

RESEARCH

Open Access

# Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus

Dmitry A Rodionov<sup>1,2\*†</sup>, Pavel S Novichkov<sup>3†</sup>, Elena D Stavrovskaya<sup>2,4</sup>, Irina A Rodionova<sup>1</sup>, Xiaoqing Li<sup>1</sup>, Marat D Kazanov<sup>1,2</sup>, Dmitry A Ravcheev<sup>1,2</sup>, Anna V Gerasimova<sup>3</sup>, Alexey E Kazakov<sup>2,3</sup>, Galina Yu Kovaleva<sup>2</sup>, Elizabeth A Permina<sup>5</sup>, Olga N Laikova<sup>5</sup>, Ross Overbeek<sup>6</sup>, Margaret F Romine<sup>7</sup>, James K Fredrickson<sup>7</sup>, Adam P Arkin<sup>3</sup>, Inna Dubchak<sup>3,8</sup>, Andrei L Osterman<sup>1,6</sup>, Mikhail S Gelfand<sup>2,4</sup>

## Abstract

**Background:** 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.

**Results:** 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. 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).

**Conclusions:** 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 constrains for an integrated model of metabolism and regulation in *S. oneidensis* MR-1.

\* Correspondence: rodionov@burnham.org

† Contributed equally

<sup>1</sup>Sanford-Burnham Medical Research Institute, La Jolla, California, USA

Full list of author information is available at the end of the article

## Background

Fine-tuned regulation of gene expression in response to extracellular and intracellular signals is a key mechanism for successful adaptation of microorganisms to changing environmental conditions. Activation and repression of gene expression in bacteria is usually mediated by DNA-binding transcription factors (TFs) that specifically recognize TF-binding sites (TFBSs) in upstream regions of target genes, and also by various regulatory RNA structures including *cis*-acting metabolite-sensing riboswitches and attenuators encoded in the leader regions of target genes. Genes and operons directly co-regulated by the same TF or by an RNA structure are considered to belong to a *regulon*. All regulons taken together form the transcriptional regulatory network (TRN) of the cell. TFs form more than 50 different protein families and constitute around 5-10% of all genes in an average bacterial genome, and their respective regulons cover a substantial fraction of bacterial TRNs [1].

Traditional experimental methods for the analysis of transcriptional gene regulation and characterization of TFBSs provided a foundation for the current understanding of regulatory interactions [2]. However, taken alone, they are limited in productivity (the scale) and feasibility (often restricted to a few model organisms). High-throughput transcriptome approaches opens new opportunities for measuring the expression of thousands of genes in a single experiment [3]. The microarray technology has been successfully used to explore transcriptional responses in several bacteria. However, convoluted regulatory cascades, multi-TF regulation of certain genes, and various indirect effects on the transcription and abundance of mRNA make the observed regulatory responses too complex for a direct top-down analysis. The chromatic immunoprecipitation approach is now increasingly used for the investigation of genome-wide DNA-binding of global TFs in bacteria [3]. At the same time, a growing number of complete prokaryotic genomes allows us to extensively use comparative genomics approaches to infer conserved *cis*-acting regulatory elements (e.g. TFBSs and riboswitches) in regulatory networks of numerous groups of bacteria ([4-15], also reviewed in [1]). These and other previous studies enabled us to define and prototype a general workflow of the “knowledge-driven” approach for the comparative-genomic reconstruction of regulons. Two major components of this analysis are (i) propagation of previously known regulons from model organisms to others and (ii) *ab initio* prediction of novel regulons (see Methods for more details). This approach is different, and in many ways complementary to the two most common alternative approaches to the TRN reconstruction: (i) the “data-driven” approach, top-down regulatory

network reconstruction from microarray data [16]; and (ii) the “computation-driven” approach, *ab initio* automated identification and clustering of conserved DNA motifs [17].

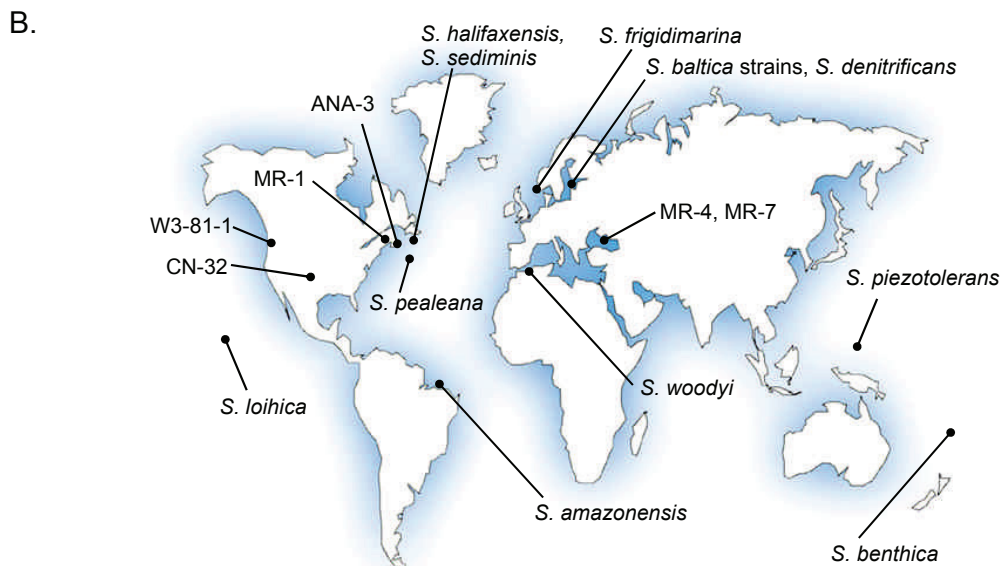
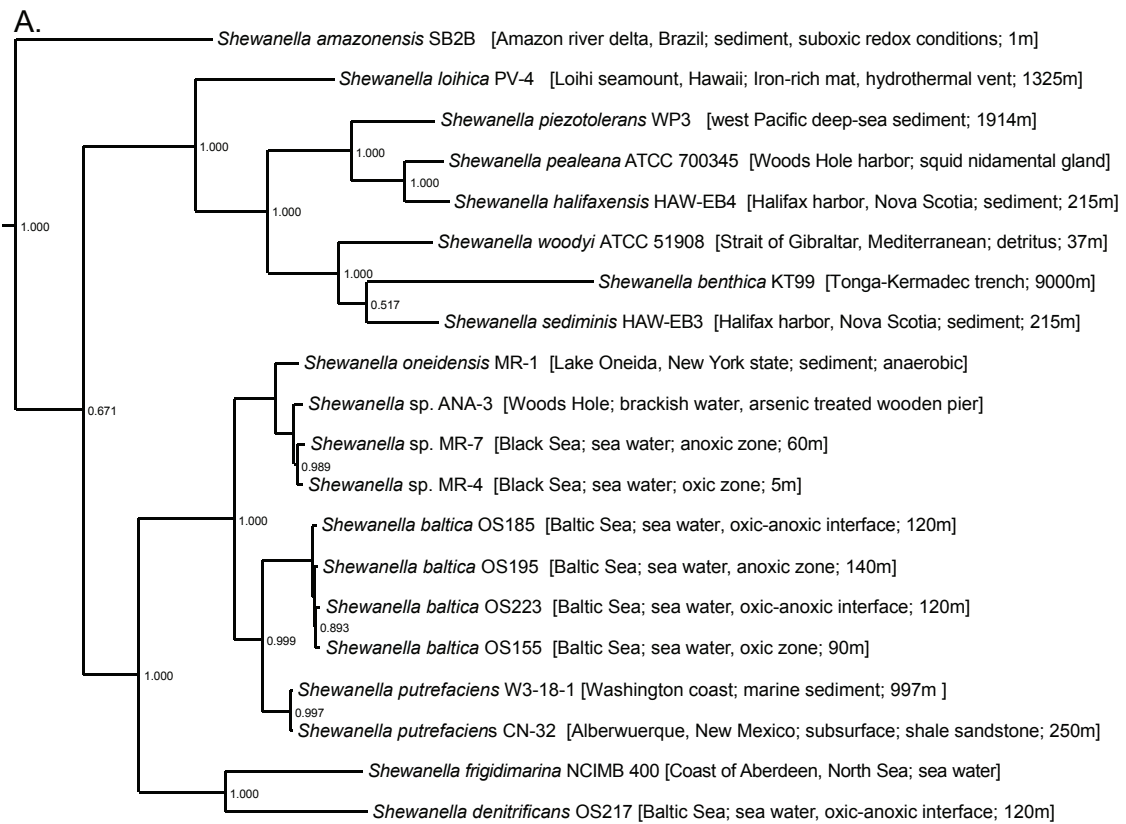
*Shewanella* spp. are Gram-negative facultative anaerobic  $\gamma$ -proteobacteria characterized by a remarkable versatility in using a variety of terminal electron acceptors for anaerobic respiration (reviewed in [18]). Isolated from various aquatic and sedimentary environments worldwide, the *Shewanella* demonstrate diverse metabolic capabilities and adaptation for survival in extreme conditions (Fig. 1) [19]. Although the model species *Shewanella oneidensis* MR-1 is a subject of extensive genetics and physiological studies, as well as genome-scale transcriptomics and proteomics approaches [18,20-22], our experimental knowledge of transcriptional regulation in *S. oneidensis* is limited to the Fur, ArcA, TorR, Crp, and EtrA (Fnr) TFs controlling iron metabolism and anaerobic respiration [23-29]. In addition, the novel NrtR regulon for NAD cofactor metabolism was inferred by comparative genomics and experimentally validated in *S. oneidensis*[11].

