

DATABASE

Open Access

RegTransBase – a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes

Michael J Cipriano¹, Pavel N Novichkov², Alexey E Kazakov², Dmitry A Rodionov^{3,4}, Adam P Arkin^{2,5}, Mikhail S Gelfand^{4,6} and Inna Dubchak^{2,7*}

Abstract

Background: Due to the constantly growing number of sequenced microbial genomes, comparative genomics has been playing a major role in the investigation of regulatory interactions in bacteria. Regulon inference mostly remains a field of semi-manual examination since absence of a knowledgebase and informatics platform for automated and systematic investigation restricts opportunities for computational prediction. Additionally, confirming computationally inferred regulons by experimental data is critically important.

Description: RegTransBase is an open-access platform with a user-friendly web interface publicly available at <http://regtransbase.lbl.gov>. It consists of two databases – a manually collected hierarchical regulatory interactions database based on more than 7000 scientific papers which can serve as a knowledgebase for verification of predictions, and a large set of curated by experts transcription factor binding sites used in regulon inference by a variety of tools. RegTransBase captures the knowledge from published scientific literature using controlled vocabularies and contains various types of experimental data, such as: the activation or repression of transcription by an identified direct regulator; determination of the transcriptional regulatory function of a protein (or RNA) directly binding to DNA or RNA; mapping of binding sites for a regulatory protein; characterization of regulatory mutations. Analysis of the data collected from literature resulted in the creation of Putative Regulons from Experimental Data that are also available in RegTransBase.

Conclusions: RegTransBase is a powerful user-friendly platform for the investigation of regulation in prokaryotes. It uses a collection of validated regulatory sequences that can be easily extracted and used to infer regulatory interactions by comparative genomics techniques thus assisting researchers in the interpretation of transcriptional regulation data.

Keywords: Transcriptional regulation, Prokaryotes, Comparative genomics

Background

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. Genes and operons directly co-regulated by the same TF are

considered to belong to a regulon. Predicting the regulon of a transcription factor that binds DNA by detecting TFBSs in most cases requires the alignment of known binding sites to create a positional weight matrix (PWM). It is very important to filter out irrelevant sites and find TFBSs that are of higher confidence, and comparative genomics is the method of choice for this.

With the advent of new and cheaper sequencing technologies and ongoing sequencing projects such as GEBA [1], which aims to close the gaps in the bacterial tree of

* Correspondence: ildubchak@lbl.gov

²Lawrence Berkeley National Laboratory, Berkeley, CA 94710, USA

⁷Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

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

life, a lot of bacterial organisms are now being sequenced [2]. Of note is that not only are organisms with no close sequenced relatives being sequenced, but specifically groups of closely related organisms and multiple strains of the same species. This trend of sequencing can be successfully exploited when using comparative analyses, and already has been used in studying and predicting transcriptional regulation [3-6].

While many transcriptional regulation experiments are performed on model organisms, the existing experimental evidence can be transferred to other organisms by comparative methods. However, even closely related organisms can have different transcriptional regulation [7], thus prediction of binding sites and regulon inference in bacteria until recently has been mostly done by careful manual analysis [8-10]. Availability of experimental data on regulation for a wider range of organisms would be very helpful in automatic verification of computationally derived predictions of regulation. These verifications require well-designed databases accessible to prediction and analysis programs.

Eukaryotic transcriptional regulation data has been summarized in both commercial and open-source databases, such as TransFac [11], Pazar [12], and ORegAnno [13], widely used by the community. There are several gene regulation databases that focus on distinct microbial organisms such as *E. coli* [14,15], *B. subtilis* [16], Mycobacterium tuberculosis [17], and corynebacteria [18]. On the other hand, PRODORIC [19], PePPER [20] and SwissRegulon [21] cover a wide range of bacterial genomes.

RegTransBase, first introduced in 2007 [22], was built with the goal to cover a wide microbial diversity and provide a collection of curated experimental data to use in external computational tools. The current advanced version of RegTransBase: (i) contains a much larger set of manually collected experimental results (Table 1); (ii) has a brand new interface with novel capabilities for multi-level data navigation such as the new Classification Browser and new data aggregation tools such as the Putative Regulons Browser; (iii) is linked to associated analytical systems.

