RegTransBase

A Database of Regulatory Sequences and Interactions in Prokaryotic Genomes

Michael Cipriano, Alexey Kazakov, Mikhail Gelfand, Adam Arkin, Inna Dubchak

Genomics Division, Lawrence Berkeley National Laboratory, The Virtual Institute of Microbial Stress and Survival, The Research and Training Center on Bioinformatics (Moscow)

Genomics Division, Lawrence Berkeley National Laboratory

The Virtual Institute of Microbial Stress and Survival

Research and Training Center on Bioinformatics

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 then 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, regular expressions), and searching using motifs or alignments (MAST).

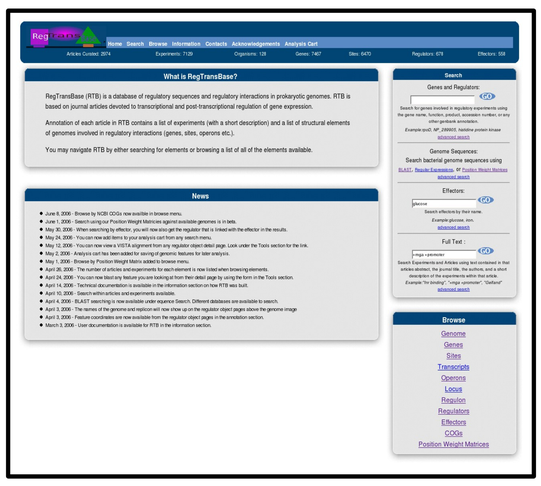
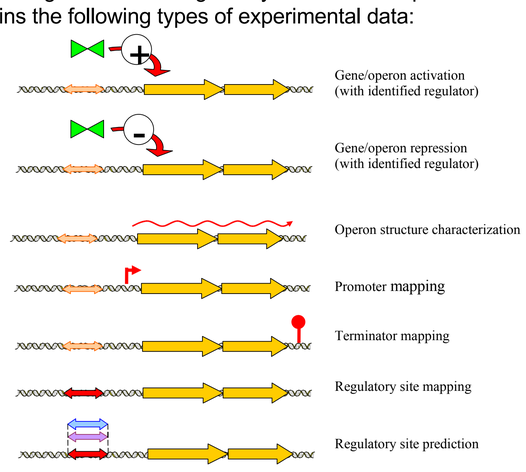


Figure 1. RegTransBase web site

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:



Experiment Types	#
Gene/operon activation	2354
Gene/operon repression	1128
Operon structure characterization	666
Promoter mapping	1410
Regulatory site maping	1670
Terminator mapping	46
Regulatory site prediction	733
Plasmid replication	16
TOTAL	**8023**

Table 1. Distribution of experiment types.

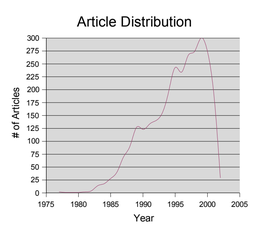


Figure 2. Number of articles annotated per year.

Taxonomy	Genes	Sites
Alphaproteobacteria	3208	1678
Betaproteobacteria	103	17
Gammaproteobacteria	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
Archea	3	4
Multi- or unknown host plasmids, transposons and phages	1331	439
TOTAL	**12817**	**6470**

Table 2. Distribution of the number of elements based on organisms.

Searching

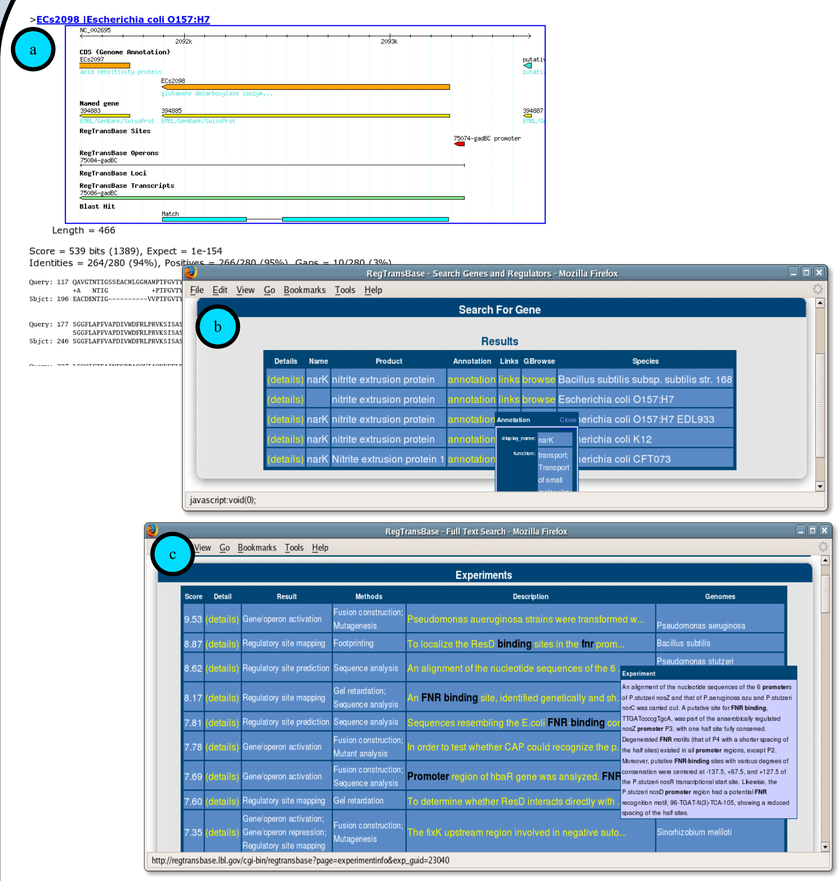


Figure 4. Searches available in RTB. a) BLAST search with genomic overview and display of experimental results, b) gene name search, c) searching through descriptions of experiments.

