DOI 10.18699/BGRS/SB-2020-024

Functional annotation of the transcription factors from *Methylotuvimicrobium alcaliphilum* 20Z^R

Semyon K. Kolmykov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
FRC Institute of Cytology and Genetics SB RAS Novosibirsk, Russia
kolmykovsk@gmail.com

Nikita V. Ivanisenko
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
n.ivanisenko@gmail.com

Ivan S. Evshin
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
ivan@developmentontheedge.com

Mikhail Kulyashov
BIOSOFT.RU, LLC
Institute of Computational Technologies SB RAS Novosibirsk State University
Novosibirsk, Russia
m.kulyashov@mail.ru

Tamara M. Khlebodarova
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk, Russia
tamara@bionet.nsc.ru

Ilya R. Akberdin
BIOSOFT.RU, LLC
FRC Institute of Cytology and Genetics SB RAS
Novosibirsk State University
Novosibirsk, Russia
akberdinir@gmail.com

***Abstract* — Methane is a promising carbon source for biosynthesis of biotechnologically useful compounds using aerobic methanotrophic bacteria as biocatalysts. Despite more than a century-long history of discovering and studying of methanotrophic microorganisms, knowledge of the molecular mechanisms of gene expression regulation by transcription factors in these bacteria is very limited with only a few isolated cases being published. Therefore, the identification of potential transcription factors for methanotrophic organisms and their target genes is not only a foreground fundamental problem in the research field of methanotrophy, but it is also especially relevant for the active development of biotechnological application of methane-oxidizing microorganisms. In this study a comparative genomics approach together with the structural modeling techniques were applied to reveal the TFs in the 20ZR genome and predict their target regulatory genes.**

Keywords — Transcription factors, haloalkaliphilic, aerobic, methanotrophs, bioinformatics

Introduction

Haloalkaliphilic aerobic methanotrophs, including *Methylotuvimicrobium alcaliphilum* 20Z^R strain (20ZR) are standing out as the promising microbial "factories" in industrial research as new sources of enzymes and protein based materials that are resistant to high salt and pH levels, and natural producers of amino acids, sugars, and osmoprotectants [1-3]. Moreover, the use of modern genetic engineering approaches, methods of genomics and system biology allows to significantly expand the scope of their application in the production of biofuels, biodegradable polymers and other valuable commercial products.

Functional annotation and search for new transcription factors (TFs) of 20ZR present enormous fundamental and applicable interest in understanding the life-cycle and biochemical pathways regulation. In the current work a comparative genomics approach together with the structural modeling techniques were applied to reveal the TFs in the 20ZR genome and predict their target regulatory genes.

Methods

Identification of potential TFs encoding in 20ZR genome

Prediction of TFs and DNA-binding proteins was carried out using tools that analyze the level of homology and similarity with known DNA-binding proteins and TFs. InterProScan package was applied to identify TFs and DNA-binding domains in the set of protein sequences [4]. To improve the quality of putative TFs identification, a list of HMM models of DNA-binding protein and TF domains for subsequent filtering of the InterProScan results was made based on information from the P2TF [5] and DBD [6] databases. Additionally to identify putative transcription factors in the set of proteins with unknown function DBD-Threader [7] and TFPredict [8] were used.

Identification of transcription factor target genes

To solve the problem of searching for TF target genes, an analysis of the scientific publications on transcriptional regulation of the genes expression of 20ZR and closely related groups of bacteria were carried out. The obtained set of regulated genes was expanded to include genes located in the same operons as the identified target genes. Information on the structure of 20ZR operons was obtained from the database of prokaryotic operons (Prokaryotic Operon DataBase; ProOpDB) [9].

In order to identify potential TF binding sites (TFBSs), promoter regions of 20ZR genes and their orthologs were analyzed using the RSAT software package [10] and frequency matrices of TF binding sites (PFM) from the PRODORIC2 [11] and RegulonDB [12] databases. Additionally a comparative analysis of putative 20ZR TFs and TFs of prokaryotes with known PFMs of TFs (DBs PRODORIC2 and RegulonDB) was conducted. Based on the obtained data on the level of homology, a consensus PFMs for the 20ZR TFBSs were constructed. The resulting consensus PFMs were used to search for potential TFBSs. The set of target genes under consideration will be expanded based on the obtained list of TFBSs.

To reconstruct 20ZR regulons based on the available genome-wide gene expression profiles (RNA-seq), correlation dependencies were analyzed and the corresponding clusters of coexpressed putative target genes and TFs were identified.