It is important to mention that we have recently developed two new resources – the RegPredict Web tool to support genomic reconstruction of transcriptional regulons in groups of closely related prokaryotic genomes [23], and the RegPrecise database to capture, visualize and analyze transcription factor regulons that were reconstructed [24]. We are working on the integration of RegTransBase, RegPredict and RegPrecise into a powerful platform for regulon reconstruction and analysis.

Construction and content

Experimental data annotation

The main objective during the article annotation phase for RegTransBase was to collect experimental evidences

Table 1 Content of RegTransBase

	2012	2006
Articles Curated:	7213	2958
Experiments:	18877	7129
Organisms:		
Phylum	11	9
Genus	224	Not reported
Species	666	128
Genes:	59754	7467
Sites:	13321	6470
Regulators:	1488	678
Effectors:	875	558
PWM	151	151
Putative Regulons		
Total	6390	N/A
Avg. Regulon/Genome	13	N/A
Median Regulon/Genome	5	N/A

of transcriptional regulation and experimentally characterized TF binding sites. The main steps of the data collection. Described in detail in our previous article [22], are the following: search for relevant articles in PubMed [25], entry of data through a specialized annotator interface, quality control, mapping sites and genes to genomes, additional manual corrections (if necessary) and presentation of the data in the final format. The entry quality is controlled by a number of consistency and completeness checks. The genomic location of a specific feature (site or gene) is then recorded by the annotator as a signature (a DNA sequence fragment of sufficient length) that is then used to map all the features in the database to a wide range of the NCBI RefSeq genomes [26,27].

Each database entry describes a single experiment that is an experimentally determined relationship between several database elements. A single entry may describe an experiment and control, identical results obtained by different methods or the results of the application of one technique to several similar objects. Only original results are recorded, normally from the 'Results' or 'Discussion' sections of an article.

The types of experimental techniques form a controlled vocabulary. The following categories of experiments were accepted: (i) regulation of gene expression by a known regulator; (ii) demonstration that a gene encodes a regulatory protein (excluding proteins that do not directly bind DNA, e.g. protein kinases); (iii) experimental mapping of DNA binding sites for known regulators; (iv) identification of mutations in regulatory genes influencing expression of regulated genes; (v) computational prediction of binding sites.

The classes of elements in the database are: regulators (regulatory proteins and RNAs directly binding to DNA, with a well-defined binding site); effectors (molecules not binding DNA or physical effects such as stress, etc.); and positional elements. The latter are described as regions in DNA sequences. Positional elements form a hierarchy: locus > operon > transcript > gene and site; an elements may be a sub-elements of other elements of the same or higher levels (e.g., a site and a gene may be a sub-element of a operon).

All elements are linked to the corresponding experiments and together they are linked to the original article. As mentioned above, positional elements are mapped to genomes, thus if two independent articles describe regulation of the same gene, the data contained in these articles will be interlinked via this gene, but sites and other experimental data will be reported as independent entries.

Our original publication on RegTransBase [22] and the Help pages at <http://regtransbase.lbl.gov> provide more details on the procedure of experimental data annotation.

Putative regulons from experimental data

The Putative Regulons section of RegTransBase provides a list of experimental sites along with a non-redundant list of target genes for each regulator. The process we undertook in developing this list of putative regulons from the manually curated data includes three steps.

First, we selected a subset of experiments using the following criteria: (i) the experiment describes a single regulator, (ii) a regulator and its regulated genes belong to the same genome, (iii) no computational predictions are included.

Second, from this subset we extracted the pairs 'regulator-regulated gene' for each genome, taking into account operon structure, that extend the list of regulated genes by adding other members of a particular operon. In some cases we see a particular pair of a regulator and an associated regulated gene in multiple entries in RegTransBase. We removed such redundant pairs from the list of regulator-regulated genes based on positional mapping.

Third, we compiled a list of putative regulons by unifying all 'regulator-gene' pairs with the same regulator.

Manually curated position weight matrices

Each record in the Manually Curated PWM section of the database comprises a TFBS training set (alignment) created by an expert curator using published experimental data and manual *in silico* analyses. The curator first gathered information about a known transcription factor where a set of binding sites was known, created a summary of a description of this transcription factor by

scanning published articles, and recorded its genomic location. The curator then annotated binding sites and their sequence, downstream gene, location in a published genome, and any published experimental evidence. In addition, curators supplied groups of organisms that they believe could be used when searching for homologous binding sites based on phylogenetic distance of organism and presence of a conserved transcription factor. Lastly, the curator recorded default scores and the expected distance a binding site would be from the start of a gene based on examination of the existing binding sites.

