

# **Integrated Genome-Based Studies of Shewanella Ecophysiology**

## **Final Scientific/Technical REPORT**

Reporting Period Start Date: 01/01/2007

Reporting Period End Date: 03/31/2010

Principal Author: ANDREI OSTERMAN

Date Report was Issued: DECEMBER 2012

DOE Award Number: DE-FG02-07ER64384

Submitting Organization : SANFORD-BURNHAM MEDICAL RESEARCH INSTITUTE

10901 North Torrey Pines Road, La Jolla, 92037 CA

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## ABSTRACT

Integration of bioinformatics and experimental techniques was applied to mapping and characterization of the key components (pathways, enzymes, transporters, regulators) of the core metabolic machinery in *Shewanella oneidensis* and related species with main focus was on metabolic and regulatory pathways involved in utilization of various carbon and energy sources. Among the main accomplishments reflected in ten joint publications with other participants of Shewanella Federation are: (i) A systems-level reconstruction of carbohydrate utilization pathways in the genus of *Shewanella* (19 species). This analysis yielded reconstruction of 18 sugar utilization pathways including 10 novel pathway variants and prediction of > 60 novel protein families of enzymes, transporters and regulators involved in these pathways. Selected functional predictions were verified by focused biochemical and genetic experiments. Observed growth phenotypes were consistent with bioinformatic predictions providing strong validation of the technology and (ii) Global genomic reconstruction of transcriptional regulons in 16 *Shewanella* genomes. The inferred regulatory network includes 82 transcription factors, 8 riboswitches and 6 translational attenuators. Of those, 45 regulons were inferred directly from the genome context analysis, whereas others were propagated from previously characterized regulons in other species. Selected regulatory predictions were experimentally tested. Integration of this analysis with microarray data revealed overall consistency and provided additional layer of interactions between regulons. All the results were captured in the new database RegPrecise, which is a joint development with the LBNL team. A more detailed analysis of the individual subsystems, pathways and regulons in *Shewanella* spp included bioinformatics-based prediction and experimental characterization of: (i) N-Acetylglucosamine catabolic pathway; (ii) Lactate utilization machinery; (iii) Novel NrtR regulator of NAD biosynthesis; (iv) HexR-controlled global regulon in central metabolism. In addition to numerous specific findings contributing to basic understanding of ecophysiology and evolution of *Shewanella*, the key components of the integrative genomic methodology of general utility for the community were optimized, validated and disseminated.

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY</b>	<b>4</b>
<b>REPORT DETAILS</b>	<b>6</b>
<b>A. EXPERIMENTAL METHODS</b>	<b>6</b>
A1. The subsystems-based approach to pathway reconstruction	
A2. Reconstruction of regulons	
A3. Pathway reconstruction workflow	
<b>B. RESULTS AND DISCUSSIONS</b>	<b>8</b>
B.1 Genomic encyclopedia of sugar utilization pathways in the <i>Shewanella</i> genus	
B2. Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the <i>Shewanella</i> genus	
B3. Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator. A Case Study in <i>S. oneidensis</i>	
B4. Genomic Reconstruction of <i>Shewanella oneidensis</i> MR-1 Metabolism Reveals a Novel Machinery for Lactate Utilization	
<b>C. CONCLUSION</b>	<b>13</b>
<b>REFERENCES</b>	<b>14</b>

## EXECUTIVE SUMMARY

In this project we combined bioinformatics and experimental techniques to map and characterize the key components (pathways, enzymes, transporters, regulators) of the core metabolic machinery in *Shewanella oneidensis* and related species in tight collaboration with other members of Shewanella Federation. Our main focus was on metabolic and regulatory pathways involved in utilization of various carbon and energy sources. The project was structured around the following specific aims:

**Aim 1. Genomic reconstruction and experimental assessment of catabolic pathways in *S. oneidensis* and other *Shewanella* spp with completely sequenced genomes.**

**Aim 2. Genome-scale reconstruction and modeling of *S. oneidensis* metabolic network.**

**Aim 3. Genomic reconstruction of transcriptional regulatory networks in *Shewanella* genus.**

The key objectives formulated above were successfully met yielding the following major accomplishments:

**Aim 1.** We have established a subsystems-based comparative approach to genomic reconstruction of sugar catabolism and applied it to a group of 19 different *Shewanella* strains with completely sequenced genomes. The key stages of our approach include: (i) a homology-based identification of gene candidates using a genomic compilation of ~500 known components of sugar catabolic pathways; (ii) functional assignment of orthologs and prediction of alternative genes and pathway variants based on genomic (operons, regulons) and functional (subsystems, pathways) context analyses; (iii) validation of bioinformatic predictions by a combination of biochemical, genetic and physiological experiments. The obtained genomic encyclopedia of sugar utilization includes ~170 protein families (mostly metabolic enzymes, transporters and transcriptional regulators) spanning 17 distinct pathways with a mosaic distribution across the *Shewanella* genus. The analysis of this distribution and comparison with other groups of bacteria provided insights into their ecophysiology and adaptive evolution. Remarkably, ~1/3 of identified genes and ~ 2/3 of reconstructed pathways represent previously unknown variants (nonorthologous gene replacements and alternative biochemical routes). Selected bioinformatic predictions as well as most of the predicted growth phenotypes were experimentally verified.

Publications: The results were reported at DOE-GTL meetings. Three articles reporting: (i) GlcNAC pathway<sup>16</sup>; (ii) providing a brief overview of all results<sup>2</sup>; and (iii) reporting the entire study<sup>15</sup> were published.

