ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 7 | Issue: 7 | July 2021|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2021: 8.047|| ISI Value: 1.188

FPMD-FUNGAL PROMOTER MOTIF DATABASE: A DATABASE FOR THE OF THE PROMOTER MOTIFS REGIONS IN FUNGAL GENOMES

Sudheer Menon[1], Shanmughavel piramanayakam[2*], Gopal Prasad Agarwal[3]
[1&2] Department of Bioinformatics, Bharathiar University, Coimbatore
[3] DBEB, Indian Institute of Technology - Delhi, India

[*]**Corresponding Author: Shanmughavel piramanayakam**

Article DOI: https://doi.org/10.36713/epra7928
DOI No: 10.36713/epra7928

ABSTRACT

Fungal promoter motif database is the collection of promoter motifs from fully sequenced fungal genomes. Promoter sequences and its frequency are analyzed by the positions of nucleotide sequence and its repetition. The fungal promoter motif database holds the promoter sequence motifs identified by genome wide motif discovery, similarity studies and clustering. These data sets are typically 6 to 10 bp long, that have been extracted from the promoter regions. These promoter regions extend from 1.5 kb upstream to 200bp downstream of a transcription start site. We believe that the availability of these promoter motifs will be a valuable resource for researchers for comparative sequence analysis and evolutionary studies.
The database is available through the World Wide Web at URL:http://sites.google.com/site/fungalpromotermotifdatabase/

KEYWORDS: ***promoter motifs; fungal genome; comparative sequence analysis; evolutionary studies.***

BACKGROUND

FPMD was a designed as a result of comparative sequence analysis or by-product of a of promoter characterization in fully sequenced fungal genome. FPMD defined as a database of promoter motifs, not as a promoter database. The availability of genomic sequence enabled scientist to compare large dataset of promoter sequences of various organisms. Comparisons of sequences between species are limited to the identification of functional regions which are evolutionary conserved. We have focused upon the promoter motifs possible on the promoter those results in transcription. The characterization of evolutionarily conserved promoter region is a potent example of putative sequence of human genome with biological activity. In order to access these results of comparison major varieties of databases have been implemented. These databases are exploited by researchers for the identification of species similarity that are conserved by particular features. Sequence comparisons between human and fungal promoters can be implemented for the identification of regulatory networks However, these are limited to the availability of fully sequences genome.

We developed FPMD by analysis of: (1) fully sequenced fungal genome; (2) extraction of whole promoter region (3) promoter motif identification (4) the comparison of match report with human genome. The latest version of FPMD contains the promoter motifs of 10 fungal species and its match reports with human genome. The importance of FPMD is that it gives complete information regarding the possible promoter regions on each chromosomes of fully sequenced fungal genome. Eventhough there are databases like TRANSFAC and COMPEL, are meant for transcription factors, FPMD is the collection of whole dataset and its matchings with human genome.

Promoter analysis involves systematic monitoring of promoter regions in specific organisms. This could be achieved by developing databases containing information similar to many other fungal promoter databases. Here, we describe the development

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 7 | Issue: 7 | July 2021|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2021: 8.047|| ISI Value: 1.188

of a database containing information for promoter analysis in fully sequenced fungal genome.

METHODOLOGY

Dataset

Sequences were downloaded from NCBI site, generated from fully sequenced fungal species were used to construct the database.

BIOINFORMATICS ANALYSIS

Database Implementation

Data is stored up loader website running in windows server. The database architecture is shown in Figure 1.

Database Interface

The database interface is implemented using HTML. The sample interface is given in Figure 1.

Database description

The Fungal promoter motif database consists of promoter motifs of fully sequenced fungal genome. The database contains promoter motifs from 7 chromosomes of *Debaryomyces hansenii, 11 chromosomes of Encephalitozoon cuniculi, 7* chromosomes of *Eremothecium gossypii, 14* chromosomes of *Filobasidiella neoformans*, 6 chromosomes of *Kluyveromyces lactis, 8* chromosomes of *Pichia stipitis,* 16 chromosomes of *Saccharomyces cerevisiae, 3* chromosomes of *Schizosaccharomyces pombe, 6* chromosomes of *Yarrowia lipolytica and 13* chromosomes of *Candida glabrata.*The promoter regions of fully sequenced fungal genome were extracted and the promoter motifs were identified. These datasets can be downloaded and extracted using winrar or winzip. These sequences are saved in text file (Fig-1).

Development of FPMD

Fungal promoter motif database is constructed using html and are available to access at http://sites.google.com/site/fungalpromotermotifdatabase/. Data were downloaded from NCBI site. Fungal promoter motif database includes Promoter motifs of 91 chromosomes; form 10 fully sequenced fungal genomes. The complete list of fungus can be found at http://sites.google.com/site/fungalpromotermotifdatabase/list. However, the list is not complete. They fungal species are alphabetically arranged corresponding to their chromosome numbers. The database can be accessed by downloading the specific chromosome of interest.