Availability of multiple closely-related genomes from the *Shewanella* genus (Fig. 1) provided a basis for the reconstruction of the metabolic and regulatory networks using comparative genomics. Recently, we have applied the comparative genomic approach to predict novel pathways and regulons for the *N*-acetylglucosamine and lactate utilization [30,31], and to reconstruct two novel regulons for the fatty acid and branched-chain amino acid utilization pathways in *Shewanella* spp. [4]. In this study, we have extended our previous analysis towards the detailed reconstruction of ~100 transcriptional regulons in 16 *Shewanella* species with completely sequenced genomes. The identified TRN contains over 450 regulated genes per genome, mostly covering the central and secondary metabolism and stress response pathways. The comparative analysis of the reconstructed regulons revealed many aspects of the metabolic regulation in the *Shewanella* that are substantially different from the established TRN model of *Escherichia coli*.

## Results

### Repertoire of transcription factors in the *Shewanella* spp

Previous comparative analysis revealed extensive gene content diversity among 10 *Shewanella* genomes [32]. To gain further insight into the scale of the TRN diversity in this lineage, we analyzed the repertoire of DNA-binding TFs encoded in 16 complete *Shewanella* genomes (Additional file 1). The total number of TFs in individual species varies broadly, from 138 TFs in *S. denitrificans* to 262 TFs in *S. woodyi*, with an average of ~200 TFs per genome (Fig. 2). 95% of all TFs of the



**Figure 1** Phylogenetic tree, isolation site characteristics and the geographic origin of 20 *Shewanella* species with available genomes. The tree was constructed using the concatenated alignment of ~78 universal prokaryotic proteins in the MicrobesOnline database <http://www.microbesonline.org/cgi-bin/speciesTree.cgi>.

*Shewanella* belong to 17 major protein families with at least two distinct members per genome. At that, the total number of TFs in most of these families varies significantly among the *Shewanella* spp. The largest TF families are LysR, OmpR, Fis, TetR, AraC, and LuxR (>10 TFs per genome on average). Among the remaining 14 families of TFs, mostly represented by single members in the genomes (without paralogs), the Fur, ArgR, BirA, LexA, MetJ, NrdR, RpiR, and TrpR families are universally conserved in the *Shewanella* (Fig. 2). A significant reduction of the TF repertoire is a unique feature of *S. denitrificans*, which has limited anaerobic growth capabilities due to massive gene loss in course of ecological specialization [32].

The 3,228 predicted TFs in 16 *Shewanella* genomes were clustered into 686 orthologous groups (Additional file 1), among which only 63 TFs (9%) were universally conserved in all genomes (the core TF set), 320 TFs (47%) were found in at least two genomes (variable TFs), whereas the remaining TFs (303 or 44%) were strain-specific (Additional file 1). Although the genomes of the *Shewanella* spp. and *E. coli* demonstrate a similar repertoire and size of TF protein families, only 73 (30%) TFs from *E. coli* have orthologs in at least one *Shewanella* genome (Fig. 2). The group of 34 TFs that are present in the *Shewanella* core TF set (Additional file 2) and conserved between *E. coli* and the *Shewanella* spp. (Additional file 3) is enriched by regulators controlling the metabolism of amino acids (ArgR, AsnC, CysB, GcvA, IlvY, MetJ, MetR, TrpR, TyrR), fatty acids (FabR,

FadR), cofactors (BirA, IscR), deoxynucleosides (NrdR), nitrogen (NtrC), phosphate (PhoB), iron (Fur), central carbohydrate metabolism (HexR, PdhR), stress responses (CpxR, LexA, NhaR, NsrR), and global regulators (ArcA, Crp, Fis, Fnr, and Lrp). The group of strain-specific *Shewanella* regulators with orthologs in *E. coli* contains 5 known regulators for local carbohydrate utilization pathways (AlgR, NanR, DgoR, GalR, GntR) that were possibly acquired together with the target metabolic pathway genes via lateral gene transfer events [33]. Near 1/2 of strain-specific TFs of the *Shewanella* spp. belong to two protein families, LysR and AraC (96 and 50 TFs, respectively), that were likely expanded via gene duplication in course of ecological adaptation of individual species.

### Comparative analysis of transcriptional regulation in the *Shewanella* spp

To infer TRNs in the *Shewanella* spp., we used the integrative comparative genomics approach that combines identification of TFs and candidate TFBSs with cross-genomic comparison of regulons and with the genomic and functional context analysis of candidate target genes. We analyzed 16 *Shewanella* genomes and inferred regulons for 82 orthologous groups of TFs that split into two groups: 41 regulators with experimentally characterized orthologs in *S. oneidensis* or other  $\gamma$ -proteobacteria (Table 1), and 41 novel regulators without characterized orthologs in any species (Table 2). The genomic and functional content of the reconstructed TF regulons from both groups, as well as of the regulons controlled by

Genome	AraC	AtsR	AsnC	CadC	CRP	DeoR	Fis	GntR	LacI	LuxR	LysR	LYTR	MarR	MerR	OmpR	TetR	XRE	Rrf2	Fur	ArgR	BirA	LexA	MetJ	NrdR	RpiR	TrpR	SdaR	Others	Total TFs	TF fraction, %	
<i>E. coli</i> K12	29	2	3	0	2	13	15	20	14	24	46	1	3	5	17	13	1	2	2	1	1	1	1	1	4	1	1	14	237	5.5	
<i>S. oneidensis</i> MR-1	9	4	5	1	4	2	14	7	2	11	53	5	6	5	24	17	10	2	1	1	1	1	1	1	1	1	1	3	193	4.5	
<i>S. putrefaciens</i> CN-32	9	6	4	1	4	2	18	12	1	9	44	4	8	7	22	14	8	2	1	1	1	1	1	1	1	1	1	5	189	4.8	
<i>S. putrefaciens</i> W3-18-1	10	7	4	1	4	2	17	12	1	9	41	4	8	7	21	15	9	2	1	1	1	1	1	1	1	1	1	4	187	4.6	
<i>Shewanella</i> sp. ANA-3	17	6	4	2	3	3	18	10	3	9	49	5	10	9	30	18	11	2	1	1	1	1	1	1	1	1	1	2	220	5.4	
<i>Shewanella</i> sp. MR-4	13	3	4	1	2	3	16	8	3	9	44	6	10	7	30	17	5	2	1	1	1	1	1	1	1	1	1	1	193	4.9	
<i>Shewanella</i> sp. MR-7	12	3	4	1	2	3	15	10	4	8	44	6	10	6	30	16	6	2	1	1	1	1	1	1	1	1	1	1	192	4.8	
<i>S. baltica</i> OS155	16	5	5	0	3	2	15	12	5	9	51	5	9	6	21	16	13	2	1	1	1	1	1	1	1	1	1	2	206	4.8	
<i>S. denitrificans</i> OS217	7	2	3	1	3	1	18	6	3	5	24	4	7	6	15	15	7	2	1	1	1	1	1	1	1	1	0	1	138	3.7	
<i>S. frigidimarina</i> NCIMB 400	6	5	4	0	2	2	20	8	4	9	50	2	10	8	24	20	9	2	2	1	1	1	1	1	1	1	1	0	195	4.8	
<i>S. amazonensis</i> SB2B	8	3	6	1	3	4	15	7	5	9	40	4	6	6	30	16	2	2	1	1	1	1	1	1	1	1	1	0	1	176	4.8
<i>S. loihica</i> PV-4	14	5	6	1	3	2	17	8	2	11	49	4	9	5	25	18	8	2	1	1	1	1	1	1	1	1	1	1	1	199	5.2
<i>S. pealeana</i> ATCC 700345	15	4	6	4	3	1	18	10	3	11	64	3	6	4	20	19	9	2	1	1	1	1	1	1	1	1	1	2	213	5.0	
<i>S. halifaxensis</i> HAW-EB4	15	4	6	4	3	1	17	8	3	15	66	3	4	5	19	20	7	2	1	1	1	1	1	1	1	1	1	1	212	5.0	
<i>S. piezotolerans</i> WP3	19	8	6	2	4	1	19	8	3	11	61	2	7	5	19	21	7	2	1	1	1	1	1	1	1	1	1	1	215	4.2	
<i>S. sediminis</i> HAW-EB3	14	6	5	2	5	2	21	6	1	18	74	3	9	5	24	22	9	2	1	1	1	1	1	1	1	1	1	1	238	5.3	
<i>S. woodyi</i> ATCC 51908	29	5	5	4	3	2	22	12	5	16	64	6	12	6	27	18	14	2	1	1	1	1	1	1	1	1	1	1	262	5.4	
Total TFs in <i>Shewanella</i> :	213	76	77	26	51	33	280	144	48	169	818	66	131	97	381	282	134	32	17	16	16	16	16	16	16	16	14	27	3228	-	
Average TF per genome:	13	5	5	2	3	2	18	9	3	11	51	4	8	6	24	18	8	2	1	1	1	1	1	1	1	1	1	-	202	-	
Overlap <i>E. coli</i> / <i>Shewanella</i>	2	2	2	0	2	1	5	7	3	3	13	1	1	3	9	4	3	2	1	1	1	1	1	1	1	1	1	3	75	-	

**Figure 2** Distribution by protein families of predicted DNA-binding transcription factors in the *Shewanella* genomes.