**Aim 2.** The SEED platform was used to cover a substantial fraction of *Shewanella* metabolism by subsystems (beyond sugar catabolism). The current collection of SEED subsystems covers ~1,600 genes of *S. oneidensis* (among them ~ 120 well-developed metabolic subsystems with connected reactions and diagrams). Our SEED-based approach was used to predict and experimentally characterize lactate utilization machinery of *Shewanella* spp. The collection of SEED subsystems was used to reconcile, improve and fill-in the gaps in the current metabolic model of *S. oneidensis*. Another contribution to this study by Osterman's group was the experimental analysis of the biomass composition.

Publications: The results were reported at DOE-GTL meetings. Two joint articles with other members of *Shewanella* Federation reporting lactate utilization gene discovery<sup>12</sup> and metabolic modeling<sup>11</sup> were published.

**Aim 3.** Using comparative genomics approach we were able to reconstruct a substantial fraction (~ 1/3) of transcriptional regulatory networks in *Shewanella oneidensis* MR-1 and 12 other species of *Shewanella* with sequenced genomes. We have identified candidate transcription factors binding sites (TFBS) for ~80 transcription factors (TF) and performed a detailed

reconstructions of respective regulons. The entire set of reconstructed regulons now includes on average ~ 1,000 genes per genome organized in ~ 250 operon units and includes the key pathways involved in central metabolism, production of energy and biomass, metal ion homeostasis and stress response. The results of this analysis were made available to the community via a newly developed web-site RegPrecise, which is our collaborative development with Dr. P. Novichkov and I. Dubchak at LBNL. Several novel TFs and their regulons were selected for experimental characterization. Among them, a discovered NrtR repressor was shown to be involved in transcriptional regulation of genes involved in NAD biosynthesis. 3D structural analysis of NrtR from *S. oneidensis* (collaboration with UT Southwestern) revealed molecular basis for DNA-binding and effector specificity of this regulator composed of HTH- and ADP-ribose pyrophosphatase-like domains. Experimental characterization of NagR regulon included gel-shift analysis and qPCR of the *nagR* KO mutant. In the most recent studies we focused on the detailed characterization of a global regulator HexR, which was predicted to control the expression of ~30 genes in central carbon metabolism (gluconeogenesis, pyruvate metabolism, deoxynucleoside and glycine utilization pathways). These predictions were confirmed in vitro (gel-shift analysis) and in vivo (by qPCR of the *hexR* KO mutant). The latter experiment also confirmed the predicted dual mode of HexR action, negative regulation (repression) for some of the target genes, and positive regulation (activation) for others. Comparison of growth phenotypes of delta-*hexR* and wild type strains on various carbon sources (N-acetylglucosamine, glycerate, inosine, and lactate) showed that *hexR* deletion leads to inability of *S. oneidensis* to grow on lactate as a single carbon source. This finding confirmed the observed positive mode of action of HexR regulator on the gluconeogenic gene *ppsA*, which is known to be essential for the growth of *E. coli* on lactate. Additional physiological studies and metabolomic profiling analyses were performed to support this interpretation and further explore the role of HexR in the regulation of CCM in *Shewanella*.

Publications: The results were reported at DOE-GTL meetings. A brief overview of reconstructed regulatory networks in *Shewanella* was included in<sup>2</sup>. Bioinformatics and biochemical analysis of NrtR regulon<sup>13</sup> and, later, 3D the structural analysis<sup>4</sup> were published. The results of global regulatory network analysis<sup>14</sup> and the integrative analysis of HexR regulon<sup>5</sup> were published. The RegPrecise web-site is publicly available (<http://regprecise.lbl.gov>) and was described<sup>7</sup>.

## REPORT DETAILS

### A. EXPERIMENTAL METHODS

Here we briefly describe the most central bioinformatics components of our methodology as applied to the metabolic and regulatory reconstruction of sugar catabolic machinery of *Shewanella*<sup>15</sup>. Experimental methods used to test some of the key functional inferences are provided in respective publications reflecting our work on this project (e.g.<sup>12,15,16</sup>).

#### ***A1. The subsystems-based approach to pathway reconstruction***

Our approach to the reconstruction of sugar catabolic pathways in a selected group of genomes was based on functional gene annotation and prediction using two principal comparative genomics techniques: (i) homology-based methods and (ii) genome context analysis. Both these methods are implemented in the SEED genomic platform (<http://theseed.uchicago.edu/FIG/>) that combines a large and rapidly growing integration of >700 complete annotated genomes (mostly bacterial) with advanced tools for comparative analysis, gene annotation, genome context analysis and functional reconstruction based on subsystems technology<sup>9</sup>. New genomes are automatically annotated by the RAST server

(<http://rast.nmpdr.org/>), a new generation of subsystem-based genome annotation tools<sup>1</sup>. Subsystems in the SEED provide the framework for further improvement of these annotations and functional predictions. They are sets of functional roles that capture the current knowledge of cellular processes and metabolic pathways including interspecific variation. Each functional role is typically associated with a set of homologous genes that implement this role in specific organisms. In addition to homology-based analysis suggesting at least general class gene functional assignments, genome context analysis provides evidence of functional coupling between genes of known and unknown functions<sup>8,10</sup>. The most common type of functional coupling evidence comes from the tendency of functionally related genes (e.g., members of the same pathway) to be clustered on the chromosome. Other important types of evidence are domain fusion events, conservation of upstream regulatory sites (i.e., regulons) and co-occurrence profiles of genes across a range of genomes. We used the tools in SEED and other public servers to compute and analyze all types of functional coupling evidence for each gene family in our analysis.

