Bioinformatics, 35(18), 2019, 3541–3543
doi: 10.1093/bioinformatics/btz093
Advance Access Publication Date: 6 February 2019
Applications Note

Databases and ontologies

RiboD: a comprehensive database for prokaryotic riboswitches

Sumit Mukherjee[1,*], Sukhen Das Mandal[2], Nikita Gupta[3], Matan Drory-Retwitzer[4], Danny Barash[4] and Supratim Sengupta [1,*]

[1]Department of Physical Sciences and [2]Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, India, [3]Department of Pharmaceutical Engineering & Technology, Indian Institute of Technology (BHU) Varanasi, Varanasi, India and [4]Department of Computer Science, Ben-Gurion University, Beer-Sheva 84105, Israel

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on November 12, 2018; revised on January 24, 2019; editorial decision on February 3, 2019; accepted on February 5, 2019

Abstract

Summary: Riboswitches are cis-regulatory non-coding genomic segments that control the expression of downstream genes by undergoing conformational change upon ligand binding. We present a comprehensive database of prokaryotic riboswitches that allows the user to search for riboswitches using multiple criteria, extract information about riboswitch location and gene/operon it regulates. RiboD provides a very useful resource that can be utilized for the better understanding of riboswitch-based gene regulation in bacteria and archaea.

Availability and implementation: RiboD can be freely accessed on the web at http://ribod.iiserkol.ac.in/.

Contact: mukherjee.sumit89@gmail.com or supratim.sen@iiserkol.ac.in

1 Introduction

Riboswitches are cis-regulatory conserved non-coding structural RNA sensors that are present in the 5′ untranslated regions (UTRs) of the bacterial mRNA, bind to specific ligands and control the expression of downstream genes (Mandal and Breaker, 2004). A riboswitch is composed of two domains; a highly conserved aptamer domain responsible for ligand binding, and the expression platform that changes conformation on ligand binding thereby regulating the expression of associated genes by a variety of mechanisms that include transcription or translation termination or a combination of those two effects (Nudler and Mironov, 2004). Riboswitches can be classified into different classes according to the type of ligand that binds to their aptamer domain (Barrick and Breaker, 2007). In bacteria, more than thirty classes of riboswitches are known (McCown *et al.*, 2017). Of these, thiamine pyrophosphate (TPP) is the only known riboswitch class that is also found in eukaryotes like fungi, plant and algae (Mukherjee *et al.*, 2018; Sudarsan *et al.*, 2003). Riboswitches provide a mechanism for bacteria to effectively respond to changes in their cellular environment and can therefore be potentially manipulated to control gene expression. This aspect has been exploited in engineering synthetic riboswitches that can bind and respond to natural metabolites (Zhou and Zeng, 2015). Since riboswitches also regulate genes responsible for virulence and/or antibiotic resistance (Blount and Breaker, 2006) the ability to target existing or synthetic riboswitches in pathogenic bacteria holds promise for the development of novel anti-bacterial therapeutics.

With the advent of new technologies such as high throughput sequencing technology, the amount of microbial genomic data has increased rapidly. In the last decade, a substantial number of ion and metabolite sensing riboswitch classes that regulate genes essential for bacterial survival (McCown *et al.*, 2017) has been discovered. In view of their important regulatory role in bacteria, it is clear that the benefits of effectively manipulating riboswitches depend on our ability to accurately identify them and construct a detailed map of their genomic location. Even though existing generic databases such as Rfam (Nawrocki *et al.*, 2015) do provide information on covariance model (CM) predicted genomic location of different classes of riboswitches, they lack important details such as information about riboswitch-regulated genes/operons. It is therefore of utmost importance to extract and compile all relevant data in a single platform to enable researchers to effectively use that information to gain new insights into the riboswitch-based gene regulations in prokaryotes. We developed a new database for prokaryotic riboswitches named RiboD that compiles all the relevant information on 31 different metabolite and ion-sensing riboswitch classes and their associated genes/operons from 1777 completely sequenced prokaryotic genomes. The database provides various analysis and search options to facilitate easy extraction of the specific information associated with riboswitches that the user is interested in.