**Table 1 Previously known TF regulons reconstructed in *Shewanella* spp**

TF name	<i>S. one- idensis</i> MR-1	<i>S. putrefa- ciens</i> CN-32	<i>S. putre- faciens</i> W3-18-1	<i>Shewa- nella</i> sp. ANA-3	<i>Shewa- nella</i> sp. MR-4	<i>Shewa- nella</i> sp. MR-7	<i>S. baltica</i> OS155	<i>S. denitri- ficans</i> OS217	<i>S. frigid- marina</i> NCIMB 400	<i>S. amazo- nensis</i> SB2B	<i>S. loihica</i> PV-4	<i>S. pealeana</i> ATCC 700345	<i>S. halifa- xensis</i> HAW- EB4	<i>S. piezoto- lerans</i> WP3	<i>S. sediminis</i> HAW-EB3	<i>S. woodyi</i> ATCCC 51908	<i>Pseudo- monas</i> spp.	<i>E. coli</i>	Regulon functional role
Agar	-	-	-	+	+	+	-	-	-	+	-	-	-	-	-	-	-	rs	GalNAc utilization
ArgR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Arginine biosynthesis
BetI	-	-	-	-	-	-	+	+	+	-	-	+	+	+	+	+	+	rs	Osmotic protection
BirA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Biotin biosynthesis
Crp	r	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Global regulon
CueR	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	rs	rs	Copper efflux
Dnr	-	-	-	-	-	-	-	+	-	+	+	-	-	-	-	-	rs	-	Denitrification
FabR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Fatty acid biosynthesis
FadR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Fatty acid degradation
Fnr	r	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Global regulon
Fur	rs	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	rs	Iron homeostasis
GalR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	rs	Galactose utilization
GcvA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Glycine metabolism
GlmR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	r	-	LPS synthesis
GntR	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	+	rs	Gluconate utilization
HexR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	+	Central sugar metabolism
HutC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	r	-	Histidine utilization
IlvY	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Isoleucine-valine synthesis
IscR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Fe-S cluster assembly
LexA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	rs	DNA damage stress
MetJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Methionine biosynthesis

**Table 1 Previously known TF regulons reconstructed in *Shewanella* spp (Continued)**

MetR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Methionine biosynthesis
ModE	+	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+	rs	Molybdenum metabolism
NanR	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	rs	Sialic acid utilization
NarP	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	rs	Nitrate/nitrite respiration
NhaR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Osmotic stress protection
NikR	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	+	rs	Nickel uptake	
NorR	-	+	+	+	+	+	-	-	+	+	+	+	+	+	+	+	+	+	rs	Nitrosative stress
NrdR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	Nucleotide metabolism
NrtR	rs	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	Nicotinamide utilization
NsrR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Nitrosative stress
NtrC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	rs	Nitrogen assimilation
PdhR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Pyruvate metabolism
PsrA	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs	-	Fatty acid degradation
RbsR	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	+	rs	Ribose utilization	
SdaR	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+	+	+	r	Glycerate utilization
SoxR	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	+	+	rs	Superoxide stress	
TorR	rs	-	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	-	rs	TMAO respiration
TrpR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Amino acid metabolism
TyrR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	rs*	rs	Amino acid metabolism
ZntR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	rs	Zinc efflux

The presence of orthologous transcription factors is shown by '+', 'r' and 'rs', whereas its absence is denoted by '-'. TFs with previously characterized target genes in model species are denoted by 'r'. TFs with previously known TFBSs at their target genes are denoted by 'rs'. Ortholog of TyrR in *Pseudomonas* spp. that was characterized as the phenylalanine and tyrosine catabolism regulator PhhR is marked by asterisk.

**Table 2 Novel TF regulons predicted and reconstructed in *Shewanella* spp**

TF name	<i>S. oneidensis</i> MR-1	<i>S. putrefaciens</i> CN-32	<i>S. putrefaciens</i> W3-18-1	<i>Shewanella</i> sp. ANA-3	<i>Shewanella</i> sp. MR-4	<i>Shewanella</i> sp. MR-7	<i>S. baltica</i> OS155	<i>S. denitrificans</i> OS217	<i>S. frigidimarina</i> NCIMB 400	<i>S. amazonensis</i> SB2B	<i>S. loihica</i> PV-4	<i>S. pealeana</i> ATCC 700345	<i>S. halifaxensis</i> HAW-EB4	<i>S. piezotolerans</i> WP3	<i>S. sediminis</i> HAW-EB3	<i>S. woodyi</i> ATCC 51908	Regulon functional role
<i>A. Regulons inferred from the analysis of metabolic pathways</i>																	
AlgR*	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	Hexuronate utilization
AraR*	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	Arabinose utilization
BglR*	-	-	-	-	-	-	+	+	+	+	-	-	-	+	-	+	Beta-glucoside utilization
HmgR*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Tyrosine degradation
HypR*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Hydroxyproline utilization
LiuR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Amino acid utilization
LldR*	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	Lactate utilization
MalR*	+	-	-	+	+	+	+	+	+	+	+	-	-	+	-	+	Maltodextrin utilization
ManR1*	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	Mannose utilization
ManR2*	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	Mannose utilization
MtlR2	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	Mannitol utilization
NagR	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	GlcNAc utilization
PflR*	-	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	Formate metabolism
PrpR*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Propionate utilization
PUR*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Purine biosynthesis
ScrR*	-	-	-	+	+	+	+	-	+	-	-	-	-	-	-	-	Sucrose utilization
TreR*	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	Trehalose utilization
XltR*	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	Xylitol utilization

**Table 2 Novel TF regulons predicted and reconstructed in *Shewanella* spp (Continued)**

*B. Regulons inferred from the analysis of chromosomal gene clusters*

AzrR*	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	Superoxide stress
CalR*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Aromatics utilization
CueR2	-	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	-	Copper efflux
DeoR*	+	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+	+	Nucleoside utilization
PnuR*	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	NAD metabolism
SO0072	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	ABC efflux transporter
SO0082	+	-	-	-	+	-	+	-	+	-	-	-	-	-	-	-	-	Benzoate degradation
SO0193	+	+	+	+	+	+	+	-	+	+	+	-	-	-	-	-	-	Phospholipid synthesis
SO0734	+	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	hypothetical transporter
SO1393	+	-	-	+	+	+	-	-	-	-	+	-	-	-	-	-	-	hypothetical
SO1415	+	-	-	-	+	-	-	-	-	-	+	+	+	-	+	-	-	flavocytochrome c
SO1578	+	-	-	+	+	+	+	-	-	+	+	-	-	-	-	-	+	Glutathione detoxification
SO1703	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	-	multidrug efflux
SO1758	+	+	+	+	+	+	+	-	-	+	+	-	-	-	-	+	+	hypothetical
SO2282	+	+	+	+	+	+	+	-	-	+	-	+	+	+	+	-	+	Amino acid efflux
SO3277	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	multidrug efflux
SO3385	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	hypothetical
SO3393	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	xenobiotic reductase
SO3494	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	multidrug efflux
SO3627	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	flavocytochrome c
SO4326	+	-	-	+	-	-	-	-	-	+	-	+	+	+	+	+	+	multidrug efflux
SO4468	+	+	+	+	+	+	+	+	-	+	-	-	-	+	+	-	-	hypothetical
SO4705	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	hypothetical

The new TF names introduced in this work are marked by asterisks.



known RNA regulatory elements (8 riboswitches and 6 transcriptional attenuators), is summarized in Additional file 4 and briefly described below. These data, in conjunction with the detailed information about DNA binding motifs and individual TFBSs, were compiled into the *Shewanella* collection of regulons that was uploaded to the RegPrecise database <http://regprecise.lbl.gov>[34].