## **A2. Reconstruction of regulons**

For identification of a candidate regulatory motif for a particular sugar catabolic pathway we started from a training set of potentially co-regulated genes participating in the pathway. Upstream regions of genes from the training set and their orthologs from multiple *Shewanella* genomes were used as an input for a DNA motif detection algorithm. A simple iterative procedure implemented in the program SignalX was used for construction of a common transcription factor-binding motif in sets of upstream gene fragments. Each genome encoding the studied transcription factor was scanned with the constructed profile using the GenomeExplorer software, and genes with candidate regulatory sites in the upstream regions were selected<sup>3,6</sup>. The threshold for the site search was defined as the lowest score observed in the training set. Among new candidate members of a regulon, only genes having candidate sites conserved in at least two other genomes were retained for further analysis. We also included new candidate regulon members that are functionally related to the reconstructed sugar catabolic pathways. Sequence logos for derived regulatory motifs were drawn using the WebLogo package (<http://weblogo.berkeley.edu>). The details of reconstructed regulons are captured and displayed in the specialized database RegPrecise (<http://regprecise.lbl.gov>)<sup>7</sup>.

## **A3. Pathway reconstruction workflow**

Reconstruction of metabolic and regulatory pathways involved in the carbohydrates utilization was performed for 19 species of the *Shewanella* genus with completely sequenced genomes uploaded from Genbank and integrated in the SEED genomic platform. First we performed a survey of all prokaryotic genes known to be involved or potentially involved in utilization of mono- and di-saccharides. A collection of ~480 FIGfams from 35 SEED sugar metabolic subsystems were classified by their general functional role (i.e. sugar transport, transcriptional regulation, biochemical transformation, and upstream/auxiliary). Each FIGfam comprises a functionally uniform group of orthologous proteins in related organisms. This extensive collection was then used for homology searches against 19 *Shewanella* genomes resulting in identification of numerous FIGfams potentially implicated in sugar metabolism. Manual curation of the identified *Shewanella* FIGfams using the SEED and other genomic resources and tools (see below) rejected many of them as well as identified some additional candidate FIGfams based on genome context analysis. As a result of this iterative process we have identified ~170 FIGfams present in at least one *Shewanella* genome and tentatively assigned a role in utilization of a particular sugar substrate. The identified FIGfams were used for metabolic reconstruction of *Shewanella* sugar catabolic pathways using the subsystems-based approach.

The refined functional annotations were combined in the aggregated SEED subsystem 'Sugar catabolome in Shewanella species'. Besides SEED, we routinely use other bioinformatic tools and databases featuring: genomes (Genbank), gene annotations (UniProt, IMG), primary literature (PubMed), reactions and pathways (KEGG, BioCyc), conserved domains and motifs (COG, PFAM, ProDom), distant homology searches and alignments (PsiBlast, FFAS, T-Coffee), genome context and occurrence profiles (STRING, Microbes on Line), transcriptional regulation (RegTransBase, RegulonDB), protein localization prediction (TMPRED, SignalP).

## B. RESULTS AND DISCUSSIONS

The main developed methods and results *obtained in the project are reflected in a series of publications listed below:*

Leyn, S. A., X. Li, Q. Zheng, P. S. Novichkov, S. Reed, M. F. Romine, J. K. Fredrickson, C. Yang, **A. L. Osterman**, and D. A. Rodionov. 2011. Control of proteobacterial central carbon metabolism by the HexR transcriptional regulator: a case study in *Shewanella oneidensis*. *The Journal of biological chemistry* 286:35782-94. PMID:21849503

Rodionov, D. A., P. S. Novichkov, E. D. Stavrovskaya, I. A. Rodionova, X. Li, M. D. Kazanov, D. A. Ravcheev, A. V. Gerasimova, A. E. Kazakov, G. Y. Kovaleva, E. A. Permina, O. N. Laikova, R. Overbeek, M. F. Romine, J. K. Fredrickson, A. P. Arkin, I. Dubchak, **A. L. Osterman**, and M. S. Gelfand. 2011. Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics* 12 Suppl 1:S3. PMID:21810205

Pinchuk, G. E., E. A. Hill, O. V. Geydebrekht, J. De Ingeniis, X. Zhang, **A. Osterman**, J. H. Scott, S. B. Reed, M. F. Romine, A. E. Konopka, A. S. Beliaev, J. K. Fredrickson, and J. L. Reed. 2010. Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: a tool for data analysis and hypothesis generation. *PLoS Comput Biol* 6:e1000822. PMID:20589080

Pinchuk, G. E., D. A. Rodionov, C. Yang, X. Li, **A. L. Osterman**, E. Dervyn, O. V. Geydebrekht, S. B. Reed, M. F. Romine, F. R. Collart, J. H. Scott, J. K. Fredrickson, and A. S. Beliaev. 2009. Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* 106:2874-9. PMID:19196979

Rodionov, D. A., J. De Ingeniis, C. Mancini, F. Cimadamore, H. Zhang, **A. L. Osterman**, and N. Raffaelli. 2008. Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. *Nucleic Acids Res* 36:2047-59. PMID:18276643

Rodionov, D. A., C. Yang, X. Li, I. A. Rodionova, Y. Wang, A. Y. Obratzsova, O. P. Zagnitko, R. Overbeek, M. F. Romine, S. Reed, J. K. Fredrickson, K. H. Nealson, and **A. L. Osterman**. 2010. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 11:494. PMID:20836887

