**Title: INTEGRATED APPROACH TO RECONSTRUCTION OF MICROBIAL REGULATORY NETWORKS**

**Award Number:** DE-SC0004999

**Sponsor / Agency:** Department of Energy

**Project Period:** 08/15/2010 – 08/14/2013

**Final Scientific / Technical Report**

---

This project had the goal(s) of development of integrated bioinformatics platform for genome-scale inference and visualization of transcriptional regulatory networks (TRNs) in bacterial genomes.

The work was done in Sanford-Burnham Medical Research Institute (SBMRI, P.I. D.A. Rodionov) and Lawrence Berkeley National Laboratory (LBNL, co-P.I. P.S. Novichkov).

The developed computational resources include: (1) RegPredict web-platform for TRN inference and regulon reconstruction in microbial genomes, and (2) RegPrecise database for collection, visualization and comparative analysis of transcriptional regulons reconstructed by comparative genomics. These analytical resources were selected as key components in the DOE Systems Biology KnowledgeBase (SBKB). The high-quality data accumulated in RegPrecise will provide essential datasets of reference regulons in diverse microbes to enable automatic reconstruction of draft TRNs in newly sequenced genomes. Below we outline our progress toward the three aims of this grant proposal.

---

**SPECIFIC AIMS:**

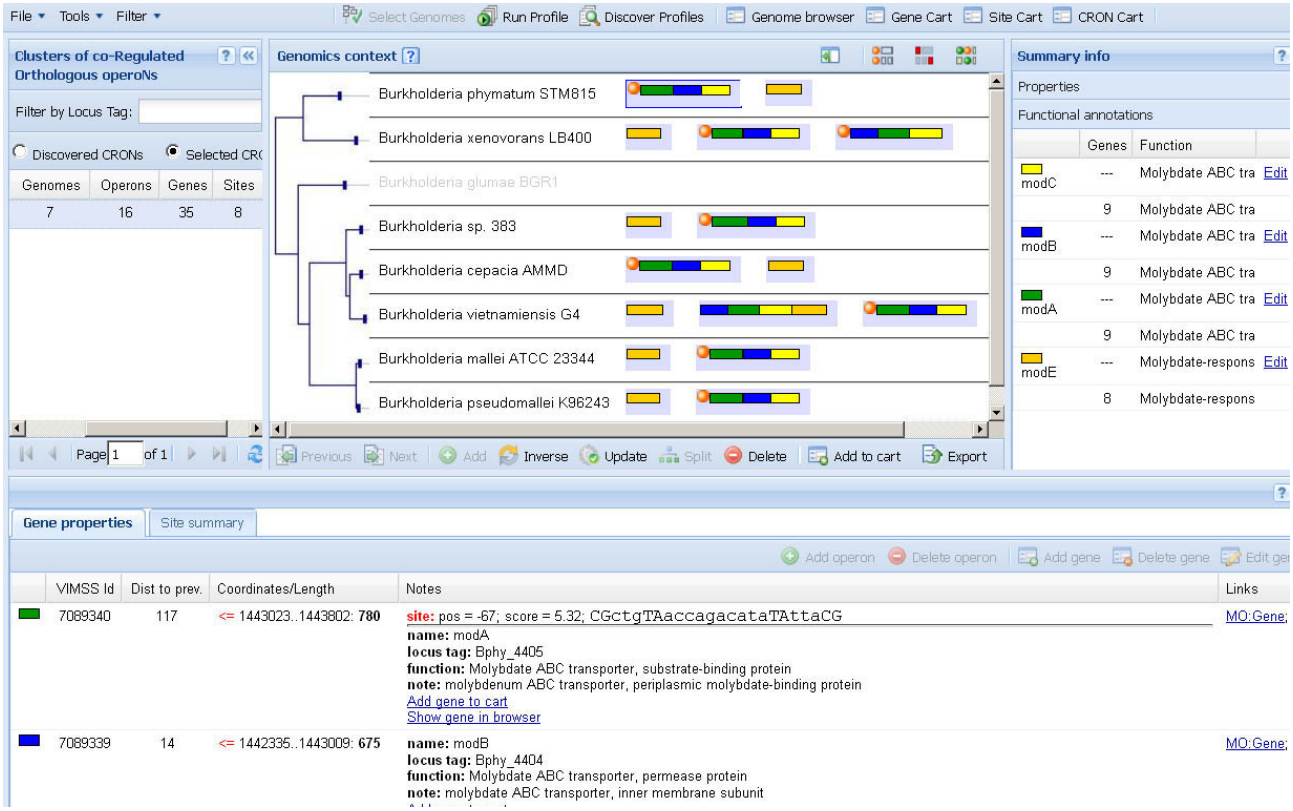
**AIM 1. Develop integrated platform for genome-scale regulon reconstruction.**

***Objective 1.a***. We developed the publically available platform RegPredict for *fast* and *accurate* inference of reference sets of regulons in well-populated groups of closely related microbial genomes. The RegPredict platform utilizes the comparative genomics approach and integrates multiple computational modules required for detailed reconstruction of regulatory interactions between transcription factors, *cis*-acting DNA binding sites and RNA regulatory elements and their target genes in microbial genomes.

A key new concept of RegPredict is a cluster of co-regulated orthologous operons (CRON) that allows prediction and analysis of regulatory motifs simultaneously in a set of taxonomically related genomes in a semi-automated way. Each CRON accumulates all pros and cons required to distinguish true regulatory sites from false positives. We

developed an automatic procedure which calculates and automatically prioritizes all predicted CRONs in a given set of genomes. A user-friendly graphical interface delivers all required information for each predicted CRON including location of predicted regulatory sites, groups of orthologous genes, operonal organization of genes, and functional gene annotations. In addition, the RegPredict interface provides a set of cross-references to respective gene pages in external web-resources, such as MicrobesOnline and the SEED, allowing a user to collect reliable information on gene function, protein domains, and gene neighbourhoods. Furthermore, in the Genomic context module of RegPredict, we developed an automatic visualization of the microbial phylogenetic species tree applied to a set of analyzed genomes. These essential features implemented in the RegPredict web-server (Fig. 1) allow fast and effective prediction of true positive regulatory interactions and regulon reconstruction.

**Fig. 1. Interactive user interface of the RegPredict web-server.**




***Objective 1.b.*** We developed and integrated into RegPredict several major *workflows* for genomics-based regulon inference: (A) *regulon reconstruction* for a known regulatory motif in a set of reference genomes from a particular taxonomic group of organisms; (B) *ab initio prediction* of novel regulons using several scenarios for generation of starting gene sets; (C) *conservative propagation* of reconstructed regulons to all other microbial genomes in the same taxonomic group.