#### **Reconstruction of regulons for previously characterized regulators**

Our general strategy of reconstructing regulons controlled by known TFs in a novel taxonomic group consists of the following steps: (i) search for orthologous TFs, (ii) collecting known target genes and TFBSs in a model genome, (iii) identifying orthologous target genes in the analyzed genomes and extracting their upstream regions, iv) application of a pattern recognition program, then constructing positional weight matrices (PWMs) and comparison of the newly identified TFBS motifs with the previously known sites/motifs in a model genome, v) search for additional sites in the analyzed genomes and consistency check or cross-species comparison of the predicted regulons (details are provided in Materials and Methods section; the strategy was also reviewed in [1]). For regulons with significantly different repertoire of target genes in the *Shewanella* spp., the above procedure was repeated starting at the third step in order to include novel candidate targets into the TFBS motif model and to revise the final gene content of the regulon.

For the *Shewanella* genomes, we performed regulon reconstruction for 41 TFs that are orthologous to previously characterized regulators (Table 1). The majority of these TFs have experimentally characterized orthologs in  $\gamma$ -proteobacteria from other lineages, such as *E. coli* (35 TFs) and/or *Pseudomonas* spp. (10 TFs), or had been previously studied in *S. oneidensis* (5 TFs) (Additional file 5). Among these regulators, there are 26 universal TFs, three strain-specific TFs and 13 TFs mosaically distributed in the *Shewanella* spp. The deduced TFBS motifs for 41 analyzed regulons in the *Shewanella* spp. were compared to previously known motifs for orthologous regulators in other  $\gamma$ -proteobacteria using the RegulonDB database for *E. coli*[35] and original publications for *Pseudomonas* spp. (Additional file 5). For three regulators with previously unknown binding sites (GlmR, HutC, and SdaR) we report, for the first time, the identity of their cognate TFBSs. The identified new motifs in *Shewanella* are conserved in upstream regions of known targets in *E. coli* (for SdaR) and *Pseudomonas* spp. (for GlmR and HutC) (data not shown). Two novel TFBS motifs (for AgaR and GcvA) in the *Shewanella* spp. are completely different from the respective motifs in *E. coli*. Five other TFBS motifs (for CueR, NhaR, PsrA, TrpR, and ZntR) in the *Shewanella*

spp. are moderately different (3-4 mismatches in the conserved positions) from the known motifs of orthologous TFs previously described in *E. coli* and/or *Pseudomonas* spp. The remaining 31 *Shewanella* TFs appear to have binding motifs that are well conserved or only slightly different (1-2 mismatches in the conserved positions) from the motifs of their previously characterized orthologs.

#### **Inference of novel regulons for metabolic pathways and chromosomal gene clusters**

To identify novel regulons in the absence of experimental data, we used two types of potentially co-regulated gene sets: i) genes that constitute functional metabolic pathways (subsystems); and ii) genes derived from conserved gene neighborhoods that include a putative TF gene. To analyze metabolic subsystems and conserved chromosomal gene clusters projected across bacterial genomes we used the SEED database [36]. Each training set of potentially co-regulated operons was collected from 16 analyzed *Shewanella* genomes, and a collection of their upstream regions was used as an input for the motif-recognition program SignalX to predict a common DNA motif allowing a limited number of sequences to be ignored. At the next step, the *Shewanella* genomes were scanned with the constructed DNA motif to reveal the distribution of similar sites that were further verified by the consistency check procedure (reviewed in [1]). Finally, the genomic context of candidate co-regulated genes was used to attribute a potential TF to each novel regulon and associated DNA motif.

As a result, we inferred 41 novel regulons in *Shewanella* spp. including: i) 18 regulons for metabolic subsystems; and ii) 23 regulons for conserved chromosomal gene clusters (Table 2). The metabolic regulons from the first group control genes from the metabolic pathways of utilization of various carbohydrates, as well as formate, lactate, propionate, hydroxyproline/proline, tyrosine, and branched chain amino acids, and the purine biosynthesis pathway. All of these metabolic regulons except the purine regulon were assigned to a TF by a combination of different evidence types such as (i) positional clustering of target genes and TFs on the chromosome; ii) autoregulation of a TF by a cognate TFBS; iii) correlation in the phylogenetic pattern of co-occurrence of TFBSs and TFs in the genomes. Each of these novel TFs was functionally annotated in the SEED database (<http://theseed.uchicago.edu>) and tentatively named using an abbreviation of the target metabolic pathway/genes. Hereinafter we mark the new names by asterisks.

Most of the novel metabolic TFs represent non-orthologous replacement of previously known TFs that control similar metabolic pathways in other lineages. For example, the propionate catabolism in the Enterobacteria is activated by the Fis-family regulator PrpR,

whereas in the *Shewanella* spp. it is predicted to be controlled by a GntR-family TF PrpR\*. The proline utilization is controlled by the Lrp-family activator PutR in the *Vibrio* spp. [37], the AraC-family activator PruR in the *Pseudomonas* spp. [38], and the predicted GntR-family regulator HypR\* in the *Shewanella* spp.. The homogentisate pathway of the tyrosine degradation is regulated by the IclR-type repressor HmgR in the *Pseudomonas* spp. [39], which is replaced by novel LysR-family regulator HmgR\* in the *Shewanella* spp.. Similar non-orthologous replacements of regulators have been detected for ten different carbohydrate catabolic pathways [33] and the lactate utilization system in the *Shewanella* spp. [30]. A novel purine-pathway regulon (named PUR\*) with hitherto unknown cognate TF was inferred in *Shewanella* instead of PurR regulon previously characterized in other  $\gamma$ -proteobacteria including *E. coli* [40] and missing in the *Shewanella* spp.. Two novel regulators PflR\* and XltR\* were predicted to control metabolic pathways of pyruvate to formate fermentation and xylitol catabolism, whose regulation have not yet been previously described in any bacteria.

Functional annotations of novel TF regulons that were deduced from the analysis of conserved gene clusters are largely hypothetical and incomplete. Most of them are local regulators controlling one or two target operons (Additional file 4). Two novel TF regulators from the Crp family, named DeoR\* and PnuR\*, control candidate phosphorylases and transporters likely involved in the nucleoside/nicotinamide ribose utilization. A novel AsnC-type regulator AzrR\* controls the *azr*-SO3586 operon, which encodes azoreductase and lactoylglutathione lyase that are likely involved in the superoxide stress protection. Novel regulator CalR\* controls expression of the coniferyl aldehyde dehydrogenase *calB* that play a role in phytochemical aromatic compound utilization. Other inferred TF regulons appear to contain various hypothetical metabolite efflux transporters or flavocytochromes potentially involved in detoxification and undescribed respiratory processes, respectively.

#### Identification of regulons for RNA regulatory elements

We used known regulatory-RNA patterns from the Rfam database [41] to scan intergenic regions in 16 *Shewanella* genomes and analyzed the genomic context of candidate regulatory RNAs (Additional file 4).

Representatives of eight metabolite-responsive riboswitch families are scattered in most *Shewanella* genomes. The lysine, glycine, thiamine, cobalamin, riboflavin, and molybdenum cofactor riboswitches control genes for the respective amino acid / cofactor biosynthetic pathways and/or uptake transporters. The purine riboswitch controls adenosine deaminase and purine transporter. The riboswitch that binds second messenger cyclic di-GMP was found to control various

subsets of genes in the *Shewanella* spp. including genes encoding extracellular proteins such as the chitin binding protein, chitinases, peptidases, and other hypothetical secreted proteins.

Six candidate attenuators that regulate operons responsible for the biosynthesis of branched chain amino acids, histidine, threonine, tryptophan, and phenylalanine in proteobacteria [42] are conserved in all analyzed *Shewanella* spp.