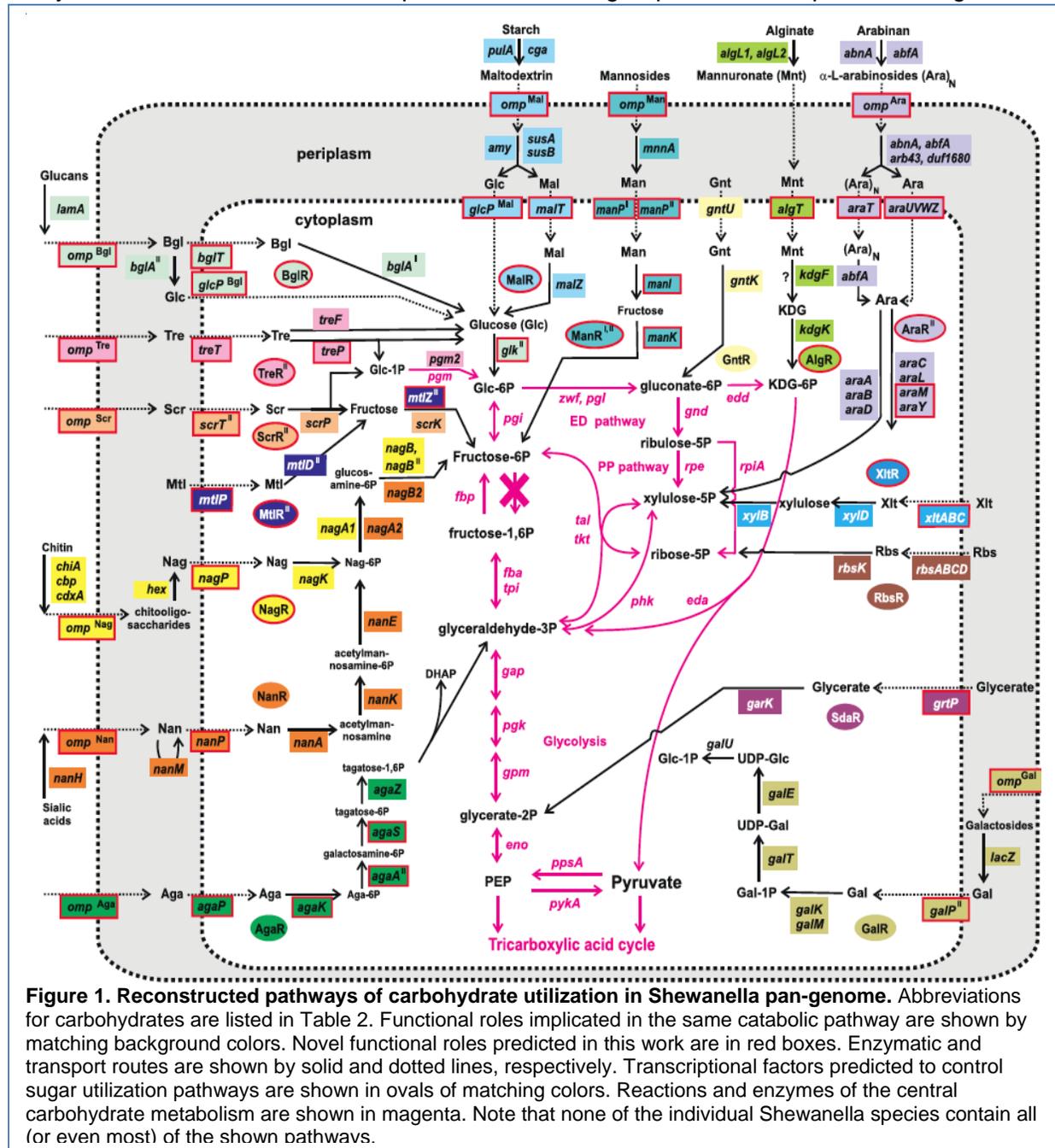
Fredrickson, J. K., M. F. Romine, A. S. Beliaev, J. M. Auchtung, M. E. Driscoll, T. S. Gardner, K. H. Nealson, **A. L. Osterman**, G. Pinchuk, J. L. Reed, D. A. Rodionov, J. L. Rodrigues, D. A. Saffarini, M. H. Serres, A. M. Spormann, I. B. Zhulin, and J. M. Tiedje. 2008. Towards environmental systems biology of *Shewanella*. *Nat Rev Microbiol* 6:592-603. PMID:18604222

*In the following subsection, we briefly outline and discuss the most important and recently published results.*

### **B.1 Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus<sup>15</sup>**

To address a practically and fundamentally important challenge of reconstruction of carbohydrate utilization machinery in any microorganism directly from its genomic sequence, we have established a subsystems-based comparative approach and applied it to a group of 19 different *Shewanella* strains with completely sequenced genomes. The key stages of our approach include: (i) a homology-based identification of gene candidates using a genomic compilation of ~500 known components of sugar catabolic pathways; (ii) functional assignment

of orthologs and prediction of alternative genes and pathway variants based on genomic (operons, regulons) and functional (subsystems, pathways) context analyses; (iii) validation of bioinformatic predictions by a combination of biochemical, genetic and physiological experiments. The obtained genomic encyclopedia of sugar utilization (Fig. 1) includes ~170 protein families (mostly metabolic enzymes, transporters and transcriptional regulators) spanning 17 distinct pathways with a mosaic distribution across the *Shewanella* genus. The analysis of this distribution and comparison with other groups of bacteria provided insights into