2 Materials and methods

All sequenced bacterial and archeal genomes were retrieved from the RefSeq Database (Tatusova *et al.*, 2015). Riboswitches listed in RiboD are predicted using the covariance model (CM) downloaded from Rfam database (Nawrocki *et al.*, 2015), and the putative riboswitches are considered legitimate if the CM score was greater than or equal to the trusted cut-off provided in Rfam for each riboswitch class. To extract the riboswitch regulated downstream genes, the genomic coordinate of the riboswitches are mapped with the corresponding gff file of this genome using BEDtools (Quinlan and Hall, 2010). All of the riboswitch-regulated genes were correlated with the dataset collected from DOOR database (a database for prokaryotic operons) (Mao *et al.*, 2014) and if those riboswitch regulated genes were found to be present in an operon, the detailed operon information was extracted. Currently, the RiboD database contains information on riboswitches for 1777 prokaryotic genomes and 31 metabolite and ion sensing riboswitch classes. The database was developed with a variety of search capabilities to facilitate easy access and utilization of the information associated with riboswitches as per user needs. The structure provided in RiboD for each detected riboswitch was generated using a combination of the covariance model alignment and energy minimization methods. First, the sequences were aligned to a covariance model taken from Rfam using cmsearch from the Infernal package (Nawrocki and Eddy, 2013). The aligned consensus structure was then passed to RNAfold from the Vienna RNA package (Lorenz *et al.*, 2011) as enforced structure constraints to obtain a local minimum energy structure. We highlight base pairs from the covariance model alignment using thick red lines while the additional minimum energy base pairs are denoted by blue lines.

The RiboD was developed on Apache HTTP server with MySQL 5.7 at the back end, and the PHP 5.5, HTML and JavaScript at the front end. CSS was used for formatting the document written in a markup language. The PHP-based web interfaces were designed to execute the SQL queries dynamically. We chose Apache, MySQL, PHP because these are open-source and platform-independent software.

3 Results and discussion

The RiboD database comprises of the genome-specific information on 31 computationally predicted metabolite and ion-sensing riboswitch classes and associated genes/operons. The database allows users to query it for riboswitches based on a variety of search criteria and generates a table listing all riboswitches matching the query along with information about their genomic location and regulated gene/operon.

In the ‘Search’ tab of RiboD, the user can search for riboswitches based on five different options accessible from a dropdown menu. The first option allows the user to carry out a search based on ‘Riboswitch Class’ by selecting one or more of the 31 different classes listed. From this option, user can get the class-specific distribution of riboswitches present in the queried genome. The second option allows for riboswitch searches based on ‘Taxonomy’, where the user can search riboswitches from 35 bacterial and 9 archeal phyla provided in the checkbox. Such a search option provides a detailed picture of the distribution (Mukherjee *et al.*, 2017) of riboswitches in a specific order of bacteria/archaea. The third option enables search based on the regulated gene. If this gene name is listed in our database, the output table lists all riboswitches found upstream to the specified gene across all species. However, in many cases, it may not be possible to find the riboswitch information based on the user-provided gene-name query since the gene may not be appropriately annotated in a genome. The fourth option allows search for riboswitch-regulated genes based on the annotated biological processes or the pathways they are present in. The fifth option allows the user to find the entire list of computationally identified Tandem riboswitches which are multiple aptamers from same or different riboswitches classes, arranged in a composite gene control system that functions as a natural Boolean logic gate (Sudarsan *et al.*, 2006). From the tandem riboswitch searches, the researcher can get a genome-wide list of tandem riboswitches, which can facilitate further experimental studies to understand how multiple aptamers can be assembled to make complex chemical-sensing fuses without involving protein factors. In RiboD, for each genome, the riboswitch-regulated gene information is extracted from the gff file provided in RefSeq. In the gff file, the locus_tag is given as the gene name, wherever the queried gene name is not annotated. To maintain consistency with NCBI annotations, we provided the gene name as per the gff file annotations and linked all the riboswitch-regulated gene/operon information to the NCBI database from where the user can find more detailed information.