#### Experimental validation of N-acetylglucosamine-responsive regulon NagR in *S. oneidensis* MR-1

A predicted transcriptional regulator NagR of the LacI family is a nonorthologous replacement of the NagC repressor from Enterobacteria. In addition to genes involved in Nag transport (*nagP* and *omp*<sup>Nag</sup>) and biochemical conversion (*nagK-nagB*<sup>II</sup>-*nagA*), the reconstructed NagR regulon contains auxiliary components that are likely involved in chemotaxis and hydrolysis of chitin and/or chito oligosaccharides (*mcp*<sup>Nag</sup>-*hex* and *cbp*). Experimental validation of the reconstructed NagR regulon in *S. oneidensis* MR-1 was performed by both *in vitro* and *in vivo* approaches. The *nagR* gene was cloned and overexpressed in *E. coli*, and the recombinant protein was purified by Ni<sup>2+</sup>-chelating chromatography. We used electrophoretic mobility shift assay to test specific DNA-binding of the purified NagR protein to its predicted operator sites in upstream regions of the *nagP* (SO3503), *nagK* (SO3507), *mcp*<sup>Nag</sup> (SO3510), *omp*<sup>Nag</sup> (SO3514) and *cbp* (SO1072) genes in *S. oneidensis* MR-1. The maximal shift of the *nagK* DNA fragment observed at 100 nM NagR was completely suppressed by the addition of 20 mM of N-acetylglucosamine, which was thus proven as a negative effector (Additional file 6A). Specific binding at 100 nM NagR protein was also confirmed for the other four tested DNA fragments. To confirm the negative regulatory effect of NagR on gene expression *in vivo*, the *S. oneidensis* *nagR* targeted deletion mutant was constructed and relative transcript levels of the predicted NagR target genes were analyzed by quantitative RT-PCR. Relative mRNA levels of the *nagP*, *nagK*, *mcp*<sup>Nag</sup>, *omp*<sup>Nag</sup>, and *cbp* genes were elevated 15-, 50-, 16- 11-, and 5-fold, respectively, in the *nagR* mutant compared to the wild-type strain when grown in the minimal medium supplied with lactate (Additional file 6B). These results confirm that NagR is a negative regulator of the chitin utilization genes that are de-repressed in response to N-acetylglucosamine.

## Discussion

### Conservation and variations in the regulatory network evolution

Conservation of 5738 regulatory interactions identified for all predicted members of the reconstructed regulons

across the *Shewanella* genus is shown in Additional file 4. Overall, the regulatory systems of the *Shewanella* spp. appears out to be considerably variable within the genus and quite distinct from other previously studied  $\gamma$ -proteobacteria. The observed variations can be classified in three distinct types: (i) “regulon expansion” in the *Shewanella* compared to other lineages that can be ranged from additions of several regulon members to larger-scale shifts in the regulated metabolic pathways (e.g., HexR, PdhR, and TyrR regulons); (ii) “fuzzy regulons” when a regulon possess a conserved core and variable periphery within the *Shewanella* group (e.g., global regulons ArgR, Crp, Fur, NarP, and Fnr); (iii) “regulon loss or acquisition” when entire regulon (including all operons from a regulated pathway) is present only in some of the *Shewanella* species (e.g., for Dnr, ModE, BetI, and 17 regulons controlling various sugar utilization pathways [33]). Of course, this distinction is very schematic and in reality these types of behavior overlap. The mostly conserved regulatory interactions occur among TF regulons that are involved in the control of essential biosynthetic pathways (e.g., BirA, FabR, GlmR, IlvY, NrdR regulons), and universal stress responses (LexA and ZntR regulons).

To estimate the relative conservation of the predicted regulatory interactions in other lineages, we searched for orthologs of the putative regulon members in *E. coli* and compared the gene contents of the regulons reconstructed in the *Shewanella* and with orthologous regulons in *E. coli* captured in the RegulonDB database (Additional file 4). Similar analysis was performed for the *Shewanella* regulons characterized in the *Pseudomonas* spp. (but not in *E. coli*), including Dnr, GlmR, HexR, HutC, and PsrA (for references see Additional file 5). Among 468 cognate operons that belong to 42 studied regulons in the *Shewanella* spp., 138 operons (30%) have orthologous known targets in *E. coli* or *Pseudomonas*, 223 operons (~50%) lack orthologous operons, whereas the remaining 107 operons (~20%) have orthologous operons that are not under control of orthologous TFs in these species. Examples of impressive variations in the content of orthologous TF regulons in the *Shewanella* and *E. coli* are discussed below.

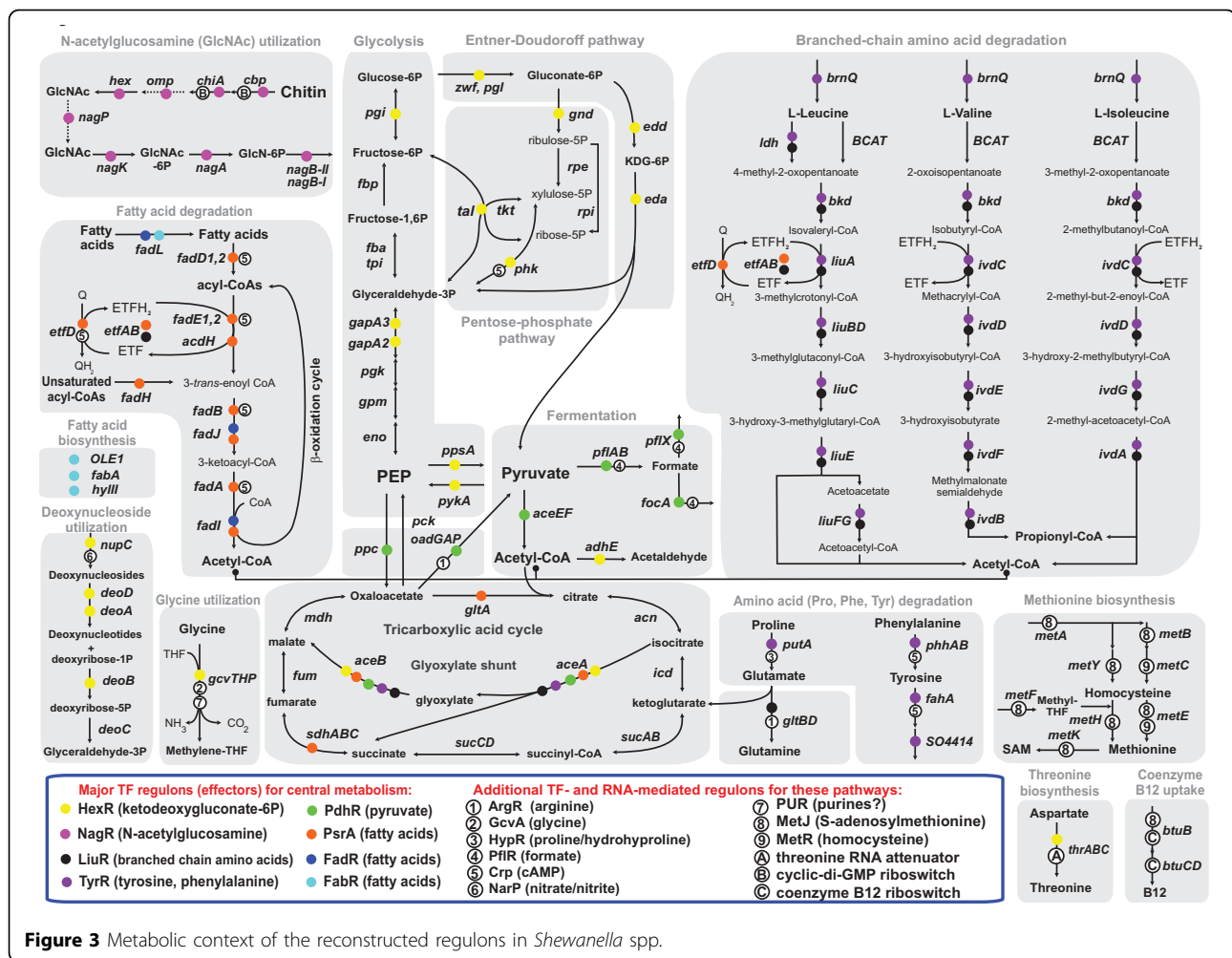
The comparison of the inferred regulons revealed striking differences in the strategies for regulation of the central carbohydrate and amino acid metabolism between the lineages comprising the *Shewanella* spp. and the Enterobacteria. In *E. coli*, two global regulators, FruR (fructose repressor/activator) and Crp (cAMP-responsive activator), control the central carbohydrate metabolism, whereas HexR (phospho-keto-deoxy-glucate-responsive repressor) and PdhR (pyruvate repressor) are local regulators of glucose-6P dehydrogenase and pyruvate dehydrogenase, respectively. By contrast,

the *Shewanella* spp. are predicted to use the HexR and PdhR regulators for the global control of the central carbohydrate metabolism and fermentation (Fig. 3). The FruR TF is absent in the *Shewanella* spp. that are not able to utilize fructose. The content and functional role of the Crp regulon is significantly different in the two lineages: the catabolism of carbohydrates and amino acids in the Enterobacteria, and the anaerobic respiration in the *Shewanella* spp. Most sugar catabolic pathways in the *Shewanella* spp. seem to be exclusively controlled by local sugar-responsive TFs that are often replaced by non-orthologous TFs (e.g., NagR vs. NagC for the N-acetylglucosamine utilization), and lack global co-regulation by Crp. Thus, the *Shewanella* spp. seem to lack many “feed-forward loops” that are characteristic for the regulation of sugar catabolism pathways in *E. coli* (when an operon is regulated by Crp and a local regulator that also is regulated by Crp) [43], thus may have a different strategy of sugar catabolism on mixed substrates.