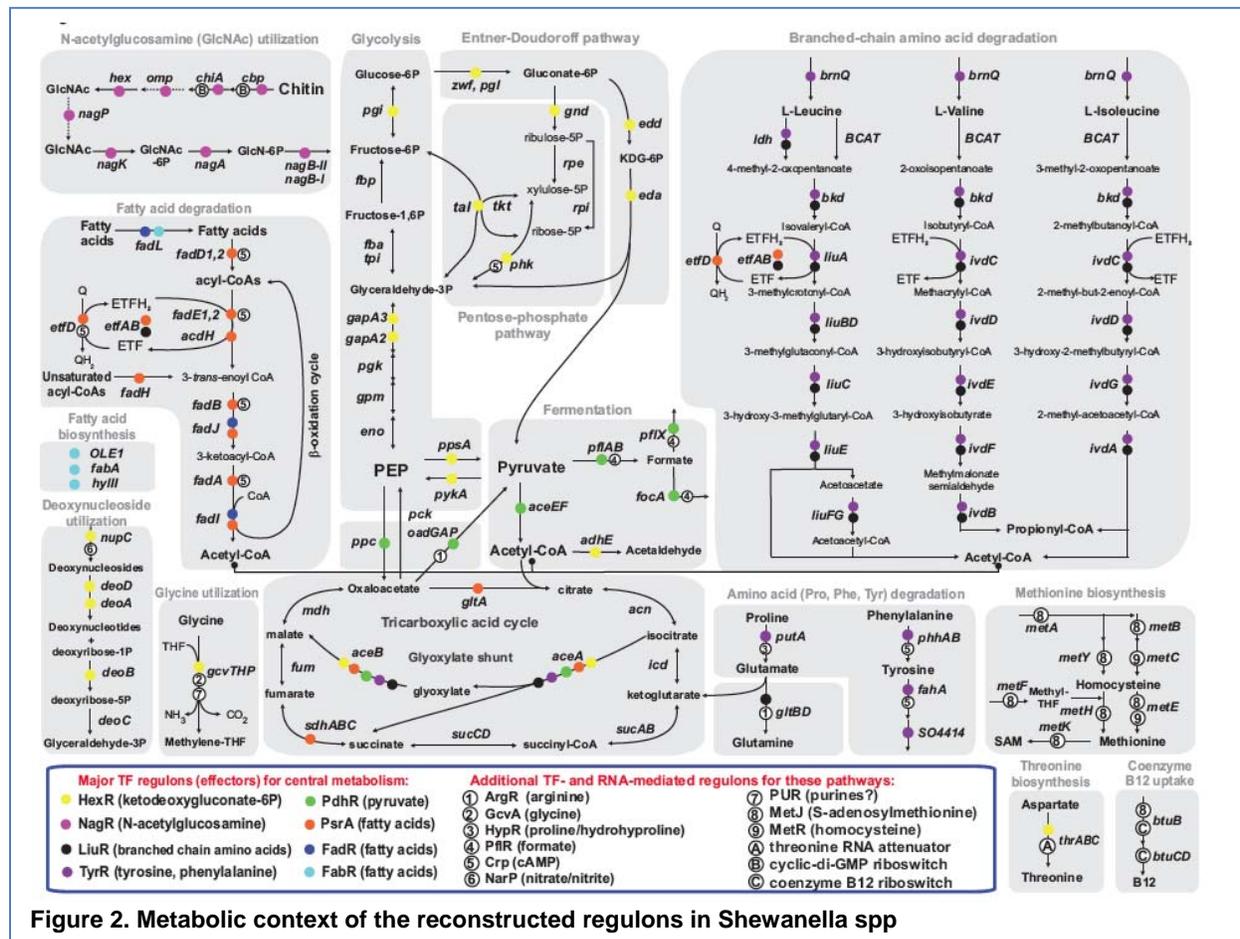


their ecophysiology and adaptive evolution. Remarkably, ~1/3 of identified genes and ~ 2/3 of reconstructed pathways represent previously unknown variants (nonorthologous gene replacements and alternative biochemical routes). Some of these bioinformatic predictions as well as most of the predicted growth phenotypes were experimentally verified. The proposed

approach is scalable and applicable to any group of microbes. A fine-grain functional annotation of all components of the carbohydrate utilization machinery in complete genomes will enable accurate recognition of corresponding functions (and pathway variants) in metagenomic samples.

## B2. Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus<sup>14</sup>

Genome-scale prediction of gene regulation and reconstruction of transcriptional regulatory networks in bacteria is one of the critical tasks of modern genomics. The *Shewanella* genus is comprised of metabolically versatile gamma-proteobacteria, whose lifestyles and natural environments are substantially different from *Escherichia coli* and other model bacterial species. The comparative genomics approaches and computational identification of regulatory sites are useful for the in silico reconstruction of transcriptional regulatory networks in bacteria. To explore conservation and variations in the *Shewanella* transcriptional networks we analyzed the repertoire of transcription factors and performed genomics-based reconstruction and