(A) In order to capture existing regulatory network information available for experimentally analyzed species, project regulons to other species in the same taxonomic group, and expand possible targets of known transcription factors we developed the workflow A. For each transcription factor (TF), we started from the construction of an initial positional weight matrix (PWM) using the training set of known TF-binding sites (TFBSs) in a model organism available from literature or specialized databases (e.g., RegTransBase, RegulonDB, TBTBS). Then all analyzed genomes that possess an orthologous TF are scanned by the initial PWM profile resulting in identification of candidate TFBSs for genes that are orthologous to the previously known target genes in a model genome. In the next step, a collection of upstream regions of genes with candidate TFBSs identified during the initial step in all genomes is used to build a refined PWM profile for a studied TF regulon. Finally, the regulon is reconstructed by applying this refined PWM to the analyzed genomes using 'Run Profile' procedure in RegPredict. As result, the inferred regulon includes set of CRONs that are curated by taking into consideration individual scores and overall conservation of TFBSs across the genomes, as well as the genomic and functional gene contexts.

(B) For *ab initio* prediction of regulons when there is no available PWM and a regulatory motif is unknown we developed the workflow B. The procedure starts with a set of potentially co-regulated genes from one or multiple related genomes from a particular taxonomic group. The input training set might be composed based on many sources including i) genes that make up a functional pathway; ii) genes homologous to known regulons from model species; iii) genes derived from conserved chromosomal loci or operons containing orthologous TF genes; and iv) genes with similar expression profile as determined by microarray experiments. Identification of candidate TFBS motifs within a set of DNA upstream regions for genes from the training set and construction of corresponding PWM profiles is implemented in the "Discover Profiles" procedure in RegPredict. This procedure uses the common approach to find profiles of different types, such as palindromes of different length, or direct repeats, using the MEME-like iterative algorithm.

(C) To enable capturing of transcriptional regulatory networks for all sequenced genomes within the analyzed reference taxonomic collections, we developed a workflow for automatic conservative propagation of regulatory interactions to any novel taxonomically related genome. To project a particular regulon to a target genome, we require the presence of a TF ortholog and use a specific PWM to search for candidate TFBSs in upstream regions of genes being orthologous to one of the previously described members of the regulon. The propagation workflow was applied to 640 genomes from 14 taxonomic groups with available genome-wide collections of reference TF regulogs (see Aim 2).

In summary, we developed a universal pipeline allowing the genomic-based strategy for reconstruction of TRNs in any taxonomic group of microorganisms (Fig. 2). This strategy includes the three above described workflows implemented in the RegPredict web-server and the RegPrecise database. The reconstructed TF-specific regulogs (sets of regulons for orthologous TFs) and PWMs are amenable for deposition into the RegPrecise database using the individual sign-on capability. The pipeline was applied to obtain large reference sets of regulons in 10 taxonomic groups (see Aim 2).