Significant shifts in the regulon content were also identified for the TyrR, FadR, and FabR regulons (Fig. 3). In *E. coli*, the tyrosine- and phenylalanine-responsive regulator TyrR represses most aromatic amino acid biosynthetic enzymes and transporters encoded by multiple *aro* and *tyr* genes scattered on the chromosome, and activates the tyrosine transporter encoded by the *mtr* gene. In the *Shewanella* spp., we identified TyrR as a master regulator of the degradation pathways for various amino acids, including phenylalanine (*phhAB* operon), tyrosine (*fahA-maiA* operon), branched chain amino acids (*ldh*, *brnQ*, *liu*, *ivd*, and *bkd* operons), proline (*putA* gene), and oligopeptides (various peptidase genes), as well as some other pathways such as the glyoxylate shunt (*aceBA* operon), and the chorismate biosynthesis (*aroA* gene). These findings are in accordance with the previously established role of PhhR, a TyrR ortholog in *Pseudomonas* spp., as an activator for phenylalanine and tyrosine degradation genes [44]. The fatty acid degradation pathway in the *Shewanella* spp. and many other  $\gamma$ -proteobacteria is controlled by PsrA, whereas in the Enterobacteria the analogous pathway is regulated by FadR [4]. The *Shewanella* spp. also have a significantly reduced in size FadR regulon, which retains only two operons shared with the orthologous regulon of *E. coli*, *fadIJ* and *fadL*[4]. Finally, the fatty acid biosynthesis regulon FabR has only one gene, *fabA*, which has conserved regulation in both *E. coli* and the *Shewanella* spp., whereas the remaining target genes were identified as a lineage-specific regulon extension.

#### Interconnections between the predicted regulons in *Shewanella* spp

The collection of the inferred *Shewanella* regulons contains at least 30 regulons (for 24 TFs and 6 regulatory



**Figure 3** Metabolic context of the reconstructed regulons in *Shewanella* spp.

RNAs) that have at least one operon under simultaneous control of at least two regulators (Additional file 4). Most of the overlapping regulons control amino acid, fatty acid, nitrogen, and central carbohydrate metabolism (Fig. 3). The glyoxylate shunt operon *aceBA* controlled by five TFs is the most regulated operon in the current TRN model (see below). The glycine utilization operon *gcvTHP* was found to be controlled by the glycine-responsive regulator GcvA, the central carbohydrate regulator HexR, and the novel purine biosynthesis regulator PUR\*. In the predicted regulons, 14 operons are under overlapping control of three regulons, whereas ~70 operons are co-regulated by two regulons. At least four regulatory cascades between various TFs were identified in the *Shewanella* spp.: LiuR for *tyrR*, NarP for *crp*, Crp for *hmgR*, and MetJ for *metR*, and only the latter cascade is conserved in *E. coli*.

The reconstructed TRN provides insight into interplay between several different TFs controlling multiple genes from the LiuR regulon (Fig. 4). LiuR is a MerR-family repressor that controls the branched chain amino acid

(Ile/Leu/Val) utilization in diverse proteobacteria [4]. In *Shewanella* spp., the predicted LiuR regulon was found to regulate Ile/Leu/Val operons (*ldh*, *liu*, *ivd*, and *bkd*) and was expanded by additional members involved in the biosynthesis of glutamate (*gltBD*) and threonine (*thrABC*), and the glyoxylate shunt (*aceBA*). Six out of nine LiuR-controlled operons are also regulated by the tyrosine/phenylalanine-responsive transcription factor TyrR [45]. Although TyrR in *E. coli* can act both as activator and repressor on its target genes, the mode of TyrR action on *Shewanella* targets is to be determined experimentally. Preliminary comparative analysis of relative positions of the TyrR- and LiuR-binding sites in *Shewanella* genomes (using multiple alignment of the promoter gene regions) suggests that TyrR probably acts as an activator for the *ldh*, *liu*, *ivd*, and *bkd* operons (data not shown). This supposition suggests that integrative effect of the LiuR and TyrR mediated control can be activation of their target genes in the simultaneous presence of Ile/Leu/Val and Tyr/Phe. Indeed, the expression data confirm strong up-regulation of the Ile/Leu/



validation of the global carbohydrate metabolism regulon HexR and the assessment of its physiological role in *Shewanella* will be published elsewhere.

This work demonstrates the power of the comparative genomics approach in application to the reconstruction of transcriptional regulons in poorly studied groups of related bacteria. The reference set of the *Shewanella* regulons is the first taxonomy-wide collection of regulons obtained by this approach. It can be assessed in the RegPrecise database [34]. We anticipate a fast growth of taxonomy-wide regulon collections for other lineages in the near future. Regulatory interactions from the reconstructed regulons will provide an additional regulatory constrains for the recently published metabolic model of *S. oneidensis* MR-1 [22], allowing one to build an integrated model of metabolism and regulation. Such integrated model can be used for phenotype prediction, functional gene assignment and understanding of organism ecology. Finally, the reconstructed regulons were useful for the genome context-based prediction of novel functions of enzymes and transporters in previously uncharacterized carbohydrate utilization pathways in *Shewanella* spp. [33]

## Methods

### Bioinformatics methods for regulon reconstruction and used databases

The *Shewanella* spp. genomes were downloaded from the Genbank [49] (Fig. 1). The set of predicted DNA-binding TFs was extracted from the DBD database [50]. The *locus\_tag* gene identifiers are used throughout. Orthologous proteins in 16 *Shewanella* genomes were defined in the previous work by the best bidirectional hits criterion [32]. Orthologous groups in *Shewanella* were named by either a common name of characterized protein, a novel name for proteins functionally annotated in this study, or by a *locus\_tag* from *S. oneidensis* genome for uncharacterized proteins. Orthologs between proteins from different taxonomic groups (e.g. *Shewanella* and other  $\gamma$ -proteobacteria) were defined as bidirectional best hits with 30% of identity threshold using the Smith-Waterman algorithm implemented in the GenomeExplorer program [51]. In dubious cases orthologs were confirmed by construction of phylogenetic trees and comparative analysis of gene neighborhoods using the MicrobesOnline tree browse tool [47]. Functional gene assignments and metabolic subsystem analysis were performed using the SEED annotation/analysis tool <http://theseed.uchicago.edu/FIG/index.cgi>, which combines protein similarity search, positional gene clustering, and phylogenetic profiling of genes [36]. In addition, the InterPro [52], and PFAM [53] databases were used to verify protein functional and structural annotations.

For *de novo* identification of a candidate regulatory motif in the training set of potential upstream regions of genes (intergenic regions up to 350 bp) we used a simple iterative procedure DNA motif detection procedure implemented in the program SignalX [54]. Weak palindromes were selected in each region. Each palindrome was compared to all other palindromes, and the palindromes most similar to the initial one were used to make a profile. The candidate site score was defined as the sum of the respective positional nucleotide weights [7]. These profiles were used to scan the set of palindromes again, and the procedure was iterated until convergence. Thus a set of PWM profiles was constructed. A profile with largest information content was used as the recognition rule [55]. Each genome encoding the studied TF was scanned with the constructed motif profile using the GenomeExplorer software [51] and genes with candidate regulatory sites in the upstream regions were selected. 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 established regulon members. Additional and more detailed description of various scenario for regulon reconstruction using comparative genomics was reviewed in [1]. Analysis of large regulons (Fur, Crp, Fnr, NarP, LexA) was carried out using the web-based tool RegPredict allowing the comparative genomics-based regulon inference <http://regpredict.lbl.gov>[56]. The details of reconstructed regulons were captured and displayed in our recently developed database RegPrecise <http://regprecise.lbl.gov>[34]. For identification of RNA regulatory motif sequences we scanned complete genomes using tools and profiles available from the Rfam database [41]. Calculation of the Pearson coefficient for the LiuR-regulated genes was done by tools available at the MicrobesOnLine resource [47].

### Experimental methods for regulon validation

The *nagR* (SO3516) gene cloned at a pET-derived vector containing the T7 promoter and His<sub>6</sub> tag [57] was kindly provided by Frank Collart (Argonne National Laboratory, IL).

**Protein purification.** Recombinant proteins of *nagR* (SO3516) from *S. oneidensis* MR-1 was overexpressed as N-terminal fusion with a His<sub>6</sub> tag in *E. coli* strain BL21/DE3. Cells were grown on LB media to OD<sub>600</sub> = 0.8 at 37°C, induced by 0.2mM IPTG, and harvested after 12 h shaking at 20°C. Protein purification was performed using rapid Ni-NTA agarose minicolumn protocol as described [58]. Briefly, harvested cells were resuspended in 20 mM HEPES buffer pH 7 containing 100 mM

NaCl, 0.03% Brij 35, and 2 mM  $\beta$ -mercaptoethanol supplemented with 2 mM phenylmethylsulfonyl fluoride and a protease inhibitor cocktail (Sigma-Aldrich). Lysozyme was added to 1 mg/mL, and the cells were lysed by freezing-thawing followed by sonication. After centrifugation at 18,000 rpm, the Tris-HCl buffer (pH 8) was added to the supernatant (50 mM, final concentration), and it was loaded onto a Ni-NTA agarose column (0.2 ml). After washing with the starting buffer containing 1 M NaCl and 0.3% Brij-35, bound proteins were eluted with 0.3 ml of the starting buffer containing 250 mM imidazole. Protein size, expression level, distribution between soluble and insoluble forms, and extent of purification were monitored by SDS-PAGE.