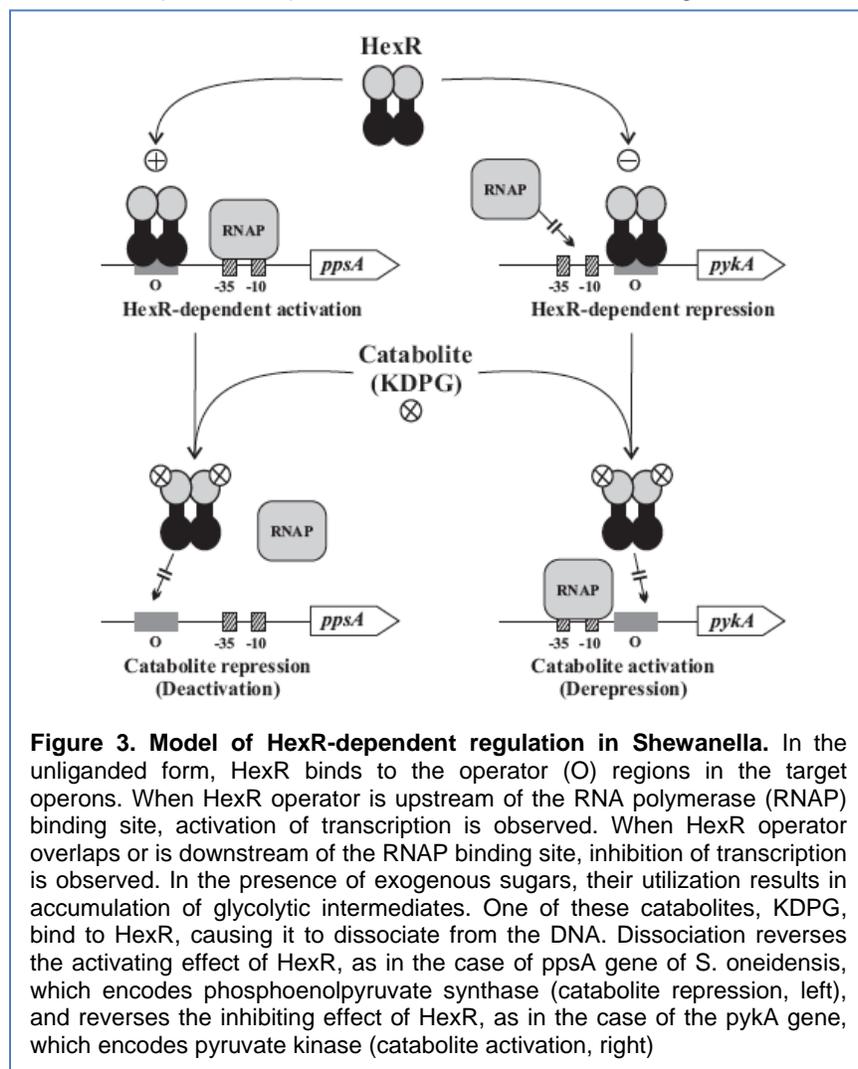


comparative analysis of regulons in 16 *Shewanella* genomes (Fig. 2). The inferred regulatory network includes 82 transcription factors and their DNA binding sites, 8 riboswitches and 6 translational attenuators. Forty five regulons were newly inferred from the genome context analysis, whereas others were propagated from previously characterized regulons in the Enterobacteria and *Pseudomonas* spp. Multiple variations in regulatory strategies between the *Shewanella* spp. and *E. coli* include regulon contraction and expansion (as in the case of PdhR, HexR, FadR), numerous cases of recruiting non-orthologous regulators to control equivalent

pathways (e.g. PsrA for fatty acid degradation) and, conversely, orthologous regulators to control distinct pathways (e.g. TyrR, ArgR, Crp). Thus, we tentatively defined the first reference collection of ~100 transcriptional regulons in 16 *Shewanella* genomes. The resulting regulatory network contains ~600 regulated genes per genome that are mostly involved in metabolism of carbohydrates, amino acids, fatty acids, vitamins, metals, and stress responses. Several reconstructed regulons including NagR for N-acetylglucosamine catabolism were experimentally validated in *S. oneidensis* MR-1. Analysis of correlations in gene expression patterns helps to interpret the reconstructed regulatory network. The inferred regulatory interactions will provide an additional regulatory constraints for an integrated model of metabolism and regulation in *S. oneidensis* MR-1.

### B3. Control of Proteobacterial Central Carbon Metabolism by the HexR Transcriptional Regulator. A Case Study in *Shewanella oneidensis*<sup>5</sup>

Bacteria exploit multiple mechanisms for controlling carbon metabolism. Previously, the RpiR



family HexR transcriptional factor has been implicated in regulation of glucose metabolism genes in *Pseudomonas putida*. Here, we used a comparative genomics approach to reconstruct transcriptional regulons controlled by HexR orthologs in 87 species of Proteobacteria. The inferred regulons include from 1-2 target operons in Enterobacteriales and up to 20 operons in Aeromonadales per genome. Overall, HexR regulates various genes from the central carbohydrate metabolism, although the gene content and cognate DNA motifs of these regulons vary between the 11 studied proteobacterial lineages. The predicted HexR regulon in *Shewanella oneidensis* was experimentally validated using *in vitro* and *in vivo* approaches. In

*Shewanella*, the HexR-binding motif is a 17-bp pseudo-palindrome with a consensus tGTAAATwwwATTACa, and the regulon includes over 30 genes organized in 15 putative operons. Electrophoretic mobility shift and fluorescence polarization assays, conducted with purified HexR protein, confirmed

recognition of all predicted binding sites and the negative effect of 2-keto-3-deoxy-6-phosphogluconate on the DNA-regulator complex formation. A dual mode of HexR action on various target promoters - repression of genes involved in catabolic pathways and activation of gluconeogenic genes – was revealed by integrated genomic and expression analysis in *S. oneidensis* using a mutant lacking *hexR*. Phenotypic profiling revealed the inability of *hexR* mutant to grow on lactate or pyruvate as a single carbon source. Flux analysis with <sup>13</sup>C-labeled lactate revealed a depressed flux to phosphoenolpyruvate in *hexR* mutant confirming that HexR operates *in vivo* as an essential activator of phosphoenolpyruvate synthase. We conclude that HexR modulates the direction of carbon flow by transcriptional activation of some gluconeogenic genes and by repression of the central glycolytic genes (Fig. 3).

