Review Article

J Integral Sci, 2023, Volume 6, Issue 2, 63-37

Journal of Integral Sciences [JIS]

[An International Open Access Journal]

Available at www.jisciences.com ISSN: 2581-5679

DATABASES AND ONLINE TOOLS THAT IS PARTICULARLY USEFUL FOR ANALYSING E. COLI TRANSCRIPTION

Kona Venkata Sri Krishna[1], Arrolla Lekha[2], Kanigiri Deepthi Sri[3], Mohammed Azharuddin[4]

[1] Jawaharlal Nehru Technological University, Hyderabad, TS, India

[2] Sreenidhi Institute of Science and Technology, Hyderabad, TS, India

[4] Vignan University, Guntur, AP, India

[3] K L University, Vijayawada, AP, India

Received: 17 May 2023 Revised: 29 May 2023 Accepted: 25 July 2023

Abstract

This paper presents general databases and webtools referred and used by the scholars and bio manufacturing research scientists to predict, analyze transcription in e.coli transcription. Over a decade various databases were generated covering biophysical models, promoters performance data, secondary structure of mRNA, polymerases working mode, inhibitors on mRNA, thermodynamics, genetic instability, toxic proteins genes, rate of intiations etc. Based on Artifical Intelligence machine learning webtools turning to be more efficient and reliable to prevalidate performance of the new genes that are expressed in e.coli.

Keywords: Biomanufacturing, e.coli transcription, mRNA, thermodynamics, Artificial Intelligence.

***Corresponding Author**
Kona Venkata Sri Krishna

DOI: https://doi.org/10.37022/jis.v6i2.59

Produced and Published by
South Asian Academic Publications

Introduction

Ecoli transcription analysis involves collecting genetic material from bacteria and using it to identify the genes responsible for a certain feature or function. This knowledge may then be applied to the study of bacterial genetics and the development of novel techniques for changing genes to produce particular effects. E.coli transcription analysis is used to explore a wide range of bacterial traits and functions, including the bacteria's adaptability to different environments, resistance to antibiotics, and capacity to create toxins. Because researchers have a stronger grasp of the genes involved in the researched processes and characteristics, they may create novel strategies for altering genes to accomplish particular effects [1].

A regulator that interacts with a particular regulatory DNA region is responsible for guiding the expression of a gene during RNA synthesis in the prokaryotic process of transcriptional regulation. A transcription factor (TF) is a protein that is essential for activating or inhibiting gene transcription but is not an enzyme. Some transcription factors (TFs) solely bind to cis-acting DNA sequences, while others bind to both DNA and other TFs. Global transcription regulators (ii) control a complex regulatory cascade by not only directly regulating the expression of specific genes but also indirectly regulating diverse cellular pathways by acting on a set of local regulators controlling only one or a few genes, and (iii) acting on target promoters that use distinct sigma factors. Global transcription regulators (GTRs) are essential transcription factors (TFs) that conduct essential physiological tasks but are excluded from this concept. Seven global transcription regulators (CRP, FNR, IHF, Fis, ArcA, NarL, and Lrp) are predicted to regulate half of all genes in E. coli, while sixty transcription factors (TFs) each control a single promoter [2].

Understanding gene regulation systems demands a comprehensive understanding of DNA-binding transcription factors (TFs). A genome that has been sequenced may be annotated with early information on the factors that encode transcription factor encoding factors (TFs). This may be performed by identifying factors that are similar to known transcription factors (TFs) or by using functional classification algorithms that categorise proteins as transcriptional regulators. Modern, more potent TF prediction methods are based on the gathering and assignment of DNA-binding motifs by computer [3]. This allows for the prediction of TFs over the whole genome of the model bacterium E. coli. Identifying transcription factors (TFs) from a broad variety of bacterial species with current whole genome sequencing the authors get the clusters of orthologous groups (COGs) involved in transcription control from the COG database. Enzyme-related COGs are subsequently eliminated in the next phase. Each of the 128 resulting groupings included orthologous transcription factors (TFs) that originated from several genomes. These COGs of known and hypothetical transcription factors account for 85 percent of the list of E. coli TFs when projected to E. coli K-12 [4]. Among the most prominent advantages of this method is that ecoli transcription analysis is a reasonably straightforward procedure. Molecular biologists simply collect genetic material from bacteria and apply the

