

Overview

RegTransBase (RTB) is a database of regulatory sequences and regulatory interactions in prokaryotic genomes. RTB is based on journal articles devoted to transcriptional and post-transcriptional regulation of gene expression.

Annotation of each article in RTB contains a list of experiments (with a short description) and a list of structural elements of genomes involved in regulatory interactions (genes, sites, transcripts, operons, loci, regulons, regulators, effectors).

RTB brings together these interactions in a user-friendly interface, allowing the user to explore and compare their genomes of interest, as well as view all experiments on a given element in one place.

RTB provides more than just a collection of articles, experiments and elements, it also provides tools for the analysis of regulation within one organism, as well as a comparison between multiple organisms. Using the combination of previous knowledge from published experiments along with computational prediction tools, a user can make informed decisions on the analysis of regulatory sites throughout genomes.

RTB contains modules for simple text searching (such as gene name, function, or experiment description), sequence based searching (BLAST), and searching using motifs or alignments (MAST).

RegTransBase is available at <http://regtransbase.lbl.gov>

Figure 1. Main page of RegTransBase.

Data

In the studies on bacterial regulation the final decision of whether to include each putative site in a particular regulon is made after detailed inspection and consultation with relevant scientific literature by a human expert. RegTransBase (RTB), a manually curated database of regulatory interactions, captures the knowledge in published scientific literature using a controlled vocabulary. RTB describes a large number of regulatory interactions reported in many organisms and contains the following types of experimental data:

Taxonomy	Genes	Sites
Alphaproteobacteria	3208	1678
Betaproteobacteria	103	17
Gammaaproteobacteria	4542	2668
E. coli	1516	997
Delta/epsilon proteobacteria	1	1
Firmicutes	3195	1459
B. subtilis	666	320
Cyanobacteria	135	196
Actinobacteria	3	3
Bacteroides/Chlorobi group	1	2
Archaea	3	4
Multi- or unknown host plasmids, transposons and phages	1331	439
TOTAL	12817	6470

Table 1. Distribution of the number of elements based on organisms.

Experiment Types	#
Gene/operon activation	2580
Gene/operon repression	1216
Operon structure characterization	782
Promoter mapping	1689
Regulatory site mapping	1922
Terminator mapping	60
Regulatory site prediction	845
Plasmid replication	17
TOTAL	9111

Table 2. Listing of the number of entries of different experiment types in RegTransBase.

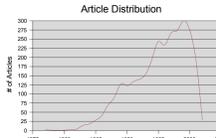
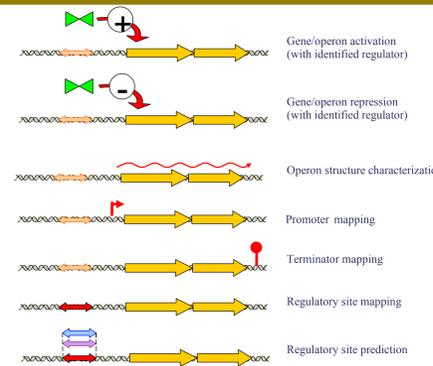


Figure 2. Number of articles annotated per year published.



Searching

Figure 3. Searches available in RTB. a) BLAST search with genomic overview and display of experimental results, b) gene name search, c) searching through descriptions of experiments.

The information contained in RTB can be thoroughly searched using a number of provided search tools. When possible, we attempt to provide the ability to scan through the information quickly using mouse-overs to provide more detailed information.

Information

RegTransBase contains structured information obtained directly from experiments explained in published literature. Articles contain multiple experiments.

Each experiment contains multiple elements that make up that experiment. Elements themselves can have a hierarchical relationship (operons--genes). Elements may be linked to other elements (sites are linked to regulators). We provide the tools to view this experiment, and then

- obtain a global view of the genomic region
- view features/elements in that region
- list effectors that act on these elements
- Provide tools for the comparisons between species.

Figure 4. The correlation between an article/experiment and how it appears in RTB. a) An actual article, b) Experiment view, c) Element view, d) Site view, e) Genome view using Gbrowse, f) GraphViz diagram based around the relationship of elements described in literature, g) View of the VISTA Genome Browser comparing the genomes of multiple species.

In addition to publication data, RTB provides its users with a growing collection of curated binding site alignments. Each alignment was created by an expert curator who provided descriptions explaining all alignments, specific sequence locations referenced to NCBI RefSeq genomes, available publications, and recommended options for using this alignment to search new genomes. This data is available for download

Prediction and Comparison

We currently have a manually curated collection of 42 position weight matrices and alignments (with plans for over 150 in the near future). We provide the ability to search sequenced genomes using these matrices or the user can supply their own alignment. Using this interface we aim to provide a tool for the following situations:

- One matrix + one genome of interest
 - Show predicted binding sites which match this matrix, while providing additional information.
- One gene + multiple genomes
 - Predict binding sites for orthologous genes using certified matrices.
- One matrix + multiple genomes
 - Compare the predicted binding sites across genomes for a particular matrix, highlighting orthologous similarities.
- Multiple matrices + multiple genomes
 - Compare the predicted binding sites across genomes for a set of matrices.

These tools allow a person to be guided through a semi-automated process which will highlight conserved transcription factor binding sites.

Figure 5. The process for comparing hits of a particular motif against multiple genomes is shown here. a) A predefined alignment is chosen to create a position weight matrix from (custom alignment option is available also), b) Genomes to compare are selected, c) Results will be filtered by the options given, d) The result is a table with rows being orthologous genes, and hits specified within each row. For each orthologous row, additional analysis tools may be accessed, such as graphical alignments, sequence logos, text alignments, phylogenetic trees and the ability to view the alignment in the feature rich application JalView.

ACKNOWLEDGEMENT

ESPP is part of the Virtual Institute for Microbial Stress and Survival supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.