#### B4. Genomic Reconstruction of *Shewanella oneidensis* MR-1 Metabolism Reveals a Novel Machinery for Lactate Utilization<sup>2</sup>

The ability to utilize lactate as a sole source of carbon and energy is one of the key metabolic signatures of *Shewanellae*, a diverse group of dissimilatory metal reducing bacteria commonly

Organism	L-LDH	Predicted L-LDH			d-LDH	Predicted d-LDH	Lactate permease	Regulators
	LldD	LldE	LldF	LldG	Dld	Dld-I	LldP	LldR
<b>Gamma-proteobacteria (90)</b>	44	30	30	30	29	28	60	31
<i>Shewanella oneidensis</i> MR-1	-	SO1520	SO1519	SO1518	-	SO1521	SO1522	SO3460 (R2)
<i>Escherichia coli</i> K12	LldD	YkgE	YkgF	YkgG	Dld	-	LldP	LldR (R1)
<i>Colwellia psychrelythraea</i>	+	+	+	+	-	+	+	R5
<i>Pseudomonas fluorescens</i>	-	+	+	+	-	+	+	R1
<b>Alpha-proteobacteria (60)</b>	37	7	7	7	5	3	9	1
<i>Roseobacter denitrificans</i>	+	+	+	+	+	-	-	-
<i>Rhodospirillum rubrum</i>	-	+	+	+	-	+	+	R1
<b>Beta-proteobacteria (32)</b>	21	21	21	21	11	2	24	19
<i>Neisseria meningitidis</i> MC58	+	+	+	+	+	-	+	R1
<i>Dechloromonas aromatica</i> RCB	-	+	+	+	-	+	+	R1
<b>Delta-proteobacteria (11)</b>	0	7	7	7	0	3	5	0
<i>Desulfobivrio vulgaris</i>	-	+	+	+	-	+	+	-
<b>Epsilon-proteobacteria (9)</b>	0	9	9	9	0	7	9	0
<i>Helicobacter pylori</i>	-	+	+	+	-	+	+	-
<b>Actinobacteria (27)</b>	18	9	9	9	3	3	7	8
<i>Propionibacterium acnes</i>	-	+	+	+	-	+	+	R4
<b>Bacillus / Clostridium (50)</b>	0	16	16	16	0	1	26	12
<i>Bacillus subtilis</i>	-	+	+	+	-	-	+	R3
<b>Bacteroidetes / Flavobacteria (11)</b>	3	3	3	3	0	1	3	0
<i>Bacteroides fragilis</i>	-	+	+	+	-	+	+	-
<b>Thermus / Deinococcus (3)</b>	0	3	3	3	0	0	1	1
<i>Deinococcus radiodurans</i>	-	+	+	+	-	-	+	R5
<b>Total:</b>	<b>123</b>	<b>105</b>	<b>105</b>	<b>105</b>	<b>48</b>	<b>48</b>	<b>144</b>	<b>72</b>

**Figure 4. Occurrence and features of genes involved in lactate utilization in representative bacterial genomes.** Representative species in several taxonomic groups of bacteria are shown as rows and the number of genomes analyzed within a group is given in parentheses. The presence or absence of genes for the respective functional roles (columns) is shown by “+” or “-”. For *E. coli* K12 and *S. oneidensis* MR-1, the gene names are indicated instead of “+”. Numbers for taxonomic group rows indicate the number of species that have a gene ortholog. Genes clustered on the chromosome (e.g., operons) are outlined by matching background colors. The genes corresponding to the lactate-specific regulators are R1 (orthologs of known LldR *E. coli* regulator), R2, R3, R4, and R5 (novel predicted regulators). Genes predicted to be regulated by one of these lactate regulators are marked in red.

found in aquatic and sedimentary environments. Nonetheless, homology searches failed to recognize orthologs of previously described bacterial d- or l-lactate oxidizing enzymes (*Escherichia coli* genes *dld* and *lldD*) in any of the 13 analyzed genomes of *Shewanella* spp. Using comparative genomic techniques, we identified a conserved chromosomal gene cluster in *Shewanella oneidensis* MR-1 (locus tag: SO1522-SO1518) containing lactate permease and candidate genes for both d- and l-lactate dehydrogenase enzymes. The predicted d-LDH gene

(*dld-II*, SO1521) is a distant homolog of FAD-dependent lactate dehydrogenase from yeast, whereas the predicted I-LDH is encoded by three genes with previously unknown functions (*lldEGF*, SO1520-19-18). Through a combination of genetic and biochemical techniques, we experimentally confirmed the predicted physiological role of these novel genes in *S. oneidensis* MR-1 and carried out successful functional validation studies in *Escherichia coli* and *Bacillus subtilis*. We conclusively showed that *dld-II* and *lldEGF* encode fully functional d- and I-LDH enzymes, which catalyze the oxidation of the respective lactate stereoisomers to pyruvate. Notably, the *S. oneidensis* MR-1 *lldEGF* enzyme is the first described example of a multi-subunit lactate oxidase. Comparative analysis of >400 bacterial species revealed the presence of *lldEGF* and *Dld-II* in a broad range of diverse species accentuating the potential importance of these previously unknown proteins in microbial metabolism.

### C. CONCLUSION

Advances in genome sequencing revolutionized our ability to model metabolism of any species as long as it has a completely sequenced genome, including those that were only marginally characterized by traditional methods. *Metabolic reconstruction technology* strongly impacts fundamental understanding of cellular organisms providing a framework for evolutionary analysis, predictive modeling of interactions of species with the environment and communities, etc. Systems-level understanding, modeling and manipulation of metabolic and associated regulatory networks drive multiple bioengineering and biomedical applications. However, presently, our capability to infer metabolic and regulatory networks from genomes is impaired by insufficient quality and coverage of our genomic *parts catalog* beyond a handful of model species. Despite the rapid progress of genomic sequencing and analysis tools, functional assignments of millions of genes in public databases are still unknown, incomplete or imprecise.