proper tools to identify the genes responsible for a particular characteristic or function. This data may then be used to investigate the genetics of bacteria and create revolutionary genetic manipulation techniques.

Databases of Transcription

A Database of Transcriptional Regulation in Escherichia coli K-12 (RegulonDB)

RegulonDB is a database that organises and integrates decades' worth of data on transcriptional regulation in Escherichia coli K-12 acquired from laboratories throughout the world. The database focuses specifically on Escherichia coli K-12. In addition, RegulonDB simulates the organisation of genes into transcription units such as operons and regulons. In all, 120 unique short RNAs control 192 genes and interact with one another in 231 various ways. The whole repertoire of 189 genetic sensory-response units (GENSOR units), as well as the signals, regulatory linkages, and metabolic processes involving these GENSOR units, have been identified. There are 78 GENSOR units that emphasise all four components, 119 GENSOR units that highlight both the genetic switch and the response, and 2 GENSOR units that highlight just the genetic switch. RegulonDB includes 103 transcription factors (TFs) known to be connected with an effector, as well as 25 two-component systems. There were sufficient sites to construct motifs for 93 transcription factors, enabling the prediction of 16,207 TF-binding sites. The graphical representation of an operon includes not only the genes that make up the operon's multiple transcription units (TUs), but also the regulatory components involved in the transcription and control of those TUs. "Operon" is a structural unit that includes all of the genes and regulatory elements in this context. An operon may contain several promoters that are situated in close proximity to one another. If so, the operon may also include dual-binding sites, suggesting that one site may activate one promoter while suppressing another. On the same page, under the operon, is a list of the multiple TUs revealed to be associated with it [5].

Table 1: Transcriptional_Regulation

TF Name	Target Gene	Binding Site	Promoter Element	Operon	Transcription Unit
CRP	lacZ	TTGACAGT AGTCA	P_lac	Lac Operon	lacZ
CRP	lacY	TTGACAGT AGTCA	P_lac	Lac Operon	lacY
CRP	lacA	TTGACAGT AGTCA	P_lac	Lac Operon	lacA
Fis	ftsZ	TGTTCTGA TAAA	P_fis	ftsZ Operon	ftsZ
Fis	sulA	TGTTCTGA TAAA	P_fis	ftsZ Operon	sulA
Fis	yafQ	TGTTCTGA TAAA	P_fis	ftsZ Operon	yafQ

- **TF_Name**: The name of the transcription factor (TF) that regulates the target gene.
- **Target_Gene**: The gene that is regulated by the transcription factor.
- **Binding_Site**: The DNA sequence where the transcription factor binds to regulate the target gene.
- **Promoter_Element**: The promoter region associated with the target gene.
- **Operon**: The operon to which the target gene belongs.
- **Transcription_Unit**: The transcription unit that includes the target gene.

Database to Study Extragenic Areas of Bacteria (ExtraTrainDB)