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 3. 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.

Prediction and Comparison

We currently have a manually curated collection of 160 position weight matrices and alignments (with plans for over 500 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 orthologus genes using certified matrices.
- One matrix + multiple genomes
  - Compare the predicted binding sites across genomes for a particular matrix, highlighting orthologus similarities.
- Multiple matrices + multiple genomes
  - Compare the predicted binding sites across genomes for a set of matrices.

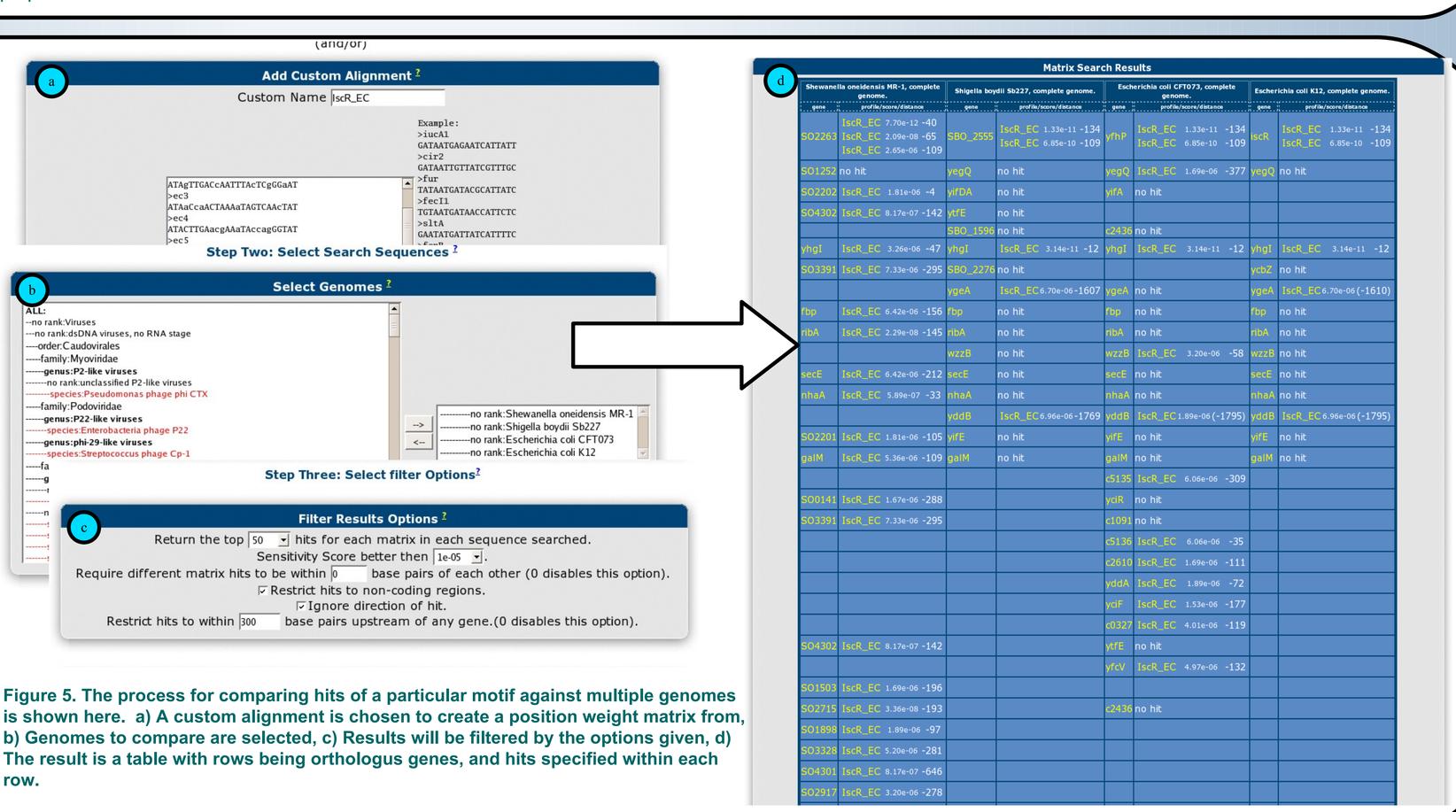


Figure 5. The process for comparing hits of a particular motif against multiple genomes is shown here. a) A custom alignment is chosen to create a position weight matrix from, b) Genomes to compare are selected, c) Results will be filtered by the options given, d) The result is a table with rows being orthologus genes, and hits specified within each row.