The integrative genomics-based analysis of *microbial machinery of carbohydrate catabolism* established and applied in this project to *Shewanella* spp provided a vivid illustration of both, challenges and opportunities in this direction of research. Indeed, sugar utilization pathways are major *feed lines* of carbon and energy for central metabolism in a large variety of heterotrophic bacteria. Although these pathways were extensively studied in model bacteria, projection of this knowledge to thousands of diverse bacteria is a major challenge exacerbated by the immense *chemical diversity of carbohydrates* in various ecosystems and a matching *variability and evolutionary plasticity* of sugar utilization machinery of inhabiting species. This variability includes *alternative biochemical routes, non-orthologous gene replacements, functionally heterogeneous families of paralogs* and other problematic issues that cannot yet be automatically resolved by annotation/reconstruction pipelines. These problems lead to innumerable misannotations rapidly propagating across public databases, obscuring the real picture and effectively preventing meaningful reconstruction and modeling.

Our work confirmed that the challenge of accurate reconstruction of metabolic and regulatory networks can be effectively addressed by *comparative genomics techniques*. Performed analysis confirmed the hypothesis that even highly divergent variants of metabolic pathways may be inferred by the presence of conserved *signature genes*. The details of these pathways were further elucidated using a combination of *genomic context* (conserved operons and regulons) and *long-range homology analysis*. The power of this approach was confirmed by the experimental verification of predicted growth phenotypes and functional assignments of novel protein families accomplished in course of our genomic reconstruction of sugar catabolic machinery in *Shewanella* species.

In addition to numerous specific findings contributing to basic understanding of ecophysiology and evolution of *Shewanella* genus, the key components of the integrative genomic methodology of general utility for the community were optimized, validated and disseminated.

## REFERENCES\*

*(Main publications directly resulting from described studies are shown in boldface)*

1. Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. PMID:18261238
2. **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. PMID:18604222**
3. Gelfand, M. S., P. S. Novichkov, E. S. Novichkova, and A. A. Mironov. 2000. Comparative analysis of regulatory patterns in bacterial genomes. *Brief Bioinform* 1:357-71. PMID:11465053
4. Huang, N., J. De Ingeniis, L. Galeazzi, C. Mancini, Y. D. Korostelev, A. B. Rakhmaninova, M. S. Gelfand, D. A. Rodionov, N. Raffaelli, and H. Zhang. 2009. Structure and function of an ADP-ribose-dependent transcriptional regulator of NAD metabolism. *Structure* 17:939-51. PMID:19604474
5. **Leyn, S. A., X. Li, Q. Zheng, P. S. Novichkov, S. Reed, M. F. Romine, J. K. Fredrickson, C. Yang, A. L. Osterman, and D. A. Rodionov. 2011. Control of proteobacterial central carbon metabolism by the HexR transcriptional regulator: a case study in *Shewanella oneidensis*. *The Journal of biological chemistry* 286:35782-94. PMID:21849503**
6. Mironov, A. A., E. V. Koonin, M. A. Roytberg, and M. S. Gelfand. 1999. Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res* 27:2981-9. PMID:10390542
7. 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 Res* 38:D111-8. PMID:19884135
8. Osterman, A., and R. Overbeek. 2003. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 7:238-51. PMID:12714058
9. Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, V. de Crecy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Ruckert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691-702. PMID:16214803
10. Overbeek, R., M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev. 1999. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* 96:2896-901. PMID:10077608
11. **Pinchuk, G. E., E. A. Hill, O. V. Geydebrekht, J. De Ingeniis, X. Zhang, A. Osterman, J. H. Scott, S. B. Reed, M. F. Romine, A. E. Konopka, A. S. Beliaev, J. K. Fredrickson, and J. L. Reed. 2010. Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: a tool for data analysis and hypothesis generation. *PLoS Comput Biol* 6:e1000822. PMID:20589080**

12. Pinchuk, G. E., D. A. Rodionov, C. Yang, X. Li, A. L. Osterman, E. Dervyn, O. V. Geydebrekht, S. B. Reed, M. F. Romine, F. R. Collart, J. H. Scott, J. K. Fredrickson, and A. S. Beliaev. 2009. Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* 106:2874-9. PMID:19196979
13. Rodionov, D. A., J. De Ingeniis, C. Mancini, F. Cimadamore, H. Zhang, A. L. Osterman, and N. Raffaelli. 2008. Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. *Nucleic Acids Res* 36:2047-59. PMID:18276643
14. Rodionov, D. A., P. S. Novichkov, E. D. Stavrovskaya, I. A. Rodionova, X. Li, M. D. Kazanov, D. A. Ravcheev, A. V. Gerasimova, A. E. Kazakov, G. Y. Kovaleva, E. A. Permina, O. N. Laikova, R. Overbeek, M. F. Romine, J. K. Fredrickson, A. P. Arkin, I. Dubchak, A. L. Osterman, and M. S. Gelfand. 2011. Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics* 12 Suppl 1:S3. PMID:21810205
15. Rodionov, D. A., C. Yang, X. Li, I. A. Rodionova, Y. Wang, A. Y. Obratsova, O. P. Zagnitko, R. Overbeek, M. F. Romine, S. Reed, J. K. Fredrickson, K. H. Nealson, and A. L. Osterman. 2010. Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 11:494. PMID:20836887
16. Yang, C., D. A. Rodionov, X. Li, O. N. Laikova, M. S. Gelfand, O. P. Zagnitko, M. F. Romine, A. Y. Obratsova, K. H. Nealson, and A. L. Osterman. 2006. Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J Biol Chem* 281:29872-85. PMID:16857666