A PWM is automatically created in the RegTransBase database based on the TFBSs alignment. We then searched all recommended bacterial genomes using MAST [28]. We recorded all hits that passed the following criteria into the RegTransBase database: e-value of $1e-5$ or better, it did not overlap coding regions and it was upstream of a predicted gene.

With each record, we provide the binding site location with a reference to a published sequence (usually NCBI RefSeq [26]), the sequence, the gene which is affected by the binding site, the evidence for the binding if any, any relevant articles pertaining to that site, and the transcription factor which binds the site. We also provide for download the sequence logo for the alignment, profiles and alignments in many different formats, and recommended options in using the profiles for searching other genomes (cut-off scores, distance from gene, taxonomy).

Database statistics

As of November 2012, RegTransBase contains information on 666 bacterial species from 224 genera. This resource allows for access to the information on 19000 different experiments from about 7200 articles from as far back as 1977 until the present day (more details in Table 1).

Utility and discussion

Our goal is to provide a comprehensive resource to the greater genomic community to allow for easy transfer of known binding site information as well as tools for discovering interesting regulatory interactions in groups of organisms. We believe that by using a comparative approach, new genomes could be more easily annotated, and this approach can help facilitate the discovery and expansion of regulons in a wide range of organisms.

Database access and features

RegTransBase is freely accessible via a user-friendly web interface at <http://regtransbase.lbl.gov>. Besides browsing, searching for various data of interest, and carrying out analytical tasks (see below), users can download the Annotators Database, which includes all of the annotated data elements and experiments as a sql dump file to perform

their own analysis, as well as the Annotators Database Schema Description, and Alignments of Binding Site through the 'Download' page.

Data navigation

We developed a new navigation interface to easily select a set of experimental records based on six categories (classifications) covering different aspects of the database.

Three categories (classifications) describe genomes that were studied in relevant experiments (Figure 1).

The 'Taxonomy' category is based on the NCBI Taxonomy [29] and describes phylogenetic relationships. A user can choose a taxon of interest starting from the super kingdom level (Bacteria or Archaea) and move down to the species level. The 'Relevance' category refers to the attributes of genome projects that provide information about the wide area of research a particular genome is a part of, such as Antibiotic production, Agricultural, etc. [30]. The 'Phenotypes' category includes attributes that describe phenotypic properties of the organisms [30].

Two categories refer to experimental methodology and the goals of experiments. The 'Experiment techniques' classification uses a controlled vocabulary of methods used in experiments. This classification has a two-level structure with the upper level containing method categories (i.e. protein analysis, RNA analysis) and lower level containing individual techniques such as Western blotting, DNAase footprinting etc. The 'Experiment result' classification describes what the experiment resulted in (i.e. promoter mapping, regulatory site mapping, gene/operon repression).

The 'Effector' classification uses a tree-like hierarchy of effectors where classes of the hierarchy are mainly based on the Chemicals and Drugs Category of MESH [31].

