

Shewregdb: Database and visualization environment for experimental and predicted regulatory information in *Shewanella oneidensis* mr-1

Mustafa H. Syed^{1*}, Tatiana V. Karpinets^{1,2}, Michael R. Leuze¹, Guruprasad H. Kora^{1,2}, Margaret R. Romine³, Edward C. Uberbacher¹

¹Oak Ridge National Laboratory, Oak Ridge, Tennessee, 37831, U.S.A.; ²University of Tennessee, Knoxville, Tennessee, 37996, U.S.A.; ³Pacific Northwest National Laboratory, Richland, Washington, 99352, U.S.A.; MH Syed – Email: syedmh@ornl.gov; *Corresponding author

Received May 27, 2009; Revised July 20, 2009; Accepted September 11, 2009; Published October 15, 2009

Abstract:

Shewanella oneidensis MR-1 is an important model organism for environmental research as it has an exceptional metabolic and respiratory versatility regulated by a complex regulatory network. We have developed a database to collect experimental and computational data relating to regulation of gene and protein expression, and, a visualization environment that enables integration of these data types. The regulatory information in the database includes predictions of DNA regulator binding sites, sigma factor binding sites, transcription units, operons, promoters, and RNA regulators including non-coding RNAs, riboswitches, and different types of terminators.

Availability: <http://shewanella-knowledgebase.org:8080/Shewanella/gbrowserLanding.jsp>.

Keywords: database; regulatory data; *Shewanella oneidensis*

Background:

Shewanella oneidensis MR-1 can use a diverse set of terminal electron acceptors including iron, manganese, nitrate, nitrite, fumarate, uranium, sulfur, trimethylamine-N-oxide, dimethyl sulfoxide [1] and thrive in the environments contaminated with heavy metals and radionuclides [2]. For efficient adaptation to such environments the organism has evolved a robust regulatory system [3]. A sophisticated and complex regulatory network exists to modulate the expression of genes that enable MR-1 to utilize different carbon and energy sources [4]. This regulatory information is available from both computational predictions [5-9] and experimental studies [10, 11], but its usefulness is limited because it is not available in a centralized, web accessible regulatory database.

Understanding regulation is key to gaining a systems level understanding of any living organism. Regulatory databases exist for many model organisms, including gram-positive bacterium, *Bacillus subtilis*, [12] and gram-negative bacterium, *Escherichia coli*. The regulatory information for *E. coli* is available in two main databases, RegulonDB [13] and EcoCyc [14]. Most of the information in these databases is manually curated. A number of eukaryotic databases [15-20] employ GBrowser from the Generic Model Organism Database Toolkit [21] to visualize the regulatory information. As a rule the type of regulators in the databases is limited, or the regulatory information is not integrated with experimentally collected data on the organism. The main objectives of the study was to develop a regulatory database for *S. oneidensis* MR-1 and a GBrowser based visualization environment that allows integration of the regulatory information in the analysis of diverse experimental data (experimental results from studies employing microarrays, proteomics, and/or gene mutagenesis) collected in the *Shewanella* knowledgebase (<http://shewanella-knowledgebase.org:8080/Shewanella/>).

Methodology:

The information collected in the Regulatory database, which is further referred as ShewRegDB, is based on computational prediction of different regulatory elements in *S. oneidensis* MR-1 and published experimental data. Different Internet resources including TractorDB, RegTransBase, Rfam, RibEx, TransTermHP, PromScan, ShewCyc, ODB, MicrobesOnline, and others were used for computational predictions of the regulatory elements. **Table1 (see supplementary material)** summarizes number of elements located by these sources in

Shewanella oneidensis MR-1. A diverse set of methods were employed by these resources/databases to find the location of the regulatory protein binding sites, operon or RNA regulators including the use of weight matrices [5], phylogenetic footprinting [5, 6] building profiles or models [7, 8], motif clustering [6], clustering co-expressed genes in order to find conserved patterns in their upstream regions, comparative genomics [22], gene/domain architecture [22], decision rule based algorithm [9], combination of above approaches [5, 6, 22, 23] or literature search by using a controlled vocabulary [24]. In addition to computational predications from the Internet resources, we also collected regulatory information based on experimental data available in literature.