Features of FPMD

Each entry in fungal promoter motif database is provided with links to the downloadable file. The fungal species row in records display a list of fungus and the representative chromosomes are highlighted. The characteristic feature of fungal promoter motif database provided in each record displays the list of characteristic feature and utility as a pop-up window indicating the importance of specific fungal species.
List of fungal species:
http://sites.google.com/site/fungalpromotermotifdatabase/list

Utility

Fungal promoter motif database pointed the importance of promoter motifs possessed by fungal species. The database finds utility to the scientific community for a quick review on the number of fully sequenced fungal species, promoter region and the motifs present in it. The database finds utility in fungal evolutionary studies. The database is freely available in public domain and the data can be accessed. The database also provides a collection of match reports to human genome. The suite of user interfaces (Fig-1) allow the user to browse the database and query for: (a) List of fully sequenced fungal genomes, (b) Total size of the genome (c) Advantage of each species (d) predicted promoter sequences (e) Match reports with that of human genome. The availability of this dataset is a useful resource for researchers studying the localization of promoter regions and in the evolutionary studies.

Future development

We frequently update and enlarge the database and experimentation as the fully sequenced data becomes available; annotate these promoter regions with experimental results of a particular fungal species; and provide users with a match score with other eukaryotes in order to study the similarity. We plan to interconnect this server to other sequence databases to study out the evolutionary conservation and to visualize these conserved regions. Since FPMD extracts and collects all promoter regions present in the species, it has the potential to overrule other databases and a major centre for the study of fungal genomics. This database provides all essential information about the promoter motifs in fungal genomes and is useful in comparative genomics. This database is up to date till this time, as there are only 10 fungal species been fully sequenced and will be expanded based upon availability of sequenced genome.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 7 | Issue: 7 | July 2021|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2021: 8.047|| ISI Value: 1.188