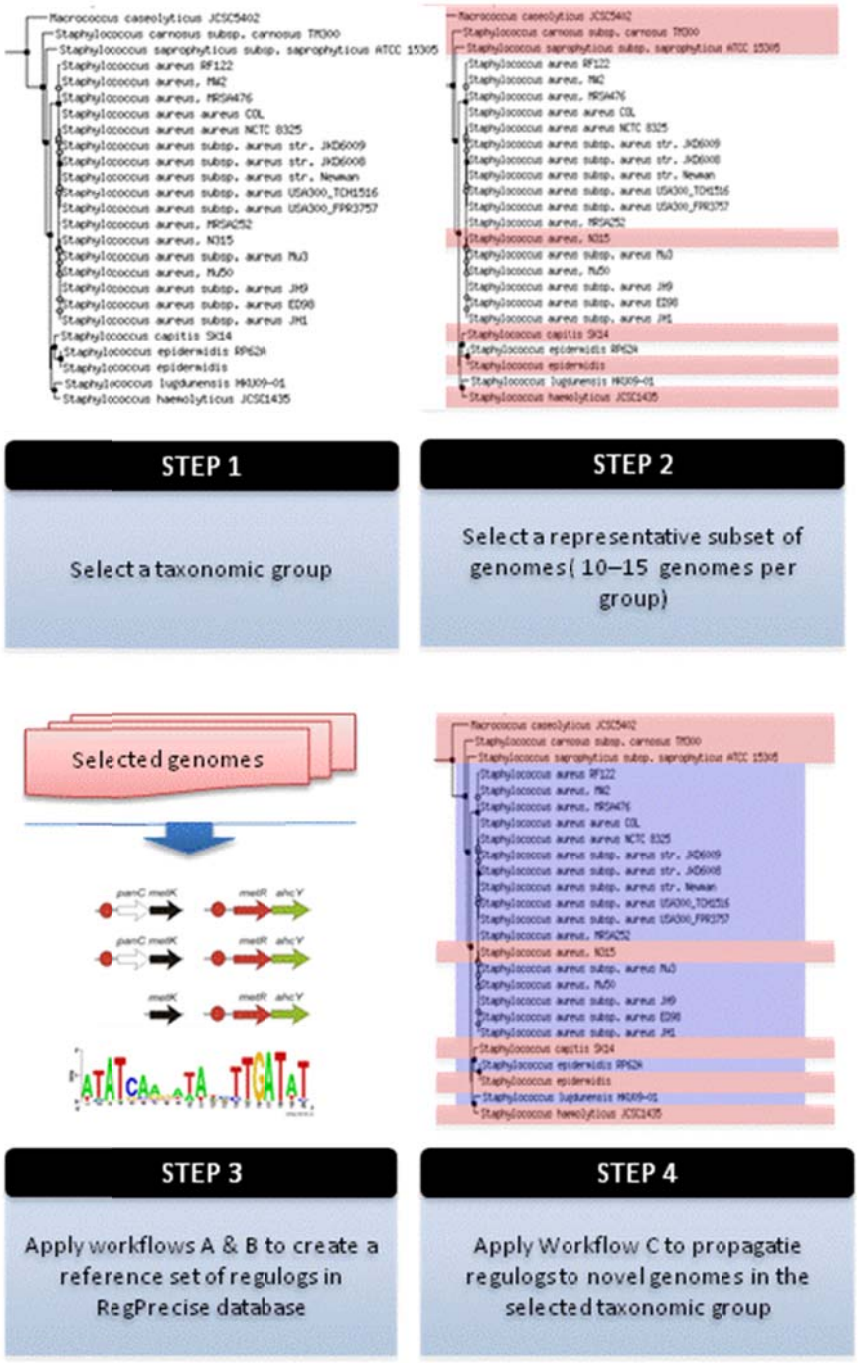
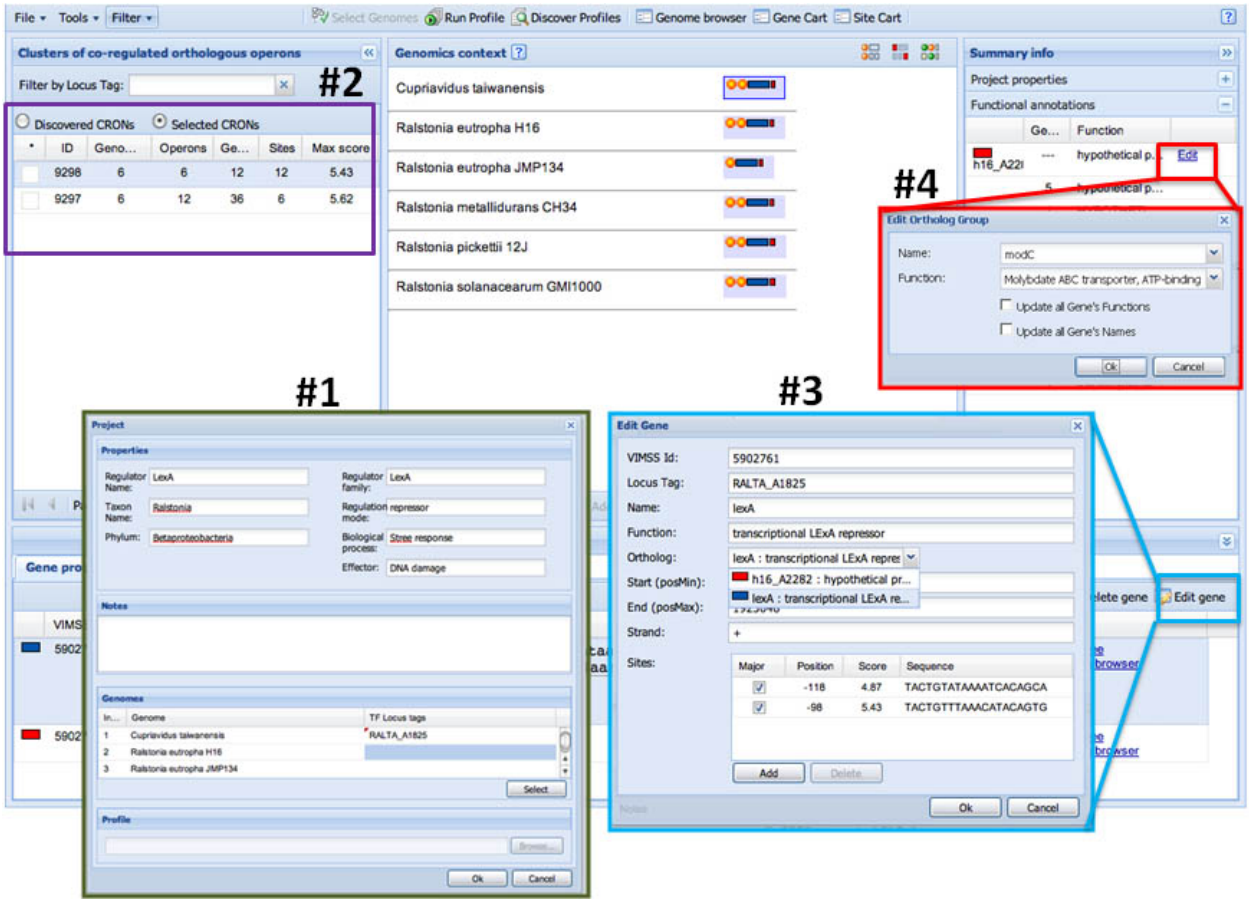



**Fig. 2. Genomics-based strategy for regulon inference in a taxonomic group.**

***Objective 1.c.*** We developed modules for the RegPredict allowing regulog inference, extensive curation, and submission to the RegPrecise database (Fig. 3). These modules allow creating a new regulog entry (#1), searching genomes and building CRONs (#2), inclusion of selected CRONs to a regulog, curation of putative operons, binding sites, and orthologs in a CRON (#3), and curation of functional annotations and names (#4). Both web resources are tightly integrated with each other: regulons inferred using the RegPredict platform are a major source of data input into the RegPrecise database; and conversely, the platform is used for curation of deposited in the database regulons.

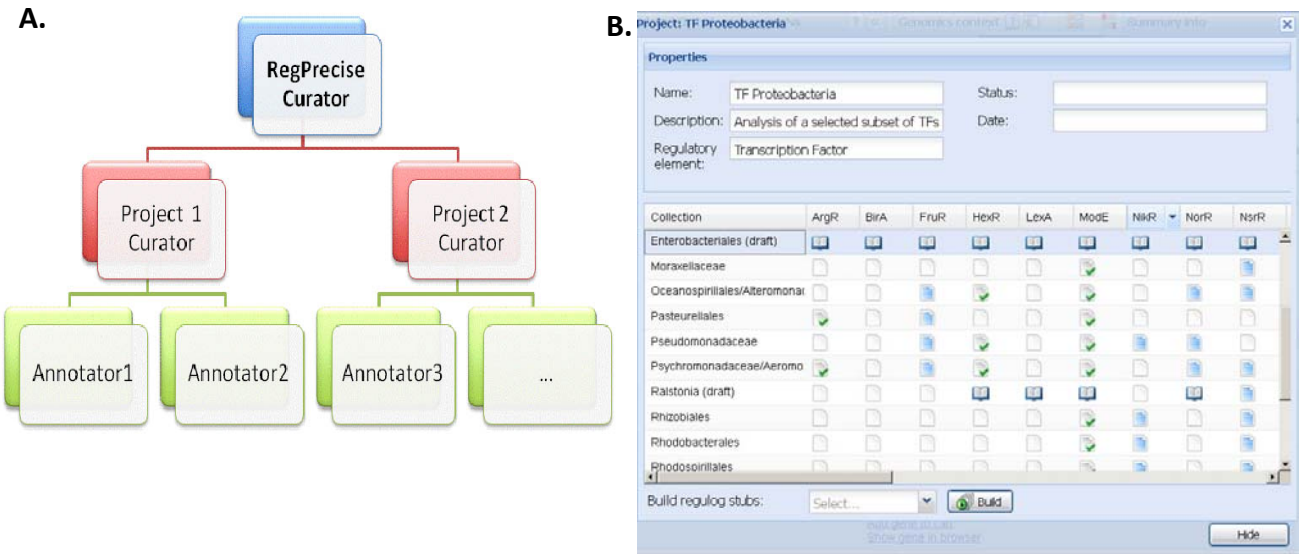
**Fig. 3. Modules in RegPredict for curation of regulons reconstructed by comparative genomics.**