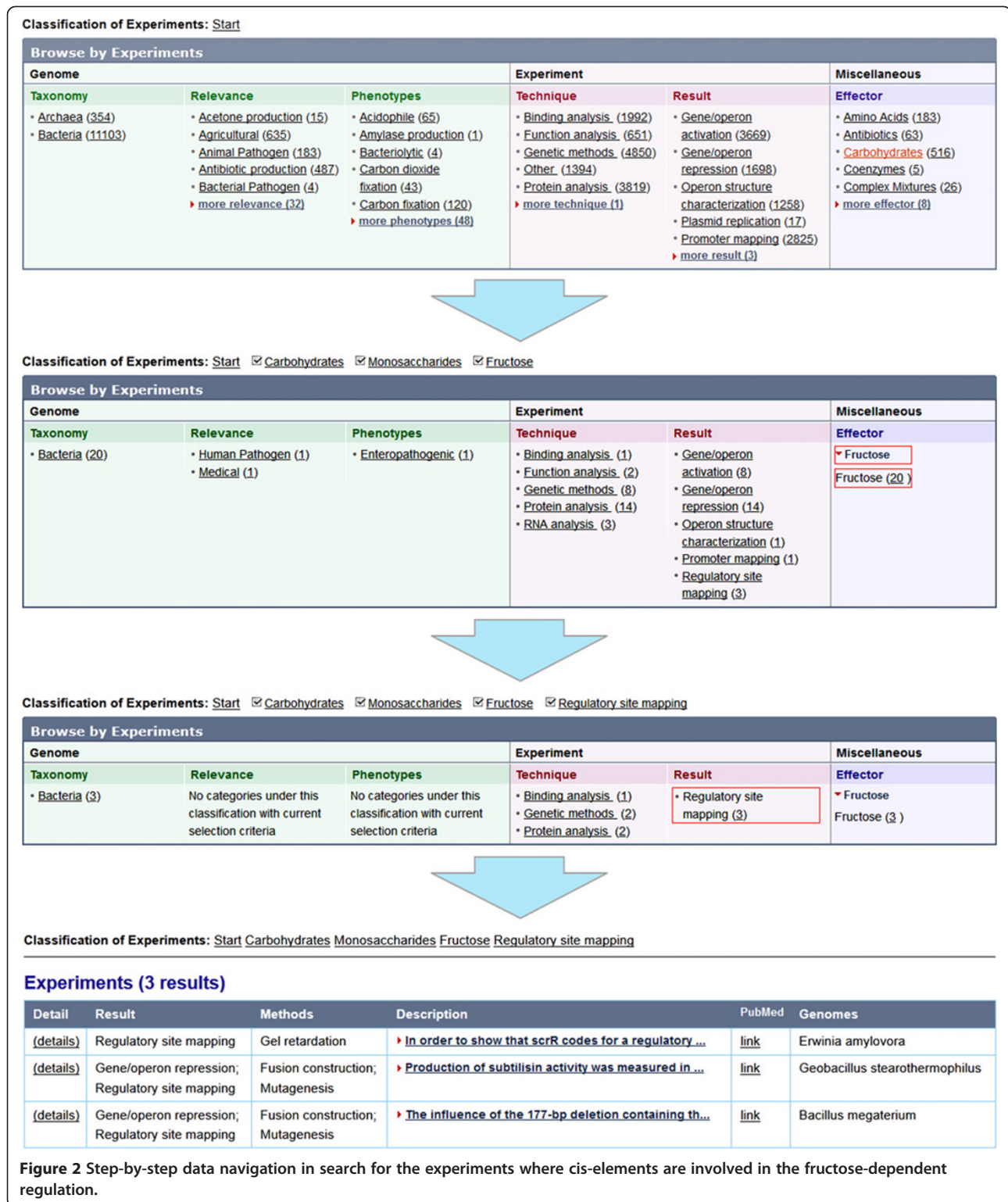
User can browse all categories in the database by choosing a term in one classification and then narrowing a result by choosing terms in other classifications as additional filters. At any time, the user can click on the number beside the classification to get articles fitting all criteria currently selected.

For example, we want to know if there is any data on experiments with cis-elements that are involved in

The screenshot shows the RegTransBase website interface. At the top right, there are links for 'RegPrecise', 'RegPredict', 'Your Cart', and 'Release v7'. Below the site logo, a navigation bar includes 'home', 'advanced search', 'browse', 'downloads', 'about us', and 'help'. A descriptive paragraph states that RegTransBase consists of two modules: a database of regulatory interactions based on literature and an expertly curated database of transcription factor binding sites. To the right, there are two featured boxes: 'Manually Curated Position Weight Matrices (PWM)' and 'Putative Regulons from Experimental Data'. The main content area is titled 'Classification of Experiments: Start' and features a 'Browse by Experiments' table. This table is organized into three columns: 'Genome', 'Experiment', and 'Miscellaneous'. The 'Genome' column lists 'Taxonomy' (Archaea: 508, Bacteria: 15195), 'Relevance' (Acetone production: 42, Agricultural: 1115, Amino acids production: 1143, Animal Pathogen: 809, Antibiotic production: 489, more relevance: 39), and 'Phenotypes' (Acidophile: 102, Alpha-hemolytic: 155, Amylase production: 5, Bacteriolytic: 4, Barophile: 9, more phenotypes: 60). The 'Experiment' column lists 'Technique' (Binding analysis: 3571, Function analysis: 889, Genetic methods: 8400, Other: 1436, Protein analysis: 5501, more technique: 1) and 'Result' (Gene/operon activation: 5675, Gene/operon repression: 3154, Operon structure characterization: 1927, Plasmid replication: 17, Promoter mapping: 4126, more result: 3). The 'Miscellaneous' column lists 'Effector' (Amino Acids: 266, Antibiotics: 109, Carbohydrates: 676, Coenzymes: 5, Complex Mixtures: 34, more effector: 8). Below the table is a 'Search by' section with three input fields: 'Genes and Regulators', 'Effectors', and 'Full Text', each with a search icon. To the right of the search section is a statistics box showing: Articles Curated: 7213, Experiments: 18877, Organisms: 658, Genes: 59754, Sites: 13321, Regulators: 1488, Effectors: 875. At the bottom right, there is a 'News' section dated 'November 12, 2012'.