Identification of regulatory proteins and DNA sequences is essential for comprehension of transcriptional control mechanisms. ExtraTrain is a unique database that may be used to study extragenic areas and give transcription information in archaea and bacteria. ExtraTrain includes all extragenic regions that correspond to fully annotated bacterial and archaeal genomes accessible through NCBI, as well as all regulatory proteins identified in UniProt that belong to the most important transcriptional regulatory protein families.Santa Cruz, California's University of California developed ExtraTrain. This database contains a tool for studying the palindromic sequence contained in bacterial genomes. ExtraTrain has 26046 regulatory proteins in total. These protein families include AraC/XylS, ArsR, AsnC, the Cold Shock Domain, CRPFNR, DeoR, GntR, IclR, LacI, LuxR, LysR, MarR, MerR, NtrC/Fis, OmpR, and TetR. ExtraTrain gives information on the genes upstream and downstream of the extragenic DNA regions. This material derives from bacteria and archaeal genomes that are publicly accessible. I An examination of the extragenic areas corresponding to a collection of strongly co-regulated genes is available to the user. The raw data may be the outcome of gene expression-measuring microarray research. Using a RefSeq or UniProt identification, the user may get the necessary extragenic regions and add them to the "working set." Both of these IDs are available. These sequences may then be submitted to Palinsight for analysis. Consequently, similar patterns may be detected among a set of co-regulated genes using an interactive step-by-step process.

i) looking for repeated characteristics in extragenic areas that relate to a particular family of regulatory proteins.

ii) ExtraTrain provides users with the means to investigate specific components of binding sites belonging to a certain family of regulatory proteins.

iii) Searching a genetic element for recurrent extragenic palindromic sequences, often known as REP sequences,

iv) The pursuit of termination proteins

v) Identification of the global regulatory site

vi) The study of insertion locations for Insertion Sequence elements

By selecting the upstream and downstream extragenic regions of many copies of an insertion sequence, the user is able to analyse and compare inverted repeats and insertion site characteristics. BLAST-like investigation of extragenic areas associated with a group of transcriptional regulatorsThe expression of bacterial transcriptional regulators is often controlled by the regulators themselves. Consequently, it is likely that identical signals will be discovered in the DNA regions upstream of a collection of similarly related transcriptionally regulating proteins. ExtraTrain enables users to see and compare the characteristics of these extragenic areas [6].

Data-Driven Approach to Collecting Transcription Factor Binding Sites in the Bacteria Domain (CollecTF)

It gathered information on naturally occurring TF-binding sites that had been experimentally validated across the Bacteria domain. It prioritised transparency in the curation process, the quality and accessibility of stored data, and fully adjustable record access. CollecTF combines several data sources in an accessible and automated manner, enabling users to dynamically modify binding motifs and the experimental support base for those motifs. Direct interactions between transcription factors and DNA play a significant role in bacterial transcription networks. CollecTF may dynamically integrate many sources of evidence by arbitrarily establishing a leader site and using two simple pair-wise distribution procedures. The evidence from two motif-linked sites is merged if the overlap is more than 75% of the combined sites' total length. We determine if evidence from non-motif-associated sites should be included in a leader site by analysing whether or not these sites completely overlap with any of the combined motif-linked sites. CollecTF was designed to standardise and streamline the data collection process for transcription factor binding sites for use in comparative genomics and machine learning techniques. CollecTF also contains a set of motif comparison tools that enable users to evaluate the degree of similarity between motifs obtained from different searches [7].

Table 2: CollecTF database

TFBS ID	Bacterial Species	Transcription Factor	Binding Site Sequence	Regulatory Effect	Experimental Evidence
TFBS001	Escherichia coli	CRP	TGTGAATCACATCA	Activator	EMSA, Footprinting
TFBS002	Bacillus subtilis	CodY	ATGTCTTTCAGATCG	Repressor	ChIP-seq
TFBS003	Pseudomonas aeruginosa	LasR	AGCCATGTGATAAG	Activator	Yeast one-hybrid
TFBS004	Staphylococcus aureus	SarA	TATTTATACGTATA	Repressor	DNaseI footprinting
TFBS005	Salmonella enterica	Fis	TTGATGACTAGGCAA	Activator	SELEX, ChIP

Database of Bacterial Operon Prediction (OperomeDB)

Using RNA-seq datasets and this database for bacterial genome operon prediction, we were able to predict bacterial genome operons. For each bacterial genome, diverse RNA-seq datasets were analysed, and computer models were used to predict operons based on these datasets. According to Rockhopper, convergent gene pairs with an overlapping head-to-head structure may be found in a broad variety of eukaryotic and prokaryotic organisms. It applies to prokaryotic genomes, where it is hypothesised that they play a role in gene regulation at both the transcriptional and post-transcriptional levels, despite the fact that the mechanisms that contribute to their frequency across genomes and the molecular basis for their occurrence remain unknown [8].

