**1518-1533.
12. **Karp, P. D., S. Paley, and P. Romero.** 2002. The Pathway Tools software. Bioinformatics. **18 Suppl 1:S225-32.:**S225-S232.
13. **Kazakov, A. E., D. A. Rodionov, E. Alm, A. P. Arkin, I. Dubchak, and M. S. Gelfand.** 2009. Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. J Bacteriol **191:**52-64.
14. **Kolker, E., A. F. Picone, M. Y. Galperin, M. F. Romine, R. Higdon, K. S. Makarova, N. Kolker, G. A. Anderson, X. Qiu, K. J. Auberry, G. Babnigg, A. S. Beliaev, P. Edlefsen, D. A. Elias, Y. A. Gorby, T. Holzman, J. A. Klappenbach, K. T. Konstantinidis, M. L. Land, M. S. Lipton, L. A. McCue, M. Monroe, L. Pasa-Tolic, G. Pinchuk, S. Purvine, M. H. Serres, S. Tsapin, B. A. Zakrajsek, W. Zhu, J. Zhou, F. W. Larimer, C. E. Lawrence, M. Riley, F. R. Collart, J. R. Yates, 3rd, R. D. Smith, C. S. Giometti, K. H. Nealson, J. K. Fredrickson, and J. M. Tiedje.** 2005. Global profiling of Shewanella oneidensis MR-1: expression of hypothetical genes and improved functional annotations. Proc Natl Acad Sci U S A **102:**2099-104.
15. **Konstantinidis, K. T., A. Ramette, and J. M. Tiedje.** 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl Environ Microbiol **72:**7286-93.
16. **Liu, J., X. Xu, and G. D. Stormo.** 2008. The cis-regulatory map of Shewanella genomes. Nucleic Acids Res **36:**5376-90.
17. **Myers, C. R., and K. H. Nealson.** 1988. Bacterial Manganese Reduction and Growth with Manganese Oxide as the Sole Electron Acceptor. Science **240:**1319-1321.
18. **Peterson, J. D., L. A. Umayam, T. Dickinson, E. K. Hickey, and O. White.** 2001. The Comprehensive Microbial Resource. Nucleic Acids Res. **29:**123-125.
19. **Riley, M., T. Abe, M. B. Arnaud, M. K. Berlyn, F. R. Blattner, R. R. Chaudhuri, J. D. Glasner, T. Horiuchi, I. M. Keseler, T. Kosuge, H. Mori, N. T. Perna, G. Plunkett, 3rd, K. E. Rudd, M. H. Serres, G. H. Thomas, N. R. Thomson, D. Wishart, and B. L. Wanner.** 2006. Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. Nucleic Acids Res **34:**1-9.
20. **Romine, M. F., T. S. Carlson, A. D. Norbeck, L. A. McCue, and M. S. Lipton.** 2008. Identification of mobile elements and pseudogenes in the Shewanella oneidensis MR-1 genome. Appl Environ Microbiol **74:**3257-65.
21. **Serres, M. H., and M. Riley.** 2006. Genomic analysis of carbon source metabolism of Shewanella oneidensis MR-1: Predictions versus experiments. J Bacteriol **188:**4601-9.
22. **Serres, M. H., and M. Riley.** 2000. MultiFun, a multifunctional classification scheme for Escherichia coli K-12 gene products. Microb Comp Genomics **5:**205-22.

23. **Serres, M. H., and M. Riley.** 2004. Structural domains, protein modules, and sequence similarities enrich our understanding of the Shewanella oneidensis MR-1 proteome. OMICS **8:**306-21.
24. **Thormann, K. M., S. Duttler, R. M. Saville, M. Hyodo, S. Shukla, Y. Hayakawa, and A. M. Spormann.** 2006. Control of formation and cellular detachment from Shewanella oneidensis MR-1 biofilms by cyclic di-GMP. J Bacteriol **188:**2681-91.
25. **Wilson, D., M. Madera, C. Vogel, C. Chothia, and J. Gough.** 2007. The SUPERFAMILY database in 2007: families and functions. Nucleic Acids Res. **35:**D308-D313.
26. **Yang, C., D. A. Rodionov, X. Li, O. N. Laikova, M. S. Gelfand, O. P. Zagnitko, M. F. Romine, A. Y. Obraztsova, K. H. Nealson, and A. L. Osterman.** 2006. Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of Shewanella oneidensis. J.Biol.Chem. **.**

6. Products developed under the award:

Publications:

Serres MH, Riley M. 2006. Genomic analysis of carbon source metabolism of Shewanella oneidensis MR-1: Predictions versus experiments. J. Bacteriol. 188(13):4601-9.

Driscoll ME, Romine MF, Juhn FS, Serres MH, McCue LA, Beliaev AS, Fredrickson, JK, Gardner TS. 2007. Identification of diverse carbon utilization pathways in Shewanella oneidensis MR-1 via expression profiling. Genome Inform. 18:287-98.

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, Zhyulin IB, Tiedje JM. 2008. Nat Rev Microbiol. 6(8):592-603.

Oral presentations:

02/13/2006 Presentation at Genomics:GTL Awardee Workshop IV, North Bethesda, MD
02/ 15/2007 Presentation at Shewanella Federation Meeting, Bethesda, MD
02/06/2008 Lecture at Medical Genetics Course for Physicians, Falmouth, MA
02/10/2008 Presentation at Shewanella Federation Meeting, Bethesda, MD
06/04/2008 Presentation at ASM 108th General Meeting, Division R Symposium, Boston, MA.
10/03/2008 Seminar at SRI, Menlo Park, CA.
10/10/2008 Presentation at Shewanella Federation Meeting, Asilomar, CA.

Posters:

Feb 2006 Genomics:GTL Awardee Workshop IV
Feb 2007 Genomics:GTL Awardee Workshop V
Apr 2007 ASM Integrating Metabolism and Genomics Meeting
Feb 2008 Genomics:GTL Awardee Workshop VI.
Sep 2008 4th International Conference of the Brazilian Association for Bioinformatics and Computational Biology, X-Meeting, Bahia, Brazil
Sep 2008 54th Brazilian Meeting of Genetics. Salvador, Bahia, Brazil

Web accessible database:

The metabolic database ShewCyc hosted at MBL: http://pathways.mbl.edu/