The collected information was analyzed to identify basic regulatory elements (entities), their characteristics (attributes) and main types of the regulatory elements (database objects or classes) that are essential to characterize regulation in *S. oneidensis*. The identified database objects include DNA regulator binding sites, RNA regulators, operons and genes. At present the class of RNA regulators includes non-coding and small RNAs, different types of terminators and riboswitches. The class of DNA regulator binding sites includes binding sites of transcription and sigma factors. An entity relationship model of the database was developed according to the identified entity classes and their characteristics, then implemented as a relational database in MySQL, and populated with the collected information.

Description of the web-server:

A web interface was designed to allow users to search and to download collected regulatory information. The search engine queries three main data objects of the database: DNA regulator binding sites, RNA regulators and operons. Users can search for a specific DNA regulator binding site from 4 different resources. RNA regulator search extracts information on seven types of RNA regulators of a specific gene queried by its name or locus tag. Operon search retrieves information on operon predictions from 6 different sources.

Search results for each query are displayed in the tabular format and can be downloaded as an EXCEL spreadsheet. Entries of the table have links to GBrowser and to the *Shewanella* knowledgebase by location of each regulatory element and by the gene locus tag. Clicking on the links provide further options on visualization of the

regulatory information by GBrowser and on association of the information with the experimental data in the knowledgebase.

A visualization environment based on GBrowser [21] was developed for accessing the collected information and for its overlaying with experimental data and other genome annotations. The browser was configured for the *S. oneidensis MR-1* genome and adjusted to the specificity of collected information. The data collected in the regulatory database on each database object were converted into standard gff-files for visualization of the regulatory elements on GBrowser tracks. The following visualization options are currently available: (a) an overview of all regulatory elements in the genome with scrolling to any selected region, (b) presentation of each type of element on different tracks, (c) different types of zooming and (d) displaying detailed information on each regulatory element including decorated (colored conserved regions) FASTA sequences. The browser has four main sections: (a) Search, navigation, download

section, (b) Overview, (c) Tracks, and (d) Track Selection. Tracks in the Genome browser are organized according to main data objects in ShewRegDB and include genes, DNA regulator binding sites, sigma factor binding sites, RNA regulators, and operons (Figure 1). Two additional tracks “Experimental Data Selection” and “ShewCyc pathways” are included to provide overlay of the regulatory information with the experimental data collected in the knowledgebase and with the *S. oneidensis* metabolic pathway annotation from ShewCyc [29].

Utility to the biological community:

ShewRegDB is unique in providing regulatory data in the framework of omics data including transcriptomics, proteomics and metabolomics datasets. This will facilitate studying the exceptional metabolic and respiratory versatility of *Shewanella oneidensis* MR-1, an important model organism for environmental research.