Database of Transcriptional regulator in Bacteria (BacTregulators)

This database compiled and consolidated data on human transcriptional regulator proteins from known families.It is possible for transcriptional regulators to fine-tune gene expression to make it more robust to environmental stresses and changes. This is done through the delicate interaction between sigma factors, which impart promoter selectivity, and transcriptional regulators, which govern RNA polymerase activity. This interaction happens at the level of the promoter. The information that BacTregulators may hold includes sequences, knowledge, and references.

a) Each protein sequence's data has been integrated, filtered, and evaluated in a unique manner. For each regulator family, we provide three non-redundant collections of full protein sequences:

(i) The NCBI maintains a collection of all protein sequences derived from the integration of SPTR sequences with bacterial genomes.

(ii) Combining the sequences from the SWISS-PROT, TrEMBL, and TrEMBL-new databases yields this collection of protein sequences.

(iii) Protein sequences were extracted from the bacterial and archaeal genomes of the NCBI. Similar to the AraC-XylS database, it includes these three groups of sequences by domains (DNA-binding domain, N-terminal domain, and C-terminal domain) in nine extra sets of sequences for each family of regulators. The arrangement of these sequences is N-terminal domain, DNA-binding domain, and C-terminal domain.

b) The knowledge extraction method used by BacTregulators is a combination of automatic data extraction and human information extraction. A portion of the data associated with each item was automatically obtained from the NCBI Microbial Resources and the SPTR Knowledgebase. The remaining information for each item was gathered manually from bibliographic sources. Due to the one-of-a-kind nature of this data collection, we developed a specialised structure, with the paragraph of text serving as the information unit. Each piece of manually obtained data is organised into text paragraphs with its own set of citations. This structure may be used to identify the origin of each individual data item. We examined experimental data to see how it supported biological characteristics while manually extracting data. These findings are referred to as "experimental evidence" in BacTregulators, and they get particular consideration. When experimental evidence supports the claims stated in a paragraph, a link to the relevant experimental data is provided. Using these basic facts, the veracity of knowledge-based data may be determined. Each paragraph of the text has a citation and a link to the abstract on Medline. Currently, the reference database includes 457 AraC-XylS and TetR family references [9].

Regulatory Networks in Prokaryotes: The PRODORIC Database

It is a database including annotated data on the regulation of gene expression in prokaryotes. It contains a vast quantity of gene regulatory information, including transcription factor binding locations, promoter architecture, and gene expression patterns. The whole dataset, which is based on the publication and extraction

of findings from scientific articles, is subject to human curation. It consists of interactive database-assisted prediction and validation of gene regulatory networks. The latest edition of the PRODORIC database contains gene regulatory information for several newly discovered bacteria. This database is accessible via one of four primary channels:

(i) Using the accessible online forms to submit a database query.

(ii) Utilizing the genome browser GBpro to go through the information

(iii) Using ProdoNet to investigate the regulatory network as a graphical graph

(iv) Web services are used to access a database (using an interface based on the Simple Object Access Protocol, or SOAP).

The Virtual Footprint application provides "regulon analysis," which is the method of analysing whole genomes with a single PWM. The second programming approach is called "promoter analysis," and it applies all of the possible patterns for a certain sequence. Similar to regulog research, a comparative examination of orthologous promoter sequences may be utilised to evaluate the evolutionary conservation of virtual footprint matching. The SMILE tool was one of the odd tools that they deployed. It is possible to examine both sequence and positional conservation within a set of orthologous matches. This method permits the examination of potential transcription factor targets and facilitates the removal of overly optimistic projections [10].