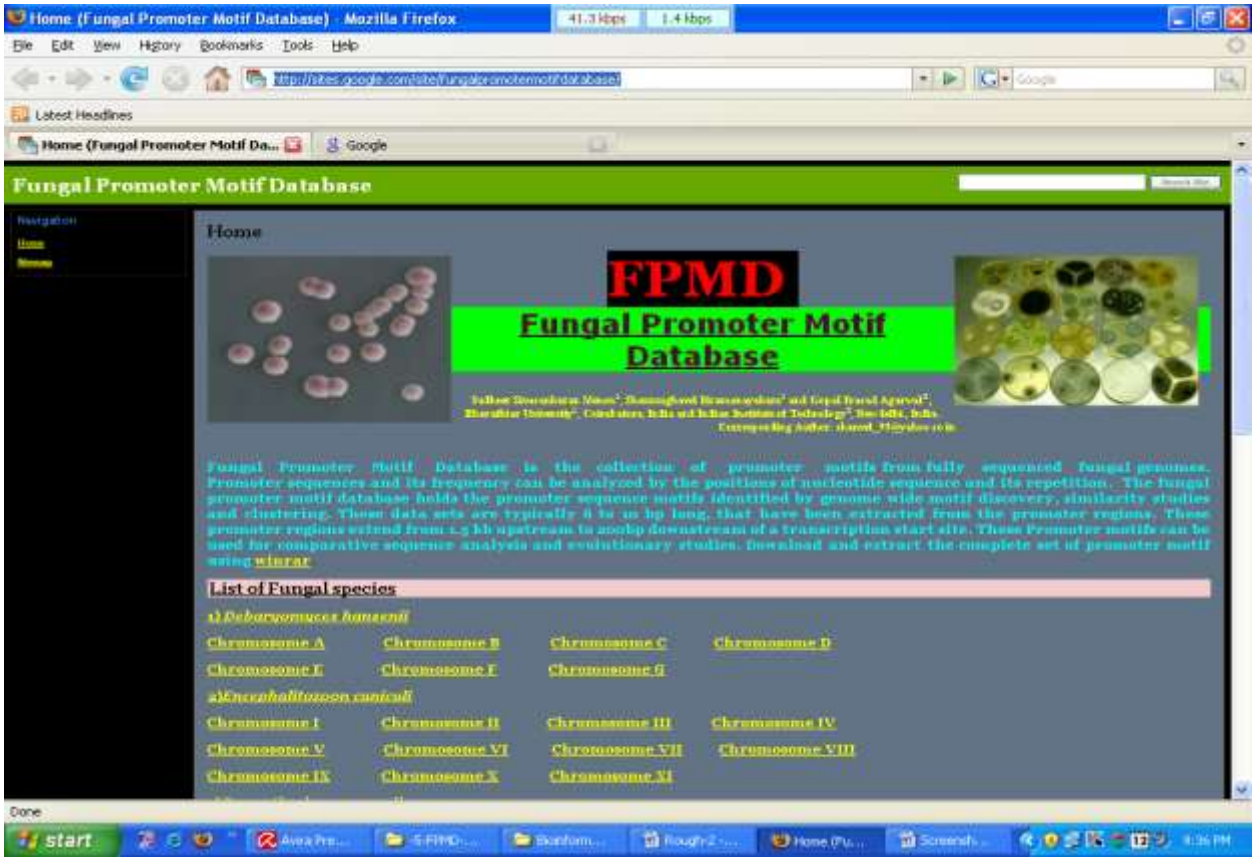

Figure 1: A screen shot of *fungal promoter motif record* entry in FPMD is shown

CONCLUSION

Day by day trillions of sequences are generated word-wide. Finding meaningful information out of it requires effort and time. These sequences can be compared with several databases for functional annotation. There are several databases available online with the collection of promoter regions but especially for the promoter motif regions and that too in fugal species are very rare. This database will help the scientist to search for the promoter motif regions in fungus or they can compare these promoter motif regions with another species.

Author Contributions

SM performed the analyses. SP and GP designed the study. SM wrote the manuscript. All authors approved the manuscript.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

The authors greatly acknowledge the facilities provided by Department of Biotechnology-Bioinformatics Infrastructure Facility (DBT-BIF), Bharathiar University, Coimbatore 641046, INDIA.

REFERENCES

1. *Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.*
2. *Bajrai, L. H., Benamar, S., Azhar, E. I., Robert, C., Levasseur, A., Raoult, D., et al. (2016). Kaumoebavirus, a new virus that clusters with faustoviruses and Asfarviridae. Viruses 8:E278. doi: 10.3390/v8110278*
3. *Benamar, S., Reteno, D. G., Bandaly, V., Labas, N., Raoult, D., and La Scola, B. (2016). Faustoviruses: comparative genomics of new Megavirales family members. Front. Microbiol. 7:3. doi: 10.3389/fmicb.2016.00003*

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 7 | Issue: 7 | July 2021|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2021: 8.047|| ISI Value: 1.188

4. *Bieda M, Xu X, Singer MA, Green R, Farnham PJ. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res. 2006;16:595–605.*
5. *Blanchette M, Bataille AR, Chen X, Poitras C, Laganiere J, Lefebvre C, Deblois G, Giguere V, Ferretti V, Bergeron D, et al. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res. 2006;16:656–68.*
6. *Boyer, M., Gimenez, G., Suzan-Monti, M., and Raoult, D. (2010). Classification and determination of possible origins of ORFans through analysis of nucleocytoplasmic large DNA viruses. Intervirology 53, 310–320. doi: 10.1159/000312916*
7. *Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proc. Natl. Acad. Sci. U.S.A. 106, 21848–21853. doi: 10.1073/pnas.0911354106*
8. *Browning, D. F., and Busby, S. J. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. 2, 57–65. doi: 10.1038/nrmicro787*
9. *Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell. 2004;116:499–509.*
10. *Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.*
11. *Elnitski L, Jin VX, Farnham PJ, Jones SJ. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. Genome Res. 2006;16:1455–64.*
12. *Favorov AV, Gelfand MS, Gerasimova AV, Ravcheev DA, Mironov AA, Makeev VJ. A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length. Bioinformatics. 2005;21:2240–5.*
13. *Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Olio V. Dall', Zardo G, Nervi C, Bernard L, Amati B. Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol. 2006;8:764–70.*
14. *Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, et al. Transcriptional regulatory code of a eukaryotic genome. Nature. 2004;431:99–104.*
15. *Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, et al. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. 2010;20(6):861–73.*
16. *Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. Cell. 2013;152(1–2):327–39.*
17. *Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. 2008;36(16):5221–31.*
18. *Marino-Ramirez L, Jordan IK, Landsman D. Multiple independent evolutionary solutions to core histone gene regulation. Genome Biol. 2006;7:R122.*
19. *Pavesi G, Mereghetti P, Zambelli F, Stefani M, Mauri G, Pesole G. MoD Tools: regulatory motif discovery in nucleotide sequences from coregulated or homologous genes. Nucleic Acids Res. 2006;34:W566–70.*
20. *Pavesi G, Zambelli F, Pesole G. WeederH: an algorithm for finding conserved regulatory motifs and regions in homologous sequences. BMC Bioinformatics. 2007;8:46.*
21. *Régnier M, Denise A. Rare events and conditional events on random strings. Discrete Math Theor Comput Sci. 2004;6:191–214.*
22. *Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, et al. Genome-wide location and function of DNA binding proteins. Science. 2000;290:2306–9.*
23. *Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–100.*
24. *Sinha S, Tompa M. YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. 2003;31:3586–8.*
25. *Staden R. Methods for calculating the probabilities of finding patterns in sequences. Comput Appl Biosci. 1989;5:89–96.*
26. *Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, et al. Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep. 2014;8(6):2015–30.*
27. *Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P, Moreau Y. A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics. 2001;17:1113–22.*
28. *Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, et al. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol. 2005;23:137–44.*
29. *Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet. 2004;5:276–87.*
30. *Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014;158(6):1431–43.*