**qPCR.** In-frame deletion mutagenesis of or *nagR* (*SO3516*) was performed using previously published method [33]. Genomic RNA was isolated from *S. oneidensis* MR-1 and  $\Delta$ *nagR* cells grown in minimal medium supplied with lactate and collected at O.D.(600) of 0.52 using the RNA purification kit from Promega (Madison, WI). Reverse transcription of total RNA was performed with random primers using iScript cDNA synthesis kit from BIO-RAD (Hercules, CA), following kits instructions. qPCR was performed using SYBR GreenER qPCR SuperMix Universal kit from Invitrogen (Carlsbad, CA). Transcript levels of the *nagP* (*SO3503*), *nagK* (*SO3507*), *mcp*<sup>Nag</sup> (*SO3510*), *omp*<sup>Nag</sup> (*SO3514*), *cbp* (*SO1072*), *SO0854*, and *zwf* (*SO2489*, used as a negative control) genes were measured and the results were normalized to the expression level of 16S mRNA. Fold change was calculated by the  $2^{-\Delta\Delta CT}$  method [59] as a ratio of normalized mRNA levels in  $\Delta$ *nagR* mutant and wild-type MR-1 strains.

**DNA-binding assay.** Interaction of purified recombinant protein NagR (*SO3516*) from *S. oneidensis* MR-1 with their cognate DNA motifs was assessed by EMSA technique using the following dsDNA segments obtained by PCR amplification or by custom synthesis of both complementary oligonucleotides (IDT, San Diego, CA), annealing and purification. One of the primers was 5'-biotinylated (IDT). By using *S. oneidensis* MR-1 DNA as the template, we amplified DNA fragments from the following upstream gene regions: *SO1072* (89 bp), *SO3507* (69 bp), *SO3510* (64 bp), *SO3514* (69 bp), *SO3503* (62 bp), *SO0854* (67 bp). For EMSA, the biotin-labeled DNA (0.1 or 1 nM) was incubated with the increasing amount of purified NagR (0-100 nM) in a total volume of 20  $\mu$ l. The binding buffer contains Tris-HCl 20mM, KCl 150mM, MgCl<sub>2</sub> 5mM, DTT 1mM, EDTA 1mM, 0.05% NP-40, 2.5% glycerol. The poly(dI-dC) (Sigma) was added as nonspecific competitor DNA at  $\sim 10^4$ -fold molar excess over labeled target DNA to reduce nonspecific binding. After 25 min incubation at room temperature, the reaction mixtures

were separated by electrophoresis on a 5% native polyacrylamide gel in 0.5  $\times$  Tris-borate-EDTA for 90 min at 90V, at 4°C. The gel was transferred by electrophoresis (30 min, at 380 mA) onto a nylon membrane (Pierce, Rockford, Ill.) and fixed by UV cross-linking. Biotin-labeled DNA was detected with the LightShift Chemiluminescent EMSA kit (Pierce, Rockford, Ill.), as recommended by the manufacturer. The effect of *N*-acetyl-glucosamine on NagR-DNA binding was tested by addition of 20 mM of *N*-acetylglucosamine to the incubation mixture.

## Additional material

**Additional file 1: Rodionov\_AF1.xls - Repertoire of DNA-binding transcription factors identified in 16 genomes of *Shewanella* spp.**

**Additional file 2: Rodionov\_AF2.xls - Transcription factors universally conserved in *Shewanella* spp.**

**Additional file 3: Rodionov\_AF3.xls - Functional classification of 73 transcription factors conserved between *E. coli* and 16 *Shewanella* strains.**

**Additional file 4: Rodionov\_AF4.xls - Functional content and conservation of reconstructed TF and RNA regulons in *Shewanella* genomes.**

**Additional file 5: Rodionov\_AF5.pdf - Comparison of predicted TFBS motifs in *Shewanella* spp. to the previously characterized orthologous regulators in model species.**

**Additional file 6: Rodionov\_AF6.pdf - Experimental validation of NagR regulon in *S. oneidensis* MR-1. (A) Electrophoretic mobility shift assays to assess NagR protein binding to their predicted DNA operators; (B) Differential regulation of NagR controlled genes determined by quantitative qPCR.**

## List of abbreviations used

TF: transcription factor; TFBS: transcription factor-binding site; TRN: transcriptional regulatory network.

## Acknowledgments

This research was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contracts DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory (ENIGMA SFA), DE-AC05-76RLO with Pacific Northwest National Laboratory (SBR FSFA); and DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Laboratory. Additional funding was provided by National Science Foundation (DBI-0850546 to D.A.R. and R.O.); Russian Foundation for Basic Research (08-04-01000 to A.E.K., 09-04-92745 and 10-04-00431 to M.S.G., 10-04-01768 to D.A.R., E.D.S. by 09-04-92742), Russian Academy of Sciences (program 'Molecular and Cellular Biology' to D. A.R. and M.S.G.); Russian Agency on Education (P2581 to E.D.S.); Russian Science Agency (2.740.11.0101 to M.S.G.). This article has been published as part of *BMC Genomics* Volume 12 Supplement 1, 2011: Validation methods for functional genome annotation. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2164/12?issue=S1>.

## Author details

<sup>1</sup>Sanford-Burnham Medical Research Institute, La Jolla, California, USA.

<sup>2</sup>Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia. <sup>3</sup>Lawrence Berkeley National Laboratory, Berkeley, California, USA. <sup>4</sup>Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia. <sup>5</sup>State Scientific Center GosNII Genetika, Moscow, Russia. <sup>6</sup>Fellowship for Interpretation of Genomes, Burr Ridge, Illinois, USA.

<sup>7</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, USA. <sup>8</sup>Department of Energy Joint Genome Institute, Walnut Creek, California, USA.

#### Authors' contributions

DARo, ALO and MSG conceived and supervised the research, and wrote the manuscript. DARo, IAR, DARa, AEK, EAP, AVG, ONL, GYK performed comparative genomic analysis to infer novel transcription factor regulons. MDK identified riboswitches. PSN developed RegPredict tool and RegPrecise database, performed propagation of regulons. DARo and PSN performed correlation analysis using expression data. EDS and RO performed computational similarity searches and gene annotation in the SEED database. MFR produced manually annotated table of orthologs and provided targeted gene knockout strains in *Shewanella*. XL carried out validation experiments for NagR regulon. JKF, ID and APA contributed to the development of the manuscript and design of the study. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Published: 15 June 2011