Biologically Similar to Conserved Operations: Databank ODB

This database enables the retrieval of information from multiple entire genomes, all of which include known operons. In addition, we present hypothetical operons that are conserved similarly to known operons. This technique integrates four kinds of associations: functional linkages in biological pathways, gene co-expression from microarray data, genomic context, and conservation of gene order across genomes. It contains information on operons described in scientific literature as well as hypothetical operons that are conservationally comparable to known operons. There are four sorts of links between genes that define an operon in this system: (i) intergenic distances; (ii) functional linkages in biological pathways; (iii) gene co-expression utilising microarray data; and (iv) gene order conservation across diverse genomes. I will discuss intergenic distances. (ii) Functional connections to biological processes(iii) Microarray data-driven gene co-expression (iv ODB also offers a method for predicting operons using the four relationships, since the conditions for identifying putative operons are stringent and do not apply to the whole genome [11].

Transcriptional Factor Database for Proteomic and Genomic Profiling: P2TF

P2TF, which stands for Predicted Prokaryotic Transcription Factors, is an integrated and exhaustive transcription factor protein database. The database includes annotation, categorization, and visualisation of transcription factor (TF) genes as well as the genomic context in which they are identified, making it a one-stop-shop for transcription factor (TF) research. In comparison to screening just predicted proteomes, the P2TF database analyses transcription factors (TFs) in both predicted proteomes and reconstituted ORFeomes, resulting in about 3% more TF proteins. The P2TF analysis then classifies DBDs as transcription factors or "other DNA-binding proteins," which include non-regulatory DNA-binding proteins such as integrases, tranposases, and histone-like proteins. Then, TFs are broken further into subcategories depending on their domain architecture. The subcategories consist of TRs, OCSs, RRs, and SFs. In the P2TF database, there are two sorts of searches: blast searches and keyword searches. Using the first search method, users may request genes based on their locus-tag, gene name, GI (GenBank Identifier), or domain ownership. Users may examine predictions by querying a particular genome or a collection of genomes from the same taxon. This allows users to limit the scope of their search. One may run a taxonomy search using the lineage name, the taxon-id, or the species name [12].

Database-Based Encoding System for Biopolymer Interactions: BIND

BIND was created to hold interactions and reactions involving biopolymers (such as protein, RNA, and DNA), in addition to tiny molecules, lipids, and carbohydrates. Utilizing a database structure designed to store all information pertinent to molecular assembly, this was accomplished. BIND is capable of encoding information on molecular processes, such as the chemical product (or products) of an enzyme process. Bind can store information on molecular interactions with atomic precision. Additionally, BIND has a very broad taxonomic reach, which means it may represent any organism having a taxon identity in the NCBI/EMBL/DDBJ taxonomy. This allows BIND to be used for a vast array of applications [13].

Webtools of Transcriptions

An Online Tool for Reconstructing Transcriptional Regulons by Comparative Genomics: RegPrecise

The prediction of gene regulation at the genome scale, as well as the reconstruction of transcriptional regulatory networks in prokaryotes, is one of the most serious problems facing contemporary genomics. Due to substantial differences in their lives and natural environments, several taxonomic groupings of bacteria have transcriptional regulation networks that are notably distinct from one another. Comparative genomics techniques are beneficial for reconstructing bacterial regulons and networks in silico. These are controlled by transcription factors (TFs) and RNA regulatory elements (RREs) (riboswitches). RegPrecise is an online tool for collecting, presenting, and analysing transcriptional regulons rebuilt by comparative genomics. Our efforts have significantly expanded a previously provided reference collection of hand-selected regulons. It allows access to reconstructed transcriptional regulons in bacterial genomes. Investigating the content, structure, and function of regulons, TF binding site motifs, and conservation and modifications in genome-wide regulatory networks across all taxonomic groupings of bacteria are analytical skills [14].