Browse by Experiments			Experiment		Miscellaneous
Genome			Technique	Result	Effector
<ul style="list-style-type: none"> Archaea (508) Bacteria (15195) 	<ul style="list-style-type: none"> Acetone production (42) Agricultural (1115) Amino acids production (1143) Animal Pathogen (809) Antibiotic production (489) more relevance (39) 	<ul style="list-style-type: none"> Acidophile (102) Alpha-hemolytic (155) Amylase production (5) Bacteriolytic (4) Barophile (9) more phenotypes (60) 	<ul style="list-style-type: none"> Binding analysis (3571) Function analysis (889) Genetic methods (8400) Other (1436) Protein analysis (5501) more technique (1) 	<ul style="list-style-type: none"> Gene/operon activation (5675) Gene/operon repression (3154) Operon structure characterization (1927) Plasmid replication (17) Promoter mapping (4126) more result (3) 	<ul style="list-style-type: none"> Amino Acids (266) Antibiotics (109) Carbohydrates (676) Coenzymes (5) Complex Mixtures (34) more effector (8)

Figure 1 Home page of RegTransBase. Data navigation panel with its major classifications in the middle of the page.



fructose-dependent regulation. By using the ‘Effectors’ classification in three steps: ‘Carbohydrates’ -> ‘Monosaccharides’ -> ‘Fructose’ we find a list of 20 experiments (Figure 2).

A subsequent choosing of the ‘Regulatory site mapping’ term in the ‘Result’ classification produces a list of 3 experiments where cis-elements involved in fructose-dependent regulation were studied.

Search methods

RegTransBase provides a user with a broad range of search options such as search by Gene name, effector name, or a full text search of an abstract. Search for genes involved in regulatory experiments can be done using the gene name, function, product, accession number, or any other GenBank annotation. Searching for effectors by their name extracts the information on regulator, experiment, and genome with all associated links. Full text search allows for running complex queries against the abstracts and experiment descriptions such as '+mga +promoter'.

Putative regulons from experimental data

Identification of transcription factor binding motifs is an important step in the computational reconstruction of regulatory elements. The 'Putative Regulons' section of RegTransBase provides sets of upstream sequences of target genes for each regulon. These sets can be used for the identification of conserved DNA motifs that may bind transcriptional regulators.

Use Case 1: use of Putative Regulon for the search of a TF binding motif

1. Find genome of interest on the Putative Regulons page.
2. Find regulon of interest based on the regulator name.
3. Get a set of upstream sequences by clicking the 'Download' link in the 'Upstream sequences column of regulons table.
4. Start RegPredict [23], select genomes of interest.
5. Open 'Discover Profiles', paste upstream sequences (at least three sequences).
6. Select profile parameters (palindrome recommended), start search.
7. Select profile with highest informational content and run search for sites in selected genomes.

This scheme was successfully tested for the TnrA regulon from *B. subtilis*.

Manually curated position weight matrices (PWM)

Positional weight matrices from RegTransBase collections can be used for computational prediction of TFBSs