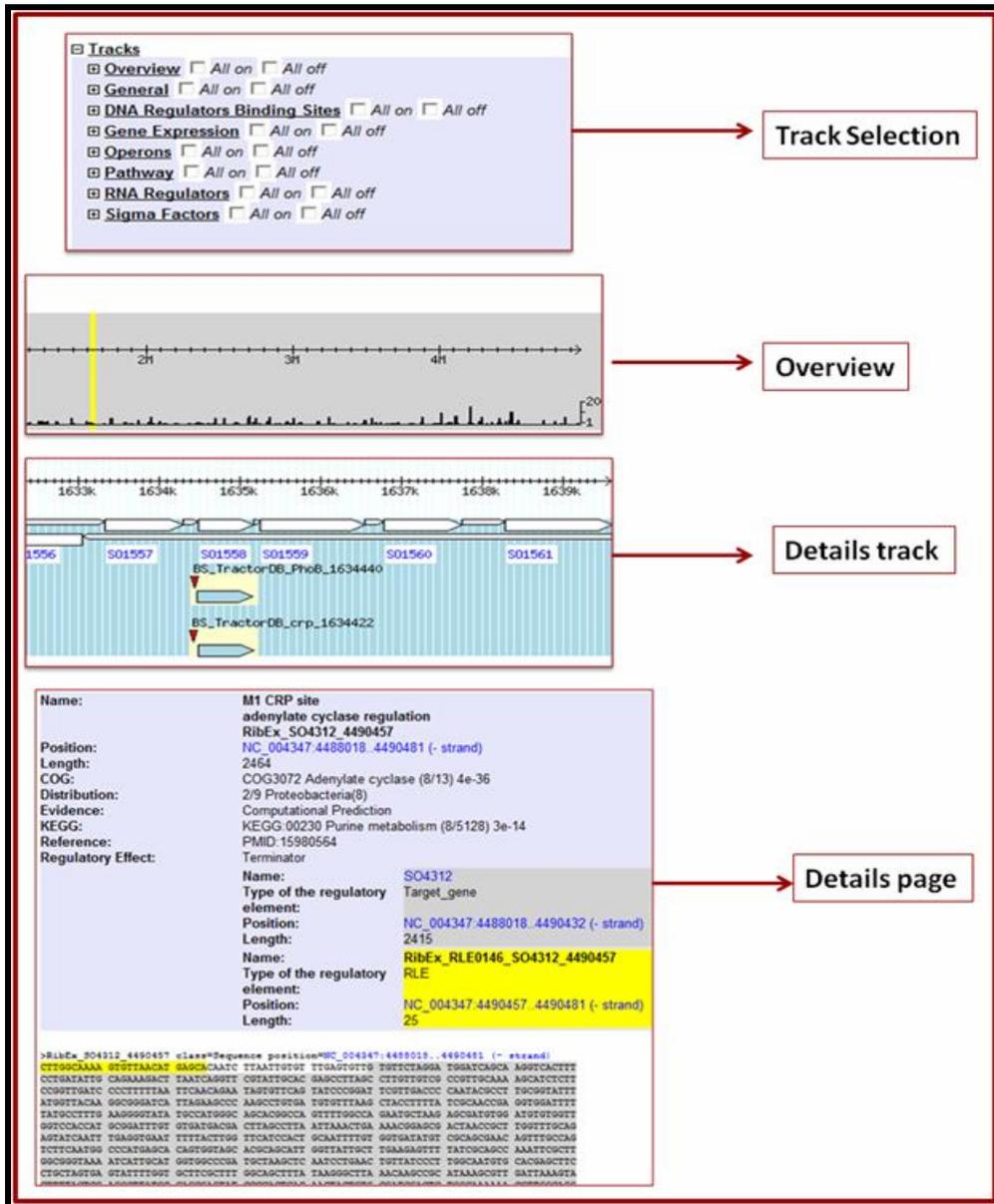


Figure 1: Visualization of the regulatory data from ShewRegDB in GBrowser

Acknowledgements:

Authors would like to thank Michael Galloway for system support and Loren Hauser for providing valuable comments and suggestions for this article. This research is sponsored by the U.S. Department of Energy, Office of Biological & Environmental Research. Oak Ridge National Laboratory is managed by the University of Tennessee-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U. S. Department of Energy under contract DE-AC05-76RL01830.

Funding: This work was supported by Department of Energy [Field Work Proposal: ERKP-558].

References:

- [1] JF Heidelberg *et al.*, *Nat Biotechnol.* **20**: 1118 (2002) [PMID: 12368813]
- [2] HH Hau, JA Gralnick, *Annu Rev Microbiol.* **61**: 237 (2007)[PMID: 18035608]
- [3] JK Fredrickson *et al.*, *Nat Rev Microbiol.* **6**: 592 (2008) [PMID: 18604222]
- [4] JA Gralnick *et al.*, *Molecular Microbiology* **56**: 1347 (2005) [PMID: 15882425]
- [5] AD González *et al.*, *Nucleic Acids Res.* **1**: D98 (2005) [PMID: 15608293]
- [6] J Liu *et al.*, *Nucleic Acids Res.* **36**: 5376 (2008) [PMID: 18701645]
- [7] S Griffiths-Jones *et al.*, *Nucleic Acids Res.* **1**: D121 (2005) [PMID: 15608160]
- [8] C Abreu-Goodger, E Merino, *Nucleic Acids Res.* **1**: W690 (2005) [PMID: 15980564]
- [9] CL Kingsford *et al.*, *Genome Biol.* **8**: R22 (2007) [PMID:17313685]
- [10] AS Beliaev *et al.*, *J Bacteriol.* **184**: 4612 (2002) [PMID: 12142431]
- [11] XF Wan *et al.*, *J Bacteriol.* **186**: 8385 (2004) [PMID: 15576789]
- [12] N Sierro *et al.*, *Nucleic Acids Res.* **36**: D93 (2008) [PMID: 17962296]
- [13] S Gama-Castro *et al.*, *Nucleic Acids Res.* **36**: D120 (2008) [PMID: 18158297]
- [14] IM Keseler *et al.*, *Nucleic Acids Res.* **1**: D334 (2005) [PMID: 15608210]
- [15] CJ Schmidt *et al.*, *Nucleic Acids Res.* **36**: D719 (2008) [PMID: 17933775]
- [16] RJ Wilson *et al.*, *Nucleic Acids Res.* **36**: D588 (2008) [PMID: 18160408]
- [17] D Swarbreck *et al.*, *Nucleic Acids Res.* **36**: D1009 (2008) [PMID:17986450]
- [18] L Kreppel *et al.*, *Nucleic Acids Res.* **1**: D332 (2004) [PMID: 14681427]
- [19] JA Blake *et al.*, *Nucleic Acids Res.* **1**: D562 (2006) [PMID: 16381933]
- [20] N de la Cruz *et al.*, *Nucleic Acids Res.* **1**: D485 (2005) [PMID: 15608243]
- [21] LD Stein *et al.*, *Genome Res.* **12**: 1599 (2002) [PMID: 12368253]
- [22] MN Price *et al.*, *Nucleic Acids Res.* **8**: 880 (2005) [PMID:15701760]
- [23] DJ Studholme, RJ Dixon, *Bacteriol.* **185**: 1757 (2003) [PMID: 12618438]
- [24] AE Kazakov *et al.*, *Nucleic Acids Res.* **35**: D407 (2007) [PMID: 17142223]
- [25] X Xu *et al.*, *PLoS Comput Biol.* **5**: e1000338 (2009) [PMID: 19343219]
- [26] S Okuda *et al.*, *Nucleic Acids Res.* **1**: D358 (2006) [PMID: 16381886]
- [27] PR Romero, PD Karp, *Bioinformatics* **22**: 709 (2004) [PMID: 14751985]
- [28] F Mao *et al.*, *Nucleic Acids Res.* **37**: D459 (2009) [PMID: 18988623]
- [29] R Caspi *et al.*, *Nucleic Acids Res.* **36**: D623 (2008) [PMID: 17965431]

Edited by P. Kanguane

Citation: Syed *et al.*, *Bioinformation* 4(4): 169-172 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Regulatory information collected in ShewRegDB

Source	Class of the Regulatory Element	Number of Elements	Number of Sites	Computational predictions(C) /Manual annotation (M)	Reference
TractorDB	DNA regulator Binding Site	20	589	C	[5]
RegTransBase	DNA regulator Binding Site	5	36	M	[24]
WUSTL	DNA regulator Binding Site	23	97	C	[6]
ShewRegDB	DNA regulator Binding Site	3	359	M	
PromScan	DNA regulator Binding Site (Sigma Factor Binding Sites)	1	383	C	[23]
Rfam	RNA regulator (ncRNA)	17	135	C	[7]
RibEx	RNA regulator (RLEs)		129	C	[8]
WUSTL (RSSVM)	RNA regulator (ncRNA)		953	C	[25]
RibEx	RNA regulator (Attenuators)		1767	C	[8]
TransTermHP	RNA regulator (rho-independent terminators)		81457	C	[9]
MicrobesOnline	Operon	2601		C	[22]
ODB	Operon	856		C	[26]
Shewybase	Operon	406		C	Michael Molla, Tim Gardner , personal communication
ShewCyc	Operon	2968		C	[27]
DOOR	Operon	2773		C	[28]
MESS	Operon	2943		M	Margaret R. Romine, personal communication