Tools to Analyze the Interaction of RNA Polymerases, Transcription Factors, and Sigma Factors with Prokaryotic Genes: PePPER

It is challenging to accurately forecast the DNA patterns with which RNA polymerases, transcription factors, and sigma factors seek to interact with factors (TFs) in prokaryotes, in part because many aspects of DNA sequences and structures in promoter regions have yet to be identified. There are currently improved ways for predicting and comparing transcription factor binding

sites (TFBSs), transcription factors (TFs), and regulon members that are associated with them. The PePPER webserver provides users with a suite of tools for extracting regulons, TFBSs, and promoters in addition to a comprehensive analytical technique [15].

Computational Prediction of Ribosomal RNA: RNAmmer

Because ribosomal RNAs (rRNAs) are required for ribosome function, their sequences and structures have remained unchanged throughout evolution. Consequently, rRNAs are used often in comparative research methods like phylogenetic inference. However, ribosomal RNA (rRNA)-encoding genes are often mislabeled or classified inconsistently. Consequently, comparative research employing rRNA genes is problematic. Therefore, we constructed computational predictors for the most important rRNA species identified in all kingdoms of life and incorporated them into the RNAmmer programme. The application utilises hidden Markov models that were trained using the 5S ribosomal RNA database. an initiative to develop a database of ribosomal RNA from Europe [16].

A Software Package to Study cis-regulatory Sequences: RSAT

Regulatory Sequence Study Tools (RSAT) is a modular software package developed to investigate cis-regulatory sequences. Elements in genomic sequences. Its principal uses include the following:

(i) discovery of motifs applicable to genome-wide data sets such as ChIP-seq;

(ii) Transcription factor binding motif quality assessment, categorization, and comparisons

(iii) in addition to comparative genomics

(iv) Examining variations in regulating processes

The Web interface contains 52 tools for performing a variety of analyses, such as obtaining sequences, discovering motifs ab initio, scanning sequences to predict transcription factor (TF) binding sites, comparing and clustering motifs, analysing TF binding site conservation and divergence, detecting inter-individual regulatory variations, and building control sets based on a wide variety of probabilistic models. These studies may be conducted by obtaining sequences, scanning sequences to predict TF binding sites, etc [17].

Conclusion

Ecoli transcription analysis is used to research a wide number of characteristics and processes in bacteria. Researchers are able to design novel strategies for modifying genes in order to attain certain results because they have a better grasp of the genes involved in the processes and processes being studied. It is predicted that around 60 transcription factors (TFs) each only control a single promoter. Annotation of a sequenced genome can provide information about the factors that encode transcription factors (TFs). This can be done by identifying factors that are homologous to known TFs or by using functional classification schemes that assign proteins to the category of transcriptional regulation. Software can be used to analyse the transcription of E. coli genes. The Transcription Factor Binding Site Tool (TFBS) can locate transcription factor binding sites in a transcriptome. RegulonDB is a database that simulates the organisation of genes into transcription units such as operons and regulons. ExtraTrain is a database that can be used to investigate the palindromic sequence found in the genome of prokaryotic organisms. ExtraTrain contains all of the DNA extragenic sections as well as information about the genes that are upstream and downstream from those areas.

ExtraTrain, CollecTF and Palinsight are tools to investigate certain aspects of the binding sites that correspond to a particular family of regulatory proteins. The aim of ExtraTrain is to make it possible to see and compare the characteristics of these extragenic areas. CollecTF was developed with the intention of standardising and streamlining the process of data collection on transcription factor binding sites for the use of comparative genomics and machine learning techniques. The database comprised a collection of information on proteins belonging to identified families of transcriptional regulators in the human genome. BacTregulators is a database that contains information that has been annotated on the regulation of gene expression in prokaryotes. It combines a vast collection of data on gene regulation, such as transcription factor binding locations, promoter architecture, and gene expression patterns. The PRODORIC database includes gene regulatory information for a number of newly discovered bacteria. This system includes four different kinds of associations between genes that determine an operon. It offers data on operons that have been recorded in the scientific literature as well as potential operons.