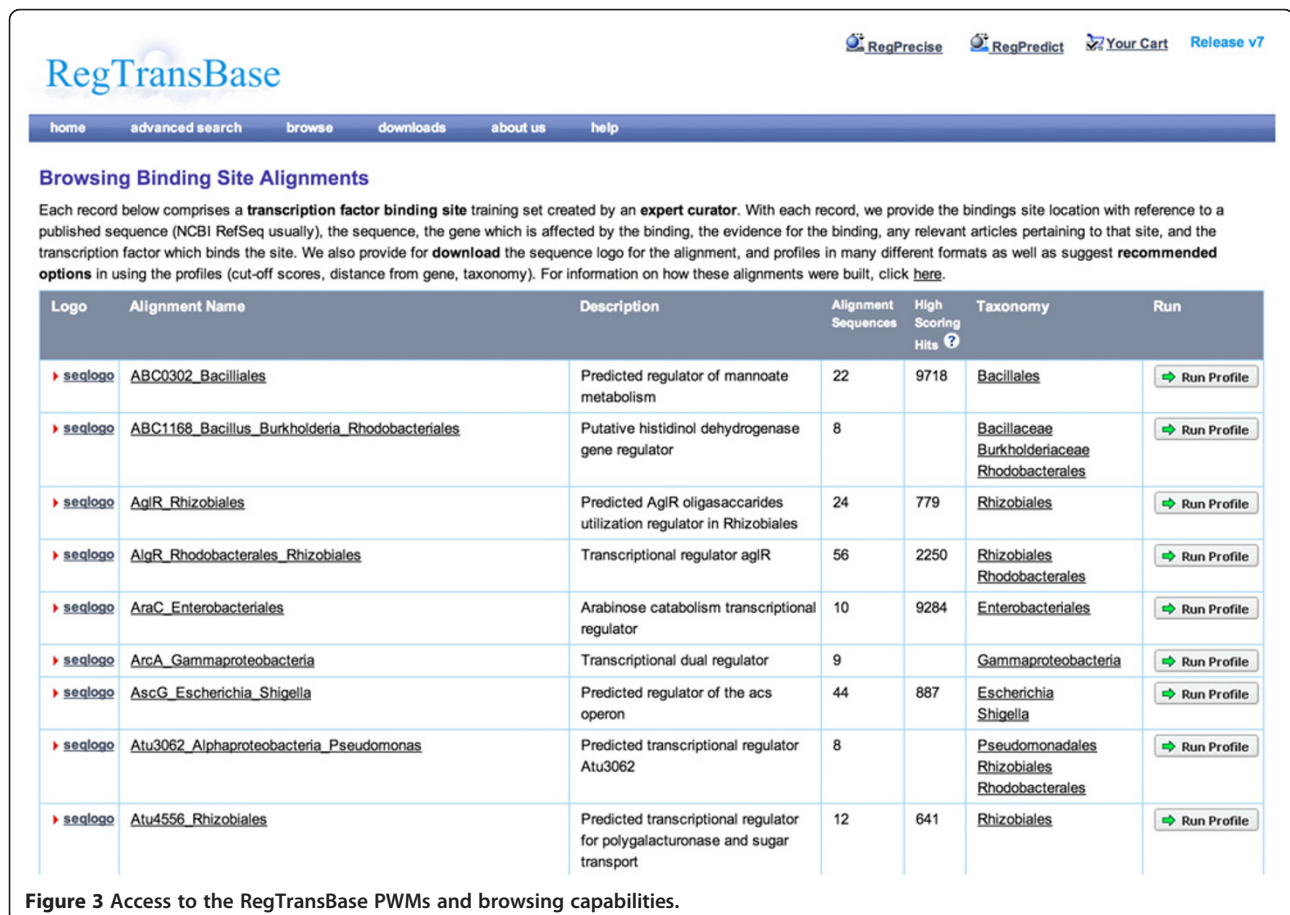


Figure 3 Access to the RegTransBase PWMs and browsing capabilities.

using RegPredict [23] or other software for PWM-based TFBS search. Figure 3 shows an access page to the RegTransBase PWMs and the associated data. A user selects a PWM of interest from the list and opens a webpage with PWM description. PWMs are available for download in different formats including a binding site alignment in FASTA format, matrices in MAST and Transfac formats and as a frequency matrix.

Use Case 2: use of manually curated PWM for computational reconstruction of a regulon

1. Open a list of the binding site alignments (http://regtransbase.lbl.gov/cgi-bin/regtransbase?page=alignment_browse).
2. Find a regulator of interest (for example, ABC0302).
3. Open the page with the ABC0302 binding sites alignment (http://regtransbase.lbl.gov/cgi-bin/regtransbase?page=show_alignment&matrix_id=95).
4. Download an alignment in FASTA format (First option in Download section at the bottom of the page).
5. Go to the RegPredict website (<http://regpredict.lbl.gov/>).
6. Start RegPredict (click 'Start Application')
7. Click 'Select genomes'.
8. Find recommended taxonomical group (Bacillales - see the 'Recommended options' section on ABC0302 page in RegTransBase) and add all genomes from that group (or as many genomes as possible).
9. Click 'Run Profile'.
10. Select the 'Sequences' tab and paste your alignment of binding sites in the FASTA format.
11. Click 'Generate profile'.
12. Set search parameters 'Position from' and 'Position to' (see 'Recommended options' section on ABC0302 page in RegTransBase).
13. Click 'Run'.