A structural annotation of TFs was carried out using homology modeling techniques, including MODELLER software [13]. Prediction and analysis of TFs interaction with target DNA fragments were analyzed using FoldX [14] and Rosetta molecular modeling software [15].

Results And Discussion

Functional and structural annotations of TFs of 20ZR were carried out for the first time. The comparative genomics approach together with molecular modeling techniques allowed to reveal the majority of TFs and assign its potential regulatory function.

The screening for DNA binding domains in the 20ZR proteome allowed us to predict the new putative TFs with unknown function. For these proteins the potential regulons in the 20ZR were predicted and ranked using *in silico* protein-dna binding energy estimations.

Obtained results provide essential data to reconstruct gene regulatory networks operating 20ZR metabolism and reveal their unique features for haloalkaliphilic aerobic methanotrophs.

REFERENCES

[1] Y.A., Trotsenko and V.N. Khmelenina, "The biology and osmoadaptation of haloalkaliphilic methanotrophs," Microbiology, vol. 71, № 2, pp.123-132, 2002.

[2] C.A. Henard and M.T. Guarnieri, "Metabolic Engineering of Methanotrophic Bacteria for Industrial Biomanufacturing," in Methane Biocatalysis: Paving the Way to Sustainability, Springer, Cham, 2018, pp. 117-132.

[3] S. Nariya and M.G. Kalyuzhnaya, "Diversity, Physiology, and Biotechnological Potential of Halo (alkali) philic Methane-Consuming Bacteria," in Methanotrophs, Springer, Cham, 2019, pp. 139-161.

[4] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, and S. Pesseat, "InterProScan 5: genome-scale protein function classification," Bioinformatics, vol. 30, №9, pp.1236-1240, 2014.

[5] P. Ortet, G. De Luca, D.E. Whitworth, and M. Barakat, "P2TF: a comprehensive resource for analysis of prokaryotic transcription factors," BMC genomics, vol. 13, № 1, p. 628, 2012.

[6] D. Wilson, V. Charoensawan, S.K. Kummerfeld, and S.A. Teichmann, "DBD—taxonomically broad transcription factor predictions: new content and functionality," Nucleic acids research, vol. 36, suppl_1, pp.D88-D92, 2007.

[7] M. Gao, and J. Skolnick, "A threading-based method for the prediction of DNA-binding proteins with application to the human genome," PLoS computational biology, vol. 5, № 11, p.e1000567, 2009.

[8] J. Eichner, F. Topf, A. Dräger, C. Wrzodek, D. Wanke, and A. Zell, "TFpredict and SABINE: sequence-based prediction of structural and functional characteristics of transcription factors," PloS one, vol. 8, № 12, p.e82238, 2013.

[9] B. Taboada, R. Ciria, C.E. Martinez-Guerrero, and E. Merino, "ProOpDB: Prokaryotic Operon DataBase," Nucleic acids research, vol. 40, № D1, pp.D627-D631, 2011.

[10] N.T.T. Nguyen, B. Contreras-Moreira, J.A. Castro-Mondragon, W. Santana-Garcia, R. Ossio, C.D. Robles-Espinoza, M. Bahin, S. Collombet, P. Vincens, D. Thieffry, and J. van Helden, "RSAT 2018: regulatory sequence analysis tools 20th anniversary," Nucleic acids research, vol. 46, № W1, pp.W209-W214, 2018.

[11] D. Eckweiler, C.A. Dudek, J. Hartlich, D. Brötje, and D. Jahn, "PRODORIC2: the bacterial gene regulation database in 2018," Nucleic acids research, vol. 46, № D1, pp.D320-D326 2017.

[12] A. Santos-Zavaleta, H. Salgado, S. Gama-Castro, M. Sánchez-Pérez, L. Gómez-Romero, D. Ledezma-Tejeida, J.S. García-Sotelo, K. Alquicira-Hernández, L.J. Muñiz-Rascado, P. Peña-Loredo, and C. Ishida-Gutiérrez, "RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12," Nucleic acids research, vol. 47, № D1, pp.D212-D220, 2018.

[13] N. Eswar, D. Eramian, B. Webb, M.Y. Shen and A. Sali, "Protein structure modeling with MODELLER," in Structural proteomicsm, Humana Press, 2008, pp. 145-159.

[14] A.D. Nadra, L. Serrano and A. Alibés, "DNA-binding specificity prediction with FoldX," in Methods in enzymology, vol. 498, Academic Press, 2011, pp. 3-18.

[15] S.J. Fleishman, A. Leaver-Fay, J.E. Corn, E.M. Strauch, S.D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon and J. Meiler, "RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite," PloS One, vol. 6, № 6, p.e20161, 2011