Researchers now have a one-stop shop to conduct their study on transcription factors (TFs). P2TF analysis sorts DBDs into transcription factors and "Other DNA-binding Proteins" BIND archives interactions and reactions that result from biopolymers such as protein, RNA, and DNA. BIND is an online tool that allows for the reconstruction of transcriptional regulatory networks in prokaryotes. Any organism that has a taxon identification in the NCBI/EMBL/DDBJ taxonomy may be represented in BIND. It provides access to the rebuilt transcriptional regulons in bacterial genomes. RNAmmer software makes use of hidden Markov models that have been trained on data taken from the 5S ribosomal RNA database. The Web interface combines 52 tools that can be used to perform various types of analyses, such as obtaining sequences and scanning sequences

Funding

No Funding

Conflict of Interest

Authors are declared that no conflict of Interest

Inform Consent & Ethical Considerations

Not Applicable

Author Contribution

All authors are contributed equally.

References

1. Levin BR, Bergstrom CTJPotNAoS. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. 2000;97(13):6981-5.
2. Seshasayee ASN, Sivaraman K, Luscombe NMJAhotf. An overview of prokaryotic transcription factors. 2011:7-23.
3. Roulet E, Bucher P, Schneider R, Wingender E, Dusserre Y, Werner T, et al. Experimental analysis and computer prediction of CTF/NFI transcription factor DNA binding sites. 2000;297(4):833-48.
4. Merino E, Yanofsky CJTig. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. 2005;21(5):260-4.
5. Tierrafría VH, Rioualen C, Salgado H, Lara P, Gama-Castro S, Lally P, et al. RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in Escherichia coli K-12. 2022;8(5):000833.
6. Pareja E, Pareja-Tobes P, Manrique M, Pareja-Tobes E, Bonal J, Tobes R. ExtraTrain: a database of Extragenic regions and Transcriptional information in prokaryotic organisms. BMC Microbiology. 2006;6(1):29.
7. Kılıç S, White ER, Sagitova DM, Cornish JP, Erill IJNar. CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. 2014;42(D1):D156-D60.
8. Chetal K, Janga SCJBri. OperomeDB: a database of condition-specific transcription units in prokaryotic genomes. 2015;2015.
9. Martínez-Bueno M, Molina-Henares AJ, Pareja E, Ramos JL, Tobes R. BacTregulators: a database of transcriptional regulators in bacteria and archaea. Bioinformatics (Oxford, England). 2004;20(16):2787-91.
10. Dudek C-A, Jahn DJNAR. PRODORIC: state-of-the-art database of prokaryotic gene regulation. 2022;50(D1):D295-D302.
11. Okuda S, Katayama T, Kawashima S, Goto S, Kanehisa M. ODB: a database of operons accumulating known operons across multiple genomes. Nucleic Acids Research. 2006;34(suppl_1):D358-D62.
12. Ortet P, De Luca G, Whitworth DE, Barakat MJBg. P2TF: a comprehensive resource for analysis of prokaryotic transcription factors. 2012;13(1):1-8.
13. Bader GD, Betel D, Hogue CWJNar. BIND: the biomolecular interaction network database. 2003;31(1):248-50.
14. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, et al. RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. 2013;14(1):1-12.
15. de Jong A, Pietersma H, Cordes M, Kuipers OP, Kok JJBg. PePPER: a webserver for prediction of prokaryote promoter elements and regulons. 2012;13(1):1-10.
16. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DWJNar. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. 2007;35(9):3100-8.
17. Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, Delerce J, et al. RSAT 2015: regulatory sequence analysis tools. 2015;43(W1):W50-W6.