We developed the community-based approach to conduct and manage large-scale projects on regulon inference (Fig. 4A). This ‘large-scale project’ module distributes the comparative analysis of multiple regulons in a large set of taxonomic collections of genomes between multiple annotators. It also provides a convenient interface for curation of regulons reconstructed by multiple users. After the regulog reconstruction is finished by annotator, the project curator can check it, and assign either recommendations for further improvements or its final acceptance. A particular regulog contains all regulatory interactions described for a group of orthologous TFs in a particular group of related genomes. Status of a regulog reconstruction is depicted by a thumbnail in the project table (Fig. 4B), which also serves as a direct link for opening a regulog page in the RegPredict.

The project curator can also change the status of an accepted regulog to a 'published' regulog resulting in its deposition to the public RegPrecise database. The community-based workflow was efficiently used for TRN inference in 4 taxonomic groups of bacteria, including Enterobacteria, Lactobacilli, Streptococcus, and Corynebacteria.

**Fig. 4. Community-based approach (A) and module (B) for management of large-scale projects on regulon inference.**




The developed module was tested for leading a large-scale research project with the goal to reconstruct draft regulatory networks for 30-40 TFs in all major taxonomic groups from the Proteobacteria phylum (21 groups, ~200 representative genomes, see Fig. 4). Another accomplished large-scale project was devoted to the reconstruction of regulons controlled by RNA regulatory elements in 11 reference taxonomic groups of bacteria (~120 genomes). During this project, we have mapped and analyzed the regulon content for ~40 known families of regulatory RNA motifs using the HMM models from Rfam database.

The controlled vocabularies of TF effectors and regulated biological processes are required for seamless integration of the RegPrecise data into SBKB in order to use regulatory constraints in building the predictive metabolic models. We developed the hierarchical classifications for TF effectors and regulated biological processes basing on the subsystems and biochemistry data in the SEED database, which became a core resource for metabolic models in SBKB. Biological processes attributed to regulons in the RegPrecise database covers a wide spectrum of the cellular metabolism. The current list of effectors of analyzed TFs includes more than 200 metabolites from the following major classes: amino acids, carbohydrates, nucleotides, lipids and fatty acids, co-enzymes, peptides and antibiotics, secondary metabolites, and inorganic chemicals. The corresponding modules to work with vocabularies of both types have been implemented in RegPredict.

**AIM 2. Infer regulatory annotations in several groups of bacteria and building of reference collections of microbial regulons**

We applied a computational genomic-based approach implemented in the RegPredict platform to infer and reconstruct transcriptional regulatory networks (TRNs) in diverse taxonomic groups of bacteria. This research allowed us to optimize and validate the platform and to provide a valuable contribution to the RegPrecise database and SBKB in a form of a reference set of regulons and regulatory annotations. RegPrecise 3.0 provides access to inferred regulatory interactions organized by phylogenetic, structural and functional properties. Taxonomy-specific collections include 781 TF regulogs inferred in more than 160 genomes representing 14 taxonomic groups of Bacteria (**Table 1**). TF-specific collections include regulogs for a selected subset of 40 TFs reconstructed across more than 30 taxonomic lineages. Novel collections of regulons operated by RNA regulatory elements (riboswitches) include near 400 regulogs inferred in 24 bacterial lineages. RegPrecise 3.0 provides four classifications of the reference regulons implemented as controlled vocabularies: 55 TF protein families; 43 RNA motif families; ~150 biological processes or metabolic pathways; and ~200 effectors or environmental signals.

**Table 1. Taxonomic collections of curated genome-wide TRNs in RegPrecise database.**

| Taxonomic group | Phylum | Reference Genomes | TF Regulogs | RNA Regulogs | TF binding sites | RNA sites | Regulated genes | Genes per genome |
|---|---|---|---|---|---|---|---|---|
| **Bacillales** | Firmicutes | 11 | 134 | 39 | 3815 | 668 | 7301 | 664 |
| **Staphylococcus** | Firmicutes | 7 | 48 | 29 | 1965 | 288 | 3329 | 476 |
| **Lactobacillaceae** | Firmicutes | 15 | 79 | 39 | 1811 | 581 | 3784 | 252 |
| **Streptococcaceae** | Firmicutes | 15 | 69 | 29 | 3118 | 400 | 5652 | 377 |
| **Clostridiaceae** | Firmicutes | 20 | 7 | 40 | 303 | 968 | 2489 | 124 |
| **Enterobacteriales** | Proteobacteria | 12 | 87 | 18 | 7365 | 188 | 9028 | 752 |
| **Shewanella** | Proteobacteria | 16 | 80 | 15 | 8450 | 291 | 10817 | 676 |
| **Ralstonia** | Proteobacteria | 6 | 24 | 10 | 574 | 66 | 1297 | 216 |
| **Desulfovibrionales** | Proteobacteria | 10 | 92 | 9 | 1942 | 72 | 3368 | 337 |
| **Thermotogales** | Thermotogae | 11 | 33 | 13 | 642 | 88 | 2153 | 196 |
| **Corynebacteriaceae** | Actinobacteria | 8 | 45 | 13 | 937 | 80 | 1624 | 203 |
| **Bacteroidaceae** | Bacteroidae | 11 | 35 | 2 | 667 | 84 | 1797 | 163 |
| **Chloroflexi** | Chloroflexi | 5 | 30 | 17 | 314 | 98 | 1014 | 203 |
| **Cyanobacteria** | Cyanobacteria | 14 | 18 | 11 | 1032 | 86 | 1442 | 103 |
| **Total:** | - | **161** | **781** | **284** | **32935** | **3958** | **55095** | **342** |

In the ***Shewanella***, the inferred complex TRN includes 82 TFs, ~500 TF binding sites, and ~600 target genes per genome. 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 strnjategies 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). 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 and HexR for central carbohydrate metabolism were experimentally validated by a combination of *in vitro* and *in vivo* approaches, and were additionally supported by observed strong correlations with massive (~200 experiments) genome-scale expression data in *S. oneidensis* MR-1. We discovered novel mechanism of activation of genes expression by HexR regulon.

