

Pro54DB: a database for experimentally verified sigma-54 promoters

Zhi-Yong Liang¹, Hong-Yan Lai¹, Huan Yang¹, Chang-Jian Zhang¹, Hui Yang¹, Huan-Huan Wei¹, Xin-Xin Chen¹, Ya-Wei Zhao¹, Zhen-Dong Su¹, Wen-Chao Li¹, En-Ze Deng¹, Hua Tang², Wei Chen^{1,3,*} and Hao Lin^{1,*}

¹ Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology and Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 610054, China.

² Department of Pathophysiology, Southwest Medical University, Luzhou 646000, China.

³ Department of Physics, School of Sciences, Center for Genomics and Computational Biology, North China University of Science and Technology, Tangshan 063009 China.

Associate Editor: Dr. John Hancock

ABSTRACT

Summary: In prokaryotes, the σ^{54} promoters are unique regulatory elements and have attracted much attention because they are in charge of the transcription of carbon and nitrogen-related genes and participate in numerous ancillary processes and environmental responses. All findings on σ^{54} promoters are favorable for a better understanding of their regulatory mechanisms in gene transcription and an accurate discovery of genes missed by the wet experimental evidences. In order to provide an up-to-date, interactive and extensible database for σ^{54} promoter, a free and easy accessed database called Pro54DB (σ^{54} promoter database) was built to collect information of σ^{54} promoter. In the current version, it has stored 210 experimental-confirmed σ^{54} promoters with 297 regulated genes in 43 species manually extracted from 133 publications, which is helpful for researchers in fields of bioinformatics and molecular biology.

Availability and Implementation: Pro54DB is freely available on the web at <http://lin.uestc.edu.cn/database/pro54db> with all major browsers supported.

Contact: Wei Chen: gretchchen@ncst.edu.cn,

Hao Lin: hlin@uestc.edu.cn

1 INTRODUCTION

The initiation of transcription is the most important stage in the regulation of gene expression determined by promoters. In prokaryotes, the σ factor of RNA holoenzyme plays key roles in recognizing and binding to the promoter sequences during gene transcription (Barrios, 1999). Thus, the types of prokaryotic promoters are defined according to the types of σ factor. Among the known seven types of σ factors, namely σ^{70} , σ^{54} , σ^{38} , σ^{32} , σ^{28} , σ^{24} and σ^{19} , the σ^{54} factor can recognize the unique regulatory elements with the consensus sequence TGC[AT][TA] around -12bp and [CT]TGGCA[CT][GA] around -24bp upstream of the transcription start site (TSS). By affecting the start site and time of gene transcription in cells, σ^{54} promoters are responsible for a series of ancillary processes and environmental responsive procedures (Bernardo, 2009), including the expression of chemotaxis transducers, assembly of motility organs (Wigneshweraraj 2008), nitro-

gen fixation (Kustu 1989), arginine catabolism (Gardan, 1995), alginate biosynthesis (Zielinski 1992), flagella assembly (Arora, 1997), and the like. Therefore, the knowledge of σ^{54} promoters is of great significance to understand the mechanism of gene transcription and to establish the network of gene expression. In recent years, with multiple approaches being employed to identify σ^{54} promoters (Bonocora, et al., 2015; Lin, et al., 2014), the construction of related database providing relevant knowledge is very important.

Although several special databases already exist for prokaryotic promoters, such as DBTBS (Sierro, 2008) and RegulonDB (Gama-Castro, et al., 2016), seldom do they specially devote to collecting and analyzing the σ^{54} promoters comprehensively. To fill this gap, we built Pro54DB (<http://lin.uestc.edu.cn/database/pro54db>), which is supposed to facilitate the identification of σ^{54} promoters and assist with further study in biology sciences.

2 METHODS

2.1 Data collection and organization

Query was made with the unquoted words “sigma54 promoter” or “ σ^{54} promoter” against all fields of PubMed indexed literatures published before Aug 31, 2016. The search hit 657 targets, each of which was then processed according to the following standards: (i) only the papers with promoter sequences were reserved; (ii) the papers without experimentally confirmed TSS were excluded. As a result, 133 published papers with available full text were obtained.

The flowchart of the construction of Pro54DB is depicted in Fig. 1A. The information of experimentally confirmed σ^{54} promoter, such as sequences, names, genes, species and TSS positions were extracted manually from the articles mentioned above. The promoter sequences were mapped into their genomes using BLAST program with default parameters. The gene product and position are mainly extracted from Uniprot and GenBank.

2.2 Database implementation

The MySQL relational database management system is used to store and manage the data. At the backend, four main tables and one joint table are created. Most table names and field names are

*To whom correspondence should be addressed.

self-evident. Also, the meanings of fields in detail are attached to the end of each data file. The core table is named as ‘Main-set’, which conjoins the other three ‘Promoter’, ‘Gene’ and ‘Species’. As the fourth ‘Product’ is only related to ‘Gene’, it is excluded in ‘Main-set’ but displayed along with gene search. Meanwhile, Pro54DB is designed to hyperlink several major external databases such as NCBI, Uniprot, DBTBS and RegulonDB. All component interfaces are coded in HTML/CSS. At the frontend, each entry of the Pro54DB has three tables ‘promoter’, ‘gene’ and ‘species’ corresponding to those at the backend for users to browse (Fig. 1C).

