Lawrence Berkeley National Laboratory

Recent Work

Title
The analysis and expansion of regulatory binding site data in a wide range of bacteria using a semi-automatic system - RegTransBase

Permalink
https://escholarship.org/uc/item/0jr5t4zi

Authors
Novichkov, Pavel
Cipriano, Michael
Kazakov, Alexey
et al.

Publication Date
2008-06-02
The analysis and expansion of regulatory binding site data in a wide range of bacteria using semi-automatic system - RegTransBase

Pavel Novichkov, Michael Cipriano, Alexey Kazakov, Dmitriy Ravcheev, Adam Arkin, Mikhail Gelfand*, Inna Dubchak* (gelfand@iitp.ru,ildubchak@lbl.gov)

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

INTRODUCTION

RegTransBase, a database describing regulatory interactions in prokaryotes, has been developed as a component of the Microbial Stress Pathway Project at Lawrence Berkeley National Laboratory and in collaboration with 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.

http://regtransbase.lbl.gov

RESULTS

RegTransBase contains curated, validated, and cataloged regulatory binding site data from sequenced genomes. The data is used to identify known transcription factor binding sites and expand their knowledge in literature using a controlled vocabulary. The database, RegTransBase (RTB), a manually curated database of regulatory interactions, captures the knowledge in literature using a controlled vocabulary. RTB contains the following types of experimental data:

1. Alignments of Binding Sites
2. Regulatory site mapping
3. Terminator mapping
4. Operon structure characterization (with identified regulator)
5. Gene/operon repression (with identified regulator)
6. Phages
7. Promoter
8. Transcription factor binding sites
9. Additional analysis tools are available, such as sequence logos, sequence alignments in graphical and text formats, phylogenetic trees and the ability to view the alignment in the feature rich application JalView.

METHODS

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

Data

Experiments - taxonomy distribution

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Number of Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria</td>
<td>9347</td>
</tr>
<tr>
<td>Proteobacteria</td>
<td>6615</td>
</tr>
<tr>
<td>Alphaproteobacteria</td>
<td>2385</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>315</td>
</tr>
<tr>
<td>Gammaproteobacteria</td>
<td>3843</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>2003</td>
</tr>
<tr>
<td>Bacilli</td>
<td>1876</td>
</tr>
<tr>
<td>Clostridia</td>
<td>187</td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td>486</td>
</tr>
<tr>
<td>Archaea</td>
<td>345</td>
</tr>
<tr>
<td>Fungi</td>
<td>538</td>
</tr>
<tr>
<td>Phages</td>
<td>162</td>
</tr>
</tbody>
</table>

SUMMARY

RTB has a manually curated collection of over 100 position weight matrices and alignments (with plans for more in the future). We provide the ability to search sequenced genomes using these matrices or the user can supply their own alignment. Using a collection of interfaces we aim to provide a tool for the following situations:

- One matrix + one genome of interest
- Multiple matrices + multiple genomes
- One gene + multiple genomes
- Multiple matrices + multiple genomes

ACKNOWLEDGEMENT

ESPP2 (MDCASE) is part of the Virtual Institute for Microbial Stress and Survival (VIMSS) 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.