In the ***Staphylococcus*** and ***Streptococcus***, we focused on the reconstruction of core TF regulons conserved in all species within each group (22 regulons), including global catabolic regulon CcpA, the nutrient limitation regulon CodY, and other regulons controlling the core metabolism and essential stress response systems. Using comparative genomics approach we identified candidate binding sites for 39 and 47 TFs in the *Streptococcus* and *Staphylococcus* groups, respectively. The inferred regulatory network in *Staphylococcus aureus* includes ~320 regulatory interactions between 47 transcription factors and ~550 candidate target genes comprising 20% of its genome. In the reconstructed *S. aureus* regulatory network, we predicted ~170 novel interactions and 24 novel regulons for the control of the central metabolic pathways. The reconstructed regulons are largely variable in the *Staphylococcaceae*: only 20% of *S. aureus* regulatory interactions are conserved across all studied genomes. Available expression data allowed the assessment of the reconstructed regulatory network in *S. aureus*.

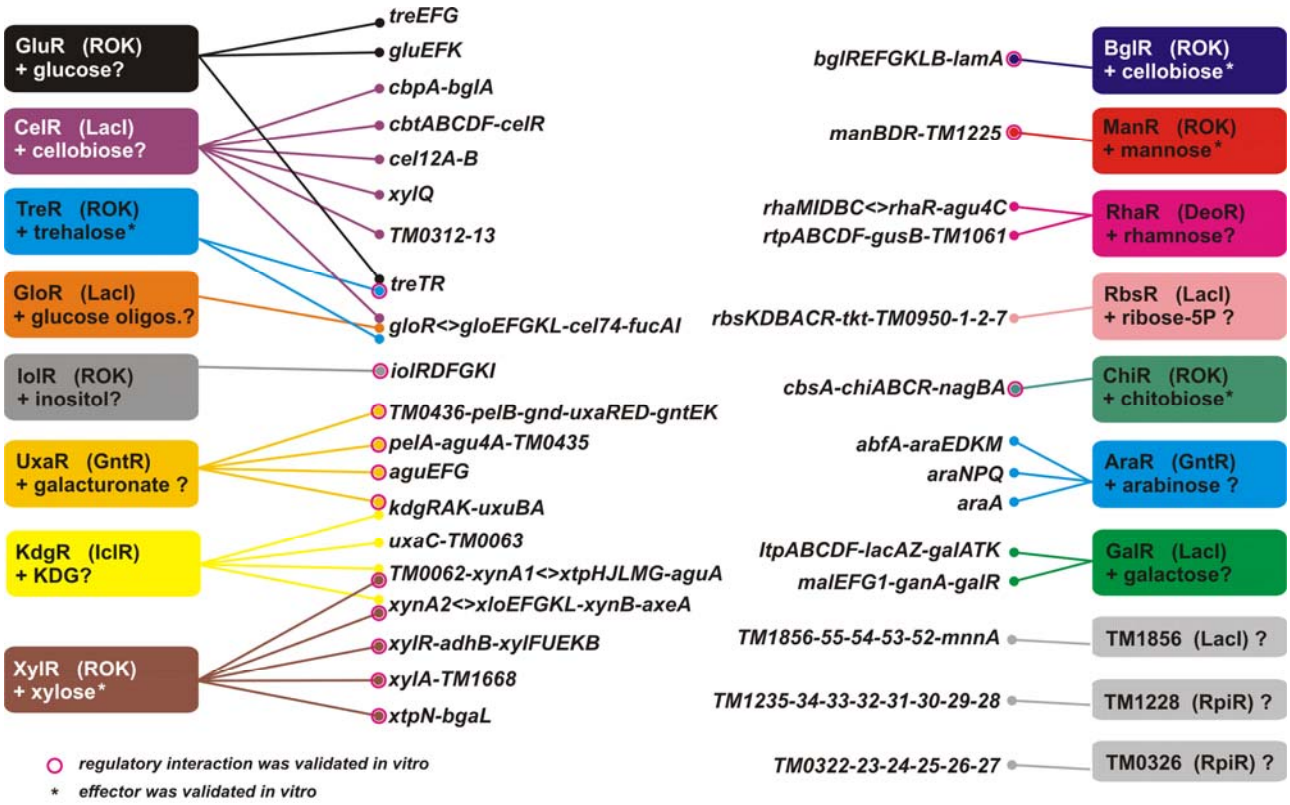
In the **Bacillales**, *Bacillus subtilis* is a model organism for studying sporulation, cell differentiation, stress response and social behavior of bacteria. According to the DBD database, *B. subtilis* genome encodes 238 DNA-binding transcription factors; many of them were studied experimentally and the respective regulatory interactions were captured in the DBTBS database. We applied both workflows A and B in the RegPredict to infer a reference collection of 121 regulons. As an input for the regulon reconstruction procedure we used any experimental information on transcriptional regulation in *B. subtilis* collected from more than 300 papers and the DBTBS database. For analysis of TF regulons, first we re-analyzed and expanded 57 TF regulons with previously known TFBS sites in *B. subtilis* and propagated them to all studied genomes, resulting in refinement of TFBS motifs and identification of novel regulon members. Second, we identified novel TFBS motifs and

described regulons for 28 experimentally studied in *B. subtilis* regulators with previously unknown binding sites. Thirdly, we discovered novel TFBS motifs and reconstructed regulons for 36 previously uncharacterized TFs. These novel regulons predicted to control genes involved in the following biological processes: utilization of various carbohydrates (alpha-galactoside, beta-glucoside, sucrose, inositol, maltodextrin, rhamnose and rhamnogalacturonan); metabolism of glutamate, histidine, and thiamine; stress responses; drug/metabolite transport. Most of the identified TFBS motifs have either palindromic or tandem repeat structure suggesting that the respective TFs bind DNA as dimers or oligomers.

