Reference collection of transcriptional regulons in Bacillales

Semen A. Leyn[1,2]*(sleyn@burnham.org), Marat D. Kazanov,[2] **Pavel S. Novichkov,[3] Dmitry A. Rodionov**[1,2]

[1]Sanford-Burnham Medical Research Institute, La Jolla, CA; [2]Institute for Information Transmission Problems RAS, Moscow, Russia; [3]Lawrence Berkeley National Laboratory, Berkeley, CA

http://regprecise.lbl.gov/

Project Goals:

1. **Develop integrated platform for genome-scale regulon reconstruction utilizing the comparative genomics approach and analysis of RNA and DNA regulatory sites.**
2. **Infer the regulatory interactions and reconstruct transcription regulatory networks in several groups of microbial species important for DOE mission.**
3. **Develop RegKnowledgeBase on microbial transcriptional regulation.**

Gram-positive facultative anaerobic bacteria from the Bacillales order were isolated from diverse habitats including soil, sea water, plants and animals. Bacillales use various strategies to respond and survive in a variety of stresses and environmental conditions including resistance to multiple antibiotics. *Bacillus subtilis* str. 168 is one of the best-characterized Gram-positive bacteria and 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 (TFs) classified in 45 protein families. Of them, 120 TFs were studied experimentally and the respective regulatory interactions were captured in the DBTBS database. However, many of the previously studied TF regulons were studied insufficiently, providing an incomplete knowledge on the range of target genes and associated TF-binding sites (TFBSs).

Continuously growing number of available complete genomes allows successful application of the comparative genomics approaches for regulon analysis. We used a "knowledge-driven" approach, which combines the accumulated experimental information from literature and databases with novel bioinformatics tools for genomic reconstruction of regulatory interactions. We perform the comparative genomics reconstruction of regulons operated by either TFs or RNA-regulatory element using the RegPredict Web-server (http://regpredict.lbl.gov). RegPredict allows prediction of TFBS and RNA motifs in a group of selected genomes, with further identification and annotation of candidate members of the respective regulons. Functional analysis of target genes was based on annotations from the SEED database that were validated by the genomic context analysis in MicrobesOnLine.

In this study we carried out large-scale comparative genomics analysis of regulatory interactions in *B. subtilis* and 10 related species from the Bacillales order. For TF regulons, we first analyzed 59 regulons with previously known TFBS motifs according to literature and the DBTBS database on transcriptional regulation in *B. subtilis*. These known regulons were expanded by prediction of novel targets in *B. subtilis* and propagated to other studied genomes of Bacillales, resulting in refinement of TFBS motifs and identification of novel regulon members. Then we predicted novel TFBS motifs and reconstructed 32 TF regulons for which target genes have been previously defined in *B. subtilis* but whose TFBSs were unknown. Finally, we discovered novel TFBS motifs and reconstructed regulons *de novo* for 34 previously uncharacterized TFs. Novel regulons involve genes from the following biological processes: utilization of various carbohydrates; metabolism of

glutamate, histidine and thiamine; stress responses; drug/metabolite transport. Totally, more than 3500 TFBSs have been predicted in the *Bacillales* group (from 200 to 600 sites per genome).

For RNA-operated regulons, we used bacterial RNA regulatory motifs collected from the Rfam database, scanned the studied genomes with these motifs to identify new occurrences of each RNA family, and annotated the respective target operons. Among 37 reconstructed RNA regulons there are 11 families of metabolite-sensing riboswitches, 17 types of aminoacyl-tRNA-responsive T-boxes, one regulon controlled by the RNA-binding protein PyrR, five predicted ribosomal protein leader RNA structures and five regulons for *in silico* predicted RNA motifs of unknown function. The reconstructed RNA motif-operated regulons in Bacillales control key metabolic pathways including biosynthesis of vitamins and cofactors (cobalamin, riboflavin, thiamine, nucleoside queuosine), biosynthesis of glucosamine, metabolism of most amino acids, biosynthesis and salvage of purines and pyrimidines, and magnesium homeostasis.

The reference collection of transcriptional regulons in the Bacillales group of bacteria is available in the RegPrecise database (http://regprecise.lbl.gov/). Currently, this collection constitutes the biggest transcriptional regulatory network among various taxonomic groups with reconstructed TRNs in RegPrecise.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), under contract DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Lab.