Conclusions

RegTransBase, a user-friendly open-access database, provides biologists involved in the investigation of microbial regulation and systems biology with convenient access to experimental data collected in thousands of original studies. It allows a user to interact with a valuable collection of manually curated data on a range of experiments related to the transcriptional regulation of bacteria. These data, with associated analytical tools, provide a valuable resource to assist in investigation of gene functions in the constantly growing number of available genome assemblies. RegTransBase collection of PWMs is currently used by various tools for TF binding prediction and motif comparison (for example, MEME-ChIP [32] and TOMTOM [33] from MEME Suite, FITBAR [34], ISGA

[35], STAMP [13]. MicrobesOnline, an integrated portal for comparative and functional genomics [36], is cross-linked with RegTransBase.

As regulon inference is of significant importance for deciphering the regulation of biological processes, we believe that a current improved and expanded version of RegTransBase is a useful tool for the research community.

Availability and requirements

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

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MJC worked on the database and interface design, general data organization and access; PSN participated in the database and interface design and construction, and lead the putative regulon collection and RegPredict access projects; AEK was responsible for data collection and manual curation; DAR proposed several critical directions of the project and actively participated in discussions; APA was involved with the MicrobesOnline integration and general discussions; MSG conceived and performed general coordination of the project. ID supervised the project and was involved with all aspects of database design, construction and implementation. MJC, AEK and ID wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors are grateful to Igor Lukashin for obtaining the genome alignment data, MicrobesOnline team for useful discussions and encouragement, and to Tatiana Smirnova for the artistic and highly functional RegTransBase Web site.

This work conducted by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231. The work was also supported by the Director, Office of Science, Office of Biological and Environmental Research, Life Sciences Division, U.S. Department of Energy under Contracts No. DE-AC02-05CH11231 and No. DE-SC0004999.

Author details

¹Department of Microbiology, University of California Davis, Davis, CA 95616, USA. ²Lawrence Berkeley National Laboratory, Berkeley, CA 94710, USA. ³Sanford-Burnham Medical Research Institute, La Jolla, CA 92037, USA. ⁴A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia. ⁵Department of Bioengineering, University of California, Berkeley, CA 94720, USA. ⁶M.V. Lomonosov Moscow State University, Vorobyevy Gory 1073, Moscow 119991, Russia. ⁷Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA.

Received: 15 November 2012 Accepted: 22 March 2013

Published: 2 April 2013

References

1. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*: A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009, **462**(7276):1056–1060.
2. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC: The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012, **40**(Database issue):D571–D579.
3. Liu J, Xu X, Stormo GD: The cis-regulatory map of *Shewanella* genomes. *Nucleic Acids Res* 2008, **36**(16):5376–5390.
4. Rodionov DA: Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* 2007, **107**(8):3467–3497.