In the **Thermotogales**, the inferred TRN includes 32 transcription factors and their DNA binding sites unevenly distributed across 11 studied genomes. A current collection of regulons is centered on *T. maritima* and includes 18 transcription factors that were predicted to control expression of ~185 genes involved in sugar catabolic machinery of this model organism. Remarkably, a large fraction of these genes and operons are controlled by multiple transcription factors pointing to a complexity of regulatory responses to changing environmental conditions (Fig. 5). For example, we established partial overlaps between the xylose, glucuronate, and galacturonate regulons (XylR, KdgR, and UxaR, respectively); the glucose, trehalose and inositol regulons (GluR, TreR, and IolR); and the cellobiose, mannose, and glucooligosaccharide regulons (CelR, ManR, and GloR). The experimental assessment of the reconstructed regulatory network included *in vitro* analysis of selected individual regulons and *in vivo* gene expression profiling of *T. maritima* on various carbohydrate substrates. We used the first approach based on gel-shift mobility assays to validate all predicted DNA targets and identify small molecule effectors for six regulators from the ROK family (BglR, IolR, XylR, ChiR, TreR, and ManR). Global gene expression profiles were obtained and analyzed for the growth on 12 different carbon sources using high-density oligonucleotide tiling arrays (Nimblegen). Gene induction patterns measured for tested mono- and disaccharides (trehalose, rhamnose, xylose, etc.) showed a strong correlation and provided additional information to refine respective regulons (TreR, RhaR, XylR, etc.) reconstructed by the genomic analysis. Finally, we have identified, reconstructed and experimentally validated novel regulon in the Thermotogales (T-Rex) controlling the central carbon metabolism and hydrogen production in response to NADH:$NAD^+$ ratio in the cell.

**Fig. 5. Reconstructed regulatory network for carbohydrate utilization genes and 18 transcription factors in *Thermotoga maritima.*** Abbreviations: TFs are shown in colored boxes; predicted regulatory interactions between a TF and its target operons are shown by lines; experimentally confirmed regulatory interactions are marked by magenta circles; predicted small molecule effectors are listed after '+'; confirmed effectors are marked by asterisks

.

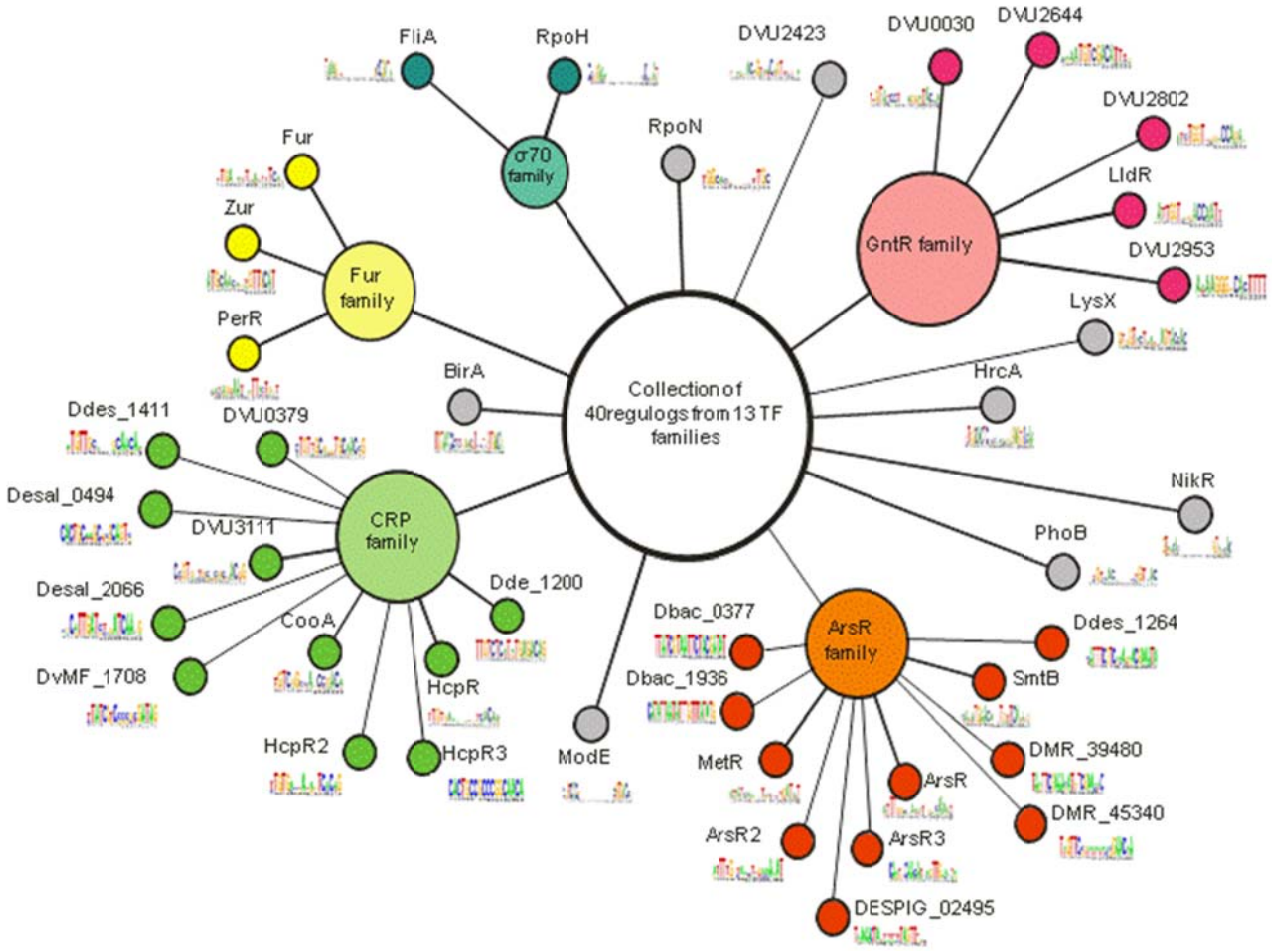



In the **Cyanobacteria**, the number of putative TFs varies substantially in the range between 30 and 134. At the current stage, the draft cyanobacterial collection contains 19 TF regulons including LexA, NrdR, NrtR, NtcA, NtcB, SphR and Zur that control essential cellular functions. For the NtcA (nitrogen metabolism), LexA (DNA damage response), Fur (iron homeostasis), SphR (phosphate metabolism), NrdR (nucleoside metabolism), and NrtR (NAD metabolism) regulons, we obtained significantly updated and extended regulatory reconstructions and associated TFBSs motif models in comparison to the regulon models inferred by comparative genomics in previous studies. Comparative genomic reconstructions of the Zur (zinc homeostasis) and NtcB (nitrogen metabolism) regulons in cyanobacteria was obtained for the first time in this work. The gene content of all reconstructed regulons except NrdR is highly variable in the studied cyanobacterial genomes, whereas the NrdR regulon was predicted to control a single gene. Analysis of the Fur regulon identified multiple regulatory cascades, in which Fur controls genes encoding putative TFs from the AraC family that likely control biosynthesis of iron siderophores. The reconstructed Zur regulon includes various putative zinc uptake transporters, putative zinc chaperones from the COG0523 family, as well as paralogs of various Zn-containing enzymes including phosphoribosyl-AMP cyclohydrolase, threonyl-tRNA synthetase, GTP cyclohydrolase I, porphobilinogen synthase (Fig. 6).