In the advanced search tab, the user can select multiple restricted search fields to refine the search query. The user should select at least two search fields and enter the keywords in the text-box to activate the advanced search function. For example, if the user wants to find if the lysC gene is regulated by the lysine riboswitch in Bacillus, the user can (i) select Genome_Name/ID from search field 1 and enter the keyword ‘bacillus’ (ii) select Riboswitch Class from search field 2 and enter the keyword ‘lysine’ and (iii) select the Gene_Name from search field 3 and enter the keyword ‘lysC’. The user can also search for all riboswitches in a genome of interest from the ‘Genomes’ tab by specifying the genome name/ID. Doing so provides a list of all riboswitches present in that genome, along with their genomic locations, regulated gene/operon and secondary structure. If the specific bacterial/archeal genome of interest is not available in RiboD, the user can upload the genome sequence in FASTA file format and genomic features in NCBI gff file format in the ‘Predict’ tab. Then our previously developed pHMM-based (Singh *et al.*, 2009) detection-method named Riboswitch Scanner (Mukherjee and Sengupta, 2016) is employed to search for riboswitches belonging to user-specified class(es) and extracted riboswitches and riboswitch-regulated gene information.

4 Conclusions

In summary, we present the first database of prokaryotic riboswitches that provides a comprehensive list of computationally predicted riboswitches, regulated gene/operon along with their genomic locations and predicted structures, from all the sequenced prokaryotic genomes available in RefSeq. It is an open-access database that allows users to easily extract and download information (as an excel file). We believe that this database will motivate and facilitate novel experimental studies on diverse aspects of these intriguing RNA regulators.

Acknowledgements

SDM thanks CSIR, India for a research fellowship.

Conflict of Interest: none declared.

References

Barrick,J.E. and Breaker,R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.*, **8**, R239.

Blount,K.F. and Breaker,R.R. (2006) Riboswitches as antibacterial drug targets. *Nat. Biotechnol.*, **24**, 1558–1564.

Lorenz,R. *et al.* (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.

Mandal,M. and Breaker,R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451–463.

Mao,X. *et al.* (2014) DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res.*, **42**, D654–D659.

McCown,P.J. *et al.* (2017) Riboswitch diversity and distribution. *RNA*, **23**, 995–1011.

Mukherjee,S. *et al.* (2017) Comparative genomics and phylogenomic analyses of lysine riboswitch distributions in bacteria. *PLoS One*, **12**, e0184314.

Mukherjee,S. *et al.* (2018) Phylogenomic and comparative analysis of the distribution and regulatory patterns of TPP riboswitches in fungi. *Sci. Rep.*, **8**, 5563.

Mukherjee,S. and Sengupta,S. (2016) Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics*, **32**, 776–778.

Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.

Nawrocki,E.P. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.

Nudler,E. and Mironov,A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.*, **29**, 11–17.

Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Singh,P. *et al.* (2009) Riboswitch detection using profile hidden Markov models. *BMC Bioinformatics*, **10**, 325.

Sudarsan,N. *et al.* (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*, **9**, 644–647.

Sudarsan,N. *et al.* (2006) Tandem riboswitch architectures exhibit complex gene control functions. *Science*, **314**, 300–304.

Tatusova,T. *et al.* (2015) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.*, **43**, 3872–3872.

Zhou,L.B. and Zeng,A.P. (2015) Engineering a lysine-ON riboswitch for metabolic control of lysine production in *Corynebacterium glutamicum*. *ACS Synth. Biol.*, **4**, 1335–1340.