#### References

- Rodionov DA: Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* 2007, **107**:3467-3497.
- Minchin SD, Busby SJ: Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* 2009, **47**:6-12.
- Grainger DC, Lee DJ, Busby SJ: Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria. *Curr Opin Microbiol* 2009, **12**:531-535.
- Kazakov AE, Rodionov DA, Alm E, Arkin AP, Dubchak I, Gelfand MS: Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. *J Bacteriol* 2009, **191**:52-64.
- Laikova ON, Mironov AA, Gelfand MS: Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria. *FEMS Microbiol Lett* 2001, **205**:315-322.
- Makarova KS, Mironov AA, Gelfand MS: Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol* 2001, **2**: RESEARCH0013.
- Mironov AA, Koonin EV, Roytberg MA, Gelfand MS: Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res* 1999, **27**:2981-2989.
- Panina EM, Mironov AA, Gelfand MS: Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res* 2001, **29**:5195-5206.
- Permina EA, Kazakov AE, Kalinina OV, Gelfand MS: Comparative genomics of regulation of heavy metal resistance in Eubacteria. *BMC Microbiol* 2006, **6**:49.
- Ravcheev DA, Gerasimova AV, Mironov AA, Gelfand MS: Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae). *BMC Genomics* 2007, **8**:54.
- Rodionov DA, De Ingeniis J, Mancini C, Cimadamore F, Zhang H, Osterman AL, Raffaelli N: Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. *Nucleic Acids Res* 2008, **36**:2047-2059.
- Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS: Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Comput Biol* 2005, **1**:e55.
- Rodionov DA, Gelfand MS: Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling. *Trends Genet* 2005, **21**:385-389.
- Rodionov DA, Mironov AA, Gelfand MS: Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* 2002, **12**:1507-1516.
- Rodionov DA, Mironov AA, Rakhmaninova AB, Gelfand MS: Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria. *Mol Microbiol* 2000, **38**:673-683.
- Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007, **5**:e8.
- Liu J, Xu X, Stormo GD: The cis-regulatory map of *Shewanella* genomes. *Nucleic Acids Res* 2008, **36**:5376-5390.
- Fredrickson JK, Romine MF, Beliaev AS, Auchtung JM, Driscoll ME, Gardner TS, Nealson KH, Osterman AL, Pinchuk G, Reed JL, et al: Towards environmental systems biology of *Shewanella*. *Nat Rev Microbiol* 2008, **6**:592-603.
- Hau HH, Gralnick JA: Ecology and biotechnology of the genus *Shewanella*. *Annu Rev Microbiol* 2007, **61**:237-258.
- Driscoll ME, Romine MF, Juhn FS, Serres MH, McCue LA, Beliaev AS, Fredrickson JK, Gardner TS: Identification of diverse carbon utilization pathways in *Shewanella oneidensis* MR-1 via expression profiling. *Genome Inform* 2007, **18**:287-298.
- Gupta N, Tanner S, Jaitly N, Adkins JN, Lipton M, Edwards R, Romine M, Osterman A, Bafna V, Smith RD, Pevzner PA: Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res* 2007, **17**:1362-1377.
- Pinchuk GE, Hill EA, Geydebrekht OV, De Ingeniis J, Zhang X, Osterman AL, Scott JH, Reed JL, Romine MF, Konopka AE, et al: Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: a tool for data analysis and hypothesis generation. *PLoS Comput Biol* 2010, **6**:e1000822.
- Gralnick JA, Brown CT, Newman DK: Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis*. *Mol Microbiol* 2005, **56**:1347-1357.
- Gao H, Wang X, Yang ZK, Palzkill T, Zhou J: Probing regulon of ArcA in *Shewanella oneidensis* MR-1 by integrated genomic analyses. *BMC Genomics* 2008, **9**:42.
- Saffarini DA, Schultz R, Beliaev A: Involvement of cyclic AMP (cAMP) and cAMP receptor protein in anaerobic respiration of *Shewanella oneidensis*. *J Bacteriol* 2003, **185**:3668-3671.
- Beliaev AS, Thompson DK, Fields MW, Wu L, Lies DP, Nealson KH, Zhou J: Microarray transcription profiling of a *Shewanella oneidensis* etrA mutant. *J Bacteriol* 2002, **184**:4612-4616.
- Bordi C, Ansaldi M, Gon S, Jourlin-Castelli C, Iobbi-Nivol C, Mejean V: Genes regulated by TorR, the trimethylamine oxide response regulator of *Shewanella oneidensis*. *J Bacteriol* 2004, **186**:4502-4509.
- Wan XF, Verberkmoes NC, McCue LA, Stanek D, Connelly H, Hauser LJ, Wu L, Liu X, Yan T, Leaphart A, et al: Transcriptomic and proteomic characterization of the Fur modulon in the metal-reducing bacterium *Shewanella oneidensis*. *J Bacteriol* 2004, **186**:8385-8400.
- Yang Y, Harris DP, Luo F, Wu L, Parsons AB, Palumbo AV, Zhou J: Characterization of the *Shewanella oneidensis* Fur gene: roles in iron and acid tolerance response. *BMC Genomics* 2008, **9**(Suppl 1):S11.
- Pinchuk GE, Rodionov DA, Yang C, Li X, Osterman AL, Deryn E, Geydebrekht OV, Reed SB, Romine MF, Collart FR, et al: Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* 2009, **106**:2874-2879.
- Yang C, Rodionov DA, Li X, Laikova ON, Gelfand MS, Zagnitko OP, Romine MF, Obraztsova AY, Nealson KH, Osterman AL: Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J Biol Chem* 2006, **281**:29872-29885.
- Konstantinidis KT, Serres MH, Romine MF, Rodrigues JL, Auchtung J, McCue LA, Lipton MS, Obraztsova A, Giometti CS, Nealson KH, et al: Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proc Natl Acad Sci U S A* 2009, **106**:15909-15914.
- Rodionov DA, Yang C, Li X, Rodionova IA, Wang Y, Obraztsova AY, Zagnitko O, Overbeek R, Romine MF, Reed S, et al: Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 2010, **11**:494.
- Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA: RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res* 2010, **38**:D111-118.
- Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penalzoza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muniz-Rascado L, Martinez-Flores I, Salgado H, et al: RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 2008, **36**:D120-124.



36. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, et al: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Res* 2005, **33**:5691-5702.
37. Lee JH, Choi SH: **Coactivation of *Vibrio vulnificus* putAP operon by cAMP receptor protein and PutR through cooperative binding to overlapping sites.** *Mol Microbiol* 2006, **60**:513-524.
38. Nakada Y, Nishijyo T, Itoh Y: **Divergent structure and regulatory mechanism of proline catabolic systems: characterization of the putAP proline catabolic operon of *Pseudomonas aeruginosa* PAO1 and its regulation by PruR, an AraC/XylS family protein.** *J Bacteriol* 2002, **184**:5633-5640.
39. Arias-Barrau E, Olivera ER, Luengo JM, Fernandez C, Galan B, Garcia JL, Diaz E, Minambres B: **The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*.** *J Bacteriol* 2004, **186**:5062-5077.
40. Ravcheev DA, Gelfand MS, Mironov AA, Rakhmanova AB: **[Purine regulon of gamma-proteobacteria: a detailed description].** *Genetika* 2002, **38**:1203-1214.
41. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A: **Rfam: updates to the RNA families database.** *Nucleic Acids Res* 2009, **37**: D136-140.
42. Vitreschak AG, Lyubetskaya EV, Shirshin MA, Gelfand MS, Lyubetsky VA: **Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis.** *FEMS Microbiol Lett* 2004, **234**:357-370.
43. Gonzalez Perez AD, Gonzalez Gonzalez E, Espinosa Angarica V, Vasconcelos AT, Collado-Vides J: **Impact of Transcription Units rearrangement on the evolution of the regulatory network of gamma-proteobacteria.** *BMC Genomics* 2008, **9**:128.
44. Panina EM, Vitreschak AG, Mironov AA, Gelfand MS: **Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria.** *J Mol Microbiol Biotechnol* 2001, **3**:529-543.
45. Palmer GC, Palmer KL, Jorth PA, Whiteley M: **Characterization of the *Pseudomonas aeruginosa* transcriptional response to phenylalanine and tyrosine.** *J Bacteriol* 2010, **192**:2722-2728.
46. Paul L, Mishra PK, Blumenthal RM, Matthews RG: **Integration of regulatory signals through involvement of multiple global regulators: control of the *Escherichia coli* gltBDF operon by Lrp, IHF, Crp, and ArgR.** *BMC Microbiol* 2007, **7**:2.
47. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, et al: **MicrobesOnline: an integrated portal for comparative and functional genomics.** *Nucleic Acids Res* 2010, **38**:D396-400.
48. Cozzone AJ: **Regulation of acetate metabolism by protein phosphorylation in enteric bacteria.** *Annu Rev Microbiol* 1998, **52**:127-164.
49. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic Acids Res* **38**:D46-D51.
50. Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA: **DBD—taxonomically broad transcription factor predictions: new content and functionality.** *Nucleic Acids Res* 2008, **36**:D88-92.
51. Mironov AA, Vinokurova NP, Gelfand MS: **GenomeExplorer: software for analysis of complete bacterial genomes.** *Mol Bio l(Mosk)* 2000, **34**:222-231.
52. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al: **InterPro: the integrative protein signature database.** *Nucleic Acids Res* 2009, **37**:D211-215.
53. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**:D211-222.
54. Gelfand MS, Koonin EV, Mironov AA: **Prediction of transcription regulatory sites in Archaea by a comparative genomic approach.** *Nucleic Acids Res* 2000, **28**:695-705.
55. Schneider TD, Stormo GD, Gold L, A E: **Information content of binding sites on nucleotide sequences.** *J Mol Biol* 1986, **188**:415-431.
56. Novichkov PS, Rodionov DA, Stavrovskaya ED, Novichkova ES, Kazakov AE, Gelfand MS, Arkin AP, Mironov AA, Dubchak I: **RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach.** *Nucleic Acids Res* 2010, **38**:W299-307.
57. Lin CT, Moore PA, Auberry DL, Landorf EV, Peppler T, Victry KD, Collart FR, Kery V: **Automated purification of recombinant proteins: combining high-throughput with high yield.** *Protein Expr Purif* 2006, **47**:16-24.
58. Osterman AL, Lueder DV, Quick M, Myers D, Canagarajah BJ, Phillips MA: **Domain organization and a protease-sensitive loop in eukaryotic ornithine decarboxylase.** *Biochemistry* 1995, **34**:13431-13436.
59. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**:402-408.

doi:10.1186/1471-2164-12-S1-S3

**Cite this article as:** Rodionov et al.: Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics* 2011 **12**(Suppl 1):S3.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