**Fig. 6. Reconstructed Zur regulog for zinc homeostasis in Cyanobacteria.** (A) gene content in Zur regulog (red circles show Zur-binding sites, arrows show the target genes; (B) Zur-binding motif logo.


In the **Desulfovibrionales**, the resulting TRN contains 40 TF regulons, >1,300 binding sites and 4,000 target genes involved in stress response, amino acid metabolism, metal homeostasis. Analysis of gene expression profiling data obtained for two TF knockout mutants, Fur (iron homeostasis) and PerR (peroxide stress response) in *Desulfovibrio desulfuricans G20*, shows good agreement with the regulons inferred by comparative genomics. The analysis of correlations between multiple microarray expression profiles provides additional information for the assessment of regulon predictions as demonstrated by analysis of a novel methionine metabolism regulon. The detailed analysis of three large TF families, ArsR, Crp, and GntR, revealed substantial variations of regulons in the *Desulfovibrionales* [7] (Fig. 7). Phylogenetic analysis of these TFs showed that 72 regulators from 9 sub-families are well conserved in the analyzed genomes, whereas 60 regulators are poorly conserved (present in less than 3 genomes) and randomly distributed in the genomes. This pattern is most likely a result of multiple evolutionary events such as horizontal gene transfer, gene loss, and gene duplication. For motif identification and regulon inference for 40 poorly conserved TFs, we found their closest homologs outside of the analyzed group of genomes, and applied motif detection procedure to a set of genes in the conserved genomic neighborhood. In summary, the inferred binding motifs for TFs in the same family show significant similarity to each other. This observation can be utilized for development of automatic approach for large-scale inference of poorly conserved regulons.

**Fig. 7. Reference set of regulons reconstructed for major TF families in Desulfoviibrionales.**




Finally, we performed a large-scale project devoted to reconstruction of regulons controlled by RNA regulatory elements (such as riboswitches) in major taxonomic groups of bacteria. Riboswitches are metabolite-sensing structures often found in bacterial mRNA leaders controlling gene expression on transcriptional or translational levels. An increasing number of riboswitches and other *cis*-regulatory RNAs have been recently classified into numerous RNA families in the Rfam database. High conservation of these RNA motifs provides a unique advantage for their genomic identification and comparative analysis.

A comparative genomics approach implemented in the RegPredict tool was used for reconstruction and functional annotation of regulons controlled by RNAs from 43 Rfam families in diverse taxonomic groups of Bacteria. The inferred regulons include ~5200 *cis*-regulatory RNAs and more than 12000 target genes in 255 microbial genomes. All predicted RNA-regulated genes were classified into specific and overall functional categories. Analysis of taxonomic distribution of these categories allowed us to establish major functional preferences for each analyzed *cis*-regulatory RNA motif family. Overall, most RNA motif regulons showed predictable functional content in accordance with their experimentally established effector ligands. Our results suggest that some RNA motifs (including thiamin pyrophosphate and cobalamin riboswitches that control the cofactor

metabolism) are widespread and likely originated from the last common ancestor of all bacteria. However, many more analyzed RNA motifs are restricted to a narrow taxonomic group of bacteria and likely represent more recent evolutionary innovations. The reconstructed regulatory networks for major known RNA motifs substantially expand the existing knowledge of transcriptional regulation in bacteria. The inferred regulons can be used for genetic experiments, functional annotations of genes, metabolic reconstruction and evolutionary analysis. The obtained genome-wide collection of reference RNA motif regulons is available in the RegPrecise database.

**AIM 3. Develop KnowledgeBase on microbial transcriptional regulation**

RegPrecise 3.0 gives access to the transcriptional regulons reconstructed in bacterial genomes. Analytical capabilities include exploration of: regulon content, structure and function; TF binding site motifs; conservation and variations in genome-wide regulatory networks across all taxonomic groups of Bacteria. RegPrecise 3.0 was selected as a core resource on transcriptional regulation of the Department of Energy Systems Biology Knowledgebase (SBKB), an emerging software and data environment designed to enable researchers to collaboratively generate, test and share new hypotheses about gene and protein functions, perform large-scale analyses, and model interactions in microbes, plants, and their communities.

To facilitate the development of various SBKB components based on the transcriptional regulatory interactions, we developed the RegPrecise Web Services API that allows programmatic access to the whole content of the RegPrecise database. The first version of API is implemented as a set of 15 RESTful web services providing data in either JSON or XML, two the most popular formats. The RegPrecise Web Services API provides a programmatic access to the regulatory interactions accumulated in the RegPrecise database: http://regprecise.lbl.gov/RegPrecise/services.jsp

**Developed Web-Resources:**

- RegPredict web-server platform integrating multiple workflows and modules for genomic-based inference of regulons in microbial genomes: http://regpredict.lbl.gov/
- RegPrecise database for capturing and representation of transcriptional regulons reconstructed using the RegPredict platform and the comparative genomics approach http://regprecise.lbl.gov/
- RegTransBase database of regulatory interactions in prokaryotes that captures the knowledge in public scientific literature using a controlled vocabulary http://regtransbase.lbl.gov/

**Publications from this proposal** (main publications are **in bold**):


1. **Rodionov DA, Novichkov PS, Stavrovskaya ED, Rodionova IA, Kazanov MA, Ravcheev DA, Gerasimova AV, Kazakov AE, Kovaleva GY, Permina EA, Laikova ON, Overbeek R, Romine MF, Fredrickson JK, Arkin AP, Dubchak I, Osterman AL, Gelfand MS. (2011) Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics,* 12(Suppl 1):S3.**

2. **Ravcheev DA, Best AA, Tintle N, Dejongh M, Osterman AL, Novichkov PS, Rodionov DA. (2011) Inference of the transcriptional regulatory network in *Staphylococcus aureus* by integration of experimental and genomics-based evidence. *J Bacteriology,* 193(13): 3228-40.**