5. Rodionov DA, Novichkov PS, Stavrovskaya ED, Rodionova IA, Li X, Kazanov MD, Ravcheev DA, Gerasimova AV, Kazakov AE, Kovaleva GY, *et al*: **Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the Shewanella genus.** *BMC Genomics* 2011, **12**(Suppl 1):S3.
6. Xu X, Ji Y, Stormo GD: **Discovering cis-regulatory RNAs in shewanella genomes by support vector machines.** *PLoS Comput Biol* 2009, **5**(4):e1000338.
7. Gelfand MS: **Evolution of transcriptional regulatory networks in microbial genomes.** *Curr Opin Struct Biol* 2006, **16**(3):420–429.
8. Gerasimova A, Kazakov AE, Arkin AP, Dubchak I, Gelfand MS: **Comparative genomics of the dormancy regulons in mycobacteria.** *J Bacteriol* 2011, **193**(14):3446–3452.
9. Suvorova IA, Tutukina MN, Ravcheev DA, Rodionov DA, Ozoline ON, Gelfand MS: **Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gammaproteobacteria.** *J Bacteriol* 2011, **193**(15):3956–3963.
10. Vitreschak AG, Mironov AA, Lyubetsky VA, Gelfand MS: **Comparative genomic analysis of T-box regulatory systems in bacteria.** *RNA* 2008, **14**(4):717–735.
11. Wingender E: **The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation.** *Brief Bioinform* 2008, **9**(4):326–332.
12. Portales-Casamar E, Arenillas D, Lim J, Swanson MI, Jiang S, McCallum A, Kirov S, Wasserman WW: **The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences.** *Nucleic Acids Res* 2009, **37**(Database issue):D54–D60.
13. Griffith OL, Montgomery SB, Bernier B, Chu B, Kasaiian K, Aerts S, Mahony S, Sleumer MC, Bilenky M, Haeussler M, *et al*: **ORegAnno: an open-access community-driven resource for regulatory annotation.** *Nucleic Acids Res* 2008, **36**(Database issue):D107–D113.
14. 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**(Database issue):D120–D124.
15. Robison K, McGuire AM, Church GM: **A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome.** *J Mol Biol* 1998, **284**(2):241–254.
16. Sierro N, Makita Y, de Hoon M, Nakai K: **DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information.** *Nucleic Acids Res* 2008, **36**(Database issue):D93–D96.
17. Sharma D, Mohanty D, Surolia A: **RegAnalyst: a web interface for the analysis of regulatory motifs, networks and pathways.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W193–W201.
18. Baumbach J: **CoryneRegNet 4.0 - A reference database for corynebacterial gene regulatory networks.** *BMC Bioinforma* 2007, **8**:429.
19. Grote A, Klein J, Retter I, Haddad I, Behling S, Bunk B, Biegler I, Yarmolinetz S, Jahn D, Munch R: **PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes.** *Nucleic Acids Res* 2009, **37**(Database issue):D61–D65.
20. de Jong A, Pietersma H, Cordes M, Kuipers OP, Kok J: **PePPER: a webserver for prediction of prokaryote promoter elements and regulons.** *BMC Genomics* 2012, **13**:299.
21. Pachkov M, Erb I, Molina N, van Nimwegen E: **SwissRegulon: a database of genome-wide annotations of regulatory sites.** *Nucleic Acids Res* 2007, **35**(Database issue):D127–D131.
22. Kazakov AE, Cipriano MJ, Novichkov PS, Minovitsky S, Vinogradov DV, Arkin A, Mironov AA, Gelfand MS, Dubchak I: **RegTransBase—a database of regulatory sequences and interactions in a wide range of prokaryotic genomes.** *Nucleic Acids Res* 2007, **35**(Database issue):D407–D412.
23. 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**(Web Server issue):W299–W307.
24. 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**(Database issue):D111–D118.
25. Coordinators NR: **Database resources of the national center for biotechnology information.** *Nucleic Acids Res* 2013, **41**(Database issue):D8–D20.
26. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2005, **33**(Database issue):D501–D504.
27. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, *et al*: **Database resources of the national center for biotechnology information.** *Nucleic Acids Res* 2005, **33**(Database issue):D39–D45.
28. Bailey TL, Gribskov M: **Combining evidence using p-values: application to sequence homology searches.** *Bioinformatics* 1998, **14**(1):48–54.
29. Federhen S: **The NCBI Taxonomy database.** *Nucleic Acids Res* 2012, **40**(Database issue):D136–D143.
30. Liolios K, Tavernarakis N, Hugenholtz P, Kyrpides NC: **The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide.** *Nucleic Acids Res* 2006, **34**(Database issue):D332–D334.
31. Rogers FB: **Medical subject headings.** *Bull Med Libr Assoc* 1963, **51**:114–116.
32. Machanick P, Bailey TL: **MEME-ChIP: motif analysis of large DNA datasets.** *Bioinformatics* 2011, **27**(12):1696–1697.
33. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS: **Quantifying similarity between motifs.** *Genome Biol* 2007, **8**(2):R24.
34. Oberto J: **FITBAR: a web tool for the robust prediction of prokaryotic regulons.** *BMC Bioinforma* 2010, **11**:554.
35. Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q: **An Ergatic-based prokaryotic genome annotation web server.** *Bioinformatics* 2010, **26**(8):1122–1124.
36. 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**(Database issue):D396–D400.

doi:10.1186/1471-2164-14-213

Cite this article as: Cipriano *et al.*: RegTransBase – a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics* 2013 **14**:213.

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