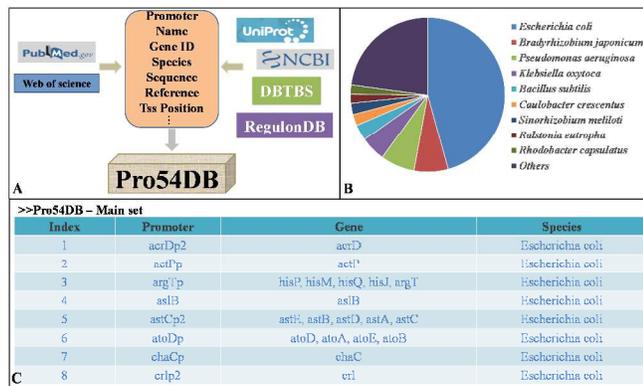


Fig. 1. The flow chart and visual process of Pro54DB. (A) The database comprises versatile information of σ^{54} promoters, including promoter name, gene ID, species, sequence, reference, TSS position, etc. (B) The pie graph indicates the data distribution across 43 species. (C) The content of Main-set in the ‘Browse’ page.

3 RESULTS

The database has stored 210 σ^{54} promoters with 297 regulated genes as well as their products across 43 different species. The main page has ten buttons ‘Home’, ‘Browse’, ‘search’, ‘BLAST’, ‘Utilities’, ‘Statistics’, ‘Download’, ‘Links’, ‘References’ and ‘Help’. All the entries included can be explored by pressing the ‘Browse’ button, where unfolds a pull-down menu offering three options ‘Promoter’, ‘Gene’ and ‘Species’. Users may achieve the information by clicking corresponding option. In the ‘Search’ page, typing a keyword will return three columns of entries with hyperlinks for the details of specific promoter(s) or gene(s) (Fig. 1 C). A BLAST program based on BALSTN 2.2.28+ is also provided for aligning query sequences in FASTA format against σ^{54} promoters present in Pro54DB database. Based on a published tool named iPro54-PseKNC (Lin, et al., 2014), we also embed the tool in the page of ‘Utilities’ to implement promoter prediction. To keep up to date and interactive, data uploads are encouraged through the ‘Submit’ channel. All data about σ^{54} promoters are packaged with three formats (xml, csv and txt) in the ‘Download’ page for batch downloading. The ‘Statistics’ page (Fig. 1 B) presents a basic statistical analysis of σ^{54} promoters in the Pro54DB. The ‘Links’ page contains a list of external web resources. The 133 published references have been listed in “References” page. For the convenience of users, a tutorial is provided in the ‘Help’ page to help users to

obtain the desired results. The Pro54DB is freely available at <http://lin.uestc.edu.cn/database/pro54db>

4 DISCUSSION AND CONCLUSIONS

Pro54DB is a conformable database covering abundant experimentally validated information of σ^{54} promoters, including promoter sequences, TSS, regulated genes, species, products, related references and brief descriptions, etc. We operated on the principle of gathering and assembling information used for σ^{54} promoter prediction and making it free for non-commercial projects. The maintenance and revision of Pro54DB database will keep on going with a cycle of half. New σ^{54} promoters being reported will be incorporated continually to comprise more essential experiments and species, as well as more detailed information about each item in Pro54DB.

Funding:

This work was supported by the Applied Basic Research Program of Sichuan Province (No. 2015JY0100 and LZ-LY-45), the Fundamental Research Funds for the Central Universities of China (Nos. ZYGX2015J144; ZYGX2015Z006), Program for the Top Young Innovative Talents of Higher Learning Institutions of Hebei Province (No. BJ2014028), the Outstanding Youth Foundation of North China University of Science and Technology (No. JP201502), China Postdoctoral Science Foundation (No.2015M582533) and the Scientific Research Foundation of the Education Department of Sichuan Province (11ZB122).

REFERENCES

- Arora, R.B., Almira EC, Lory S, Ramphal R (1997) A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in *Pseudomonas aeruginosa* in a cascade manner, *J Bacteriol*, **179**, 5574-5581.
- Barrios, H., B. Valderrama, and E. Morett (1999) Compilation and analysis of sigma(54)-dependent promoter sequences, *Nucleic Acids Res*, **27**, 4305-4313.
- Bernardo, J.L., Skårstød E, Shingler V (2009) sigma54-promoter discrimination and regulation by ppGpp and DksA, *J Biol Chem*, **284**, 828-838.
- Gama-Castro, S., et al. (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond, *Nucleic Acids Research*, **44**, D133-143.
- Gardan, R., G. Rapoport, and M. Debarbouille (1995) Expression of the rocDEF operon involved in arginine catabolism in *Bacillus subtilis*, **249**, 843-856.
- Kustu, S.E., Keener J, Popham D, Weiss D (1989) Expression of sigma 54 (ntrA)-dependent genes is probably united by a common mechanism, *Microbiol Rev*, **53**, 367-376.
- Lin, H., et al. (2014) iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition, *Nucleic acids research*, **42**, 12961-12972.
- Sierro, M.Y., de Hoon M, Nakai K (2008) DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information, *Nucleic Acids Res*, **36**, 93-96.
- Wigneshweraraj, B.D., Burrows PC, Joly N, Schumacher J, Rapps M, Pape T, Zhang X, Stockley P, Severinov K, Buck M. (2008) Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor, *Mol Microbiol*, **68**, 538-546.
- Zielinski, M.R., Roychoudhury S, Danganan CE, Hendrickson W, Chakrabarty AM (1992) Alginate synthesis in *Pseudomonas aeruginosa*: environmental regulation of the algC promoter, *J Bacteriol*, **174**, 7680-7688.