3. Suvorova I, Tutukina MN, Ravcheev DA, Rodionov DA, Ozoline ON, Gelfand MS. (2011) Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gamma-proteobacteria. *J Bacteriology*, 193(15): 3956-63.
4. **Leyn SA, Li X, Zheng Q, Novichkov PS, Reed S, Romine MF, Fredrickson JK, Yang C, Osterman AL, Rodionov DA. (2011) Control of Proteobacterial central carbon metabolism by the HexR transcriptional regulator. A case study in *Shewanella oneidensis*. *J Biol Chem*, 286(41): 35782-94.**
5. Zhang L, Leyn SA, Gu Y, Jiang W, Rodionov DA, Yang C. (2012) Ribulokinase and transcriptional regulation of arabinose metabolism in *Clostridium acetobutylicum*. *J Bacteriology,* 194(5): 1055-64.
6. Brune I, Götker S., Schneider J, Rodionov DA, Tauch A. (2012) Negative transcriptional control of biotin metabolism genes by the TetR-type regulator BioQ in biotin-auxotrophic *Corynebacterium glutamicum* ATCC 13032. *J Biotechnology,* 159(3): 225-34.
7. **Ravcheev DA, Li X, Latif H, Zengler K, Leyn SA, Korostelev YD, Kazakov AE, Novichkov PS, Osterman AL, Rodionov DA. (2012) Transcriptional regulation of central carbon and energy metabolism in Bacteria by redox-responsive repressor Rex. *J. Bacteriology,* 194(5): 1145-57.**
8. Chen X, Kohl TA, Rückert C, Rodionov DA, Li L, Ding J, Kalinowski J, Liu S. (2012) Phenylacetic acid catabolism and its transcriptional regulation in *Corynebacterium glutamicum*. *Appl. Environ. Microbiol.,* 78(16): 5796-804.
9. **Novichkov PS, Brettin TS, Novichkova ES, Dehal PS, Arkin AP, Dubchak I, Rodionov DA. (2012) RegPrecise web services API: programmatic access to the transcriptional regulatory interactions reconstructed by comparative genomics in bacteria. *Nucleic Acids Res*, 40(Web Server issue):W604-8.**
10. **Leyn S, Gao F, Yang C, Rodionov DA. (2012) N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J Biol Chem*., 287(33): 28047-56.**
11. Rodionova IA, Scott DA. Grishin NV, Osterman AL, Rodionov DA. (2012) Tagaturonate-fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in *Thermotoga maritima. Environ. Microbiol.* 14(11): 2920-34.
12. Rodionova IA, Yang C, Li X, Kurnasov OV, Best AA, Osterman AL, Rodionov DA. (2012) Diversity and versatility of the *Thermotoga maritima* sugar kinome, *J. Bacteriol*, 194(20): 5552-63.

13. Antunes A, Camiade E, Monot M, Courtois E, Barbut F, Sernova N, Rodionov DA, Martin-Verstraete I, Dupuy B (2012) Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res.*, 40(21): 10701-10718.
**14. Kazakov AE, Rodionov DA, Price MN, Arkin AP, Dubchak I, Novichkov PS. (2013) Transcription factor family-based reconstruction of singleton regulons: study of the Crp/Fnr, ArsR and GntR families in Desulfovibrionales genomes. *J Bacteriology*, 195(1): 29-38. PubMed PMID: 23086211.**
**15. Kazanov MD, Li X, Gelfand MS, Osterman AL, Rodionov DA. (2013) Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum. *Nucleic Acids Research*, 41 (2): 790-803. PubMed PMID: 23209028.**
**16. Leyn SA, Kazanov MD, Sernova NV, Ermakova EO, Novichkov PS, Rodionov DA. (2013) Genomic Reconstruction of Transcriptional Regulatory Network in *Bacillus subtilis. J Bacteriology*. 195 (11): 2463-73. PubMed PMID: 23504016.**
17. Rodionova IA, Leyn SA, Burkart MD, Boucher N, Noll KM, Osterman AL, Rodionov DA. (2013) Novel inositol catabolic pathway in *Thermotoga maritima*. *Environ Microbiol*. 15(8): 2254-2266. PubMed PMID: 23441918.
**18. Ravcheev DA, Best AA, Sernova NV, Kazanov MD, Novichkov PS, Rodionov DA. (2013) Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics*.14(1):94. PubMed PMID: 23398941.**
19. Cipriano MJ, Novichkov PN, Kazakov AE, Rodionov DA, Arkin AP, Gelfand MS, Dubchak I (2013) RegTransBase--a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics.*14(1):213. PubMed PMID: 23547897.
**20. Rodionov DA, Rodionova IA, Li X, Ravcheev DA, Tarasova E, Portnoy VA, Zengler K, Osterman AL. (2013) Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. *Frontiers in Microbiology*,4: 244. PubMed PMID: 23986752.**
21. Kazakov AG, Rajeev L, Luning E, Zane G, Siddartha K, Rodionov DA, Dubchak I, Arkin AP, Wall J, Mukhopadhyay A, Novichkov PS (2013) New family of tungstate-responsive transcriptional regulators in sulfate-reducing bacteria. *J Bacteriology*, 195 (19): 4466-75. PubMed PMID:23913324.
**22. Novichkov PS, Li X, Kuehl JV, Deutschbauer AM, Arkin AP, Price MN, Rodionov DA. (2013) Control of methionine metabolism by the SahR transcriptional regulator in Proteobacteria. *Environ Microbiol*. Oct. 9. doi: 10.1111/1462-2920.12273. PubMed PMID: 24118949.**
**23. Sun E, Leyn SA, Kazanov MD, Saier MH, Novichkov PS, Rodionov DA. (2013) Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in Bacteria. *BMC Genomics,* 14(1): 597. PubMed PMID: 24060102.**
**24. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA. (2013) RegPrecise 3.0 - a resource for genome-scale exploration of transcriptional regulation in Bacteria. *BMC Genomics,* 14(1): 745. PubMed PMID: 24175918**
25. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA. (2013) Polysaccharide utilization in *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics*, in press.

## **Students and Postdocs** trained during this work:

Postdoctoral Associates:

1. Marat D. Kazanov, PhD (SBMRI, now at *A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences)*
2. Dmitry A. Ravcheev, PhD (SBMRI, now at University of Luxembourg)
3. Alexey E. Kazakov, PhD (LBNL)
4. Elena Satvrovskaya, PhD (LBNL, now in Moscow State University)

Graduate Students:

1. Semen A. Leyn (SBMRI, now at *A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences)*

Undergraduate Trainees:

1. Ilia A. Zharov (SBMRI, now at *A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences).*